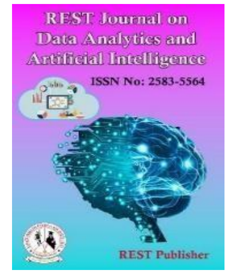
REST Journal on Data Analytics and Artificial Intelligence
Vol: 4(1), March 2025
REST Publisher; ISSN: 2583-5564
Website: http://restpublisher.com/journals/jdaai/
DOI: https://doi.org/10.46632/jdaai/4/1/88



Sentiment Analysis

*** Shaik Ashraf, Satyam. Abhinav, Dyaga Vinay, Rajesh Myakala**
Anurag University, Hyderabad, India.
*Corresponding Author Email: 21eg107b41@anurag.edu.in

Abstract: *With the growing popularity of online shopping, it has become essential to have reliable systems for evaluating products. This project creates a system that uses advanced natural language processing (NLP) techniques like Sentiment Analysis, Opinion Mining, and Aspect-Based Sentiment Analysis (ABSA) to provide accurate scores for products. It uses the Distil BERT model to analyze reviews and give ratings from 1 to 5 based on what customers write. ABSA helps the system break down feedback into specific areas, such as quality, price, and durability, offering more detailed insights. The system also detects emotions like happiness, anger, or sadness in reviews, giving a clearer picture of user opinions. It supports multiple languages, making it useful for analyzing reviews globally. Built with tools like Python, Beautiful Soup, NLTK, and Transformers, this system processes online reviews and converts them into easy-to-understand scores, helping buyers make better decisions.*

Key Words: *Sentiment Analysis, Aspect-Based Sentiment Analysis (ABSA), Emotion Detection DistilBERT, Multilingual Sentiment Analysis.*

1. INTRODUCTION

The rapid growth of e-commerce platforms like Amazon and Flipkart has transformed how customers make purchasing decisions, with online reviews becoming a critical source of information. While these reviews provide rich qualitative insights, traditional systems often reduce them to basic star ratings or keyword analysis. This oversimplification misses key emotions and specific product aspects that matter to customers. Furthermore, the lack of support for multiple languages limits the ability to analyze feedback from a global audience [1-4]. This gap between detailed reviews and simplified ratings creates significant challenges in capturing customer sentiment accurately. For example, a seemingly neutral review might express dissatisfaction, while subtle negative remarks in positive reviews often go undetected. Existing tools primarily focus on English reviews, ignoring valuable insights from other languages. As a result, traditional systems fail to provide a complete understanding of global customer feedback [5-9]. To address these limitations, this project develops an advanced system using NLP and machine learning models. It integrates Aspect-Based Sentiment Analysis (ABSA) to evaluate specific product attributes like quality and price and detects emotions for a deeper understanding of reviews. Multilingual analysis ensures inclusivity by processing reviews in different languages. By combining these features, the system bridges the gap between numerical ratings and rich, detailed customer insights [10-12].

2. EXISTING SYSTEM

Sentiment Analysis in E-commerce: The rapid growth of e-commerce platforms like Amazon and Flipkart has led to the emergence of online reviews as a vital source of information. These reviews significantly influence purchasing decisions, providing insights into customer satisfaction [13]. Traditional sentiment analysis methods, while effective to some extent, often fail to capture the richness of review data. They tend to oversimplify feedback into star ratings or broad sentiment categories, neglecting nuanced emotions and specific product aspects. The complexity of human language, including the use of sarcasm, implicit sentiment, and multilingual feedback, poses additional challenges to existing systems [14-17]. Challenges in Traditional Sentiment Analysis: Conventional sentiment analysis systems struggle with several limitations, such as their inability to detect specific emotions or provide granular feedback on aspects like price, quality, or durability. They often lack multilingual support, excluding valuable insights from reviews written in non-English languages. Moreover, traditional systems rarely address fake reviews or opinion spam, which can skew sentiment analysis outcomes. The absence of fine-grained analysis leaves businesses with an incomplete understanding of customer sentiment, limiting their ability to make informed decisions [18]. Advancements with NLP and Machine Learning: This project leverages advanced natural language processing (NLP) techniques and machine learning models to overcome these limitations. By integrating. Aspect-Based Sentiment Analysis (ABSA), the system analyzes specific product features, providing detailed insights into customer feedback. Emotion detection models further enrich the analysis by identifying sentiments like happiness, anger, or sadness. The inclusion of multilingual support ensures a global perspective by analyzing reviews in diverse languages. Through these innovations, the system addresses the gaps in traditional sentiment analysis, offering a comprehensive and accurate reflection of customer feedback [19]. Importance of Emotion Detection in Customer Feedback: Customer emotions play a crucial role in understanding the deeper context of reviews. A simple "neutral" sentiment may mask frustration, while "positive" feedback might include subtle negative remarks. Emotion detection goes beyond binary classifications to identify feelings such as happiness, sadness, anger, or surprise within review texts. This capability provides businesses with actionable insights into customer satisfaction and dissatisfaction, enabling more targeted product improvements and marketing strategies [20]. The particularities of IIoT environments are needed to address these issues. Examples include investigating sophisticated anomaly detection methods and adaptable security measures. Role of Aspect-Based Sentiment Analysis (ABSA): Aspect-Based Sentiment Analysis (ABSA) offers a more granular approach to sentiment analysis by evaluating specific attributes of a product, such as durability, price, or design. This allows businesses to understand which aspects customers appreciate or criticize. For instance, a review may highlight excellent battery life but poor camera quality in a smartphone. ABSA captures these distinctions, providing detailed insights into which features resonate with customers and which require improvement. Advances in Multilingual Sentiment Analysis: As e-commerce continues to expand globally, analyzing reviews in multiple languages has become essential. Many existing systems are limited to English, excluding significant feedback from international customers. By incorporating multilingual sentiment analysis using models like MarianMT, this system can translate and analyze reviews in various languages, ensuring inclusivity. This broadens the system's applicability, allowing businesses to gain a more comprehensive understanding of their global customer base and make data-driven decisions tailored to diverse markets.

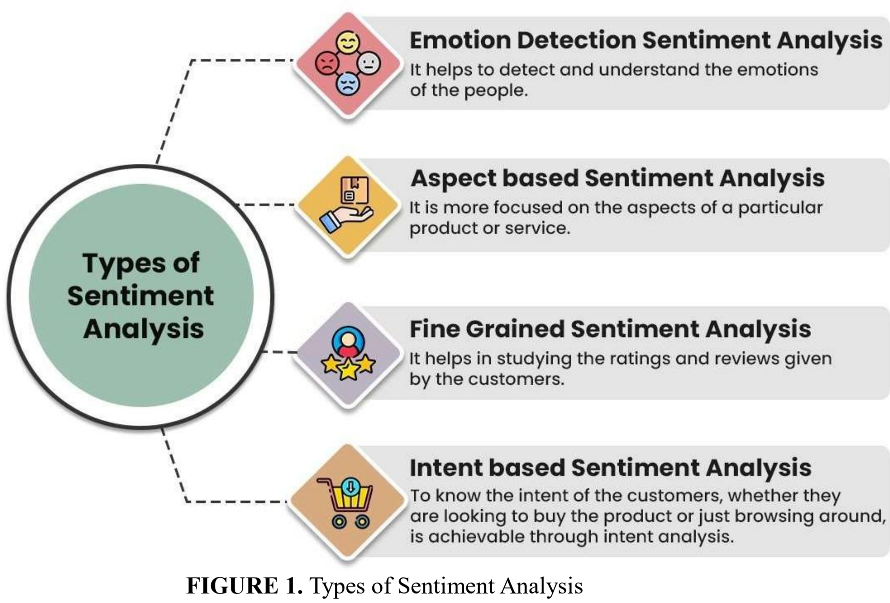


FIGURE 1. Types of Sentiment Analysis

3. LITERATURE REVIEW

Fang and Zhan (2015) - Sentiment Analysis Using Product Review Data: Their system focused on sentiment polarity categorization for Amazon product reviews, utilizing algorithms for sentiment score computation and negation phrase identification. However, it primarily concentrated on overall sentiment, without addressing specific product aspects. Li Yang et al. (2020) - Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning: They proposed the SLCABG model, which combined CNNs, GRUs, and sentiment lexicons. This system significantly improved sentiment feature extraction for Chinese reviews but was limited to one language. Mayur Wankhade et al. (2022) - A Survey on Sentiment Analysis Methods, Applications, and Challenges: Their work emphasized that existing systems mostly perform document or sentence-level sentiment analysis, which lacks the granularity needed for aspect-based sentiment analysis. Jurek et al. (2021) - Lexicon-Based Sentiment Analysis: They introduced a sentiment analysis method using sentiment lexicons but pointed out challenges with sarcasm and irony detection, which complicates accurate sentiment classification. Subhashini et al. (2021) - Comprehensive Review of Opinion Mining in Product Reviews: Their review identified that existing sentiment analysis systems often fail to handle fake reviews and opinion spam, leading to less reliable sentiment assessments.

4. METHODOLOGY

Web Scraping: The user gives ScraperAPI an API key and a URL of the product page on some e-commerce platform. Process: The system sends an HTTP request to ScraperAPI to get the HTML content of the product page specified in the ENVIRONMENT variables. Data Preprocessing: It cleans up the reviews for unecessary characters and other words, whitespace normalization, and emoji to text conversion using Demoji. The cleaned reviews are then tokenized, stop words are removed and the entries are lemmatized using NLTK to get a list of processed user reviews for analysis. Sentiment Classification: The system is using a pre-trained sentiment analysis model from the Hugging Face Transformers library. SentimentEach review is classified as Positive (encoded to 1) if sentiment > 3, else Negative (encoded to 0). Aspect Based Sentiment Analysis(ABSA): Product-Variation Identifies different aspect (e.G := Battery, camera, performance)‖ It also looks for these features in the reviews and categorizes its sentiment on each field of the product. Emotion Detection: Each review is then analyzed by a pre-trained emotion detection model. The emotions—joy, anger, sadness—and sentiments expressed in the reviews are identified and assigned with a category. Translation: The process uses the MarianMT model in the translation of reviews to English. It splits the long review into shorter pieces for efficient translation. Visualization: The system applies Matplotlib and Seaborn for visualizing the distribution of the emotions. A pie chart is generated based on the proportion of each emotion found in the review.

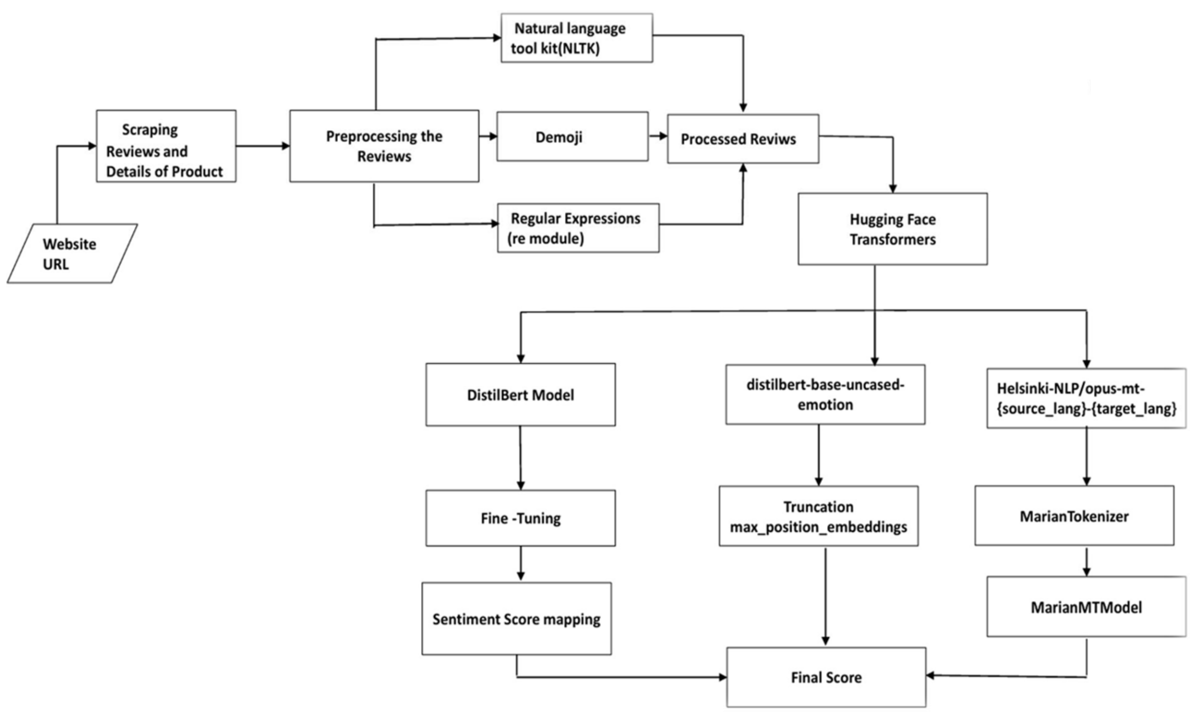


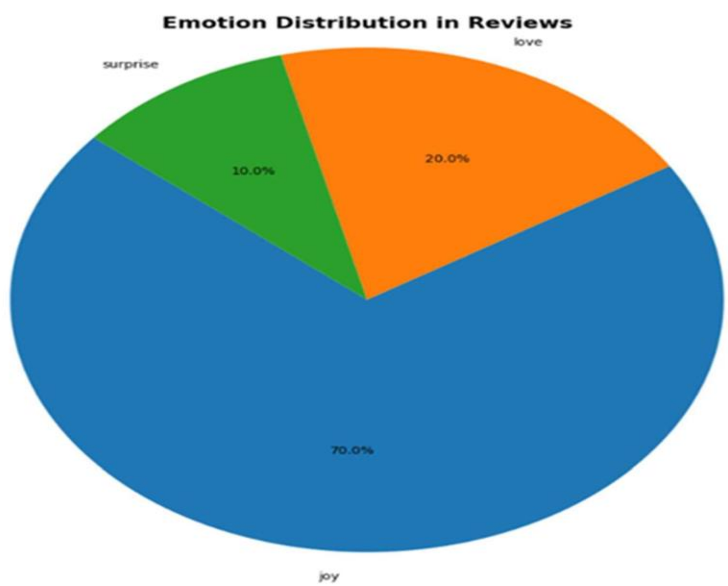
FIGURE 2. Flow of Work

Web scraping is the foundational step in the system, enabling the extraction of product reviews from e-commerce platforms. The user provides an API key from ScraperAPI and the URL of the desired product page. ScraperAPI facilitates efficient data retrieval by managing proxy connections and bypassing potential blocks by the target websites. Upon receiving the input, the system sends an HTTP request to ScraperAPI to fetch the HTML content of the product page. This process ensures the collection of raw review data, including text and metadata, which serves as the input for subsequent analysis. By specifying the ENVIRONMENT variables, the system dynamically adapts to different platforms, ensuring flexibility and scalability in data collection. The raw reviews obtained from web scraping often contain noise, such as special characters, emojis, and redundant spaces. The preprocessing module addresses these issues by cleaning the text. It removes unnecessary characters, normalizes whitespace, and converts emojis into text descriptions using the Demoji library. Tokenization then breaks down the text into smaller units like words, making it easier to process. Stop words— common but insignificant words—are removed to focus on meaningful content. Finally, lemmatization, powered by the NLTK library, reduces words to their base forms (e.g., "running" becomes "run"), standardizing the text for analysis. The output is a list of clean, structured reviews ready for sentiment classification and emotion detection. Sentiment classification determines the overall sentiment of each review. Using a pre-trained sentiment analysis model from the Hugging Face Transformers library, the system processes the pre-cleaned reviews. Each review is assigned a sentiment score, typically on a scale of 1 to 5. Reviews with scores greater than 3 are classified as positive (encoded as 1), while those with scores of 3 or lower are classified as negative (encoded as 0). This classification provides a binary understanding of the review's overall tone, enabling quick aggregation of positive and negative sentiments. Aspect-Based Sentiment Analysis (ABSA) delves deeper into reviews to evaluate specific product attributes.

The system identifies key aspects such as battery life, camera quality, and performance. It analyzes the sentiment associated with each aspect, categorizing them as positive, negative, or neutral. For instance, a review might praise the battery life but criticize the camera quality. ABSA ensures a granular understanding of customer feedback by focusing on specific product features rather than general sentiments, providing valuable insights into areas that require improvement. Emotion detection goes beyond basic sentiment classification by identifying specific emotions expressed in the reviews. Using a pre-trained emotion detection model, the system categorizes emotions such as joy, anger, and sadness. For example, a review stating, "I love this product!" is identified with joy, while "This product is frustrating to use" is associated with anger. This nuanced understanding of emotions helps businesses comprehend the emotional impact of their products on customers, enabling them to make data-driven improvements. To ensure inclusivity and global applicability, the system incorporates a translation module. Using the MarianMT model, it translates reviews written in various languages into English. For efficiency, long reviews are split into smaller chunks before translation, ensuring coherence and preserving context. This multilingual capability allows the system to analyze reviews from diverse linguistic backgrounds, making it a valuable tool for global e- commerce platforms. The insights derived from the analysis are made actionable through data visualization. The system uses Matplotlib and Seaborn to create visual representations of the emotion and sentiment distribution. For example, a pie chart illustrates the proportion of emotions (e.g., joy, anger, sadness) detected in the reviews. These visualizations provide an intuitive understanding of customer feedback, making it easier for stakeholders to interpret trends and take informed actions.

5. RESULTS

```
Product Name: Boult Drift BT Calling 1.69" HD Display, 140+ Watchfaces, 475Nits Brightness, I
Number of Reviews: 10
Sample Reviews: ['always forget drink water wearing watch bad habit gone reminds drink water
Sentiment Predictions (1-5 scale): [5, 3, 4, 4, 4, 5, 5, 4, 4, 5]
Aspect-Based Sentiment Analysis: {'battery': [5, 4], 'camera': [], 'display': [], 'performanc
Emotions Detected: [{'label': 'joy', 'score': 0.9884846806526184}, {'label': 'joy', 'score':
```




```
Translated Reviews: ['always forget drink water weathering watch bad habit goone reminds (
Final Rating: 4.3
```

6. FUTURE DIRECTION

Multimodal Analysis: Future research could integrate other data sources like images and videos related to products for a more comprehensive multimodal analysis. For instance, analyzing product photos or unboxing videos alongside textual reviews can reveal customer sentiments more holistically. Visual data can provide context that text alone might miss, such as product quality or packaging. This approach could also detect discrepancies between advertised and actual product features. Machine learning models, like image recognition and natural language processing, could be combined for this purpose. Such integration enhances the depth and accuracy of insights into customer perceptions.

Real-Time Review Analysis: Developing a real-time review analysis feature could help businesses stay updated on customer sentiment as it evolves. This system would involve continuous scraping and processing of new reviews, providing instant insights, particularly useful during product launches or promotional events. Businesses could track trends, quickly identify issues, and respond proactively. The implementation could use APIs or web scraping tools for data collection and machine learning models for analysis. Real-time feedback loops would improve decision-making and foster better customer engagement by addressing concerns swiftly. Sentiment Analysis for Different Domains: Expanding sentiment analysis to cover diverse domains such as hospitality, electronics, and clothing can enhance the versatility of the project. Each domain has specific terminology and sentiment expressions, necessitating domain-specific fine-tuning of models. For example, phrases indicating satisfaction in electronics may differ from those in clothing or hotels. Tailored training datasets and models for each domain could improve classification accuracy. This diversification also broadens the potential user base, making the system applicable to a variety of industries. Data Visualization: Incorporating data visualization tools, such as Matplotlib or Plotly, can make sentiment analysis results more user-friendly and actionable. Graphical representations like bar charts, pie charts, and trend lines can display review sentiments, emotional tones, and evolving trends over time. These visuals can help businesses easily identify patterns, such as rising dissatisfaction or frequent mentions of specific issues. Interactive dashboards could allow users to explore data dynamically, offering deeper insights. Visualization bridges the gap between complex analysis and intuitive understanding for decision-makers.

7. CONCLUSION

This project brings together different technologies to scrape, analyze, and visualize customer reviews from online platforms. It accurately identifies customer sentiment, emotions, and product-specific feedback, giving businesses a clear view of what people like or dislike about their products. By scoring reviews from 1 to 5 and focusing on key product features, companies can make improvements and better meet customer needs. The system also supports reviews in different languages, making it accessible to a wider audience, and presents insights in a way that’s easy to understand. Additionally, the inclusion of emotion detection provides deeper insight into how customers feel about specific aspects of products. These insights allow businesses to refine their marketing strategies and stay competitive in the marketplace. Overall, this system offers a well-rounded solution for enhancing customer satisfaction, product quality, and business decisions.

REFERENCES

[1]. M. U. Rana, M. S. Sajjad, and A. H. Khan, “A Deep Convolutional Neural Network Based Approach for Handwritten Digit Recognition,” 2019 IEEE 16th International Symposium on Autonomous Systems (ISAS), Tokyo, Japan, 2019, pp. 120-125, doi: 10.1109/ISAS.2019.00029.

[2]. Reddy, B.S., Mallikarjuna Reddy, A., Sradda, M.H.D.S., ... Mounika, S., Meghana, K. A Comparative Study on Object Detection Using Retinanet, MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference, 2022

[3]. Navyasree, V., Surarchitha, Y., Reddy, A.M., ... Anuhya, A., Jabeen, H. Predicting the Risk Factor of Kidney Disease using Meta Classifiers MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference, 2022

[4]. Rao, B.H., Reddy, A.M., Ramya, V., Khanam, F., Sirisha, K.S.MTESSERACT: An Application for Form Recognition in Courier Services, 3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings, 2022, pp.848–853

[5]. Silpa, P.S., Reddy, A.M., Durga, C.B.V., Priya, C.H.H., Mounika, J.Designing of Augmented Breast Cancer Data using Enhanced Firefly Algorithm, 3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings, 2022, pp.759–767

[6]. Reddy, A.M., Reddy, K.S., Jayaram, M., ... Kumar, V.V., Stalin Alex, D. An Eficient Multilevel Thresholding Scheme for Heart Image Segmentation Using a Hybrid Generalized Adversarial Network, Journal of Sensors, 2022, 2022, 4093658

[7]. Manoranjan Dash, N.D. Londhe, S. Ghosh, et al., “Hybrid Seeker Optimization Algorithm-based Accurate Image Clustering for Automatic Psoriasis Lesion Detection”, Artificial Intelligence for Healthcare (Taylor & Francis), 2022, ISBN: 9781003241409

[8]. Manoranjan Dash, Design of Finite Impulse Response Filters Using Evolutionary Techniques - An Efficient Computation, ICTACT Journal on Communication Technology, March 2020, Volume: 11, Issue: 01

[9]. Manoranjan Dash, “Modified VGG-16 model for COVID-19 chest X-ray images: optimal binary severity assessment,” International Journal of Data Mining and Bioinformatics, vol. 1, no. 1, Jan. 2025, doi: 10.1504/ijdmb.2025.10065665.

[10]. Manoranjan Dash et al.,” Effective Automated Medical Image Segmentation Using Hybrid Computational Intelligence Technique”, Blockchain and IoT Based Smart Healthcare Systems, Bentham Science Publishers, Pp. 174-182,2024

[11]. Manoranjan Dash et al.," Detection of Psychological Stability Status Using Machine Learning Algorithms", International Conference on Intelligent Systems and Machine Learning, Springer Nature Switzerland, Pp.44-51, 2022.

[12]. Samriya, J. K., Chakraborty, C., Sharma, A., Kumar, M., & Ramakuri, S. K. (2023). Adversarial ML-based secured cloud architecture for consumer Internet of Things of smart healthcare. IEEE Transactions on Consumer Electronics, 70(1), 2058-2065.

[13]. Ramakuri, S. K., Prasad, M., Sathiyanarayanan, M., Harika, K., Rohit, K., & Jaina, G. (2025). 6 Smart Paralysis. Smart Devices for Medical 4.0 Technologies, 112.

[14]. Kumar, R.S., Nalamachu, A., Burhan, S.W., Reddy, V.S. (2024). A Considerative Analysis of the Current Classification and Application Trends of Brain–Computer Interface. In: Kumar Jain, P., Nath Singh, Y., Gollapalli, R.P., Singh, S.P. (eds) Advances in Signal Processing and Communication Engineering. ICASPACE 2023. Lecture Notes in Electrical Engineering, vol 1157. Springer, Singapore. https://doi.org/10.1007/978-981-97-0562-7_46.

[15]. R. S. Kumar, K. K. Srinivas, A. Peddi and P. A. H. Vardhini, "Artificial Intelligence based Human Attention Detection through Brain Computer Interface for Health Care Monitoring," 2021 IEEE International Conference on Biomedical Engineering, Computer and Information Technology for Health (BECITHCON), Dhaka, Bangladesh, 2021, pp. 42-45, doi: 10.1109/BECITHCON54710.2021.9893646.

[16]. Vytla, V., Ramakuri, S. K., Peddi, A., Srinivas, K. K., & Ragav, N. N. (2021, February). Mathematical models for predicting COVID-19 pandemic: a review. In Journal of Physics: Conference Series (Vol. 1797, No. 1, p. 012009). IOP Publishing.

[17]. S. K. Ramakuri, C. Chakraborty, S. Ghosh and B. Gupta, "Performance analysis of eye-state charecterization through single electrode EEG device for medical application," 2017 Global Wireless Summit (GWS), Cape Town, South Africa, 2017, pp. 1-6, doi:10.1109/GWS.2017.8300494.

[18]. Gogu S, Sathe S (2022) autofpr: an efficient automatic approach for facial paralysis recognition using facial features. Int J Artif Intell Tools. https://doi.org/10.1142/S0218213023400055

[19]. Rao, N.K., and G. S. Reddy. "Discovery of Preliminary Centroids Using Improved K-Means Clustering Algorithm", International Journal of Computer Science and Information Technologies, Vol. 3 (3), 2012, 4558-4561.

[20]. Gogu, S. R., & Sathe, S. R. (2024). Ensemble stacking for grading facial paralysis through statistical analysis of facial features. Traitement du Signal, 41(2), 225–240.