10.48047/jocaaa.2025.34.08.10

Multimodal Deep Learning Models for Unstructured Data Integration in Enterprise Analytics

Suresh Sankara Palli
Independent Researcher, USA.

Abstract

The necessity for sophisticated processing systems that can effectively extract valuable insights has increased due to the growth of multi-modal unstructured data. Despite the potential, current studies show significant limitations, such as poor multi-modal data combining, insufficient big data volume or variation management, and a lack of thorough taxonomies for classifying AI-based ISs. In order to provide a more thorough understanding and forecast of consumer behaviour, this study proposes a multimodal framework for learning that combines multiple information sources, including user location, behaviour data, and product attributes. It does this by integrating multimodal data analysis or big data technology. Big data analytics alone are not enough to handle a threat landscape that is expanding quickly and getting more complex every day. Unstructured big data can take any form, including text, audio, photos, and video, and has no set organisation or structure. The issues of emotions and sentiment modelling resulting from unstructured large data with many modalities are discussed in this work. First, we provide a current overview of emotion and sentiment modelling, encompassing the most advanced methods. Next, we provide a novel architecture for large data sentiment and multimodal emotion modelling. Data collection, multimodal data aggregation, multimodal data feature extraction, fusion and decision, and application are the five key components of the suggested architecture. The multimodal data feature extraction module in the design is suggested to use two new feature extraction methods: divide-and-conquer linear discriminant analysis (Div-ConLDA) and divide-and-conquer principal component analysis (Div-ConPCA). To verify the effectiveness of the suggested methods, tests are conducted on a multicore computer architecture.

Keywords: - Multi-Modal, Unstructured Data, AI-Based, Component Analysis, Divide-And-Conquer, Sentiment Modeling, Data Aggregation Module, Novel Feature, Proposed Techniques.

I. INTRODUCTION

In today's cutthroat business world, emotion and sentiment modelling techniques are crucial tools for companies to gauge how customers feel about their goods and services as well as how they stack up against their rivals [1]. Organisations with excellent business analytics capabilities saw notable performance increase, according to the authors' study [1, 2]. Big data analytics may be used to obtain valuable insights about the sentiment of customers and products.

10.48047/jocaaa.2025.34.08.10

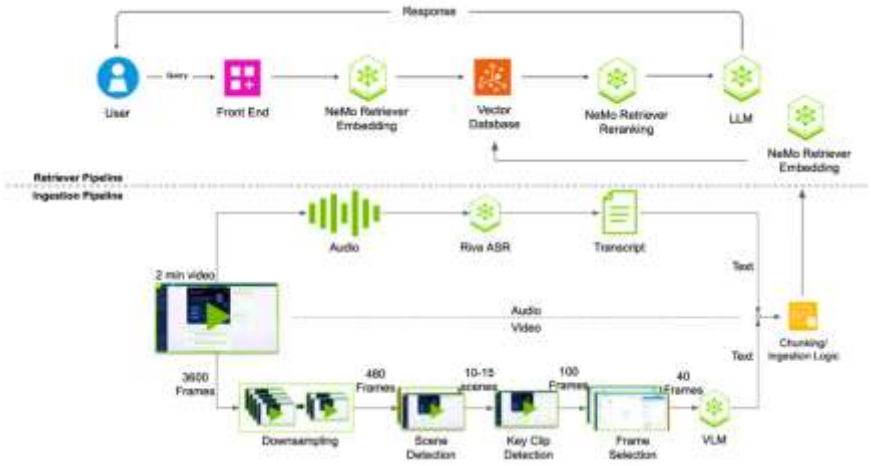

Fig. 1 A blog page that displays many modalities (written, audio, and visual). [8]

Although some recent advances have been made, the collection and analysis of emotion & sentiment knowledge gathered from multimodal data, such as audio and video, is still relatively unexplored by academics compared to the well-developed field of text-based mining. The main challenge is that the word associations for text-based data do not include emotion and sentiment information [2, 3]. As a result, sentiment and emotion modelling must be categorised straight from the voice and video data. Audio characteristics such as frequency (pitch), length, intensity, and [3, 5] must be used to directly classify speech, and facial expressions must be used to classify videos.

There is an urgent demand for advanced processing methods as a result of the growth of unstructured data in a variety of industries, including social media, health care, and finance. An important area of research is the efforts to handle and analyse such data, which can take many different forms, including text, photos, and videos [4, 5]. Finding relevant information and insights from unstructured data is extremely difficult due to its inherent complexity and variability. This study presents UDNet, a novel framework based on deep reinforcement learning, as a solution to these issues. The careful design of UDNet to parse multiple ways of learning unstructured data [4, 6] makes it easier to choose algorithms that are appropriate for a given task. Modularity and extensibility define its architecture, guaranteeing [6, 7].

The foundation of UDNet is in the collective knowledge gained from several ground-breaking investigations, each of which made a distinct contribution to the comprehension and development of unstructured information processing and analysis [8, 9]. The thorough evaluation ushered in a new age of systematic research by shedding light on the complex field of unstructured data analysis [8, 9]. The team's further efforts produced creative solutions, such

10.48047/jocaaa.2025.34.08.10

as an unsupervised noise detection technique, opening the door for automated data processing by detecting and reducing abnormalities [8, 9]. A wide range of data kinds and formats were easily accommodated by their construction of a flexible parsing pipeline, which greatly increased processing flexibility.

Additionally, an innovative adaptive structure to support deep reinforced semantic parsing was presented, which dynamically improves parsing techniques to successfully handle the complexities that data brings complexity [9, 10]. The potential for tailored data analysis in the banking industry was highlighted by additional research into deep reinforcement learning for obtaining individualised insights from financial transactions.

The technology, people, and procedures that collaborate to gather, evaluate, store, and distribute data in order to manage institutions and make better choices are known as information systems (IS). To facilitate the flow of information within and among enterprises, these models make use of a variety of technologies, including databases, networks, computer hardware, and software [9, 10]. By providing users with relevant and timely information, ISs primarily aim to support industry operations, increase efficacy, and advance strategic goals [11]. The solutions offered by ISs include Business Intelligence (BI) models for data analysis and decision support, Customer Relationship Management (CRM) models for customer interactions, and Enterprise Resource Planning (ERP) systems for combining industrial operations. By using smart technology and information management, ISs are crucial in today's businesses and industries inside institutions, encouraging innovation, efficiency, and competition [12]. Artificial Intelligence (AI) technology is used by AI-based ISs to process data, automate procedures, and make defensible judgements without requiring direct human involvement.

These models are employed in a variety of industries, including manufacturing, healthcare, finance, and retailing [13]. Computer vision, Machine Learning (ML), processing of natural languages, and other AI techniques are used by AI-based ISs to extract information from vast volumes of data, forecast results, improve procedures, and personalise user experiences [14]. Self-driving models for decision-making, recommendation engines for tailored content, chatbots for customer support, and predictive analytics are all examples of AI-based ISs [15]. In the digital age, these technologies give businesses the ability to use AI to enhance efficacy, creativity, and decision-making, giving them a competitive advantage and industry success [16].

10.48047/jocaaa.2025.34.08.10

The process of analysing large and intricate datasets to find hidden patterns, connections, and discoveries that might support strategic business decisions and increase the efficacy of the organisation is known as big data analytics [16]. This subject involves controlling vast volumes of organised and unstructured data from several sources by utilising contemporary analytical tools like machine learning, statistical analysis, and data mining.

Big data analytics has numerous and important uses in a variety of industries, including manufacturing (for supply chain optimisation and predictive maintenance), healthcare (for patient outcomes and predictive analytics), retail (for customer fragmentation and personalised marketing), and finance (for risk management and fraud detection) [17, 18]. Big data analytics may help companies make sense of their data, improve operational efficiency, create new products and services, and gain a competitive edge in today's data-driven business environment [19].

Text, sensor data, images, videos, and signals are examples of heterogeneous sources or modalities [20], which are combined and synthesised in a process known as multi-modal data fusion [21] to provide comprehensive and relevant insights. In addition to improving the precision, dependability, and depth of information processing and interpretation, this integration of various data types allows for a more comprehensive understanding of intricate processes [22].

Improved algorithms and techniques from fields such as machine learning, signal processing, and computer vision are used in multi-modal data fusion strategies to combine data representations in a synergistic way, extract additional data, and address the inherent challenges of handling a range of data formats and features [21]. Multi-modal data fusion applies the collective intelligence found in several data sources to improve knowledge discovery, problem-solving, and decision-making.

The Internet will be the source of the unstructured big data, including blogs, forums, consumer reviews, news, Twitter, videos, and more [21]. Three data spouts—video, audio, and textual—will be created by disassembling these sources; each spout will only include one modality type (for example, a video spout will only hold video data) [22]. In addition to this new architecture, two innovative divide and rule feature extraction strategies for big data analytics are presented in this work.

10.48047/jocaaa.2025.34.08.10

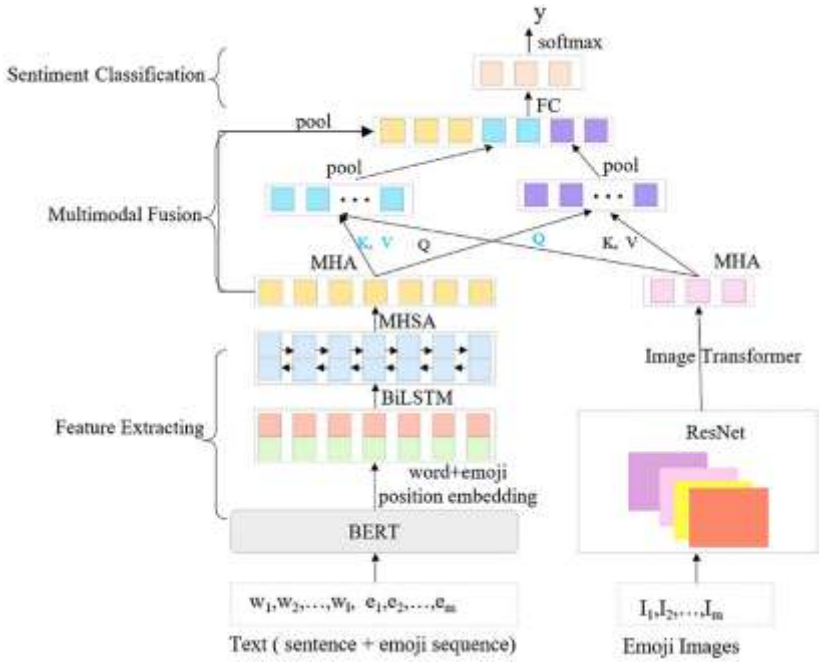

Fig. 2 An overview of sentiment and multimodal emotion modelling using unstructured large data. [21]

Similar to this, a Big DL-distributed Data-Deep learning framework for Apache is introduced in a work that is used by a number of customers of companies that deploy Deep learning applications on production Big Data platforms [1]. Additionally, this paradigm enables deep learning applications to run on either the Apache-Hadoop or Spark clusters. This would make it possible to handle production-level Big Data, create a deployment pipeline, and manage data analysis.

The distribution model of execution, training scalability, real-time use cases, and computing process performance are all included in this study's review of the Big DL distributed data deep learning framework for Apache. Deep learning algorithms are used to create these kinds of effective Big Data apps. Additional uses of the research include helping to construct security-primitives and demonstrating the deep-learning foundation architecture of the feature extraction procedure [8, 9]. According to this, auto-encoders should be used to transfer conventional-state variables to lower dimensions. The use of deep learning in the development of security primitives is being presented in another work [9, 10].

The decision-making process based on several criteria is also essential for overcoming the challenges posed by big data analytics [11]. These procedures would choose to use new machine-learning techniques, such the decision-making process that produces Big-data insights, to find the answer. The article's shift from analytics to AI (artificial intelligence) is

10.48047/jocaaa.2025.34.08.10

another aspect of its focus. The article [12] provides a quick overview of the several methods for assessing analytical skills, the development of business strategies, and the company's goal in the AI stream. The research demonstrates how the AI-stream affects the organisation, the current capabilities of the business, and the need for an effective plan.

Manufacturing companies have disorganised data-processing problems because as industries and businesses grow, the more data they handle, the more computer power and efficient network speed they can achieve [13]. However, IoT-Internet of Things research has adapted data manipulation of bigger datasets by using Deep Learning techniques in Big Data analytics. A few of the current methods that have been added to Deep Learning patterns and Deep Learning applications across several fields were demonstrated. The deep learning approach has improved the capacity forecasts of computer devices one step at a time. The existence of big data and the assistance of superior learning model algorithms enable this [13]. Thus, the superior, effective performance of deep-learning processes and the trustworthy performance analysis have captured the attention of research studies in every discipline in order to rectify remedies to the aforementioned difficulties [14].

Building models that allow computers to learn from several modalities and promote information sharing among them is known as multimodal machine learning, or MMML. Applications of multimodal learning have proven successful in a variety of domains, including gaming, smart hardware, healthcare, and education, and they have a lot of room to grow [14].

Multimodal technology in the educational field can provide students with more engaging experiences and richer learning materials [15]. Medical imaging data may be intelligently analysed and interpreted through the use of multimodal approaches, which combine image recognition, audio recognition, and natural language processing. This helps physicians make precise diagnoses. Multimodal features in smart hardware improve the perception and interaction capabilities of devices; by combining picture and speech recognition technologies, these devices may provide a wider range of functionalities and comprehend human commands more precisely [16].

The game business produces more immersive virtual reality gaming experiences by combining picture, voice, and gesture detection technology. Furthermore, multimodal features allow for more complex character emotions and action exchanges, which improves game pleasure and involvement [18]. From the standpoint of modality fusion, this encompasses both feature and data fusion [19]. Feature fusion techniques train a model for every modality, make choices, and

10.48047/jocaaa.2025.34.08.10

then combine these choices into a final, all-inclusive conclusion via an attention mechanism. Using the fused data for model training, data fusion directly connects feature data from several modalities.

Enterprise digital transformation depends on data empowerment [12]. Data empowerment has been greatly improved by developments in analytical methods and data technology. By connecting the supply chain's ends, intelligent sales allow for customised manufacturing at the back end and differentiated demand mine at the front end. The incorporation of artificial intelligence provides novel management models for the fashion manufacturing system, which is advantageous for the clothing supply chain, which is characterised by quick changes, short cycles, and adaptability [22]. Supply chain intelligence may be greatly improved by utilising machine learning algorithms and corporate big data technologies [23]. Supply chain transformation is fuelled by artificial intelligence in domains including advantage rebuilding, ecosystem reshaping, and platform reconstruction.

As cross-border e-commerce grows quickly, sales organisations want to forecast sales performance precisely and create items that make sense in order to increase profitability, draw in more capital, and improve customer satisfaction [23]. Online shoppers frequently utilise text, photos, and videos to convey their wants to customer support, which generates a substantial volume of unstructured data.

Text and visual data are examples of multimodal data that are essential to e-commerce customer support. However, the intricacy of consumer behaviour is frequently not completely reflected by conventional unimodal approaches, which are limited to capturing data from a single dimension [22]. Multimodality is a major problem for e-commerce customer support systems because of this constraint. Text and picture modalities are used in a multimodal late fusion technique for classifying e-commerce products. Their study showed that in multimodal product categorisation tasks, the suggested approach performed better than conventional unimodal approaches [21].

In order to estimate the demand for e-commerce products, a neural network model was constructed using a spatial feature fusion and grouping technique based on multimodal data. The efficacy and superiority of the suggested algorithm were validated by the experimental findings. A methodology for multimodal analysis to forecast product return rates in e-commerce live streaming [20]. Research utilising actual data from Taobao Live showed that multimodal signals from anchors and items may accurately forecast return rates [23].

10.48047/jocaaa.2025.34.08.10

Expanding on current knowledge, understanding the current state-of-the-art, and identifying opportunities for additional research all depend on the analysis of prior pertinent studies. It allows researchers to better advise and guide their present investigations by drawing on insights and lessons learnt from earlier studies [22]. To identify potential approaches and modify successful tactics for creating cutting-edge AI-driven systems that effectively harness a range of data sources, it is essential to acknowledge earlier relevant studies in AI-enabled ISs that leverage integrated big data analytics and multi-modal data fusion.

This gave a thorough rundown of recent developments in multimodal representation of information and fusion techniques. Through the integration of many data kinds to improve accuracy and resilience, it demonstrated the importance of multimodal data in updating ISs. Their study looked at techniques and methods for overcoming data heterogeneity, with applications in autonomous systems, healthcare, and communication technologies (PES University) [24]. However, in order to increase the quality and evolution of information, it looked at ways to improve the synthesis of data from different sources [21]. Their paper addressed the benefits and limitations of each approach while examining a wide range of data representations and fusion techniques. It provided a comprehensive taxonomy of the fundamental challenges and solutions in the field and emphasised the need of model-agnostic and model-enabled approaches in multimodal signal processing.

II. EMOTION & SENTIMENT MODELLING: A REVIEW

A review and synopsis of sentiment and emotion modelling are provided in this section. Both single and many modalities of study in this topic are covered in this review [23]. Some recent and previous representative research efforts on sentiment modelling and single-modality emotion modelling are included in Tables 1 and 2, respectively.

When it comes to research on single modalities (text, picture, audio, video, and physiological inputs), Table 1 displays the representative emotion and sentiment modelling works in chronological order, whereas Table 2 displays the representative works for research on multiple modalities [21].

Table 1 Representative research works for emotion and sentiment modeling using five single modalities (text, image, audio, video, physiological). [23]

Represented Research Work	Year	Database	Classification technique	No. of sentiments or emotions	Outcome

10.48047/jocaaa.2025.34.08.10

Taboada et al. [8]	2011	Amazone	SVM	2	Accuracy 86.96%
Pang et al. [4]	2008	Movies reviews	Dictionary	2	--
K Lucykx et al. [6]	2012	2011 medical NLP challenge	SVM	2	Accuracy 86.94%
Richard et al. [5]	2013	Stanford sentiment Tree Bank	Deep Learning	2	Accuracy 80.96%
Kherwa et al. [1]	2014	Senti WordNet	Multimap structure	2	--

Table 2 Five distinct modalities are used in representative research projects for emotion and sentiment modelling (text, picture, audio, video, physiological). [22]

Representative research work	**Year**	**Database**	**Fusion technique**	**Classification technique**	**Number of sentiments or emotions**	**Outcome**
Kanluan et al. [12]	2008	Vam corpus recorded from the German TV talk show veram Mittag	Model	Support vector regression for 3 Continous dimension's	3	The correlation between the prediction and ground truth increased by 12.3% and 9.6%, respectively, while the average error of the fused result was 17.9% and 13.5% lower than the individual errors of the acoustic and visual modalities.
Nicolaou et al. [11]	2010	Sensitive artificial listener database	Model	HMM and like hood Space via SVM	2	The best fusion results from mixing

10.48047/jocaaa.2025.34.08.10

						shoulder, auditory, and face emotions are 94%, whereas the best mono-cue results from facial expressions over ten-fold cross-validations are 91.69%.
Poria et al. [13]	2015	International survey of emotional antecedents & reactions (ISEARI) Database & CK++& eNTERFACE datasets	Feature	K-nearest neighbour (KNN), Artificail Neural Network (ANN), Extreme Learning Machine (ELM) & SVM	7	It achieved a relative error rate reduction of 56%, or almost 10%, over the best state-of-the-art system, with an overall accuracy of 89.69%.

III. ARCHITECTURE

The specifics of the suggested multimodal emotion & sentiment modelling architecture for unstructured large data are shown in this section. The five key elements of this design are shown in Fig. 3. Data collection module [14], multimodal data aggregation module, multimodal data extraction of features module, fusion and decision module, and application module [25] are the modules in question. The following describes these modules and their features. Both online (real-time) and offline data types are taken into account by this design.[26]

10.48047/jocaaa.2025.34.08.10

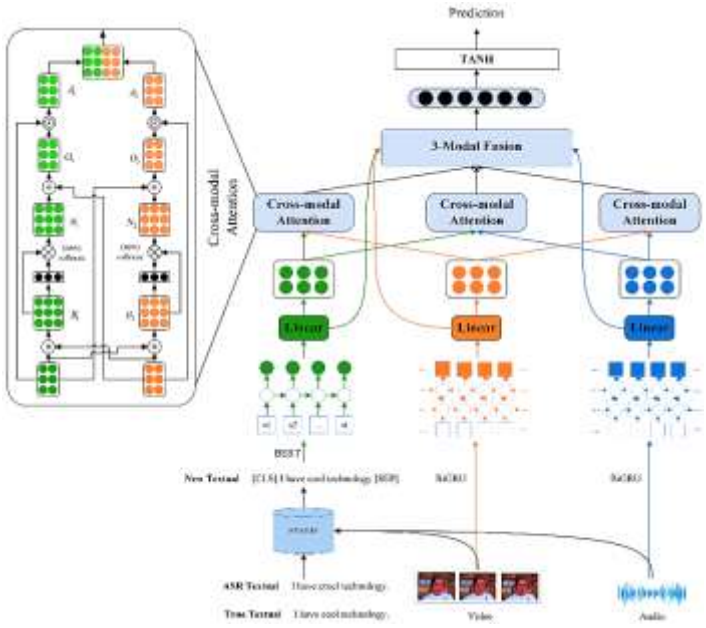

Fig. 3 The multimodal big data architecture that has been suggested for sentiment and emotion modelling. [16]

The dimensionality reduced dataset Y is first split into two subsets, Y1 and Y2, each of which contains half of the dimensionality reduced samples in Y. The two Divide LDA modules then analyse these subsets in parallel as part of the Divide and Conquer LDA [27]. The Conquer LDA module fuses the outputs from the two modules [28]. The description of the Div-ConLDA pseudocode is provided in Algorithm 2. The Div-ConLDA decomposition modules are displayed in Fig. 4. The Conquer Class stage of Div-ConLDA employs the class matrix information, which is one way that it differs from Div-ConPCA, as seen in Fig. 4 [11].

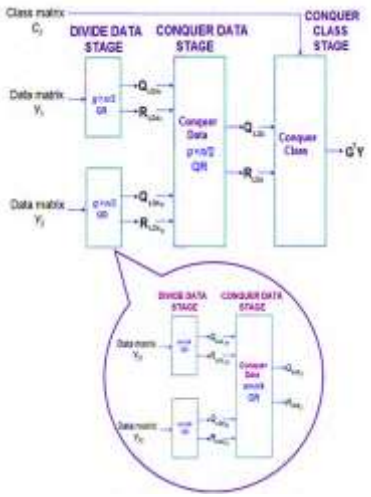

Fig. 4 The LDA known as Divide & Conquer (Div-Con LDA). [19]

10.48047/jocaaa.2025.34.08.10

IV. EXPERIMENTAL RESULTS

Using three real-world datasets commonly used for research in emotion and sentiment modelling and recognition, this section compares the classification performance of the cascade of Div-ConPCA or DivConLDA with other conventional techniques that do not employ the divide and conquer strategy [19, 20]. These include a large video dataset gathered from YouTube [20], the Japanese Female Facial Expression (JAFFE) Database, which is a visual emotion detection database, and the eNTERFACE'05, which is an audio-visual emotion identification database.

The cascade DivConPCA-LDA was compared to the conventional PCA and PCA + LDA (Fisherface) approaches in terms of performance [23]. Figure 5 compares the recognition rates that were found for the different methods. [29].

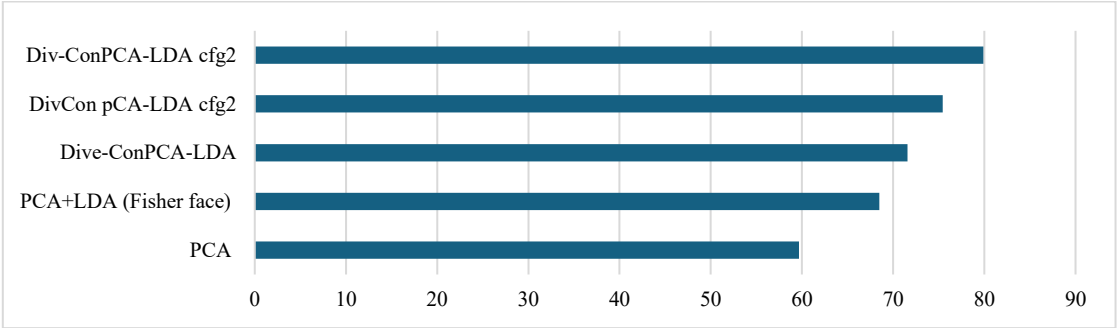

Fig. 5 Rates of JAFFE database recognition. [19]

The comparison of recognition rates acquired for the different methods is displayed in Fig. 6. The visual modality's identification rate for the eNTERFACE database [18] was 66% for the Div-ConPCA-LDA, 42% for PCA, and 65% for PCA + LDA (the Fisher face approach). An enhanced identification rate of 74% was obtained by the Div-ConPCA-LDA by combining the visual and aural senses using a fusion approach. [31].

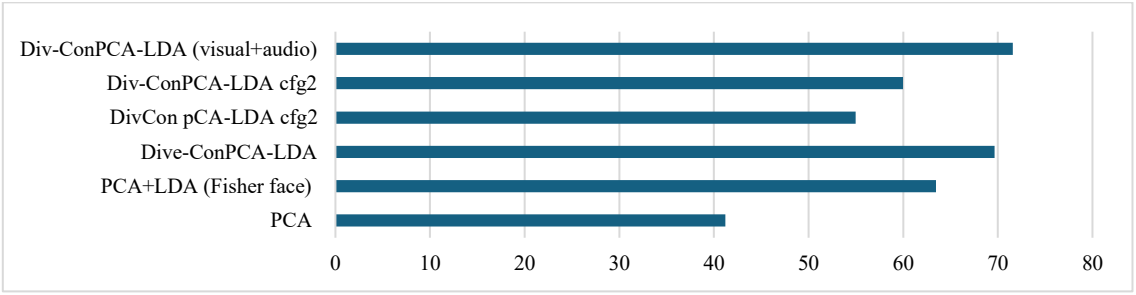

Fig. 6 Rates of recognition for the eNTERFACE'05 database. [28]

10.48047/jocaaa.2025.34.08.10

The Div-ConPCA-LDA approaches were used in a final experiment [23] that used simulations on a multicore machine architecture and a large dataset for sentiment modelling that was gathered from the internet. The dataset included over 50,000 frames of real-world YouTube videos. The dataset includes user films that were manually tagged using three emotion classifications (positive, neutral, or negative) after being gathered from the social media platform YouTube [30]. A selection of the YouTube sentiment modelling video dataset is displayed in Fig. 7 [14].

Fig. 7 Sentiment modelling examples from the YouTube dataset. [22]

The resultant performance comparisons (execution time in seconds) on multiple cores architecture implementations are presented in Table 3, along with the recognition rates that were achieved for four distinct classifiers (SVM, k-NN, classification tree, and boost tree) [19]. In comparison to the other classifiers for sentiment modelling, the SVM provided a recognition rate of 76.9% [32].

Table 3 Performance comparisons on multicore architecture implementations and recognition rates for large datasets of YouTube videos. [17]

Classifier	Recognition rate (%)	Execution time (1-core)	Execution time (4-cores)
SVM	79.89%	796s	286s
K-NN	75.09%		
Classification tree	78.96%		
Boosted tree	59.98%		

V. CONCLUSION

Business analytics-focused emotion and sentiment modelling for customer retention and product marketing improvements are crucial instruments for increasing company efficiency and cutting expenses. The difficulties of emotion and sentiment modelling for unstructured big data have been discussed in this paper. A review of emotion and sentiment modelling with single and multiple modalities was followed by a new architecture of emotion and sentiment modelling to process and analyse multimodal unstructured big data. The architecture can

10.48047/jocaaa.2025.34.08.10

recognise various data sources and organise the data blocks according to their modalities for a computing approach known as "divide and conquer.

In order to demonstrate the use of the divide and conquer method for feature extraction, the study also suggested the Div-ConPCA-LDA. These methods work well for unstructured big data analytics and are computationally efficient. The eigenvector decompositions employing SVD for conventional implementations can be substituted with the less expensive QR decompositions, as demonstrated by the Div-ConPCA-LDA cascade. Additionally, the method enables the QR decompositions to be recursively split into smaller sub-matrices for parallel processing, which are subsequently recombined to provide the precise reconstruction that is equivalent to the complete eigenvector decompositions.

As the datasets become accessible for use on high-speed computing platforms and GPU clusters, future research will apply the methodology to ever-larger emotion and sentiment datasets.

VI. REFERENCES

[1] P. Kherwa, A. Sachdeva, D. Mahajan, N. Pande, and P. K. Singh, ‘‘An approach towards comprehensive sentimental data analysis and opinion mining,’’ in Proc. IEEE Int. Adv. Comput. Conf. (IACC), Feb. 2014, pp. 606–612.

[2] H. Ha, W. Hwang, S. Bae, H. Choi, H. Han, G. N. Kim, and K. Lee, ‘‘CosMovis: Semantic network visualization by using sentiment words of movie review data,’’ in Proc. 19th Int. Conf. Inf. Vis., vol. 19, Jul. 2015, pp. 436–443.

[3] E. Guzman and W. Maalej, ‘‘How do users like this feature? A fine-grained sentiment analysis of app reviews,’’ in Proc. IEEE 22nd Int. Requirements Eng. Conf., Aug. 2014, pp. 153–162.

[4] B. Pang and L. Lee, ‘‘Opinion mining and sentiment analysis,’’ Found. Trends Inf. Retr., vol. 2, pp. 1–135, Jul. 2008.

[5] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, ‘‘Recursive deep models for semantic compositionality over a sentiment treebank,’’ in Proc. Conf. Empirical Methods Natural Lang. Process., 2013, pp. 1631–1642.

[6] K. Luyckx, F. Vaassen, C. Peersman, and W. Daelemans, ‘‘Fine-grained emotion detection in suicide notes: A thresholding approach to multi-label classification,’’ Biomed. Inform. Insights, vol. 5, no. 1, pp. 61–69, 2012.

[7] M. Hasan, E. Rundensteiner, and E. Agu, ‘‘EMOTEX: Detecting emotions in twitter messages,’’ in Proc. ASE Bigdata/Socialcom/Cybersecu. Conf., 2014, pp. 27–31.

[8] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, ‘‘Lexiconbased methods for sentiment analysis,’’ Comput. Linguistics, vol. 37, no. 2, pp. 267–307, 2011. Doi

[9] Y. Sawakoshi, M. Okada, and K. Hashimoto, ‘‘An investigation of effectiveness of ‘opinion’ and ‘fact’ sentences for sentiment analysis of customer reviews,’’ in Proc. Int. Conf. Comput. Appl. Technol. (CCATS), 2015, pp. 98–102.

[10] Z. Hu, J. Hu, W. Ding, and X. Zheng, ‘‘Review sentiment analysis based on deep learning,’’ in Proc. IEEE 12th Int. Conf. E-Bus. Eng., Oct. 2015, pp. 87–94.

10.48047/jocaaa.2025.34.08.10

[11] A. Nicolaou, H. Gunes, and M. Pantic, ‘‘Audio-visual classification and fusion of spontaneous affective data in likelihood space,’’ in Proc. 20th Int. Conf. Pattern Recognit., Aug. 2010, pp. 3695–3699.

[12] Kanluan, M. Grimm, and K. Kroschel, ‘‘Audio-visual emotion recognition using an emotion space concept,’’ in Proc. 16th Eur. Signal Process. Conf., Aug. 2008, pp. 1–5.

[13] S. Poria, E. Cambria, and A. Gelbukh, ‘‘Deep convolutional neural network textual features and multiple kernels learning for utterance-level multimodal sentiment analysis,’’ in Proc. Conf. Empirical Methods Natural Lang. Process., Oct. 2015, pp. 2539–2544.

[14] S. Poria, I. Chaturvedi, E. Cambria, and A. Hussain, ‘‘Convolutional MKL based multimodal emotion recognition and sentiment analysis,’’ in Proc. IEEE 16th Int. Conf. Data Mining (ICDM), Dec. 2016, pp. 439–448.

[15] Liang, P.P.; Liu, Z.; Zadeh, A.; Morency, L.-P. Multimodal language analysis with recurrent multistage fusion. arXiv 2018, arXiv:1808.03920.

[16] Gavric, A. (2023, November). Enhancing process understanding through multimodal data analysis and extended reality. In PoEM Companion.

[17] Oviatt, S. (2022). Multimodal interaction, interfaces, and analytics. In Handbook of Human Computer Interaction (pp. 1-29). Cham: Springer International Publishing.

[18] Zhun Liu, Ying Shen, Varun Bharadhwaj Lakshminarasimhan, Paul Pu Liang, AmirAli Bagher Zadeh, and Louis-Philippe Morency. 2018. Efficient lowrank multimodal fusion with modality-specific factors. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2247–2256. Association for Computational Linguistics.

[19] Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, and Hal Daume III. 2015. Deep unordered compo- ´ sition rivals syntactic methods for text classification. In Association for Computational Linguistics.

[20] Philippe, S.; Souchet, A.D.; Lameras, P.; Petridis, P.; Caporal, J.; Coldeboeuf, G.; Duzan, H. Multimodal teaching, learning and training in virtual reality: A review and case study. Virtual Real. Intell. Hardw. 2020, 2, 421–442.

[21] Shyam Sundar Rajagopalan, Louis-Philippe Morency, Tadas Baltrusaitis, and Goecke Roland. 2016. Ex- ˇ tending long short-term memory for multi-view structured learning. In European Conference on Computer Vision.

[22] Fang, X. Tao, Z. Tang, R. Qiu, and Y. Liu, ‘‘Dataset, groundtruth and performance metrics for table detection evaluation,’’ in Proc. 10th IAPR Int. Workshop Document Anal. Syst., Mar. 2012, pp. 445–449.

[23] D. N. Tran, T. A. Tran, A. Oh, S. H. Kim, and I. S. Na, ‘‘Table detection from document image using vertical arrangement of text blocks,’’ Int. J. Contents, vol. 11, no. 4, pp. 77–85, Dec. 2015.

[24] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, ‘‘Multimodal machine learning: A survey and taxonomy,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, Feb. 2019. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, ‘‘Attention is all you need,’’ in Proc. Adv. Neural Inf. Process. Syst., vol. 30, 2017, pp. 5998–6008.

[25] Syed, S. Big Data Analytics In Heavy Vehicle Manufacturing: Advancing Planet 2050 Goals for A SustainableAutomotive Industry.

[26] Dilip Kumar Vaka. (2019). Cloud-Driven Excellence: A Comprehensive Evaluation of SAP S/4HANA ERP. Journal of Scientific and Engineering Research.

[27] Yale Song, Louis-Philippe Morency, and Randall Davis. 2013. Action recognition by hierarchical sequence summarization. In Proceedings of the IEEE Conference on Computer V.

10.48047/jocaaa.2025.34.08.10

[28] Thandaga Jwalanaiah, S. J., Jeena Jacob, I., & Mandava, A. K. (2023). Effective deep learning based multimodal sentiment analysis from unstructured big data. Expert Systems, 40(1), e13096.

[29] Jain, S., & Fallon, E. (2024). UDNet: A Unified Deep Learning-Based AutoML Framework to Execute Multiple ML Strategies for Multi-Modal Unstructured Data Processing. IEEE Access, 12, 77959-77975.

[30] Lopez, L. (2022). Multi-Modal Graph Representation Learning for Unified Analysis of Structured and Unstructured Enterprise Data.

[31] Kalisetty, S., & Lakkarasu, P. (2024). Deep Learning Frameworks for Multi-Modal Data Fusion in Retail Supply Chains: Enhancing Forecast Accuracy and Agility. American Journal of Analytics and Artificial Intelligence (ajaai) with ISSN 3067-283X, 2(1).

[32] Shrestha, Y. R., & He, V. F. (2023). Integrating multimodal data and machine learning for entrepreneurship research. Strategic Entrepreneurship Journal.