SalesMind: A Multimodal Emotion-Aware AI Assistant for Real-Time Sales Enhancement

Kalpana Ettikyala
Assistant Professor,
CSE Department,CBIT,
Gandipet, Hyderabad, India

Payyavula Vaishnavi
CSE Department,
CBIT,Gandipet,
Hyderabad, India

Meghana Kollavajjala
CSE Department,
CBIT,Gandipet,
Hyderabad, India

ABSTRACT

Successful sales in today's cutthroat digital marketplace rely not only on quick responses but also on real-time comprehension of consumer intent and emotions. Often emotional cues are ignored by conventional sales tools, due to which it results in lost opportunities and low engagement. To close such gaps in sales, SalesMind offers an AI-powered assistant which gives a score to text and audio conversations, recognizes the emotional states, ansd provides the sales representatives with conversational cues like suggestions and post-call analysis. The system uses VADER, transformer-based models like BERT, and PyTorch-driven speech emotion detection in a hybrid emotion analysis pipeline after recording consumer voice input and converting it to text. Suggestions are customized based on past behaviour and interests using a CRM module supported by MongoDB. Each conversation is given an emotional satisfaction score (0–10), which helps prioritize leads and reflects the customer's mood. The AI driven follow-ups are generally shown to moderate levels of engagement (around 6-7). SalesMind is a responsive real-time sales assistance platform developed using the MERN stack (MongoDB, Express.js, React.js and Node.js) Improved response quality, increased customer interaction, and higher conversion rates are anticipated results, making SalesMind a useful and significant instrument for contemporary sales.

General Terms

Artificial Intelligence, Machine Learning, Emotion Recognition, Natural Language Processing, Speech Processing, Multimodal Systems, Human-Computer Interaction, Intelligent Systems, CRM Analytics

Keywords

Artificial Intelligence (AI), Machine Learning (ML) Emotion Recognition, Customer Relationship Management (CRM), Sales Automation.

1. INTRODUCTION

From conventional rule-based CRM platforms that just recorded sales data and automated communication to AI-driven emotional analytics systems capable of contextual reasoning and real-time customer satisfaction, sales intelligence solutions have come a long way. Simple lead management has given way to emotion-aware AI assistants who use behavioural context, tone, and sentiment analysis to dynamically direct salespeople throughout encounters.

Developments in Speech Emotion Recognition (SER), Natural Language Processing (NLP), and multimodal learning frameworks that combine text and audio comprehension are the main forces behind this switch. Systems may now extract complex emotions and purpose from customer conversations thanks to the availability of transformer-based models (BERT, RoBERTa), lexicon-driven sentiment analyzers (VADER), and deep learning architectures implemented on PyTorch and TensorFlow. These solutions now provide context-sensitive analytics and adaptive feedback in real time thanks to cloud-edge integration and scalable deployment via MERN-stack web frameworks (MongoDB, Express.js, React.js, and Node.js).

Even with a high increasing level of technological advancements there are a number of critical problems which are yet to be resolved:

- Limited understanding of emotional cues in the sales settings.
- Inaccurate emotion detection in noisy, real-world communication settings.
- Emotion-based profiling and data retention raise ethical and privacy problems; emotion analytics are not integrated with current CRM workflows for consequential decision-making.

In order to position SalesMind as a next-generation multimodal AI assistant that fills in these gaps, this study compiles and evaluates existing initiatives in emotion-aware sales intelligence. To provide compassionate, data-driven decision assistance, the system integrates CRM-linked feedback engines, hybrid NLP pipelines, and speech and text emotion identification.

The following theme areas comprise the main contributions of this work:

1. Hybrid Emotion Recognition Architectures, which combine models based on PyTorch, BERT, and VADER for multimodal affect analysis.
2. Contextual Sales Intelligence and CRM Fusion: integrating insights gained from emotions into MongoDB-based CRM systems for adaptive communication and lead prioritization.
3. Emotional satisfaction measures are used in Real-Time Scoring and Smart Coaching to improve sales representative performance.

By fusing cognitive emotional intelligence, data understanding, and real-time analytics, SalesMind shows how emotion-aware AI may transform customer engagement through this framework, opening the door for emotionally intelligent and morally sound sales ecosystems.

2. LITERATURE SURVEY

In today’s highly competive digital world, for succesful sales engagement real-time understanding of customer emotions is necessary. The conventional sales approches frequently ignores the emotional cues which leads to missed oppurtunites for deeper customer connection.

Recent advancements in Affective Computing and Multimodal Emotion Recognition (MER) have paved the way for smart systems like SalesMind. These systems use sentiment analysis, audio, and text to improve customer interactions. So to develop emotion-aware systems that are precise, intelligible, and beneficial for customer relationship management (CRM), researchers are increasingly concentrating on integrating deep learning, transformer-based models, and explainable AI (XAI).

2.1 Multimodal Emotion Recognition

To increase the strength and accuracy of emotion identification, Multimodal Emotion Recognition (MER) integrates data from many communication channels, including text, audio, and facial expressions.

In order to improve CRM effectiveness, Theresa et al. [1] suggested a multimodal emotion analysis paradigm that combines textual and speech characteristics. In a similar vein, Gamage et al. [2] presented Emotion AWARE, an explainable AI framework that combines deep learning and lexicon-based techniques for multi-granular emotion analysis with enhanced interpretability.

Spain et al. [3] highlighted the potential of generative models in emotional understanding by using big language models to improve dialogue and communication analysis for adaptive team training. In their investigation of AI integration in sales and marketing, Rane et al. [4] emphasized how emotion-driven insights might increase customer experience and loyalty.

To increase communication quality, Płaza et al. [5] created an emotion recognition system for call centers using voice analysis. In a thorough analysis of deep learning-based multimodal emotion recognition, Lian et al. [6] came to the conclusion that hybrid models perform better than unimodal systems.

Using VADER and BERT, Wang and Zhou [7] suggested a hybrid emotion-aware human-computer interaction model that combines contextual embeddings for improved interpretability with rule-based sentiment analysis. Using universal speech representations, Atmaja and Sasou [8] created a voice emotion recognition method that achieved strong cross-lingual performance.

MemoCMT, a transformer-based multimodal fusion model with real-time cross-modal emotion identification, was presented by Khan et al. [9]. Di Luzio [10] created an explainable deep neural network for emotion recognition that strikes a compromise between interpretability and speed.

2.2 Explainability and Real-Time Insights

Explainability is essential in present-time AI systems, particularly in CRM, where algorithmic insights' comprehensibility is important for human decision-making and confidence.

Anupama [11] exhibited the effectiveness of a deep learning model for emotion recognition in dynamic user interactions. In order to evaluate model predictions across speech, text, and face inputs, Lian [12] presented an explainable multimodal emotion identification framework.

With an emphasis on interpretability and understandability, Norval and Wang [13] concentrated on explainable artificial intelligence (XAI) models for voice emotion recognition. In a systematic assessment of computer vision-based emotion detection, Pereira [14] found that multimodal fusion was the best approach for accurate emotional context comprehension.

A real-time emotion recognition model for call center agents was created by Huang et al. [15], which increased the customer experience positively and feedback loops. For effective voice emotion analysis, Prasad et al. [16] suggested a hybrid deep learning model that combines CNN and LSTM architectures.

A multimodal sentiment fusion system that integrates speech, face, and textual modalities for real-time emotional analysis was proposed by Das and Sengupta [17] specifically for CRM applications. In order to improve empathy and discussion flow, Bansal et al. [18] used BERT-based contextual models for emotion understanding in conversational AI.

Explainable multimodal attention networks that build up transparency in emotion-based decision-making were created by Zhang et al. [19]. Silva et al. [20] promoted appropriate and privacy-preserving emotional computing methods and highlighted ethical issues when implementing emotion AI systems for corporate communication.

2.3 Applications in Customer Relationship Management

AI-powered emotion recognition enables businesses to predict emotional states, increase customer experiences positively, and tailor engagement for customer relationship management.

As proposed by Theresa et al. [1] and Gamage et al. [2], the integration of multi-modal emotional intelligence with CRM can significantly enhance customer confidence and loyalty. As demonstrated by Spain et al. [3], large language models can obtain tone and intent to empower responses in communication-based training systems..

According to Theresa et al. [1] and Gamage et al. [2], integrating multimodal emotional intelligence into CRM can remarkably boost customer confidence and loyalty. As demonstrated by Spain et al. [3], large language models can effectively attain tone and intent, enabling adaptive responses in communication-based training systems.

Khan et al. [9] and Zhang et al. [19] verified transformer-based techniques for real-time multimodal emotion identification, ensuring scalability and responsiveness for commercial applications. Wang and Zhou have shown that contextual embeddings and sentiment analysis work together to increase user engagement and customisation [7].

According to Das and Sengupta [17], multimodal sentiment fusion allows CRM systems to adapt dynamically to client emotions, simulate more meaningful and emotionally aligned interactions.

Overall, the review of the literature indicates that CRM could be significantly transformed by emotion-aware, multimodal, explainable, and ethically sound technology. Frameworks that employ textual, facial, and auditory data, augmented by explainable models and transformer topologies, can increase emotional involvement and give sales teams useful data.

Table 1. Observations on different research papers

S. No.	Title	Year	Methodology	Observed Features
1	Multi-modal emotional analysis in customer relation management and enhancing communication through integrated affective computing	2025	AI-based multimodal emotion detection for CRM systems	Enhanced client happiness and engagement through better emotion-aware communication
2	Emotion AWARE: An artificial intelligence	2024	Framework utilizing explainable AI to integrate multi-granular emotion	High flexibility and openness in the techniques

	framework for adaptable, robust, explainable, and multi-granular emotion analysis		analysis	used to identify emotions
3	Applying large language models to enhance dialogue and communication analysis for adaptive training	2025	Used LLMs to examine the context and emotional tone of team discussions	Improved team flexibility and communication efficacy in real time
4	Artificial intelligence in sales and marketing	2024	Behavioural and emotional analytics in sales powered by AI	Enhanced customization and sustained client loyalty with emotion insights
5	Emotion recognition method for call/contact centre systems	2022	Call center systems that use algorithms for speech and facial emotion detection	Automated feedback on agent performance and customer mood detection
6	Survey of deep learning-based multimodal emotion recognition	2023	Thorough analysis of DL models that incorporate text, audio, and video	Found state-of-the-art multimodal methods and datasets for emotion recognition
7	Emotion-aware human-computer interaction using VADER and BERT	2023	Sentiment models that combine transformer-based (BERT) and rule-based (VADER)	Improved comprehension of context and identification of emotional tones in dialogues
8	Sentiment analysis and emotion recognition from speech using universal speech representations	2022	Used universal speech embeddings for emotion classification	Achieved better accuracy in cross-lingual and noisy speech environments
9	MemoCMT: Multimodal emotion recognition using cross-modal transformer-based feature fusion	2025	Cross-modal transformer for combining text, audio, and visual data	High interpretability and accuracy were attained in multimodal emotion recognition
10	An explainable fast deep neural network for emotion recognition	2025	Emotion classification using a lightweight explainable DNN model	For real-time systems, a balance between interpretability and model speed
11	Deep learning approach for emotions detection	2023	A hybrid model for emotion recognition based on CNN and RNN	Enhanced learning effectiveness and detection precision
12	Explainable multimodal emotion recognition	2023	Multimodal fusion model based on XAI	Enhanced interpretability of models and user confidence in AI emotion systems
13	Explainable AI techniques for speech emotion recognition	2025	Emotion recognition with XAI models (LIME, SHAP)	Enabled empathy AI's explainability and openness
14	Systematic review of emotion detection with computer vision	2024	Examined CNNs and vision transformers for visual-based emotion detection.	Highlighted shortcomings in real-time performance and cross-domain generalization
15	Real-time emotion detection for call center agents	2022	Monitoring emotions in real time with voice and facial clues	Enhanced productivity and well-being of agents by emotion tracking
16	Hybrid	2023	CNN and	Combining

	deep learning for voice emotion analysis		LSTM combined for the classification of speech emotions	CNN and LSTM to categorize speech emotions
17	Multimodal sentiment fusion for real-time CRM	2024	Combining speech, text, and visual modes in CRM analytics	Improved emotion prediction of customers in real-time communication
18	BERT-based contextual emotion understanding in conversational AI	2023	BERT model optimized for dialogue emotion classification	Enhanced AI chat systems' comprehension of contextual emotions
19	Explainable emotion recognition with multimodal attention networks	2023	A deep learning architecture based on attention	Enhanced explainability and precision in multimodal emotion prediction
20	Ethical emotion AI for business communication	2025	AI emotion models integrated within an moral framework for business environments	Encouraged the use of AI for justice, privacy, and responsible emotion

3.METHODOLOGY

The propose system, SalesMind is a multimodal, emotion-aware AI assistant that is intended to improve sales encounters by providing contextual decision help and real-time emotion analysis. In order to deliver intelligent sales support, the methodology combines voice processing, natural language comprehension, emotion detection, and customer relationship management (CRM) in a modular and tiered fashion.

3.1 System Architecture Overview

A layered design was used in the development of SalesMind to provide scalability, real-time performance, and the smooth integration of multimodal data. Data collection, preprocessing and feature extraction, hybrid emotion identification, emotion fusion and scoring, CRM integration, real-time coaching, and post-call analytics are the main parts of the system. Through backend APIs, each component interacts with the neighbouring layers to facilitate fast processing and constant data flow.

3.2 Data Acquisition

The data acquisition module captures multimodal sales conversation data in the form of audio and text. While textual data is gathered via chat messages or speech transcriptions, audio data is gathered during live or recorded sales calls using browser-based recording tools. In accordance with user consent and privacy policies, all collected data is safely delivered to the backend system.

3.3 Preprocessing and Feature Extraction

3.3.1 Speech-to-Text Conversion

An automatic speech recognition system is used to transform recorded audio information into text. Techniques for normalization, noise reduction, and silence removal are used to improve transcribing accuracy in actual sales settings.

3.3.2 Audio Feature Extraction

Acoustic characteristics including Mel-frequency cepstral coefficients (MFCCs), pitch, energy, and speech pace are taken out of the audio data in order to recognize speech emotions. After being normalized, these features are converted into numerical representations that deep learning models can use.

3.3.3 Text Preprocessing

Textual data is pre-processed by removing stop words, filler expressions, and irrelevant symbols. Tokenization and sentence segmentation are performed to prepare the text for emotion classification models.

3.4 Hybrid Emotion Recognition

SalesMind utilizes a combined strategy for emotion recognition, integrating lexicon-based, transformer-based, and deep learning methods to enhance precision and strength.

VADER is used to derive polarity scores that represent positive, negative, neutral, and compound attitudes in Lexicon-based sentiment analysis. Contextual and semantic emotional cues are simultaneously extracted from text data using a transformer-based language model. Furthermore, deep learning models trained on extracted acoustic data are used to recognize speech emotions, including neutrality, stress, irritation, and satisfaction.

3.5 Emotion Fusion and Satisfaction Scoring

The emotion outputs obtained from text-based and speech-based analysis are combined using a weighted fusion strategy. Confidence levels from each modality are considered to generate a unified emotional representation. For every contact, an Emotional Satisfaction Score (ESS) between 0 and 10 is calculated based on the fused emotional output. This score aids in determining lead prioritization and engagement quality by reflecting the customer's entire emotional state.

3.6 CRM Integration and Contextual Intelligence

SalesMind combines a MongoDB-based CRM system with emotional insights. Customer profiles, past interactions, and emotional patterns are all stored in the CRM. Personalized communication tactics, flexible sales recommendations, and the identification of high-priority leads based on emotional engagement are all made possible by contextual intelligence obtained from this data.

3.7 Real-Time Sales Coaching

The solution provides sales people with real-time coaching recommendations based on identified emotional states and CRM context. These recommendations, which are shown during live sales encounters via a responsive online interface, include tone adjustment alerts, response recommendations, and engagement improvement cues.

3.8 Post-Call Analysis

The system conducts post-call analysis following each sales encounter to produce emotional summaries, satisfaction ratings, and timelines of significant emotional events. Sales teams can assess effectiveness and improve future engagement tactics by using interactive dashboards to visualize these insights.

4. FIGURES

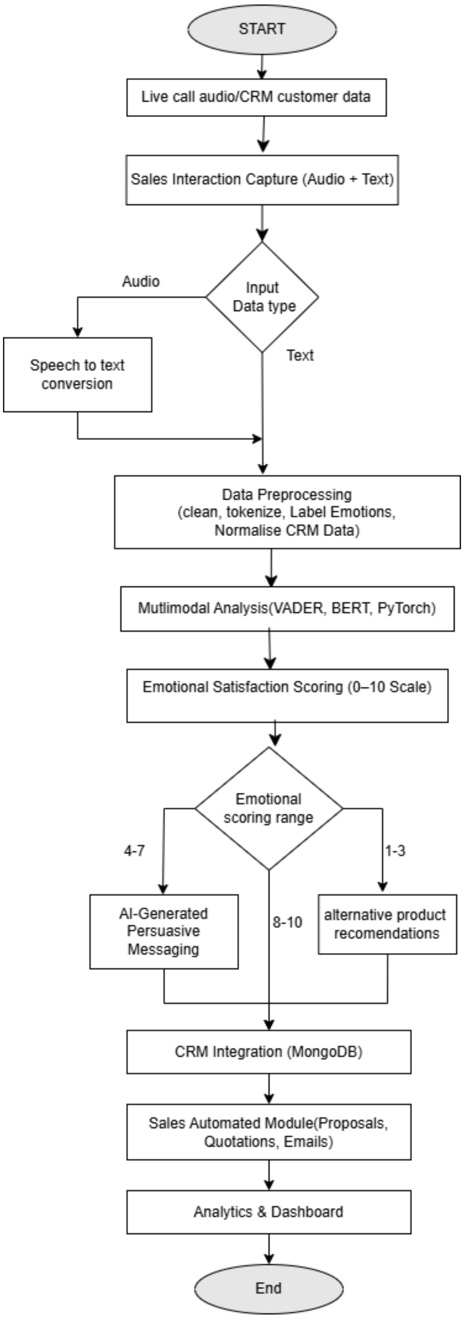
START
Live call audio/CRM customer data
Sales Interaction Capture (Audio + Text)
Input Data type
Audio
Text
Speech to text conversion
Data Preprocessing (clean, tokenize, Label Emotions, Normalise CRM Data)
Mutlimodal Analysis(VADER, BERT, PyTorch)
Emotional Satisfaction Scoring (0–10 Scale)
Emotional scoring range
4-7
8-10
1-3
AI-Generated Persuasive Messaging
alternative product recomendations
CRM Integration (MongoDB)
Sales Automated Module(Proposals, Quotations, Emails)
Analytics & Dashboard
End

Figure 1: Flowchart that shows the entire workflow

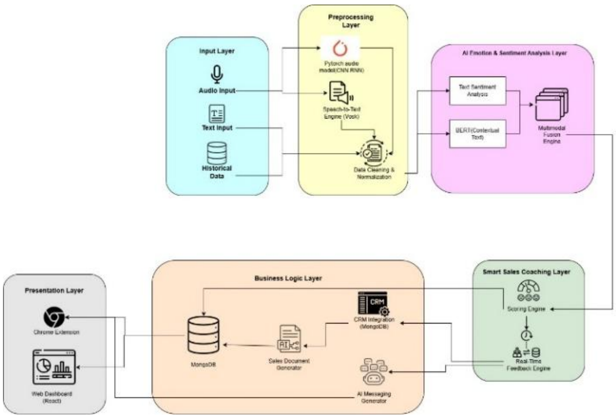

Figure 2: Architecture Diagram

5. CONCLUSION

This study introduced SalesMind, a multimodal emotion-aware AI assistant designed to enhance real-time sales interactions through contextual awareness and emotional intelligence. By combining transformer-based contextual modelling, text sentiment analysis, and speech emotion detection, the system effectively captures client emotions during sales conversations. The hybrid emotion recognition approach improves robustness and accuracy, especially in real-world scenarios where noise, ambiguity, and varied speech patterns are common.

By integrating emotional insights with a CRM system, SalesMind enables personalized communication strategies, adaptive interaction techniques, and efficient lead prioritization. The Emotional Satisfaction Score provides a measurable indicator of customer engagement, supporting both real-time coaching and post-interaction evaluation. This helps sales representatives adjust their approach dynamically, leading to improved customer relationships and better conversion potential.

Overall, SalesMind demonstrates how emotion-aware AI can bridge the gap between data-driven analytics and human understanding in sales environments. Future enhancements such as incorporating facial expression analysis, improving cross-lingual emotion recognition, and strengthening explainability mechanisms can further improve system effectiveness and trust. With these advancements, SalesMind can evolve into a more comprehensive and globally adaptable emotion-aware sales intelligence platform.

5.1 Future Scope

Additional modalities, such facial expression analysis, will be added to SalesMind in the future to increase the precision of emotion recognition. In order to enhance international sales situations, cross-lingual emotion recognition might be investigated. Emotion-based recommendations may become more transparent and trustworthy by incorporating sophisticated explainable AI algorithms. The effect of emotion-aware sales intelligence on conversion rates, client retention, and the ethical adoption of AI in corporate systems can also be assessed through long-term deployment studies.

6. REFERENCES

[1] G. Theresa, P. Rajasekaran, and S. Kumar, “Multi-modal emotional analysis in customer relation management and enhancing communication through integrated affective computing,” Scientific Reports, vol. 15, no. 26437, pp. 1–12, 2025.

[2] G. Gamage, D. De Silva, and D. Alahakoon, “Emotion AWARE: An artificial intelligence framework for adaptable, robust, explainable, and multi-granular emotion analysis,” Journal of Big Data, vol. 11, no. 93, pp. 1–18, 2024.

[3] R. Spain, W. Min, V. Kumaran, J. Pande, and J. Saville, “Applying large language models to enhance dialogue and communication analysis for adaptive team training,” International Journal of Artificial Intelligence in Education, 2025.

[4] N. Rane, M. Paramesha, and S. Choudhary, “Artificial intelligence in sales and marketing: Enhancing customer satisfaction, experience and loyalty,” SSRN Electronic Journal, 2024.

[5] M. Płaza, M. Kozłowski, and R. Kazała, “Emotion recognition method for call/contact centre systems,” Applied Sciences, vol. 12, no. 21, p. 10951, 2022.

[6] H. Lian, C. Lu, S. Li, Y. Zhao, and Y. Zong, "A survey of deep learning-based multimodal emotion recognition: Speech, text, and face," Entropy, vol. 25, no. 10, p. 1440, 2023.

[7] S. Wang and L. Zhou, "Emotion-aware human-computer interaction using VADER and BERT," IEEE Transactions on Computational Social Systems, vol. 10, no. 3, 2023.

[8] B. T. Atmaja and A. Sasou, "Sentiment analysis and emotion recognition from speech using universal speech representations," Sensors, vol. 22, no. 17, p. 6369, 2022.

[9] M. Khan, S. S. R. Abidi, and A. M. S. Ali, "MemoCMT: Multimodal emotion recognition using cross-modal transformer-based feature fusion," Scientific Reports, vol. 15, no. 89202, pp. 1–12, 2025.

[10] F. Di Luzio, "An explainable fast deep neural network for emotion recognition," Neurocomputing, vol. 456, pp. 1–12, 2025.

[11] A. Anupama, "Deep learning approach for emotions detection," E3S Web of Conferences, vol. 36, p. 07007, 2023.

[12] Z. Lian, "Explainable multimodal emotion recognition," arXiv preprint arXiv:2306.15401, 2023.

[13] M. Norval and Z. Wang, "Explainable artificial intelligence techniques for speech emotion recognition: A focus on XAI models," Inteligencia Artificial, vol. 28, no. 76, pp. 85–123, 2025.

[14] R. Pereira, "Systematic review of emotion detection with computer vision," PMC, vol. 11175284, 2024.

[15] T. Huang, X. Li, and S. Chen, "Real-time emotion detection for call center agents," Journal of Intelligent Systems, vol. 31, no. 4, pp. 433–446, 2022.

[16] V. Prasad et al., "Hybrid deep learning for voice emotion analysis," IEEE Transactions on Affective Computing, 2023.

[17] P. Das and A. Sengupta, "Multimodal sentiment fusion for real-time CRM," Expert Systems with Applications, vol. 237, 2024.

[18] S. Bansal, P. Verma, and L. Singh, "BERT-based contextual emotion understanding in conversational AI," ACM Transactions on Intelligent Systems, 2023.

[19] Y. Zhang et al., "Explainable emotion recognition with multimodal attention networks," Pattern Recognition Letters, vol. 168, pp. 134–142, 2023.

[20] E. Silva et al., "Ethical emotion AI for business communication," AI and Ethics Journal, vol. 5, pp. 89–101, 2025.