ISSN: 3107-6513

Multi-Modal Dissatisfaction: Integrating Voice Tonality and Textual Semantics for High-Fidelity Complaint Priority

Ananya A P, Dr. M. Kayalvizhi
Department of Information Technology, Sri Krishna Adithya College of Arts and Science, Coimbatore, Tamil Nadu, India
ananya09ap@gmail.com
Department of Information Technology, Sri Krishna Adithya College of Arts and Science, Coimbatore, Tamil Nadu, India.
kayalmp.02@gmail.com

Abstract:
This study presents a concise overview of research focused on improving customer complaint management. In modern organizations, complaints are submitted through multiple channels, including emails, chat messages, and voice calls. Existing complaint analysis systems predominantly rely on text-based sentiment analysis, which often fails to capture emotional urgency. Emotions such as anger, frustration, and stress are frequently communicated more effectively through voice intonation than through words alone.

To address this, we propose a ***Multi-Modal Dissatisfaction Analysis System*** that integrates both textual semantics and voice emotion detection to assign complaint priorities with greater accuracy. The system leverages Natural Language Processing (NLP) for analyzing textual content and Speech Emotion Recognition techniques to extract emotional cues from audio complaints. Acoustic indicators such as pitch variation, energy levels, and speech intensity are combined with semantic sentiment scores generated by advanced language models.

A fusion mechanism merges insights from both text and audio, enabling the system to distinguish low-risk complaints from high-urgency ones more effectively. Experimental results demonstrate that the multi-modal approach significantly outperforms text-only models in prioritizing complaints. This system provides organizations with a tool to respond promptly to critical issues, enhancing service quality and customer satisfaction. The study emphasizes the importance of incorporating emotional intelligence into modern customer support systems.

Introduction:

Customer complaint management is a critical component of service-oriented organizations. Complaints provide valuable insights into customer dissatisfaction and areas requiring improvement. With the digital transformation of customer support services, complaints are now submitted through diverse platforms such as online forms, emails, chatbots, and call centers. This multi-channel nature of complaints presents new challenges in accurately understanding customer emotions and urgency.

Traditional complaint analysis systems rely primarily on textual data and sentiment polarity, classifying complaints as positive, negative, or neutral. While effective to some extent, these systems often fail to capture emotional intensity. For instance, a customer might use polite language while expressing extreme frustration through voice tone. Ignoring such cues may lead to delayed responses to high-priority complaints.

Voice tonality carries rich emotional information such as anger, stress, and anxiety, which plays a crucial role in determining complaint urgency. Recent advancements in speech processing and NLP have made it possible to extract and analyze these emotional cues automatically. Integrating these technologies can significantly enhance complaint prioritization accuracy.

This research focuses on designing a ***multi-modal complaint analysis system*** that combines textual semantic understanding with voice emotion recognition. By leveraging both modalities, the system aims to provide a more realistic assessment of customer dissatisfaction. The proposed system is especially relevant for industries such as banking, healthcare, e-commerce, and telecommunications, where delayed responses to critical complaints can result in serious consequences.

The introduction sets the foundation for the study by outlining the problem, identifying limitations of existing systems, and motivating the

ISSN: 3107-6513

need for a multi-modal approach to complaint prioritization.

Literature Review :

The literature review examines existing research related to complaint analysis, sentiment analysis, speech emotion recognition, and multi-modal learning. Early complaint management systems used rule-based approaches and keyword matching to identify complaint categories. Although simple to implement, these systems lacked scalability and adaptability to linguistic variations [3].

With the rise of machine learning, statistical models such as Naïve Bayes, Support Vector Machines (SVM), and Decision Trees were applied to sentiment classification tasks. These models improved accuracy but still struggled with contextual understanding. The introduction of deep learning and transformer-based models such as BERT revolutionized text analysis by capturing contextual and semantic relationships within sentences [1][3].

Parallel research in speech emotion recognition focused on extracting acoustic features such as Mel Frequency Cepstral Coefficients (MFCC), pitch, speech rate, and energy. These features were used to classify emotions like anger, happiness, sadness, and neutrality. However, most speech-based systems were developed independently of textual analysis [2].

Recent studies explored multi-modal sentiment analysis by combining text, audio, and visual data, mainly for social media reviews and video content. While these studies demonstrated improved performance, they did not specifically address complaint prioritization in enterprise systems [4].

Moreover, many existing systems convert voice complaints into text using speech-to-text models, losing emotional depth in the process. This research addresses this gap by preserving and analyzing both modalities independently before fusion. The literature review highlights the lack of focused research on multi-modal complaint priority systems and justifies the need for the proposed approach.

Significance of the Study:

The significance of this study lies in its practical and technological contributions to modern customer support systems. In highly competitive service industries, timely resolution of customer complaints is crucial for customer retention and brand reputation. Traditional text-based systems often fail to recognize emotional urgency, leading to inefficient complaint handling.

By integrating voice tonality with textual semantics, this research introduces emotional intelligence into automated complaint management. The system enables organizations to identify high-risk complaints quickly, ensuring faster responses to emotionally distressed customers. This reduces escalation, improves service-level agreements, and enhances overall customer satisfaction.

From an academic perspective, the study contributes to the growing field of multi-modal machine learning by demonstrating its application in complaint analysis. It also encourages further research into emotion-aware IT systems.

For students and developers, the proposed system serves as a practical example of applying NLP, speech processing, and machine learning concepts in real-world applications. The study is especially relevant for UG IT students as it aligns with emerging technologies and industry needs.

Limitations of the Existing System :

Existing complaint analysis systems primarily rely on ***textual sentiment analysis***, which presents several critical limitations. These systems classify complaints based on keywords, sentiment polarity, or rule-based logic, often failing to capture emotional depth and urgency.

A major limitation is the ***inability to recognize emotional intensity***. Text-based systems cannot differentiate between a mildly negative complaint and an emotionally distressed customer if both use similar words. This results in misclassification of high-risk complaints as low or medium priority.

Another limitation is the ***loss of emotional cues in voice complaints***. Many existing systems convert voice inputs into text using speech-to-text models and analyze only the transcribed content. This process removes essential acoustic features such as tone, pitch, and stress, which are crucial for understanding dissatisfaction.

Existing systems also struggle with ***sarcasm, politeness bias, and indirect language***. Customers often express dissatisfaction politely or indirectly, which misleads sentiment classifiers. Furthermore, keyword-based models lack contextual understanding and are sensitive to linguistic variations.

ISSN: 3107-6513

Scalability is another challenge. As complaint volume increases, manual prioritization becomes inefficient, and automated text-based systems generate inaccurate results. Additionally, existing systems lack adaptability across languages, accents, and cultural expressions.

These limitations highlight the need for a more intelligent, emotion-aware system capable of accurately identifying complaint urgency beyond textual analysis.

Proposed System :

The proposed system, titled ***"Multi-Modal Dissatisfaction: Integrating Voice Tonality and Textual Semantics for High-Fidelity Complaint Priority"***, is designed to overcome the limitations of conventional text-based complaint analysis systems by incorporating ***emotional intelligence*** *through* ***multi-modal data processing***. The system analyzes customer complaints received in both **textual** and **voice** formats and assigns accurate priority levels based on emotional intensity and semantic meaning.

Overall Architecture

The system follows a ***modular and layered architecture***, ensuring scalability, flexibility, and ease of integration with existing customer relationship management (CRM) platforms. The architecture consists of six major components: Input Acquisition, Textual Semantics Analysis, Voice Tonality Analysis, Multi-Modal Fusion, Priority Classification, and Decision Support Interface.

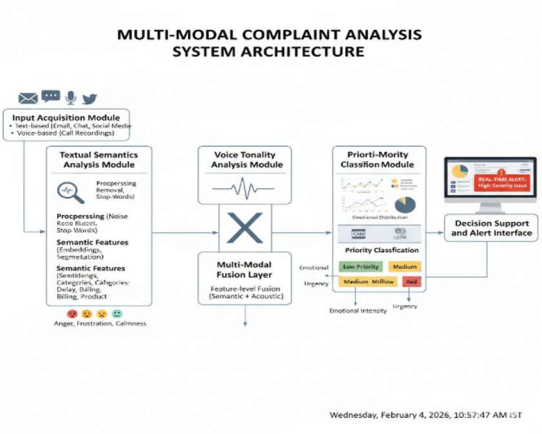

Fig: 1 Architecture

1. Input Acquisition Module

This module serves as the entry point of the system and collects customer complaints from multiple sources:

- Text-based complaints such as emails, chat transcripts, feedback forms, and social media messages
- Voice-based complaints including call center recordings and voice assistant interactions

The system automatically detects the input type and routes it to the appropriate processing pipeline. Voice complaints are stored in audio format, while textual complaints are stored in structured text databases.

2. Textual Semantics Analysis Module

The textual analysis module is responsible for understanding the ***meaning and emotional intent*** behind written complaints. Initially, the text undergoes preprocessing steps including noise removal, stop-word elimination, punctuation removal, and lemmatization.

After preprocessing, semantic features are extracted using advanced NLP models. Contextual embeddings capture the underlying meaning of words in relation to surrounding text. Sentiment polarity (positive, negative, neutral) and ***sentiment intensity scores*** are computed to measure dissatisfaction levels.

This module also identifies key complaint categories such as service delay, billing issues, product defects, or customer support behavior. These semantic features form a high-dimensional representation of the complaint content.

3. Voice Tonality Analysis Module

The voice tonality analysis module captures ***emotional cues embedded in speech***, which are often absent in text. Voice data undergoes preprocessing steps such as noise reduction, silence trimming, and segmentation to isolate meaningful speech segments.

Acoustic features are extracted, including:

- Pitch and pitch variation
- Speech energy and intensity
- Speech rate and pause duration
- Mel Frequency Cepstral Coefficients (MFCC)

These features are used to classify emotions such as anger, frustration, stress, sadness, or calmness using trained emotion recognition models. The output of this module is an emotion score vector representing the emotional state of the customer during the complaint.

4. Multi-Modal Fusion Layer

The fusion layer is the core component of the proposed system. It integrates features extracted from both the textual semantics and voice tonality modules. Feature-level fusion is employed to

ISSN: 3107-6513

combine semantic embeddings and acoustic emotion vectors into a unified representation.

This combined feature vector ensures that the system evaluates both ***what the customer says*** *and* ***how the customer says*** **it**. The fusion mechanism reduces ambiguity and enhances robustness, especially in cases of polite language masking strong emotional distress.

5. Priority Classification Module

The fused feature vector is fed into the priority classification module. This module uses machine learning and deep learning classifiers such as CNN and LSTM to categorize complaints into predefined priority levels:

- **Low Priority** – Minor issues with low emotional intensity
- **Medium Priority** – Moderate dissatisfaction requiring timely response
- **High Priority** – Critical complaints with strong emotional distress

The classifier learns complex patterns between emotional intensity and complaint urgency, ensuring accurate prioritization.

6. Decision Support and Alert Interface

The final module presents the prioritized complaints through a **dashboard interface**. High-priority complaints trigger real-time alerts to customer support teams. Visualization tools display complaint trends, emotional distribution, and response performance.

This interface enables support agents and managers to make informed decisions, allocate resources efficiently, and improve customer satisfaction.

Advantages of the Proposed System:

The proposed multi-modal complaint analysis system offers several significant advantages over traditional approaches. By integrating ***voice tonality and textual semantics,*** the system introduces emotional intelligence into automated complaint management.

One of the primary advantages is ***high-fidelity complaint prioritization***. The system accurately identifies high-urgency complaints by analyzing emotional cues such as anger, frustration, and stress in voice data, even when textual sentiment appears neutral.

The system improves ***response efficiency*** by enabling customer support teams to focus on critical issues first. Automated prioritization reduces manual intervention, operational workload, and response delays, thereby improving service-level agreement (SLA) compliance.

Another advantage is ***robustness and reliability***. The fusion of multiple modalities reduces misclassification caused by ambiguous language, sarcasm, or incomplete information. The system adapts well to diverse communication styles.

From a technological perspective, the modular architecture ensures ***scalability and flexibility***. The system can be easily integrated with existing CRM and helpdesk platforms. It is applicable across industries such as banking, healthcare, e-commerce, and telecommunications.

The proposed system also enhances ***customer satisfaction and trust*** by ensuring emotionally distressed customers receive timely attention. This proactive approach helps organizations prevent complaint escalation and reputational damage.

Research Methodology :

The research methodology adopted for this study follows a ***systematic experimental and analytical approach*** to design, implement, and evaluate a multi-modal complaint prioritization system. The methodology is structured into multiple stages to ensure accuracy, reliability, and reproducibility of results.

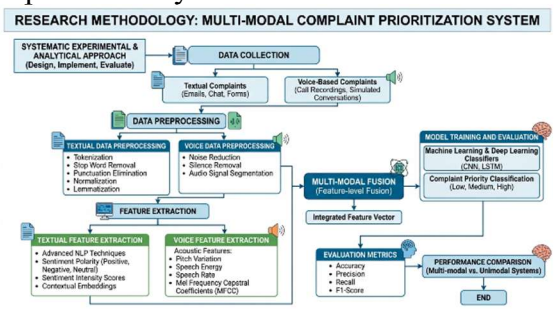

Fig 2: Methodology

Data Collection

The dataset used in this research consists of both ***textual complaints*** *and* ***voice-based complaints***. Textual data includes customer emails, chat transcripts, and online complaint forms. Voice data comprises call recordings obtained from publicly available speech emotion datasets and simulated customer service conversations. This combination ensures diversity in language usage, emotional expression, and complaint intensity.

Data Preprocessing

Textual data undergoes preprocessing steps such as tokenization, removal of stop words, punctuation elimination, and normalization. Lemmatization is applied to reduce words to their

ISSN: 3107-6513

base forms, improving semantic understanding. Voice data preprocessing involves noise reduction, silence removal, and segmentation of audio signals to isolate meaningful speech components.

Feature Extraction

For textual analysis, semantic representations are generated using advanced NLP techniques. Sentiment polarity (positive, negative, neutral) and sentiment intensity scores are computed to capture emotional depth. Contextual embeddings enable the system to understand complaint meaning beyond surface-level keywords.

For voice analysis, acoustic features such as pitch variation, speech energy, speech rate, and Mel Frequency Cepstral Coefficients (MFCC) are extracted. These features are critical for identifying emotional states such as anger, frustration, stress, or calmness.

Multi-Modal Fusion

The extracted textual and audio features are combined using ***feature-level fusion techniques***. This fusion allows the system to simultaneously analyze what the customer says and how it is expressed. The integrated feature vector is then passed to the classification model.

Model Training and Evaluation

Machine learning and deep learning classifiers such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are trained on the fused dataset. The system classifies complaints into ***Low, Medium, or High priority*** categories.

Evaluation metrics include accuracy, precision, recall, and F1-score. The performance of the multi-modal system is compared against unimodal (text-only) systems to validate its effectiveness

Results And Discussion of the Proposed Complaint Prioritization System :

The table summarizes the experimental results comparing text-only and multi-modal approaches, highlighting improvements in accuracy, precision, and recall achieved through integrating voice emotion analysis with textual sentiment features.

Approach	Evaluation Metric	Observed Outcome	Discussion
Text-Based Sentiment Analysis	Accuracy	Moderate performance in identifying sentiment polarity	Since the model relies only on textual information, it cannot detect emotional intensity when customers express dissatisfaction politely or indirectly.
Text-Based Sentiment Analysis	Recall (High-Priority Complaints)	Lower recall rate	Several urgent complaints were not detected because emotional cues conveyed through voice were not analyzed.
Voice Emotion Recognition	Emotion Detection Accuracy	Effective detection of emotions such as anger, stress, and frustration	Acoustic features including pitch variation, speech rate, and speech energy helped identify emotional distress in voice complaints.
Multi-Modal Model (Text + Voice)	Overall Accuracy	Higher accuracy compared to text-only systems	Combining textual semantics with vocal emotion cues improved the model's ability to detect complaint urgency more reliably.
Multi-Modal Model (Text + Voice)	Precision	Improved precision for high-priority classification	The integration of multiple modalities reduced false positive predictions and increased reliability in identifying urgent complaints.

ISSN: **3107-6513**

Approach	Evaluation Metric	Observed Outcome	Discussion
Multi-Modal Model (Text + Voice)	Recall	Higher recall for emotionally intense complaints	Emotional signals present in speech allowed the system to detect urgent cases even when textual sentiment appeared neutral.
Multi-Modal Fusion Approach	Robustness	Better handling of ambiguous complaints	The fusion model effectively addressed challenges such as sarcasm, politeness bias, and incomplete textual information.
Proposed System	Practical Impact	Improved complaint prioritization and response efficiency	Organizations can identify and respond to critical customer issues more quickly, enhancing customer satisfaction and service quality.

Conclusion:

This research concludes that integrating ***voice tonality analysis with textual semantic understanding*** significantly enhances complaint priority classification. Traditional text-based systems fail to capture emotional intensity, leading to inaccurate prioritization and delayed responses.

The proposed multi-modal dissatisfaction analysis system addresses these limitations by combining NLP-based sentiment analysis with speech emotion recognition. Experimental results validate that emotional cues present in voice data play a critical role in identifying urgent complaints.

The findings demonstrate that the multi-modal approach provides higher accuracy, robustness, and reliability compared to unimodal systems. By incorporating emotional intelligence, the system aligns more closely with human judgment in assessing complaint urgency.

This research contributes to the field of intelligent customer support systems and multi-modal machine learning. It also provides a practical framework for real-world deployment in service-oriented industries.

Future enhancements may include real-time processing, multilingual support, accent-independent emotion recognition, and advanced transformer-based fusion models. Overall, the proposed system represents a significant step toward emotionally aware and intelligent complaint management solutions.

References

1. *Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.*
2. *Schuller, B., Batliner, A. (2013). Computational Paralinguistics: Emotion, Affect and Personality in Speech.*
3. *Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis.*
4. *Poria, S., et al. (2017). A review of affective computing: From unimodal to multimodal analysis.*
5. *Jain, M., & Kulkarni, P. (2020). Automated customer complaint analysis using NLP techniques.*