iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VIII Month of publication: August 2025

DOI: https://doi.org/10.22214/ijraset.2025.73790

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VIII Aug 2025- Available at www.ijraset.com

Emotion Recognition from Text and Voice: A Multimodal AI Approach to Understanding Human Feelings

Mansara Sharma[1], Dr. Umadevi Ramamoorthy[2]
MCA, CMR University, Bangalore, India

Abstract: Now, in our global online world, technology acts as an interface for communications within our lives, from intelligent assistants to internet customer support and even social networking. Humans, being proficient at interpreting the feelings that come with communication, are now done by machines. The coating of emotions conveying words in real time continues to be an enigma still unsolved. This project tries to answer the question of how to solve machines interpreting emotions. Presenting a multimodal AI system that interprets emotions based on the simultaneous examination of speech and text content with NLP and speech analysis algorithms and a fusion-based deep learning system is what our work revolves around. Using cutting-edge NLP and speech processing, we are creating systems that decode for content, as well as conduct and ethics. The backbone of the system for voice content decoding is the emotional cues of the language processing, or L through. Our model utilizes the transformer models Distil BERT and Roberta. Emotions are also present in the voice with MFCCs and speech as a series of frames via chroma spectrograms, which are processed by CNN-LSTM hybrids for emotion recognition from voice. More sophisticated models are also built for further fusion. Our performance on model as well as on individual streams enhances emotion detection systems. We accomplish in this project combining various models. The voice and the text are processed independently and different outputs are provided by each model.

Keywords: Emotion Recognition, Multimodal AI, Natural Language Processing (NLP), Speech Analysis, Text Sentiment Analysis, Acoustic Features, Voice Emotion Detection, Hybrid AI Model, Deep Learning, Transformer Models (Distil BERT Roberta), CNN-LSTM, MFCC (Mel-Frequency Cepstral Coefficients), Chroma Features, Spectrograms, Late Fusion, Ensemble Learning, Accuracy, Precision, Recall, F1-Score, Confusion Matrix, Real-time Emotion Detection, Affective Computing, Emotion-Aware Systems, Sentiment Monitoring, Human-Computer Interaction, Facial Expression Recognition, Multilingual Emotion Analysis, Temporal Emotion Tracking, Mobile-Optimized Emotion AI,

I. INTRODUCTION

Emotions are those intangible strings that guide us in understanding how to connect, make decisions, and connect with others. Happiness, anger, confusion, or empathy – emotions weigh more heavily on our communication than any words do. However, even with all the technology, most computers remain functioning on language levels without understanding such lower emotional levels—something that comes as a natural feature to man. In our modern day and age, where computerized communication is so enormous a part of our lives—whether in the form of smart assistants, mobile apps, virtual learning spaces, or online aid systems—it is no longer sufficient for technology to simply "hear" what we say. It has to understand how we feel when we utter it. Imagine a virtual assistant that can recognize when you're upset and soften its response, or a study guide that can pick up on a student's aggravation and react sympathetically. These are the kinds of emotionally intelligent systems the world should have in the future. Thus far, emotion detection has been about analyzing text. That gets some way, but it fails to describe the whole emotional picture. Words don't always speak entirely truthfully—especially where sarcasm, humor, or cultural context enters the frame. In contrast, looking at just the voice can be insufficient, especially where the emotional tone is the reverse of the literal meaning of words. That’s where this research branches off. We believe that a grasp of emotion does indeed involve attention to the content and delivery of communication. By combining text analysis with Natural Language Processing (NLP) and voice-emotion recognition, we offer a multimodal solution—a system that combines the best of both worlds to offer a deeper and more human-like interpretation of emotions. Our vision is to create a smart system that not only listens to what people say but also how they say it—melting textual sentiment with vocal intonation for greater emotional accuracy.

This kind of model can power real-world technology like affective customer service robots, virtual therapy platforms personalized to a person's needs, and intelligent learning platforms that adapt based on a learner's emotional state.Finally, we're headed towards a world where machines don't just crunch numbers—but interact with us in a more human manner. Here, we attempt to close the gap toward emotionally intelligent AI that can listen, understand, and reply with empathy.

II. PROBLEM STATEMENT

A. Text-Based Emotion Detection Challenges

Text can typically fool one's actual emotions. NLP has gotten much better at sentiment recognition but fails when identifying nuances like sarcasm, irony, or emotionally charged terms that are highly context-dependent. For example, the words "Great, just what I needed!" can express annoyance or genuine enthusiasm depending on the tone. Text models will interpret it at face value and lose the affective connotation, perhaps generating incorrect results in situations where they are entirely text-based.

B. Limitations of Voice-Only Emotion Detection

Audio-based emotion-detection systems read pitch, volume, and speech rate to determine the emotional state of a person. Without access to what is said, voice-only systems can be misled. For instance, the high tone could mean excitement, stress, or even fear—depending on the situation. Without understanding the words spoken, those systems may make incorrect conclusions. That can lead to misunderstanding, especially in emotionally complex conversations like therapy or customer support.

C. Hybrid Approach Requirement

Communication is never voice or text alone—feelings have a tendency to get communicated through a combination of both. Most emotion detection systems today, though, operate on these channels in isolation, ignoring the entire emotional context. The person may be speaking in a soothing voice but in words that express anguish, or in a quivering voice but saying plain words. This difference signifies the limitation of mono-modal models and proposes the need for a hybrid model that can process both modalities in parallel in order to build a more accurate interpretation of emotional intent.

D. Consequences of Misreading Emotions in Real-World Applications

Misreading of emotions has its impact not only on system performance but also influences people directly. For example, in online therapy, failure to recognize signs of sadness or anxiety may lead to inappropriate behavior that worsens the user's emotional situation. In education, frustration or discouragement among students may be overlooked, reducing learning effectiveness. In customer service, failure to recognize emotional cues may escalate complaints instead of resolving them.These are examples of why more precise and empathetic emotion recognition is very important in delicate fields.

E. The Case for a Multimodal Emotion Recognition System

To address these concerns, this study does come up with a single framework encompassing both text-based sentiment analysis and voice emotion detection. By combining these two potent sources of emotional data, the model is designed to comprehend what users say and how they say it. Through such a combination, the emotional meaning ends up being much more detailed, rendering the interactions of AI more sensitive, responsive, and human-like.

III.OBJECTIVES

A. Identify Emotions from Written Text and Speech

The primary target of this study is to actually identify human emotions based on written text as well as speech. Since emotions are usually conveyed not only in what one is saying but also in how one is saying something, the aim is to create a system that can recognize and comprehend linguistic content along with vocal tone. The two-channel analysis will be able to capture a richer impression of the user's emotional state beyond the mere use of text or voice alone.

B. Develop a Hybrid AI Model Incorporating NLP and Acoustic Features

This work proposes to develop and deploy a hybrid artificial intelligence system with the integration of Natural Language Processing (NLP) methods for text and acoustic feature extraction for voice so that it can study both these kinds of data in one system. By integrating these two kinds of data, the system will be able to evaluate emotion in a more contextual and accurate manner. It is not merely a matter of merging the two but allowing the model to learn how the two inputs interplay and lead to emotional significance.

C. Boost Accuracy and Contextual Understanding

Improving the accuracy of emotion recognition systems in daily life situations that are typical and where context determines the decision is one of the main objectives. The system must be able to recognize emotional cues even under difficult contexts like sarcasm, unclear emotions, or indeterminate expression. Being advantaged with the twin strengths of both voice and text, the system will be in a better position to understand intent, tone, and sentiment.

D. Facilitate Real-Time Emotion Perception for Human-Computer Interaction

Another most critical goal is to build a real-time inference pipeline that enables the system to detect and react to emotions in real time. The capability will be of greatest significance for virtual assistant, mental wellness bots, and adaptive learning systems use, where real-time emotional intelligence can greatly enhance user experience and interaction. The model will be fine-tuned to respond immediately and scale up rapidly, thus being appropriate for real-time deployment.

IV. LITERATURE REVIEW

This article describes sentiment analysis (SA) and text emotion detection (TED) models' influence on affective computing to facilitate machines processing emotions in text. While the majority of machine learning models are very good, no single model is ever tested in a real environment. The article centers on under-exploited design parameters such as emotion models, training data, and test procedures, and recommends future research activities attempt to minimize bias and maximize real-world applicability.[1]

Social media communications provide us with rich information about personality trends in the modern digital age. Trends of introversion and extroversion are seen in the use of language and behavior online, and researchers use this information to inform mental health applications, marketing, hiring, and user experience. The research confirms the increasing relevance of personality identification in data-driven areas.[2]

Social media such as X (formerly known as Twitter) are popularly used but frequently victimized by bots that propagate disinformation or harmful information. While there are beneficial update bots, most manipulate online discussions. To rectify this, explainable AI techniques such as SHAP and LIME make detecting bots more precise, transparent, and reliable.[3]

This paper describes how individuals share views about health tools via social media. It examines how sentiment analysis facilitates the study of these views. Some approaches apply word lists, while others employ machine learning. The aim is to identify improved methods for harnessing online views in healthcare.[4]

Individuals read reviews of hotels before making reservations, but too much of everything is perplexing. This research verifies hotel reviews in Jordan for whether they are positive or negative. It makes use of simple tools to enable travelers and hotels to make wiser decisions.[5]

With the extensive utilization of social media, hate speech, propaganda, and extremist content have increased, particularly during crises such as the COVID-19 lockdown. This research addresses the problem by developing an Arabic tweet dataset annotated by humans to train a neural network to identify radical and hate content. Through the analysis of 100,000 tweets for the last ten years, the model detects patterns of abusive language, which assists in preventing the diffusion of online extremism. The research also identifies the particular linguistic difficulties of the Arabic language in detecting hate speech.[6]

The majority of people currently communicate their feelings and emotions on social media. Emojis are used in a bid to make these communications emotional and interpretable. Nevertheless, it is still hard to understand emotions when using the Arabic language as there are a lot of rules and different types of speech. The current study develops a new tool, which is an Emoji Sentiment Lexicon (Emo-SL), for Arabic. It uses such a tool using machine learning to enhance insight into human emotions using word and emoji analysis used in posts.[7]

A study utilizes tweets to investigate mothers' attitudes towards breastfeeding. It utilizes sentiment analysis to categorize opinions based on them being positive or negative. The determinants that advantage or disadvantage breastfeeding are discovered. Such determinants can be used to inform improved support for mothers across the world.[8]

The last decade's digital revolution has transformed the way humans interact, communicate, and relate to the world. Facebook, Twitter, and Instagram are now necessities as they have become a means by which billions use them to get information, communicate, and contribute their voices to global discourse. Their universal use has made these platforms powerful vehicles for the dissemination of ideas, facilitation of arguments, and even detection of misinformation. Their increasing strength bears witness to their power to drive public opinion and bridge cultural divides.[9]

To solve the ever-increasing need for reliable sentiment analysis of the Thai language, this research proposes SETAR—a machine learning ensemble model that integrates the most effective state-of-the-art NLP and machine learning approaches. SETAR integrates specially trained Thai datasets with hybrid text representation to considerably improve the accuracy of short social media text sentiment classification. The model was contrasted with various algorithms and datasets and established a new standard for the area of sentiment analysis studies in Thai.[10]

Sentiment analysis is applied in real life because opinions are conveyed by means of reviews, social media, and internet posting. Machine learning and optimization methods are utilized to analyze these emotions even in complicated or unclear language. This research explores sophisticated models such as deep RNNs and employs algorithms such as gradient descent and genetic techniques to enhance performance on different datasets. The aim is to assist machines in better identifying human emotions in a range of real-world contexts.[11]

Experts examined comments made on Reddit in Saskatoon to learn about important public issues like housing, healthcare, and education. They applied AI models such as BERTopic and SiEBERT to cluster conversations and determine sentiment. These findings assist regional leaders in planning more effectively and addressing residents' needs.[12]

The research addresses the increasingly prevalent problem of hate speech on Twitter by drawing on real-time tweet data and sophisticated NLP methods. With tools such as TextBlob, TF-IDF, and word cloud visualization, the authors constructed and studied a tailored tweet dataset, using conventional machine learning as well as deep learning models such as LSTM. Their proposed model using LSTM was 97% accurate, which demonstrated its competence to identify hateful content and provide a scalable method to monitor sentiment in real-time on social media.[13]

This research examines the ways in which sophisticated AI methods, machine and deep learning specifically, are transforming depression detection on social media. Conventional approaches such as questionnaires tend to overlook initial symptoms, but analyzing what people post, do, and share can provide more precise, timely mental health data. The article surveys emerging detection models, data available for these models, ethical issues, and suggests a comprehensive framework for future research and applications of mental health monitoring.[14]

This paper introduces a novel Arabic dataset named ArabSis, constructed especially for the analysis of emotions in text. In contrast to most other datasets, which consider solely whether something is positive or not, ArabSis consists of five types of different emotions, providing a more comprehensive picture of the feelings of people. Various models were tried out by the researchers and concluded that simple methods can provide quality results. This data is particularly valuable for enhancing sentiment analysis in those languages that lack much material, such as Arabic.15]

V. METHODOLOGY

A. Text-Based Emotion Analysis

Emotion recognition from text begins with the process of turning the raw text into clean, machine-understandable text. The system first does data preprocessing—this involves splitting sentences into individual words (tokenization), filtering out words in the text that are not helpful such as "is" or "the" (stop-word removal), and shrinking words to their root form (stemming or lemmatization). After the text is purified, meaningful features are extracted by the system.

It can be achieved with TF-IDF (which analyzes how significant a word is in a text), or word embeddings such as GloVe or BERT, which comprehend words in context. In classification, the study employs sophisticated transformer-based models such as RoBERTa or DistilBERT, which are capable of grasping subtle emotions despite the presence of sarcasm, ambiguity, or cultural subtleties in the text.

B. Voice-Based Emotion Analysis

The audio input, such as a sentence spoken, is initially tidied up with signal preprocessing methods such as noise removal and silence stripping—this means that only useful audio is processed. From this tidied-up audio, the system extracts emotional signals through the use of features such as: MFCCs (Mel-Frequency Cepstral Coefficients) – assists in capturing the way the voice actually sounds to the human ear.

Chroma Features – picks up musical tones and pitch aspects of speech. Spectrogram Images – images of sound that may be used to identify changes in pitch or volume over time. These are then input into a hybrid deep learning network—CNN-LSTM—that uses convolutional layers (to process spatial patterns in spectrograms) and recurrent layers (to learn how the emotion evolves over time in the speech).

C. Fusion Model: Merging Text and Voice

In order to have a better view of what emotions are, the system does not depend on a single input. It employs a fusion model that combines the predictions from both text and voice. The most common approach is Late Fusion, which makes the system initially process both inputs individually and then combine their predicted probabilities in order to make a final decision. To further enhance the reliability and consistency, ensemble learning methods are used. That is, multiple models are combined so that the detected emotion is more robust and reliable—particularly in real-world applications where inputs may be noisy or ambiguous.

VI.RESULT& EVALUATION

A. Evaluation Measures

To quantify how well the emotion recognition system is performing, several standard metrics are used:

1) *Accuracy:* This measures the ratio of correct predictions made out of all the predictions. It gives a rough estimate of how well the model is doing.
2) *Precision:* Precision measures the number of emotions which the model correctly predicts. It's very important when wrong positive (incorrect emotional labels) are costly.
3) *Recall:* Recall measures the capacity of the model to capture all the true emotional events. For example, how many times it accurately recognized sadness or anger when actually present.
4) *F1-Score:* This is the intersection of precision and recall. It's especially useful when dealing with imbalanced datasets so the model doesn't compromise precision and reliability.
5) *Confusion Matrix:* A line-by-line analysis that shows where the model is picking up on things correctly or not—i.e., how many times it's getting anger mixed up with frustration or disappointment mixed up with sadness.

B. Experimental Results and Hypothetical Results

Whereas ultimate conclusions will be subject to the dataset's quality and size, the following are expected outcomes considering preliminary tests and existing models:

1) *Text-Only Model Accuracy (~85%):* The model that looks only at the text written should be very good. Text emotion detection has been greatly improved with language models like BERT and RoBERTa, especially when using sarcasm or mild emotions.
2) *Voice-Only Model Accuracy (~80%):*The voice-alone model, which picks up tone, pitch, and speed, is somewhat less accurate on its own since it may misdiagnose emotions without having the aid of words. However, it still picks up on major emotional signals, especially regarding stress, happiness, or nervousness.
3) *Combined Multimodal Model Accuracy (~90–92%):*If voice and text inputs are compared side by side, the accuracy improves remarkably. The combining of these helps the system not just know what was said but also how it was said. The combining reduces misinterpretation and creates a better-rounded view of the mood of the speaker.

C. Significance of Results

These findings show that a multimodal model far surpasses single-input models. It shows how human emotions are best interpreted when both speech and vocal signals are examined together—similar to how humans interpret each other in real life. This increased accuracy is more than just a statistic—it means better assistance in real-world applications such as therapy bots, learning software, or emotionally-intelligent assistants that must respond empathetically and intelligently.

VII. APPLICATIONS

Understanding human feelings from what we say and how we say it opens the way to a new generation of emotionally intelligent technologies. A few of the key domains where this work can impart real-world value and emotional intelligence to computer systems are discussed below:

A. Enhancing Mental Health Support Systems

Our tone and words are likely to carry emotional significance even when we do not explicitly state how we are feeling. A tone change, or sustained negative words, may betoken underlying emotional challenges like anxiety, burnout, or depression. Emotion detection technology can help therapists and mental health apps pick up such early warning signals and step in accordingly. That isn't replacing professionals—but helping them become more proactive and responsive. Consider a journaling program that prompts a user to catch up with a counselor once it has detected consistent emotional breakdown—that's where this technology shines.

B. Emotion-Aware Digital Learning Platforms

Every learner learns according to their method, and feelings greatly influence the way we absorb new information. Frustration, boredom, or confusion stand in the way of learning, especially in virtual or digital classes. With such emotions picked up through text or voice responses, AI-based e-learning or tutoring platforms can react adaptively. A disoriented tone of voice may be answered back with a clearer explanation, or a discouraged tone may be answered with motivational support. This would create online learning as more personalized and emotionally nurturing as an effective human educator.

C. Sentiment Monitoring on Social Media

Social sites are full of emotion—humans release happiness, anger, sadness, and hope in the moment. By analyzing both written and spoken intonation (in audio/video media), emotion recognition systems can be used to help mark hate speech, identify crises in real-time, or even take the pulse of public mood during events. As an example, in a disaster situation, identifying widespread fear would help direct emergency response more effectively. The technology acts as an emotional radar that gives a glimpse of the collective mood of the internet.

D. Designing Empathetic Virtual Assistants

Modern digital assistants can now trigger alarms and provide trivia responses—but the human touch, no yet. Think if your assistant could sense frustration in your tone and react less frantically, or provide assistance if you sound sad or stressed. Emotionally smart assistants might then speak with less stilted and more natural voice tones. It's not attempting to be human, but becoming emotionally more suitable and sensitive in the way they interact.

E. Emotional Intelligence Enables Better Customer Support

Nobody wants to have to repeat themselves to a clearly not frustrated chatbot. With emotional intelligence added in, customer support software can detect tension or frustration building and adjust the interaction as a result—either by providing more empathetic answers or passing the issue on to a live agent. That translates into quicker resolutions, contented customers, and a support system that's responsive.

F. Emotionally Intelligent Caregiving Robots

Robots are gradually finding their way into our households—particularly for the elderly or kids who would be in need of constant care. Robots that are emotionally aware can sense loneliness, sadness, or excitement with words and speech. To illustrate, an elderly robot companion can start a conversation or notify family members if it senses sadness in the tone of the person's voice. This gives machines warmth and empathy, in turn making them more useful and human-friendly.

G. Adaptive Gaming and Entertainment

Gaming is no longer solely high scores—it's emotional, immersive experiences. With emotion recognition, games can adapt their narratives, their pace, or difficulty to how the player feels. If the player appears bored, the game can add more action. If they are stressed, it might provide less challenging challenges. This creates more engaging and individualized gameplay, responding not only to decisions—but to emotions.

H. Workplace Wellness and HR Insights

Staff tend to hide their stress, particularly in telecommuting environments. But voice tone patterns during a phone call or writing styles may betray deeper emotional states. AI systems may assist HR departments in catching signs of burnout or disengagement early—enabling supportive interventions before things get out of hand. Even in interviews, small emotional cues may provide a better idea of how candidates really feel—giving a more holistic picture than words.".

I. Emergency Services and Crisis Support

Under high-stakes conditions, individuals will sometimes say something other than what they feel. Emergency responders usually trust tone and urgency to gauge risk. Emotion-sensitive systems applied in emergency call centers or crisis chats can detect anxiety, terror, or distress sooner than words may indicate. This allows responders to prioritize and respond to calls with greater sensitivity and velocity—possibly saving lives when seconds count.

VIII. FUTURE ENHANCEMENTS

Although our current multimodal system—merging voice and text—has performed well in emotion detection, emotional AI is an evolving area. In order to further improve the power, inclusiveness, and preparedness of the system for broader applications in real-world situations, the following can be improved in future work.

A. Enabling Facial Expression Recognition

Although the tone of voice and words used say a great deal about how someone feels, facial expressions have the ability to convey emotions that verbal speech cannot. The faintest expression in the form of a smirk, raised eyebrow, or tightened lips may express underlying feelings that may be concealed through spoken or written text. Adding facial expression recognition with computer vision and deep learning to the current system will allow us to create a tri-modal approach that simultaneously processes text, speech, and facial expressions. This would be particularly effective in cases where emotional intent is hard to figure out or remains ambiguous. This feature could be a tremendous boon to teletherapy, video conferencing, online education, and robotics applications, where visual input is involved and real-time emotion understanding is of significance.

B. Multicultural and Multilingual Flexibility

Although emotions are global, one's expression of them is both culture- and language-specific. Tone, idioms, and words will express different levels of emotions or sometimes different emotions across languages. A seemingly objective-sounding sentence in one language can convey frustration in another. To make the system workable for the world, the future models would have to be trained on multilingual and culturally rich corpora, including both large languages (English, Spanish, Mandarin) and low-resource or regional languages. Additionally, cultural emotion mapping included would allow the model to make its interpretation based on cultural norms so that there is minimal likelihood of misinterpretation and bias in international uses.

C. Temporal Emotion Tracking

Real human emotions are dynamic—they evolve and shift during a dialogue. Someone can begin a conversation tranquil but become angered, or begin depressed and escalate to hope. Existing methods process emotions in isolated chunks, with no record of the emotional path over time. Future systems will also include temporal emotion modelling, keeping tabs on the emotion progression over the entire interaction and not input by input. This would create a fuller and more accurate emotional portrait, which could be invaluable for long-term therapy, educational return systems, and AI friends that tailor responses to the user's emotional trajectory.

D. Light, Mobile-Optimized Applications

To make emotion recognition ubiquitous, it has to run on regular devices—not servers with high power. Having an offline-enabled, low-power version of the system available to operate on mobile devices would enable the technology to be presented to users in low-connectivity regions, offer privacy for sensitive applications like tracking mental health, and offer mobile professionals real-time response. These could involve reducing the size of the AI model or using distilled variants which are still accurate but use fewer resources. Some possible uses are emotion-sensitive journaling programs, voice diaries, and real-time mood tracking by conversation, so the user can have an emotionally intelligent assistant in their pocket.

IX. CONCLUSION

Human interaction is far more complex than the words we speak or type — it's a dense blend of tone, silence, facial expression, and implicit emotional cues. This research set out to bridge the chasm between the kinds of things machines can sense and the kinds of things humans actually experience. Combining Natural Language Processing (NLP) for text analysis with acoustic properties of speech, we created a multimodal system that could identify and read emotions with greater accuracy and subtlety. The most surprising thing about our results is this pattern: when both voice and text are analyzed, the system is more reliable at recognizing emotions than when analysed alone. This is because certain emotions might be hidden within word choice, and other emotions only with pitch, rate, or volume. Blending these perspectives provides the system with a richer emotional context so that it may respond in a way that sounds concerned and sensitive. The implications of this work extend beyond the laboratory. In psychiatric care, such a system could identify emotional distress early on and deliver interventions early enough. In education, it could adjust classrooms to accommodate the emotional needs of the students. In customer support, it could de-escalate frustration before it reaches anger. Most importantly, it allows for the introduction of AI systems that interact not as cool data processors, but as acquaintances that are attuned to — and appropriately respond to — the human experience.

While results are promising, this is just the beginning. Emotions have many dimensions, are complex and subject to individual, social, and cultural variables. Future steps might include visual features like faces, extend to multilingual contexts, and track changes across time in emotions to build truly adaptive, empathetic systems. Fundamentally, this book is a step toward giving machines not merely the ability to "hear" and "read" us, but to listen — to the heart behind the words — and respond in such a way as is natural, human, and responsive to the feelings.

REFERENCES

[1] A. D. L. Languré and M. Zareei, "Breaking Barriers in Sentiment Analysis and Text Emotion Detection: Toward a Unified Assessment Framework," in IEEE Access, vol. 11, pp. 125698-125715, 2023, doi: 10.1109/ACCESS.2023.3331323

[2] A. Naz, H. U. Khan, S. Alesawi, O. Ibrahim Abouola, A. Daud and M. Ramzan, "AI Knows You: Deep Learning Model for Prediction of Extroversion Personality Trait," in IEEE Access, vol. 12, pp. 159152-159175, 2024, doi: 10.1109/ACCESS.2024.3486578

[3] D. Javed, N. Z. Jhanjhi, N. A. Khan, S. K. Ray, A. Al-Dhaqm and V. R. Kebande, "Identification of Spambots and Fake Followers on Social Network via Interpretable AI-Based Machine Learning," in IEEE Access, vol. 13, pp. 52246-52259, 2025, doi: 10.1109/ACCESS.2025.3551993

[4] J. Polisena, M. Andellini, P. Salerno, S. Borsci, L. Pecchia and E. Iadanza, "Case Studies on the Use of Sentiment Analysis to Assess the Effectiveness and Safety of Health Technologies: A Scoping Review," in IEEE Access, vol. 9, pp. 66043-66051, 2021, doi: 10.1109/ACCESS.2021.3076356

[5] K. Alemerien, A. Al-Ghareeb and M. Z. Alksasbeh, "Sentiment Analysis of Online Reviews: A Machine Learning Based Approach with TF-IDF Vectorization," in Journal of Mobile Multimedia, vol. 20, no. 5, pp. 1089-1116, September 2024, doi: 10.13052/jmm1550-4646.2055

[6] K. T. Mursi, M. D. Alahmadi, F. S. Alsubaei and A. S. Alghamdi, "Detecting Islamic Radicalism Arabic Tweets Using Natural Language Processing," in IEEE Access, vol. 10, pp. 72526-72534, 2022, doi: 10.1109/ACCESS.2022.3188688

[7] M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. AlSobeh, K. Mansour and H. Migdady, "Emo-SL Framework: Emoji Sentiment Lexicon Using Text-Based Features and Machine Learning for Sentiment Analysis," in IEEE Access, vol. 12, pp. 81793-81812, 2024, doi: 10.1109/ACCESS.2024.3382836

[8] O. Oyebode, R. Lomotey and R. Orji, "“I Tried to Breastfeed but…”: Exploring Factors Influencing Breastfeeding Behaviours Based on Tweets Using Machine Learning and Thematic Analysis," in IEEE Access, vol. 9, pp. 61074-61089, 2021, doi: **10.1109/ACCESS.2021.3073079**

[9] O. Berjawi, G. Fenza and V. Loia, "A Comprehensive Survey of Detection and Prevention Approaches for Online Radicalization: Identifying Gaps and Future Directions," in IEEE Access, vol. 11, pp. 120463-120491, 2023, doi: 10.1109/ACCESS.2023.3326995

[10] P. Thiengburanathum and P. Charoenkwan, "SETAR: Stacking Ensemble Learning for Thai Sentiment Analysis Using RoBERTa and Hybrid Feature Representation," in IEEE Access, vol. 11, pp. 92822-92837, 2023, doi:10.1109/ACCESS.2023.3308951

[11] P. Durga and D. Godavarthi, "Deep-Sentiment: An Effective Deep Sentiment Analysis Using a Decision-Based Recurrent Neural Network (D-RNN)," in IEEE Access, vol. 11, pp. 108433-108447, 2023, doi:10.1109/ACCESS.2023.3320738

[12] S. Kumi, C. Snow, R. K. Lomotey and R. Deters, "Uncovering Concerns of Citizens Through Machine Learning and Social Network Sentiment Analysis," in IEEE Access, vol. 12, pp. 94885-94913, 2024, doi:10.1109/ACCESS.2024.3426329

[13] S. S. Roy, A. Roy, P. Samui, M. Gandomi and A. H. Gandomi, "Hateful Sentiment Detection in Real-Time Tweets: An LSTM-Based Comparative Approach," in IEEE Transactions on Computational Social Systems, vol. 11, no. 4, pp. 5028-5037, Aug. 2024, doi: 10.1109/TCSS.2023.3260217

[14] W. B. Tahir, S. Khalid, S. Almutairi, M. Abohashrh, S. A. Memon and J. Khan, "Depression Detection in Social Media: A Comprehensive Review of Machine Learning and Deep Learning Techniques," in IEEE Access, vol. 13, pp. 12789-12818, 2025, doi: 10.1109/ACCESS.2025.3530862

[15] Z. Doughan, S. Itani and S. Itani, "ArabSis: Arabic Corpus Sentiment Analysis," in IEEE Access, vol. 13, pp. 81083-81095, 2025, doi: 10.1109/ACCESS.2025.3567755