Abstract

Sentiment analysis is a vital area of research with applications in various fields such as security, modern healthcare, and business strategy. It plays a key role in addressing serious issues, for example, in helping to identify people with suicidal tendencies, enabling timely intervention and support. It also helps to detect and mitigate the effects of cyberbullying, where individuals can lose social standing due to negative online interaction. In addition, sentiment analysis contributes to improving the quality of products, food, and services by analyzing customer feedback and insights. In today's world, its influence is especially significant in the modern healthcare system, where it supports better patient care and good service.

This thesis explores different innovative approaches to analyze pain sentiments among patients. This study incorporates linguistic and behavioral signals captured from textual, audio, and visual datasets. Here, we have developed text-based pain analysis systems using machine learning approaches as well as deep learning approaches. The machine learning approaches are pillared on feature extraction methods like Bag of Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) and classification methods. Deep learning models include Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) architectures to gain deeper insight into the emotional context conveyed in the text.

Text-based systems suffer from some inherent limitations. To overcome these issues, we next design audio-based pain sentiment analysis systems. Vocal signals are rich sources of emotion, picking up subtle changes in pitch, speech rhythm, and instability in the voice that are omitted or hard to detect when written. The feature extraction module of the audio-based system computes statistical features, Mel-frequency cepstral coefficients (MFCCs), and spectral features. Each feature set is initially

classified using traditional machine learning models; subsequently, the features are concatenated, and a deep learning architecture with a fully connected network is implemented for classification.

Despite the significant benefits provided by audio-based pain analysis, it also poses serious challenges, especially in terms of ambiguity in data from various aspects such as due to tone, language barriers, and cultural differences. This fundamental issue prompted us to incorporate facial expression analysis as a complementary modality because facial movements are governed by consistent neuromuscular patterns and are difficult to consciously suppress, offering a more stable reflection of true pain labels. To resolve the issue, we developed image-based systems. From the facial image, features like the Local Binary Pattern (LBP) and the Histogram of Oriented Gradients (HOG) are derived and then classified. On the other hand, we designed four CNN deep architectures.

Image-based analysis has its own limitations, primarily its static nature, which prevents the assessment of temporal dynamics such as the duration and progression of pain expressions. To overcome this limitation, we extend the framework by incorporating video, a dynamic signal. The image-based system using the deep learning approach has much superior performance in comparison to the machine learning approach, so a video-based pain analysis system is designed following the deep architecture only. Here, features are derived using the two well-performing CNN models along with five well-known pretrained models. For classification, we propose two architectures, *PainCapsule* and *AttentionPainCapsule*.

The individual system based on different types of data cannot achieve sufficient accuracy. Finally, we design multimodal systems to fuse different systems. In this work, we apply the post-classification fusion technique and achieve significant improvement.