**Research questions:**

1. What is the overall accuracy rate for emotion recognition from conversational speech, and how does it vary across different studies and populations?
2. What are the most used methods and techniques for emotion recognition from conversational speech, and how effective are they compared to each other?
3. What are the most reliable and valid measures of emotion recognition from conversational speech, and how do they compare to each other?
4. How do different factors (such as speaker gender, culture, language, or emotional state) affect the accuracy of emotion recognition from conversational speech?
5. Are there any consistent patterns or trends in the results of studies on emotion recognition from conversational speech, and what factors may explain any inconsistencies or conflicting results?
6. What are the most promising areas for future research on emotion recognition from conversational speech, based on the findings of existing studies?

**Keywords:**

**Emotion recognition in conversation & speech & physiological signals & video & deep learning &machine learning**

**Datasets:**

1. **IEEE**
2. **Elsevier**
3. **Springer**
4. **Ebsco**
5. **Web of science**
6. **Others**

**Inclusion and exclusion criteria:**

**Inclusion:**

1. **Conversation dataset in any language (Poria et al. 2019)**
2. **Modalities: speech, video, physiological signals (IoT), wearables**
3. **Deep learning and machine learning**
4. **Natural, or acted conversation**
5. **Conversation between humans only**
6. **Years from 2010-2023**
7. **Categorical or dimensional models**

**Exclusion:**

1. **Conversation with an avatar (i.e. Semaine)**
2. **Studies that do not use any deep learning or machine learning**
3. **Older than 2010**
4. **Stress and anxiety studies**
5. **Textual conversation**
6. **Preprints, no peer-reviewed articles (e.g. arXiv)**

**Evaluation parameters (measures-scores):**

**Accuracy, AUROC, F-score, sensitivity, specificity, CCC, precision or recall.**

**Subgroup Analysis:**

1. Language: subgrouping studies based on the language of the conversation (e.g., English, Mandarin, Arabic, etc.).
2. Culture: subgrouping studies based on the cultural context of the conversation (e.g., Western, Eastern, African, etc.).
3. Modality: subgrouping studies based on the modality of the input data (e.g., audio, video, or audiovisual).
4. Methodology: subgrouping studies based on the methodology used for emotion recognition (e.g., machine learning, deep learning, rule-based systems, etc.).
5. Dataset: subgrouping studies based on the dataset used for training and evaluation (e.g., AffectNet, IEMOCAP, MSP-IMPROV, etc.).
6. Emotion categories: subgrouping studies based on the categories of emotions recognized (e.g., basic emotions, complex emotions, or sentiment analysis).