

# Sentiment Analysis Approach in Healthcare Data Assessment

**Objective:** Convert unstructured patient feedback into structured, high-confidence insights including overall sentiment, domain-specific entity extraction, and entity-level sentiment to improve operational and clinical decision-making.

## 1 Core Techniques - click here to access the code : [Github](#)

Hybrid Transformer-Based NLP Architecture:

- BioBERT for medical entity recognition
- Clinical NER for healthcare terminology
- Rule-based entity patterns for domain precision
- BERT multilingual sentiment classifier for overall feedback
- FLAN-T5 for aspect-based sentiment analysis
- Zero-shot classification for adaptive entity detection
- Lexicon fallback for robustness

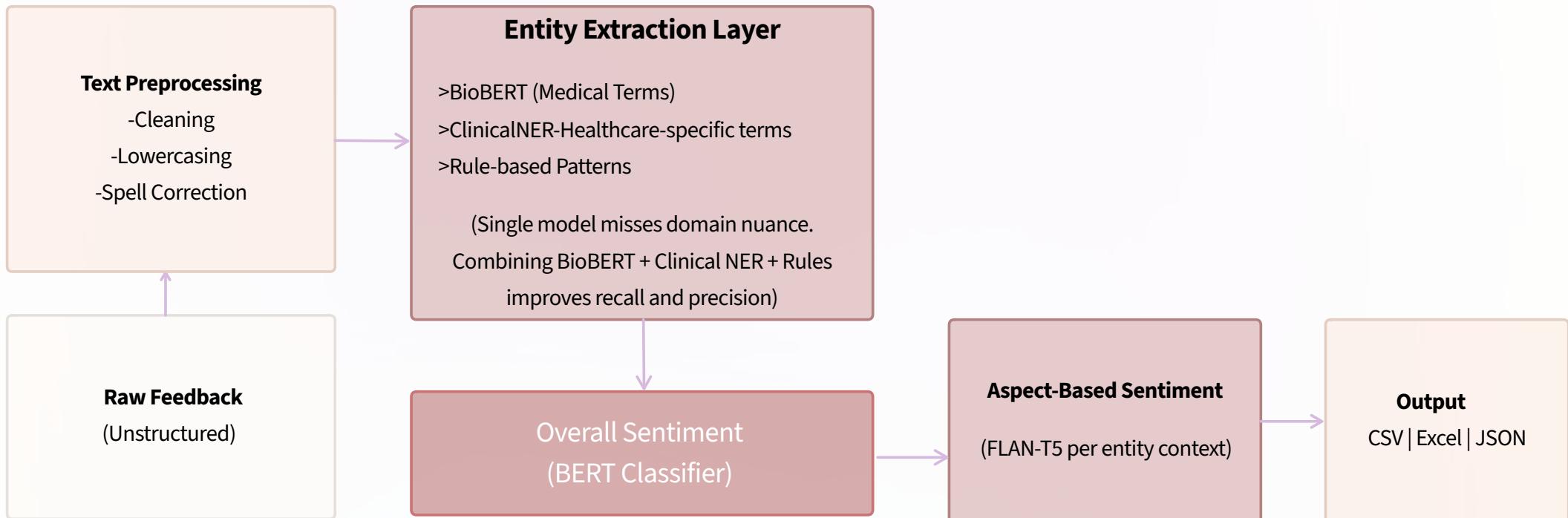
## 2 Entity-level sentiment detection for:

- Doctor
- Nurse
- Facility
- Surgery / Procedure
- Appointment
- Parking

## 3 Key Challenges

- No labeled training dataset
- Mixed sentiment within same feedback
- Healthcare terminology variability
- Misspellings
- Overlapping entity mentions
- Need for confidence-aware outputs

# System Architecture: NLP Pipeline Flow



# Entity Extraction Layer – Hybrid Detection Architecture

## BioBERT – Medical Term Recognition

- **What it does**

Uses BioBERT transformer to detect medical procedures, treatments, and clinical terminology.

- **Why it helps sentiment analysis**

Ensures clinical entities (e.g., “spinal fusion”, “biopsy”) are correctly identified so sentiment can be linked to the right medical context.

- **Input**

Preprocessed feedback text

- **Output**

List of medical entities with type and confidence score

## Clinical NER – Healthcare Entity Detection

- **What it does**

Uses a healthcare-specific NER model to detect structured entities like Doctor, Facility, Procedure, Treatment.

- **Why it helps sentiment analysis**

Separates operational entities (e.g., facility, staff) from clinical terms, enabling aspect-level sentiment analysis.

- **Input**

Preprocessed feedback text

- **Output**

Structured healthcare entities with entity type labels

## Rule-Based Pattern Layer – Precision Capture

- **What it does**

Applies regex and keyword patterns to detect domain-specific phrases such as “Dr. Name”, “appointment rescheduled”, “parking was difficult”.

- **Why it helps sentiment analysis**

Captures short operational complaints that transformer models may miss, improving recall and interpretability.

- **Input**

Preprocessed feedback text

- **Output**

High-precision domain entities

## Entity Merging & Deduplication (Ensemble Layer)

- **BioBERT, Clinical NER, and Rule-Based detectors run independently in parallel** on the same input text.

Their outputs are aggregated in this layer.

- **What it does**

Combines entities from all detectors and Removes duplicates and overlaps by Normalizing entity names and Assigning confidence scores

- **Why it matters:**

Improves recall through ensemble detection Reduces redundancy before sentiment scoring

- **Input**

Entity lists from three independent detectors

- **Output**

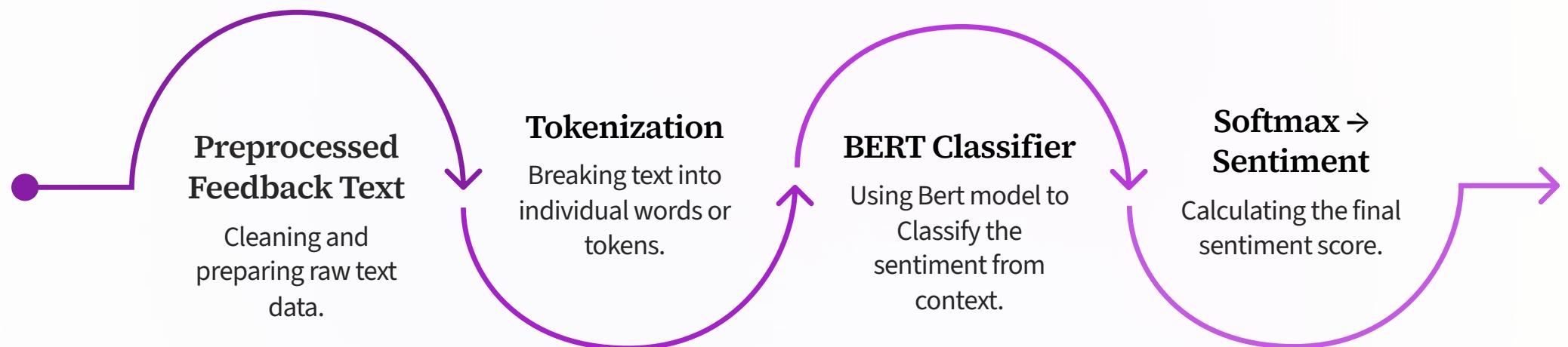
Unified, deduplicated, confidence-aware entity set

# Overall Sentiment Layer (Transformer-Based)

Uses a pretrained BERT-based sequence classification model to analyze the full feedback text and predict overall sentiment.

**What happens here:**

Preprocessed text → Tokenization → BERT classifier → Softmax probabilities → Sentiment label (Positive / Neutral / Negative) with confidence score



**Why it matters**

- Captures global emotional tone of the feedback
- Handles contextual language (e.g., “handled professionally despite delays”)
- Provides probability-based confidence score

**Input :** Preprocessed full feedback text

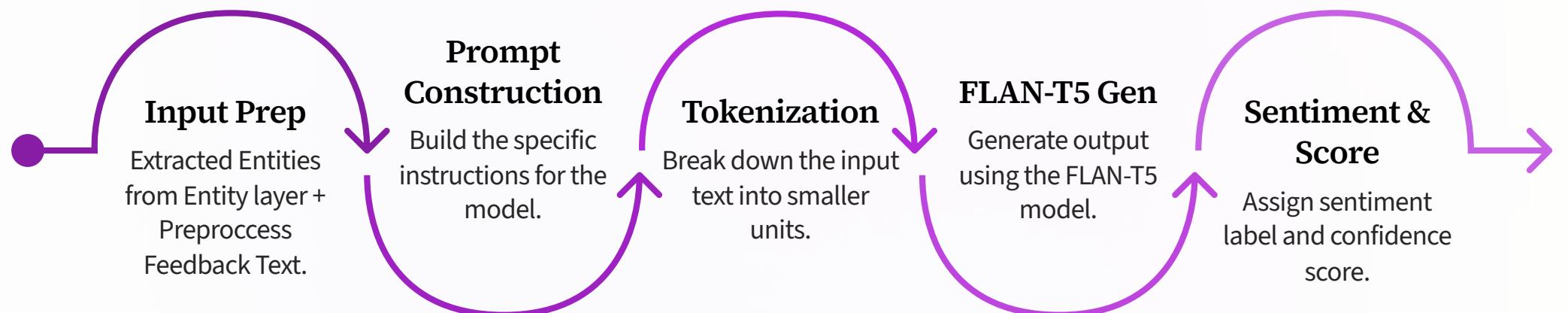
**Output :** Overall sentiment label + confidence score

# Aspect-Based Sentiment Layer(FLAN-T5 Prompt-Based)

Uses a pretrained sequence-to-sequence transformer (FLAN-T5) to assign sentiment to each extracted entity.

**What happens here:**

**Extracted Entities (from Entity Extraction & Merging Layer) + Preprocessed Feedback Text** → **Prompt Construction** → **Tokenization** → **FLAN-T5 Sequence Generation** → **Sentiment Label (Positive / Neutral / Negative)** → **Confidence Score**



**Example Prompt:** “Analyze the sentiment about Doctor ‘Dr. Riya Patel’ in the following text...”

**Why it matters:**

- Enables fine-grained entity-level sentiment
- Handles mixed feedback within a single comment
- Works without labeled aspect-based dataset

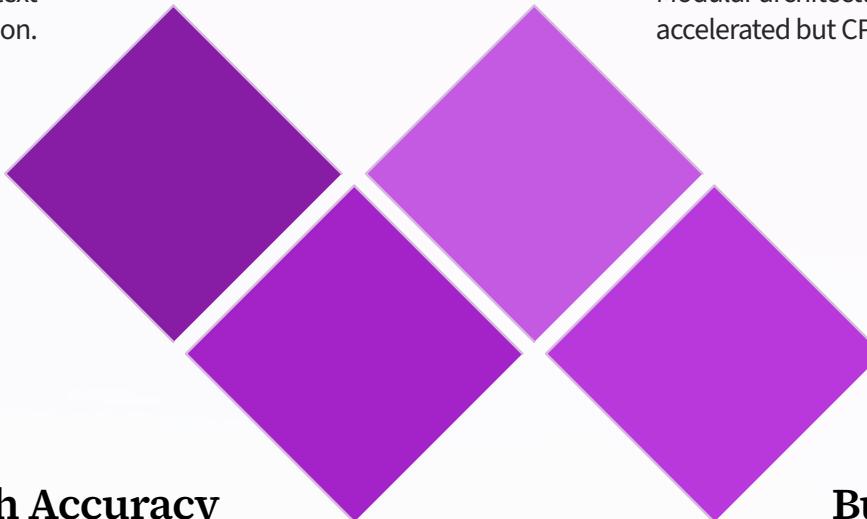
**Output:** Entity-level sentiment label + confidence score

# Production-Ready & Robust Architecture

## Robustness

Handles Real-World Variability:

Supports variable input lengths and text normalization.



## High Accuracy

Intelligent & Reliable:

Ensemble entity extraction combining transformers and rule-based logic.

## Operational Ease

Deployment Friendly:

Modular architecture that is GPU-accelerated but CPU-compatible.

## Business Value

Business-Ready Output:

Generates normalized entities with confidence metrics for monitoring.