

## Project Description

- To develop an automated text-to-SQL system for querying Electronic Health Records (EHR) databases using natural language.
- To enable clinicians without SQL expertise to retrieve patient information efficiently through intuitive questioning.
- To bridge the gap between clinical expertise and database management in healthcare settings.

## Objectives

- To convert natural language queries into accurate SQL statements for EHR databases.
- To handle complex multi-table queries involving patient demographics, diagnoses, prescriptions, and lab results.
- To achieve high query accuracy and minimize errors in SQL generation.
- To implement robust error handling mechanisms for unanswerable or ambiguous queries.

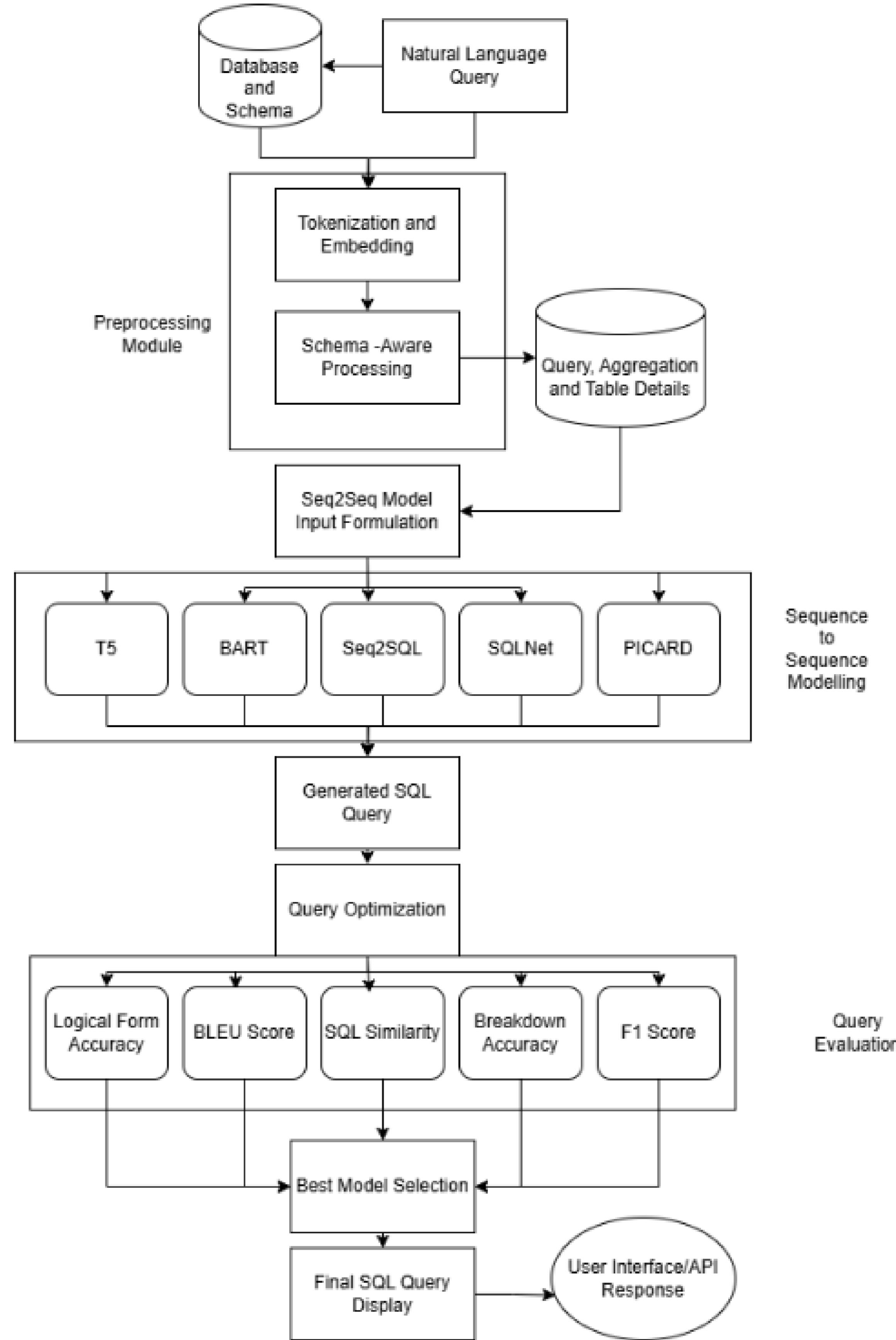
## Novelty

- Integration of constrained decoding techniques to ensure syntactically valid and executable SQL queries.
- Schema-aware processing that adapts to complex medical database structures with multiple relational tables.

## Deliverables

- Web-based user interface for intuitive query input and result visualization.
- Comprehensive evaluation framework with multiple metrics (BLEU, F1-Score, Component Accuracy).

## System Architecture



## Results

Metric	T5	BART	Seq2SQL	SQLNet	PICARD
Logical Form Accuracy (%)	72.4	74.8	68.2	70.1	<b>79.3</b>
BLEU Score	0.54	0.56	0.50	0.52	<b>0.59</b>
SQL Similarity Score	0.74	0.76	0.70	0.72	<b>0.80</b>
Aggregation Operator Accuracy (%)	69.5	71.2	65.3	67.8	<b>74.6</b>
Aggregation Column Accuracy (%)	66.8	69.1	62.7	65.2	<b>72.5</b>
Table Selection Accuracy (%)	75.1	77.3	71.9	74.0	<b>81.4</b>
Condition Column and Operator Accuracy (%)	64.2	66.7	60.4	62.9	<b>70.1</b>
Condition Value Accuracy (%)	61.5	64.0	58.3	60.8	<b>68.5</b>
Average Breakdown Accuracy (%)	67.2	69.8	63.8	66.2	<b>73.0</b>
Precision	0.68	0.70	0.64	0.66	<b>0.74</b>
Recall	0.65	0.68	0.62	0.64	<b>0.71</b>
F1-Score	0.67	0.69	0.63	0.65	<b>0.73</b>

## Future Development

- Integration with real-time EHR systems (Epic, Cerner, Allscripts) through standardized API interfaces.
- Expansion to support voice-based queries for hands-free clinical workflows.

## Outcome of the Project Impact on Industry/Society

- Increased Efficiency:** Reduces query formulation time from minutes to seconds, allowing clinicians to focus on patient care rather than database navigation.
- Democratized Data Access:** Enables nurses, residents, and non-technical staff to access critical patient information without SQL training.
- Cost Reduction:** Eliminates need for dedicated database specialists for routine queries.
- Improved Clinical Decision-Making:** Faster access to patient histories, lab results, and treatment outcomes leads to more informed clinical decisions.
- Enhanced Patient Safety:** Quick access to medication histories, allergies, and past diagnoses reduces risk of adverse events and medical errors.

## Conclusion and Outcomes

- Successfully developed an automated text-to-SQL system achieving 79.3% accuracy on medical EHR queries using the PICARD architecture.
- Demonstrated significant performance improvements through schema-aware processing and constrained decoding techniques.
- Comprehensive evaluation across five models established PICARD as the optimal choice for medical database querying.