RAG-Based Breast Cancer Nursing Care

Technical Overview

Version: 1.0
Author: Deeksha P
Date: May 21, 2025
Classification: Public Documentation

Executive Summary

This document provides a technical overview of the Retrieval-Augmented Generation (RAG) based Question and Answer system developed at the Quality of Life Technology Laboratory to support nursing staff in providing care for breast cancer patients. The system combines the knowledge retrieval capabilities of advanced search algorithms with the contextual understanding of large language models to provide accurate, evidence-based responses to nursing queries in real-time.

---

1. Introduction

1.1 Purpose

The RAG-Based Nursing Q&A System addresses the challenge of providing consistent, evidence-based information to nursing staff caring for breast cancer patients. By integrating with established clinical guidelines and peer-reviewed research, the system serves as an on-demand knowledge resource that:

- Reduces information retrieval time for busy nursing staff
- Ensures consistent, evidence-based responses to common questions
- Provides citations to source materials for verification
- Adapts to individual patient contexts while maintaining clinical accuracy

1.2 System Overview

The system utilizes a Retrieval-Augmented Generation (RAG) architecture that combines:

1. **Knowledge Base:** A curated collection of breast cancer care guidelines, protocols, and peer-reviewed research
2. **Retrieval System:** Advanced search algorithms that identify relevant information from the knowledge base
3. **Generation Model:** A fine-tuned large language model that synthesizes retrieved information into coherent, contextually appropriate responses
4. **Clinical Verification Module:** A rule-based system that ensures all responses comply with established clinical guidelines

Figure 1: High-level architecture of the RAG-Based Nursing Q&A System

---

2. Technical Architecture

2.1 Knowledge Base

The knowledge base consists of:

Content Type	Source	Update Frequency	Format
Clinical Guidelines	National Comprehensive Cancer Network (NCCN)	Quarterly	JSON
Medication Information	FDA and pharmaceutical databases	Monthly	Structured XML
Research Literature	PubMed and medical journals	Weekly	Vector embeddings
Hospital Protocols	Internal hospital documentation	As needed	Markdown

The knowledge base is stored in a secure MongoDB database with the following schema structure:

{

"document_id": "unique_identifier",

"document_type": "guideline|medication|research|protocol",

```
 "title": "Document title",
 "content": "Full document text",
 "metadata": {
  "source": "Source information",
  "publication_date": "YYYY-MM-DD",
  "last_updated": "YYYY-MM-DD",
  "verification_status": "verified|pending"
 },
 "vector_embedding": [0.23, 0.45, ...],
 "sections": [
  {
   "section_id": "section_identifier",
   "heading": "Section heading",
   "content": "Section content",
   "vector_embedding": [0.21, 0.42, ...]
  }
 ]
}
```

2.2 Retrieval System

The retrieval system uses a hybrid approach combining:

1. **Dense Retrieval:** Document embeddings generated using MedicalBERT to capture semantic relationships
2. **Sparse Retrieval:** BM25 algorithm for keyword matching to ensure relevant clinical terms are prioritized
3. **Metadata Filtering:** Allows filtering by document type, recency, and verification status

Key metrics for the retrieval system:

- Mean Reciprocal Rank (MRR): 0.87
- Recall@5: 0.92
- Precision@5: 0.86

2.3 Generation Model

The generation component utilizes a fine-tuned domain-specific large language model trained on:

- 25,000 nursing question-answer pairs related to breast cancer care
- Specialized medical terminology and breast cancer treatment vocabularies
- Clinical reasoning patterns from experienced oncology nurses

The model was fine-tuned using Reinforcement Learning from Human Feedback (RLHF) with oncology nursing specialists providing the evaluation criteria.

2.4 Clinical Verification Module

All generated responses pass through a verification module that:

1. Checks for contraindications or potentially dangerous advice
2. Ensures all medication dosages match established guidelines
3. Adds appropriate disclaimers for experimental treatments
4. Flags responses that require human verification based on question sensitivity

---

3. User Interface and Workflow

3.1 Interface Design

The system is accessible through:

- Web interface integrated with the hospital's Electronic Health Record (EHR) system
- Mobile application for on-the-go access
- Integration with nursing communication systems

Figure 2: Web interface of the RAG-Based Nursing Q&A System

3.2 Typical Workflow

1. Nurse enters a question in natural language (e.g., "What are the recommended interventions for managing radiation-induced skin reactions for a 65-year-old patient with diabetes?")

2. System identifies key clinical parameters from the question (treatment: radiation, complication: skin reactions, comorbidity: diabetes, age: 65)
3. Retrieval system pulls relevant guidelines and research considering all parameters
4. Generation model synthesizes information into a concise, actionable response
5. Verification module ensures the response accounts for diabetes as a complication factor
6. Response is presented with citations and confidence level
7. Nurse can request more details, simplification, or alternative approaches

4. Performance and Evaluation

4.1 Response Metrics

Metric	Target	Current Performance
Response Time	<3 seconds	2.4 seconds (average)
Accuracy	>90%	94.2%
Citation Accuracy	100%	100%
Nurse Satisfaction	>4.5/5	4.7/5

4.2 Evaluation Methodology

The system undergoes continuous evaluation through:

1. **Automated Testing:** Daily comparison against a test set of 1,000 validated Q&A pairs
2. **Clinical Review:** Weekly random sampling of 50 responses reviewed by oncology specialists
3. **User Feedback:** In-app rating and feedback system after each interaction
4. **Quarterly Audit:** Comprehensive review of system performance and accuracy

5. Security and Compliance

5.1 Data Protection

The system implements:

- End-to-end encryption for all data in transit
- HIPAA-compliant data storage and access controls
- De-identification of patient information in logging and analytics
- Role-based access controls for system maintenance and updates

5.2 Regulatory Compliance

The system maintains compliance with:

- HIPAA and HITECH Act requirements
- FDA guidelines for Clinical Decision Support (CDS) systems
- Joint Commission standards for knowledge resources
- Hospital-specific information security policies

---

6. Future Development Roadmap

Phase	Timeline	Key Features
2.0	Q3 2025	Integration with patient-specific treatment plans
2.5	Q1 2026	Multi-modal input (including image analysis of skin conditions)
3.0	Q3 2026	Extension to additional cancer types
3.5	Q1 2027	Predictive alerting for potential complications

---

7. Appendix

7.1 Technical Integration Specifications

Integration with existing hospital systems is achieved through:

```
# Example API call structure for EHR integration
import requests


def query_system(patient_context, question):
```

```python
    """
    Query the RAG-Based Nursing Q&A System with patient context

    Parameters:
    - patient_context: Dict containing relevant patient information
    - question: String containing the nurse's question

    Returns:
    - Dict containing the system response and metadata
    """
    api_endpoint = "https://api.hospital-domain.org/rag-qa/query"

    payload = {
        "patient": {
            "age": patient_context.get("age"),
            "diagnosis": patient_context.get("diagnosis"),
            "comorbidities": patient_context.get("comorbidities"),
            "current_treatments": patient_context.get("current_treatments"),
            "allergies": patient_context.get("allergies")
        },
        "question": question,
        "response_type": "detailed",  # Options: "brief", "detailed", "protocol"
        "include_citations": True
    }

    response = requests.post(
        api_endpoint,
        json=payload,
        headers={
            "Authorization": f"Bearer {get_auth_token()}",
```

```
        "Content-Type": "application/json"
    }
)


return response.json()
```

7.2 Glossary of Technical Terms

Term	Definition
RAG	Retrieval-Augmented Generation, a hybrid AI approach that combines information retrieval with text generation
Vector Embedding	A numerical representation of text that captures semantic meaning in a high-dimensional space
RLHF	Reinforcement Learning from Human Feedback, a technique for aligning AI models with human preferences
Dense Retrieval	Information retrieval method using semantic understanding rather than keyword matching
BM25	A bag-of-words retrieval function used for sparse retrieval based on term frequency

8. References

1. Zhang et al. (2024). "Retrieval-Augmented Generation for Medical Question Answering." Journal of Medical AI Applications, 12(3), 234-251.
2. National Comprehensive Cancer Network. (2024). "Clinical Practice Guidelines in Oncology: Breast Cancer." Version 2.2024.
3. Johnson, K. & Smith, P. (2023). "Improving Clinical Decision Support in Oncology Nursing." Oncology Nursing Forum, 50(2), 112-128.
4. World Health Organization. (2024). "WHO Guidelines for Breast Cancer Treatment and Management."

For questions regarding this documentation, please contact: research@qualityoflifetech.edu

End of Document