

## ### In-Depth Report for ScanAI RAG Agent Integration: CeliScan Diagnostic Tool

### #### Overview

CeliScan is a cutting-edge diagnostic tool designed to revolutionize the identification and management of celiac disease, an autoimmune disorder affecting approximately 1 in 100 individuals globally. Despite its prevalence, celiac disease remains underdiagnosed, with up to 83% of cases unrecognized or misdiagnosed due to limitations in traditional diagnostic methods. CeliScan leverages a multi-modal machine learning (ML) framework to provide a highly accurate, efficient, and accessible diagnostic solution. This report outlines the tool's functionality, technical specifications, and integration considerations for the Retrieval-Augmented Generation (RAG) agent, ScanAI, to be embedded within the CeliScan app.

---

### ### Purpose and Objectives

CeliScan addresses the shortcomings of conventional diagnostic approaches—such as serological tests, genetic screening, and endoscopic biopsies—which suffer from false negatives, dependency on gluten exposure, high costs, and patient discomfort. The tool aims to:

1. Enhance diagnostic accuracy through multi-modal data integration.
2. Reduce reliance on invasive procedures.
3. Improve early detection rates.
4. Lower healthcare costs and improve patient comfort.
5. Provide an accessible, user-friendly interface for both patients and healthcare providers.

---

### ### Technical Foundation

#### #### Machine Learning Model

CeliScan is powered by a Support Vector Machine (SVM) algorithm, selected for its superior performance in a comparative analysis of various ML models. The model was trained and validated on a dataset comprising 2,206 patient records, achieving:

- **Diagnostic Accuracy**: 97.51%
- **False Negative Rate**: 0%
- **Outperformance**: Significantly surpasses traditional single-method diagnostics.

#### #### Training Data

The SVM model integrates diverse data modalities, including:

1. **Serological Markers**: Levels of anti-tissue transglutaminase (tTG-IgA), anti-endomysial antibodies (EMA), and deamidated gliadin peptide (DGP) antibodies.
2. **Genetic Information**: Presence of HLA-DQ2 and HLA-DQ8 haplotypes, which are strongly associated with celiac disease.
3. **Clinical Symptoms**: Patient-reported symptoms such as chronic diarrhea, abdominal pain, bloating, fatigue, and unexplained weight loss.

4. **Anthropometric Measurements**: Metrics like short stature, body mass index (BMI), and body composition changes.
5. **Additional Contextual Data**: Patient history, dietary patterns, and family history of autoimmune disorders.

This multi-modal approach mitigates the limitations of individual tests (e.g., false negatives in serology due to gluten-free diets) by synthesizing a comprehensive diagnostic profile.

---

### ### Functionality of CeliScan

#### #### User Interface

CeliScan offers a dual-purpose interface tailored to:

- **Patients**: Self-assessment by inputting symptoms, dietary habits, and optional test results.
- **Healthcare Providers**: Detailed diagnostic input including lab results, genetic data, and clinical observations.

#### #### Input Modalities

The tool accepts a variety of data types across multiple categories:

1. **Text-Based Inputs**:
  - Symptom descriptions (e.g., frequency and severity of diarrhea or fatigue).
  - Medical history (e.g., prior diagnoses, family history).
  - Dietary information (e.g., gluten exposure duration).
2. **Numerical Data**:
  - Serological test results (e.g., tTG-IgA levels in U/mL).
  - Anthropometric measurements (e.g., height, weight, BMI).
3. **Categorical Data**:
  - Genetic screening results (e.g., HLA-DQ2 positive/negative).
  - Binary symptom presence (e.g., yes/no for abdominal pain).
4. **Optional Uploads**:
  - PDFs or text files of lab reports.
  - Images of physical symptoms (e.g., dermatitis herpetiformis rashes, if applicable).

#### #### Processing Workflow

1. **Data Collection**: Users input data via a structured form or upload relevant files.
2. **Preprocessing**: The system standardizes and normalizes inputs (e.g., converting units, encoding categorical variables).
3. **Model Inference**: The SVM processes the multi-modal data to generate a diagnostic prediction.
4. **Output Generation**: A detailed report is produced, including:
  - **Diagnosis**: Positive or negative for celiac disease.
  - **Certainty Score**: A percentage indicating the likelihood of the result's accuracy (e.g., 92.3% certainty).

- **Supporting Evidence**: Breakdown of contributing factors (e.g., elevated tTG-IgA, presence of HLA-DQ2).

#### #### Output Report

The diagnostic report is concise yet comprehensive, designed for both patient understanding and clinical use:

- **Header**: Patient ID (if applicable), date, and tool version.
- **Diagnosis**: “Positive for Celiac Disease” or “Negative for Celiac Disease.”
- **Certainty Score**: Expressed as a percentage (e.g., “97.8% likelihood of accuracy”).
- **Key Indicators**: List of inputs most influential to the diagnosis (e.g., “High tTG-IgA: 45 U/mL,” “Chronic diarrhea reported”).
- **Recommendations**: General guidance (e.g., “Consult a gastroenterologist for confirmatory biopsy if positive” or “Monitor symptoms if negative with low certainty”).

---

#### #### Integration with ScanAI (RAG Agent)

ScanAI, as a Retrieval-Augmented Generation agent, will enhance CeliScan by providing contextual explanations, answering user queries, and retrieving relevant information based on the diagnostic process. Below are integration specifications:

##### ##### Role of ScanAI

1. **Patient Education**:
  - Explain diagnostic results in layman’s terms (e.g., “A 95% certainty score means the tool is highly confident you may have celiac disease”).
  - Provide information on celiac disease symptoms, management, and gluten-free diets.
2. **Clinical Support**:
  - Offer detailed breakdowns of how inputs influenced the diagnosis for healthcare providers.
  - Suggest follow-up tests or procedures based on the certainty score and input data.
3. **Interactive Queries**:
  - Respond to user questions like “Why does my genetic result matter?” or “What does a false negative mean?”
  - Retrieve external resources (e.g., latest research or guidelines) if permitted by app settings.

##### ##### Data Flow to ScanAI

- **Input Sharing**: ScanAI receives the full diagnostic report, including raw inputs, processed data, and output (diagnosis + certainty score).
- **Memory Utilization**: Leverage xAI’s memory feature to recall prior user interactions (e.g., “Last time, you reported diarrhea but no genetic data—adding it increased certainty from 85% to 95%”).
- **Real-Time Processing**: ScanAI processes queries in parallel with CeliScan’s diagnostic output, ensuring seamless user experience.

##### ##### Customization for ScanAI

- **Tone and Style**: Adopt a supportive, clear, and empathetic tone for patients; a precise, technical tone for providers.
- **Guardrails**: Avoid speculative diagnoses beyond CeliScan's output (e.g., "I can't diagnose other conditions, but here's what CeliScan found").
- **Image Handling**: If users upload symptom images, ScanAI can describe them contextually (e.g., "This rash resembles dermatitis herpetiformis, a celiac-related condition") but should not generate or edit images unless explicitly requested and confirmed.

#### #### Example Interaction

**User**: "Why did I get a positive result with 92% certainty?"

**ScanAI**: "CeliScan analyzed your inputs: elevated tTG-IgA (38 U/mL), HLA-DQ2 positivity, and chronic bloating. These strongly suggest celiac disease. The 92% certainty reflects the model's confidence based on 2,206 patient records it was trained on. I recommend discussing a biopsy with your doctor to confirm."

---

#### ### Benefits and Impact

- **Accuracy**: 97.51% diagnostic precision with no false negatives, outperforming traditional methods.
- **Accessibility**: Non-invasive and cost-effective, usable in diverse settings (e.g., rural clinics, home use).
- **Scalability**: Multi-modal framework can adapt to additional data types or diseases with retraining.
- **Patient Empowerment**: Equips users with actionable insights and reduces diagnostic delays.

---

#### ### Considerations for ScanAI Integration

1. **Privacy**: Ensure compliance with HIPAA or equivalent standards; anonymize data shared with ScanAI.
2. **Memory Management**: Allow users to delete prior diagnostic sessions via the app's "Data Controls" (aligned with xAI's memory feature).
3. **Error Handling**: If inputs are incomplete, ScanAI should prompt users (e.g., "Adding genetic data could improve accuracy—do you have it?").
4. **Updates**: Sync ScanAI's knowledge base with CeliScan's evolving dataset as new patient records are incorporated.

---

#### ### Conclusion

CeliScan represents a paradigm shift in celiac disease diagnosis, combining advanced ML with user-friendly design. By integrating ScanAI, the tool gains an intelligent, interactive layer that

enhances usability and education. This synergy positions CeliScan as a transformative solution, bridging diagnostic gaps and improving outcomes for millions of undiagnosed individuals worldwide.

---

This report provides ScanAI with a comprehensive blueprint of CeliScan's functionality, enabling seamless integration and optimal performance within the app.