BreastCareAI: An Intelligent Multimodal System for Personalized Breast Cancer Guidance and Medical Image Analysis

Hugo Iglesias Pombo
Dept. of Intelligent Systems Engineering
Universidad Intercontinental de la Empresa (UIE)
A Coruña, Spain
hugo.iglesias.01@uie.edu

***Abstract*—Breast cancer remains one of the most prevalent forms of cancer worldwide, and timely access to personalized, accurate, and understandable medical information remains a persistent challenge for patients. BreastCareAI is an intelligent application that aims to support breast cancer patients by integrating advanced AI tools such as Retrieval-Augmented Generation (RAG), explainable medical imaging segmentation and analysis, and natural language understanding to provide comprehensive guidance and decision support.**

***Index Terms*—Breast cancer, artificial intelligence, RAG, medical imaging, segmentation, NLP, LLM, explainability, conversational AI**

I. Introduction

Breast cancer remains one of the most urgent global health concerns, representing the most frequently diagnosed cancer among women worldwide. In 2022 alone, there were approximately 2.3 million new cases and 670,000 deaths reported, making it responsible for one in every four cancer diagnoses among women globally [1]. Beyond its epidemiological impact, breast cancer imposes a profound emotional and informational burden on patients. Upon diagnosis, individuals often experience an urgent need to understand their condition and treatment options. However, this need frequently clashes with persistent information gaps throughout the care journey, leading to confusion, anxiety, and difficulty in making informed medical decisions.

Existing informational resources—such as brochures, websites, or standalone mobile apps—tend to lack personalization and rarely integrate with the patient's clinical context. Critical services like psychological counseling, medication reminders, and imaging explanations are typically dispersed across multiple platforms, requiring patients to coordinate their own care. For instance, imaging results are usually reported using the BI-RADS standard [2], a system well-understood by clinicians but often incomprehensible to patients without medical training.

In response to these challenges, there is a growing call for integrated, intelligent digital solutions capable of delivering continuous, personalized, and understandable support. Previous efforts, such as the Vik breast cancer chatbot, have demonstrated that conversational AI can enhance patient satisfaction and treatment adherence [3]. However, many of these tools remain narrow in scope and are not integrated with broader aspects of cancer care.

To address this gap, we introduce *BreastCareAI*, an AI-powered platform designed to provide holistic, patient-centric support throughout the breast cancer journey. The system integrates a conversational agent based on LLaMA-family large language models (LLMs), enhanced with retrieval-augmented generation (RAG) techniques to deliver context-aware, evidence-grounded responses tailored to individual user needs. By automatically detecting medical terms, incorporating patient profiles, and generating explainable outputs, *BreastCareAI* bridges the gap between medical complexity and user accessibility.

Complementing its conversational capabilities, the platform includes an AI-driven medical imaging analysis pipeline. Using an attention-based U-Net architecture, the system performs automated segmentation of breast ultrasound images. These segmentations are then interpreted through a rule-based BI-RADS classification engine, producing clinically standardized outputs. To promote transparency, a multimodal vision-language model (LLaMA3-Vision) [4] translates the imaging findings into patient-friendly reports that mirror the explanatory style of a radiologist.

Beyond diagnostic assistance, *BreastCareAI* offers practical care tools such as medication reminders and calendar integration. These features help patients manage appointments, treatments, and daily routines—improving not only adherence but also their overall sense of control and engagement.

The platform's architecture combines several state-of-the-art technologies: LLaMA-based LLMs for dialogue [5], multimodal AI approaches inspired by works such as LLaVA [6], and Whisper [7] for robust voice interaction. It adheres to established medical standards such as BI-RADS [8], ensuring clinical relevance, and places special emphasis on transparency and explainability—key pillars of trustworthy medical AI.

Altogether, this project aims to demonstrate how AI-driven, multimodal platforms can significantly improve patient education, engagement, and diagnostic insight within breast cancer care. Drawing on successful precedents in digital health—such as improved treatment adherence via app-based reminders [9] and the clinical viability of cancer support chatbots—*BreastCareAI* embodies the intersection of artificial intelligence, medical imaging, and patient empowerment.

II. Related Work

A. Conversational AI for Cancer Support

Conversational agents have been explored as a means to support cancer patients by providing information and psychosocial assistance. In oncology, chatbots can offer 24/7 support, answer questions, and help monitor patient needs in a personalized manner [10]. One prominent example is Vik, a chatbot designed for breast cancer patients and survivors. Vik engages users via text-based dialogue, providing medically vetted information about the disease, treatments, side effects, and lifestyle (e.g., nutrition, exercise) in an empathetic tone [11]. Studies on Vik have demonstrated positive outcomes. In a one-year deployment with 4,737 breast cancer patients, Vik exchanged over 130,000 messages per month and was associated with improved medication adherence—patients who interacted more frequently with the chatbot showed higher adherence to therapy regimens. User satisfaction with Vik was reported at approximately 94%, with 88% of patients affirming that the chatbot provided valuable support and helped them manage their treatment. Another trial compared Vik's responses to those of physicians in answering common questions of breast cancer patients. In a randomized non-inferiority study with 142 patients, the chatbot's answers achieved a satisfaction score above the acceptable threshold in 69% of cases, versus 64% for a panel of physicians, confirming that the AI was non-inferior to human doctors in informing patients [12].

Beyond Vik, a variety of conversational AI systems have emerged in cancer care: for example, Watson for Oncology assists clinicians with treatment planning by analyzing medical records [10], while patient-facing apps like Babylon Health and Ada act as symptom checkers for early warning signs [10]. Mental health and counseling chatbots (e.g., Vivobot, Wysa) have also been applied to support cancer patients' emotional well-being [10]. Overall, these conversational agents leverage advances in natural language processing (NLP) to simulate empathetic dialogues and provide tailored guidance. Early-generation healthcare chatbots largely relied on rule-based or retrieval-based NLP with intent classification and entity recognition pipelines [11]. For instance, Vik uses supervised machine learning to classify user intents (questions about symptoms, medications, etc.) and to detect specific entities (e.g., a drug name or time) in the query, which allows it to fetch appropriate responses from a curated knowledge base and even personalize replies to whether the user is a patient or caregiver [11]. This ensures the answers are medically correct and contextually relevant. Such approaches have proven effective for addressing "frequently asked questions" and routine support in oncology. However, they are constrained by the scope of their predefined knowledge and lack the open-ended conversational ability of large language model (LLM) based chatbots. Recent advances in LLMs (e.g., GPT-3/4) promise more fluid and flexible interactions, but they introduce new challenges regarding accuracy and reliability of the information provided. In critical domains like cancer care, an incorrect or misleading answer can have serious consequences. This has led researchers to investigate techniques to ground generative models with trusted medical knowledge, as discussed next.

B. Retrieval-Augmented Generation in Medical QA

Unbounded generative AI models such as GPT-4 have shown an ability to converse naturally on medical topics, yet they often exhibit "hallucinations" (confidently producing factually incorrect or irrelevant statements) [13]. In the context of cancer information, this risk is pronounced: a recent study observed that about one-third of ChatGPT's suggested treatment recommendations for various cancers deviated from established clinical guidelines, with answers varying significantly based on phrasing of the query [10]. To mitigate such issues and improve the relevance and accuracy of AI responses, the medical AI community has embraced Retrieval-Augmented Generation (RAG). RAG is a strategy that combines an LLM's language generation capabilities with an external knowledge retrieval component. Instead of relying solely on static knowledge internal to the model's weights, a RAG system will query external databases or documents (such as medical literature, guidelines, or patient records) for each user question, and provide the retrieved facts as context for the LLM to condition its answer. This approach effectively grounds the model's output in up-to-date, domain-specific information [13].

In the healthcare domain, RAG has been shown to significantly boost factual correctness and reduce hallucinations. For example, Miao et al. developed a RAG-empowered clinical QA assistant tailored to nephrology guidelines, and demonstrated that the system could deliver specialized, guideline-consistent advice for kidney disease management [13]. These examples underscore how retrieval augmentation anchors the generative output in validated sources, yielding answers that are not only more correct but also easier to verify (an important aspect for building trust in AI-driven decision support). An additional benefit is that RAG systems can cite their sources, allowing clinicians or patients to trace back the origin of the information provided (e.g., a link to a journal article or clinical protocol), thus increasing transparency. Given these advantages, our approach for BreastCareAI employs RAG to ensure that any textual responses (for instance, answers to patient questions about breast cancer) are backed by relevant clinical knowledge bases or literature. This helps bridge the gap between open-ended conversational AI and the stringent reliability requirements of medical applications.

C. Deep Learning for Breast Ultrasound Segmentation

Breast ultrasound imaging is a key modality for diagnosing and monitoring breast tumors, but it presents challenges for automated analysis due to issues like speckle noise, heterogeneous textures, and low contrast boundaries. In recent years, deep learning methods—particularly convolutional neural networks (CNNs)—have become the state-of-the-art for segmenting lesions in ultrasound images. The prevailing architecture for medical image segmentation is the U-Net and its many variants. U-Net's encoder-decoder structure with skip connections enables effective localization of structures even with limited data. Building on this, researchers have introduced attention mechanisms into the U-Net to improve focus on the regions of interest (tumors) while suppressing irrelevant background. The seminal Attention U-Net by Oktay et al. adds attention gates that learn to weight the skip-connections' features, essentially filtering for tumor-relevant activations before merging decoder and encoder feature maps [14]. This helps the network concentrate on salient tissue regions (e.g., lesion edges) and ignore confounders, mimicking how a radiologist would "zoom in" on suspicious areas.

Attention U-Net and related designs have proven beneficial for ultrasound segmentation, which often requires distinguishing subtle lesion boundaries from surrounding glandular tissue. For example, Lee et al. incorporated a Channel Attention Module with multiscale pooling into a U-Net, which improved the extraction of both global and local context and yielded more accurate segmentation of breast masses in ultrasound images [14]. More recently, adaptive attention techniques have set new benchmarks. Chen et al. proposed an Adaptive Attention U-Net (AAU-Net) that replaces standard convolutions with a Hybrid Adaptive Attention Module combining channel-wise and spatial self-attention [15]. By dynamically re-weighting features across both feature channels and spatial locations, this model can better handle the variability in tumor appearance (sizes, shapes, textures) and the complexity of ultrasound backgrounds. In evaluations on three public breast ultrasound datasets, AAU-Net outperformed several prior segmentation models, achieving higher Dice similarity scores and more robust generalization across benign and malignant lesions [15].

The use of attention-based architectures has thus significantly advanced the performance of breast lesion segmentation. Other innovations include dual-attention networks that combine different attention mechanisms (e.g., spatial + channel attention) [14], and cascaded or multi-scale approaches that capture both fine detail and global context of the ultrasound images. In our work, we leverage an Attention U-Net backbone for segmenting breast ultrasound images. This choice is motivated by the consistent gains in accuracy reported in the literature when attention mechanisms are used to refine U-Net's focus on tumor regions [14]. Precise segmentation of the tumor area provides downstream benefits: it allows computation of lesion size and shape features, and it can inform subsequent classification modules (for instance, by providing a cropped lesion image or mask as input to a diagnostic classifier). Moreover, having a segmented contour is valuable for explainability, as it can be visualized to clinicians and patients, delineating where the AI believes the tumor lies. This visual interpretability complements the numeric categorization (such as a BI-RADS score) that the system might output, as discussed next.

D. Interpretability of BI-RADS Classification

After localizing a lesion in an ultrasound image, the next step is often to characterize it and assess the likelihood of malignancy. Radiologists do this by applying the BI-RADS (Breast Imaging-Reporting and Data System) lexicon – a standardized set of descriptors and an overall category (1 to 5) indicating level of suspicion. BI-RADS categories guide clinical management (e.g. Category 2 = benign finding, routine follow-up; Category 5 = highly suggestive of malignancy, biopsy recommended). A challenge for AI systems is not only to predict such categories accurately, but also to make the results interpretable and patient-friendly. Simply outputting "BI-RADS 4" may alarm a patient without context, or may not clearly explain to a clinician what imaging features led to that assessment. Recent research has therefore focused on explainable AI approaches that align algorithmic outputs with the BI-RADS framework used by human experts. Zhang et al. introduced BI-RADS-Net, an explainable deep learning model for classifying breast ultrasound lesions with accompanying explanations [16]. The network was trained to perform two tasks simultaneously: (1) predict whether a lesion is benign or malignant, and (2) predict the high-level BI-RADS descriptors that a radiologist would note (such as shape, margin, echo pattern, etc.). By learning these intermediate descriptors, the model's decisions become more transparent – for any given case, BI-RADS-Net can output statements like "the lesion is irregularly shaped with spiculated margins," which corresponds to BI-RADS criteria that increase suspicion. These human-interpretable features are then used to support the model's final malignancy prediction or BI-RADS category. In Zhang's study, this multitask approach yielded improved diagnostic accuracy and produced explanations in familiar radiological terms [16]. In other words, the system's "reasoning" was exposed through attributes that map to the BI-RADS lexicon, enhancing trustworthiness. This aligns with the need in healthcare AI for transparent decision support: clinicians are more likely to trust and adopt an algorithm if it provides understandable justifications for its outputs [16]. For patient-facing explanations, additional care is needed to convey BI-RADS assessments in plain language. Projects exist to translate radiology reports into layperson summaries – for example, explaining that "Category 3" means an "almost certainly benign finding – a very small chance of cancer, and follow-up in six months is recommended as a precaution." An automated system could generate similar easy-to-understand explanations alongside the BI-RADS category, which is a feature we envision for BreastCareAI. By leveraging the intermediate BI-RADS features that our model detects (e.g. "the mass appears oval with a smooth margin" suggesting benign characteristics), we can generate conversational explanations to the user about why a certain category was assigned. This approach draws from both the medical ontology (BI-RADS definitions) and NLP techniques to present the information in a reassuring and comprehendible manner. Thus, interpretability is woven into the system's imaging pipeline: from segmentation masks that highlight where the model is looking, to BI-RADS feature descriptors that clarify what the model sees, and finally verbal or written summaries that tell the patient what it means for them.

This comprehensive literature review underlines the key components and design choices for BreastCareAI. In the following sections, we describe how these insights shape our system's architecture and implementation, detailing the integration of conversational agents, knowledge retrieval, image analysis, and explainability features in a unified framework.

III. Methods

A. Overview of BreastCareAI

BreastCareAI is a comprehensive application designed to provide personalized information and analysis for breast cancer patients, survivors, and the general public. It integrates state-of-the-art AI technologies to address critical challenges in breast cancer care, such as limited access to trustworthy information, difficulty interpreting medical imaging, and treatment regimen management.

The system combines Retrieval-Augmented Generation (RAG), medical image analysis, and natural language processing to deliver evidence-based information, diagnostic assistance, and personalized guidance. This integration enables BreastCareAI to offer holistic support throughout the breast cancer journey—from diagnosis through treatment and into survivorship.

The application's main innovations lie in its modular architecture, which integrates specialized components tailored to the unique demands of breast cancer care:

1) AI-powered conversations using RAG technologies deliver contextually relevant, evidence-based responses tailored to patient needs.
2) Advanced breast ultrasound analysis using segmentation and BI-RADS classification assists in diagnostic interpretation.
3) Interactive scheduling and medication management tools support treatment adherence.
4) Automated detection and explanation of complex medical terminology bridge the gap between clinical language and patient understanding.

BreastCareAI emphasizes privacy, transparency, and ethical AI by ensuring all data processing occurs locally and by including source attribution and medical disclaimers with every response.

B. System Architecture Overview

BreastCareAI follows a modular architecture designed for flexibility, integration, and user-centered design. The complete system architecture is depicted in Appendix A, which illustrates how all modules interact with the Streamlit-based user interface.

The system is composed of the following core components:

1) **Frontend Interface:** A Streamlit-based web interface that presents users with a tabbed navigation structure, enabling access to multiple tools including conversational AI, ultrasound analysis, calendar integration, medication tracking, and a medical glossary.
2) **Document Processing and RAG Engine:** Medical literature is processed through LangChain loaders and chunking strategies, and stored in a FAISS vector database using Ollama-generated embeddings. Relevant content is retrieved and passed to the LLM to generate answers grounded in real-world medical sources.
3) **LLM Integration:** Local inference is performed with Ollama using a specialized 8B LLaMA 3 model, fine-tuned for breast cancer guidance. This ensures data privacy and provides fast, relevant natural language generation.
4) **Image Analysis:** A customized Attention U-Net performs ultrasound image segmentation. Its encoder-decoder structure, augmented with attention gates, enables accurate delineation of lesions and artifacts.
5) **BI-RADS Classification:** A rule-based classifier evaluates segmented lesion features—shape, margins, echogenicity, etc.—to assign BI-RADS categories using a Weighted Rule-Based System (WRBS), offering transparent and interpretable assessments [8].
6) **Explainability Components:** A vision-language module based on LLaMA3.2-Vision generates visual and textual explanations of segmentation results, offering patient-friendly radiology-style summaries.
7) **Supporting Modules:**
   - *Calendar Integration:* Generates medical questions and synchronizes appointments using Google Calendar API.
   - *Medication Reminders:* Automates scheduling and notification of treatment doses.
   - *Glossary and Terminology Detection:* Identifies and explains complex clinical terms in real time.
   - *Voice Interface:* Enables speech-to-text and text-to-speech capabilities for greater accessibility.
   - *Web Scraper Module:* Collects current breast cancer guidelines and research updates from trusted online sources.

C. Conversational AI Engine

Purpose and Role. This component enables users to engage in natural conversations with BreastCareAI, asking questions about breast cancer diagnosis, treatment options, side effects, and emotional support. It serves as the primary interface for information retrieval, providing evidence-based responses tailored to the user's specific situation and knowledge needs.

Evolution from Pattern-Based to Neural Approaches. Our development process began with a rule-based system named MammaELIZA, inspired by the classic ELIZA model. This initial prototype utilized regular expression pattern matching to identify key topics in user queries and respond with pre-programmed information about breast cancer. While MammaELIZA demonstrated satisfactory performance for anticipated questions—particularly for specific queries about symptoms, treatments, and emotional concerns—it exhibited significant limitations in generalization, contextual awareness, and conversation coherence. This led us to transition to a more flexible and knowledge-rich approach using retrieval-augmented generation.

RAG Architecture. The current system implements a sophisticated RAG pipeline built with LangChain that combines the knowledge accuracy of document retrieval with the natural language capabilities of large language models. This architecture addresses several critical requirements for healthcare information systems: factual correctness, contextual awareness, citation of medical sources, and personalization. The implementation utilizes FAISS for efficient vector similarity search and Ollama for local LLM inference, ensuring both performance and privacy.

Document Processing Pipeline. The foundation of our RAG system is a carefully curated document repository containing breast cancer guidelines, research publications, and patient education materials.We obtain these documents thanks to the Web Scraping Module, designed to collect, process, and organize high-quality breast cancer guidelines and patient information from authoritative sources on the web. It supports Retrieval-Augmented Generation (RAG) systems by intelligently identifying, filtering, and categorizing resources, while enriching metadata for enhanced retrieval and generation capabilities. Key features include multi-source scraping, resource classification, audience detection, PDF validation, and metadata enrichment. The scraper employs a multi-stage architecture for collection, filtering, enrichment, validation, and persistence, ensuring the creation of a structured and metadata-rich repository optimized for healthcare AI applications.

After that, these documents undergo a multi-stage processing workflow:

1) *Document Loading:* PDF documents are ingested using either PyPDFLoader (for speed) or UnstructuredPDFLoader (for complex formatting), with metadata preservation.
2) *Chunking:* Documents are segmented into manageable chunks (500-2000 tokens) using RecursiveCharacterTextSplitter with strategic overlap (50-200 tokens) to maintain context across boundaries.
3) *Embedding Generation:* Text chunks are transformed into dense vector representations using domain-appropriate embedding models (nomic-embed-text or all-minilm via Ollama).
4) *Vector Indexing:* FAISS indexes are constructed for efficient similarity search with cosine distance metrics.

Our empirical testing revealed that chunk sizes of 1000 tokens with 100-token overlaps provided optimal balance between context preservation and retrieval precision for breast cancer literature.

Query Processing and Retrieval Enhancement. A significant innovation in our system is the enhanced retrieval function, which improves upon standard similarity search. This approach implements a dual retrieval strategy that combines: (1) an enriched query incorporating conversation history and detected medical terminology, and (2) a direct query using only the user's question. Results are combined and deduplicated, significantly outperforming conventional retrieval in our evaluations, particularly for complex queries involving medical terminology and references to previous conversation turns.

Contextual Prompt Engineering. Another key innovation is our contextualized prompt generation. Rather than passing retrieved documents directly to the language model, we construct a comprehensive prompt incorporating patient profile information (age, cancer stage, information preferences), conversation history with automatically generated summaries, identified medical terms and concepts, and specific instructions for response structure and medical accuracy. This approach ensures responses are not only factually accurate but also appropriate to the user's specific situation and information needs.

Multimodal Interaction: Speech Integration

To enhance accessibility and user engagement, the conversational engine supports full voice interaction. This functionality enables users to speak their queries and listen to the AI's responses, facilitating use by individuals with visual impairments, reading difficulties, or a preference for spoken communication.

The speech-to-text (STT) module leverages OpenAI's Whisper model, which provides robust transcription capabilities even in noisy environments and supports a wide range of accents and speech patterns [7]. Once transcribed, the query follows the standard RAG pipeline for document retrieval and response generation.

The response is then converted into spoken output using Google Text-to-Speech (gTTS), which synthesizes natural-sounding audio directly from the generated text. Users can optionally listen to the answer through an embedded audio player in the Streamlit interface, making the system more inclusive and user-friendly. By combining cutting-edge NLP with audio processing, we significantly broaden the accessibility and usability of the system across diverse patient populations.

Models and Inference. We evaluated several language models for the final response generation, comparing performance across medical accuracy, empathy, and appropriate detail level. First, we attempted LoRA fine-tuning on the Microsoft Phi-2 model to tailor it for processing breast cancer medical guidelines. The fine-tuning attempted to train and integrate low-rank adaptation layers into the base model. However, the effort faced significant challenges: the dataset was too limited, leading to overfitting, and the computational demands of the fine-tuning surpassed our resources, making the approach inefficient. Consequently, we abandoned the fine-tuning strategy and shifted our focus to prompt engineering and retrieval-augmented generation (RAG) techniques, which were more resource-efficient and practical for our objectives. We enhanced the llama3:8b model by integrating a specialized system prompt tailored for providing breast cancer information with accuracy, sensitivity, and hallucination prevention. The prompt ensures the model distinguishes between general medical information and user-specific queries, explicitly avoiding assumptions about the user's personal situation. It also provides a clear response structure—starting with a direct answer, followed by expanded context with verified citations, and concluding with practical, non-personalized recommendations. By clearly defining the scope and limitations of its responses, the model effectively minimizes hallucinations, increasing trust and reliability in medical AI assistance.

Post-Processing and Quality Assurance. Before presenting responses to users, we implement several quality assurance measures:

1) *Medical Accuracy Verification:* Responses are analyzed for absolute claims, personal medical advice, and appropriate limitations disclaimers.
2) *Medical Term Highlighting:* Specialized terminology is automatically identified and highlighted with hover-over definitions.
3) *Source Attribution:* Retrieved document sources are presented in an expandable interface for transparency.
4) *Confidence Assessment:* Responses receive a confidence rating (Low/Medium/High) with appropriate disclaimers.

Design Challenges and Solutions. Several challenges emerged during development. First, hallucination control was addressed by implementing stricter retrieval-dependence in prompting and the medical verification system. Second, context management issues with long conversations were solved through automatic conversation summarization triggered after specific message counts. Third, confusion with medical terminology in user testing led to the implementation of automatic term detection and explanation with interactive definitions.

Performance and Evaluation. The conversational component was evaluated through both technical metrics and user studies. To evaluate the performance of the conversational models integrated into BreastCareAI, we conducted a systematic comparison using a set of 30 breast cancer–related questions spanning diagnostic, emotional, and treatment-oriented topics. Each model's response was rated across three key dimensions—accuracy, empathy, and appropriateness. The breast-cancer-llama3 model outperformed the baselines with mean scores of 0.89 in accuracy, 0.84 in empathy, and 0.86 in appropriateness. In contrast, LLaMA3-8B achieved 0.79, 0.72, and 0.75 respectively, while phi2-breast scored significantly lower (0.11, 0.14, and 0.13). These findings highlight the effectiveness of retrieval-augmented generation. A full breakdown of question-wise results can be found in Appendix B.

The transition from pattern-matching approaches to retrieval-augmented generation demonstrates how hybrid architectures combining knowledge retrieval with language generation can create conversational systems that maintain both factual accuracy and natural interaction in sensitive medical domains. The integration of LangChain, FAISS, and locally-deployed language models via Ollama provides a robust foundation for healthcare information systems that prioritize both accuracy and accessibility.

D. Vision explainer: Ultrasound Segmentation and Analysis

Purpose and Role. The Vision Explainer module in BreastCareAI focuses on analyzing and interpreting breast ultrasound images with a patient-centered approach, aiming to help individuals understand their diagnostic tests and the rationale behind their results. It provides radiological insights by generating reports based on segmentation results and integrates seamlessly with the conversational AI engine to deliver explanations in clear, patient-friendly language. This component ensures that users receive accurate, accessible, and comprehensible information about their ultrasound findings, fostering better understanding and engagement in their care process.

Segmentation Pipeline: From Baseline to Optimized Attention U-Net. Our segmentation pipeline evolved through three key models and two datasets, systematically improving performance at each stage. Initial experiments with a standard U-Net on CBIS-DDSM mammography images achieved only a 0.04 Dice coefficient. Switching to the BUSI ultrasound dataset proved decisive, enabling development of two Attention U-Net variants: a GA-optimized version achieving 0.562 Dice with 40 percent faster training, and our final architecture with manually tuned hyperparameters reaching 0.728 Dice coefficient. A breakdown of segmentation results can be found in Appendix C.

Basic U-Net and CBIS-DDSM dataset. For the initial segmentation model, we used the CBIS-DDSM dataset, a curated subset of the DDSM mammography dataset containing annotated lesions. The architecture implemented was a basic U-Net, featuring an encoder-decoder structure with skip connections to preserve spatial information. The network consisted of four downsampling blocks with convolutional layers (each with 64, 128, 256, and 512 filters respectively), followed by a bottleneck layer with 1024 filters, and four symmetric upsampling blocks using transposed convolutions. All convolutional layers used ReLU activations and batch normalization, with max pooling for downsampling and dropout (rate = 0.5) applied in the deeper layers to reduce overfitting. The pipeline involved preprocessing grayscale mammograms, resizing inputs and masks to 224×224 pixels, normalizing intensities, and training the model using binary cross-entropy loss and the Dice coefficient as evaluation metric. While the training loss decreased and the Dice score improved, the validation metrics remained unstable, and the predicted masks were overly smooth and failed to accurately delineate the lesion areas, often producing rectangular artifacts. This lack of generalization and segmentation precision motivated the transition to the BUSI dataset, which offers more varied and clearly annotated ultrasound images, along with the exploration of an enhanced architecture in the next phase of the project. detailed visualization of the training and validation loss and Dice score evolution is provided in Appendix D.

Attention U-Net with GA hyperparameter optimization. For the second experiment, we adopted the Breast Ultrasound Images (BUSI) dataset, which offers clearer lesion contours and greater annotation consistency compared to CBIS-DDSM, making it more appropriate for segmentation. We implemented an Attention U-Net architecture optimized through a Genetic Algorithm (GA), aiming to automatically fine-tune the model's hyperparameters for improved performance. The GA searched for the optimal number of convolutional filters per layer (ranging from 16 to 256), dropout rates (from 0.1 to 0.5), attention gate placements (on skip connections), and the learning rate (between 1e-5 and 1e-3). The GA used a population of 20 individuals, evaluated over 10 generations using the average Dice score on a validation subset as the fitness function. Selection was performed via tournament method, with crossover probability of 0.8 and mutation probability of 0.2. The resulting architecture consisted of an encoder-decoder U-Net with four downsampling stages (filters: 32, 64, 128, 256), batch normalization, and ReLU activations. Attention gates were applied to all skip connections to enhance focus on salient regions, especially useful in BUSI where lesion textures can vary.

The network was trained in two stages: an initial training of 50 epochs with early stopping, followed by 20 epochs of fine-tuning with a reduced learning rate. As shown in Appendix E, the model achieved stable training with minimal overfitting. Post-processing techniques such as morphological closing and small-object removal were tested to improve mask quality. However, as illustrated in Appendix F, these methods did not significantly improve the overall performance, with the best average Dice score of 0.5623 achieved in the standard configuration.

Due to the stochastic nature of evolutionary optimization, it is likely that the model did not reach a global optimum. Nonetheless, it significantly reduced training time and required fewer manual adjustments, making it a promising option for low-resource environments or real-time applications. In the context of this research, where computational resources were not a limiting factor, we ultimately favored the model that achieved the highest segmentation accuracy.

Final model: Attention U-Net. In the final stage of the project, we continued working with the BUSI dataset due to its favorable image quality and clear lesion annotations, but replaced the GA-optimized model with a manually configured standard Attention U-Net [15], [17]–[19]. The motivation behind this shift was to eliminate the variability introduced by evolutionary tuning and to analyze performance under a more reproducible and interpretable training regime. The architecture followed a symmetric encoder-decoder U-Net structure, enriched with attention gates at each skip connection to suppress irrelevant background features and enhance focus on the lesion regions. The encoder consisted of four convolutional blocks with 32, 64, 128, and 256 filters respectively, each block containing two 3×3 convolutional layers followed by batch normalization, ReLU activation, and max pooling. The decoder mirrored this setup with transposed convolutions for upsampling and concatenation with attention-gated skip connections.

Training was conducted for 55 epochs using the Adam optimizer with a learning rate of 1e-4, batch size of 8, and Dice loss as the objective. Early stopping with patience of 10 epochs was applied based on validation Dice to prevent overfitting. As shown in Appendix G, the model exhibited smooth convergence and reached a best validation Dice of 0.715, significantly outperforming the baseline U-Net and the GA-optimized version. For post-training analysis, we examined predicted probability maps and their corresponding binary masks after thresholding. While a default threshold of 0.5 worked reasonably well, a lower value (0.3) yielded better visual alignment with the ground truth in some ambiguous cases by recovering finer lesion borders (see Appendix H).

This final model not only delivered the highest Dice score, but also produced coherent and visually realistic masks, suggesting that attention mechanisms alone (when combined with careful training and thresholding) can outperform more complex optimization pipelines. These results support the adoption of attention-based architectures for robust and interpretable breast lesion segmentation in ultrasound images.

BI-RADS Classification using Fuzzy Weighted Rule-Based System. To complement the segmentation pipeline with clinically meaningful output, we developed a fuzzy weighted rule-based system for assigning BIRADS (Breast Imaging Reporting and Data System) categories based on the morphological and acoustic characteristics of the lesions. This approach replicates the reasoning process followed by radiologists, but formalized through a set of fuzzy logic rules, where each descriptor contributes with a degree of certainty rather than a binary decision. The system evaluates parameters such as lesion shape (round, oval, irregular), margins (circumscribed, indistinct, microlobulated, angular, spiculated), orientation (parallel or non-parallel to the skin), internal echo pattern (anechoic, hypoechoic, isoechoic, complex) and posterior acoustic features (enhancement, shadowing, no effect).

Each input is mapped to a linguistic variable and processed through fuzzy inference rules, which capture expert-level knowledge in a flexible way. These rules are weighted according to the relative diagnostic importance of each feature (for instance, an irregular margin has more weight in increasing malignancy suspicion than a hypoechoic echo pattern alone). The fuzzy logic framework allows intermediate reasoning in cases where features are ambiguous or partially present. The system aggregates the outputs of all rules into a final risk score, which is then mapped to a BIRADS category from 1 (negative) to 5 (highly suggestive of malignancy). This method ensures both transparency and adaptability, and while it is not learned from data, it offers explainable decision-making useful in clinical assistance, education, or low-resource deployment.

Model Explainability through Grad-CAM. To enhance the transparency and interpretability of the deep learning segmentation models, we integrated Gradient-weighted Class Activation Mapping (Grad-CAM) into the analysis pipeline, that has been widely adopted in medical imaging to provide interpretable heatmaps that enhance trust and validation of deep learning models [20]–[22]. Grad-CAM generates visual heatmaps that highlight the regions in the input image which most strongly influenced the model's prediction, by computing the gradient of the target class with respect to the final convolutional layer's feature maps. While originally designed for classification models, Grad-CAM can be adapted to segmentation architectures such as U-Net by focusing on the activations that contribute to the mask prediction.

In our implementation, we applied Grad-CAM to the output of the encoder's deepest convolutional block, allowing us to visualize which areas the model focused on when detecting potential lesions. The resulting heatmaps were overlaid on the original ultrasound images, providing an intuitive explanation of the model's attention. This is particularly relevant in medical contexts, where trust and accountability are critical [20]. The Grad-CAM maps showed that models with attention mechanisms tended to focus more accurately on the lesion areas, while simpler architectures sometimes highlighted irrelevant or noisy regions. Thus, Grad-CAM not only served as a post-hoc interpretability tool, but also helped validate the design choices of the attention-based networks, reinforcing their alignment with radiological intuition.

Radiologist-Inspired Visual Explanation System. To unify the outputs of the different AI modules and provide a clinically coherent summary, we developed a visual explanation system that mimics the reporting style of a radiologist. Rather than learning directly from images, this module takes as inputs the intermediate outputs already generated by previous components of the pipeline: the original ultrasound image, the binary segmentation mask, the probability map (before thresholding), the Grad-CAM heatmap, and a set of quantitative morphological features computed from the segmented lesion. These include area, eccentricity, solidity, elongation, boundary irregularity, and relative position.

The system uses a vision model, Llama3.2-vision with a specialized system prompt, to interpret these inputs and generate a natural language report. It evaluates whether the mask geometry suggests benign or malignant traits (e.g., regular vs. spiculated contours), cross-references the probability map to assess internal lesion consistency, and checks whether the Grad-CAM heatmap is aligned with the mask — an essential step to verify that the model is indeed attending to the correct region. Based on this multi-source analysis, the system generates descriptive sentences that characterize the lesion's shape, margins, orientation, and attention profile, concluding with a probabilistic BIRADS category.

This final component transforms raw AI outputs into human-readable summaries, closing the loop between model prediction and explainable medical reasoning. It enhances interpretability, facilitates clinician review, and demonstrates how structured vision outputs can be converted into narrative diagnostics.

E. Calendar Integration

Purpose and Role. The *Medication Reminders* and *Google Calendar Integration* modules were designed to extend the system's functionality beyond image interpretation, reinforcing continuity of care through personalized scheduling and proactive support. Their main goal is to ensure that clinical recommendations are not only understood but also acted upon in a timely and organized manner. This is particularly critical in breast cancer contexts, where delays in medication, follow-up imaging, or biopsies can have serious consequences. By linking AI-driven diagnosis with real-world logistics, these modules elevate the system from a passive assistant to an active coordinator in the patient's care journey.

Module Operation. The *Medication Reminder* module allows users or clinicians to input prescribed medications, dosages, and administration schedules. It then creates automated, recurring alerts tailored to the user's treatment plan, with delivery options such as in-app or email notifications, helping users stay compliant and enabling future adherence analysis. The *Calendar Integration* module connects directly to the user's Google Calendar via the official Google Calendar API. Beyond basic scheduling, this module also plays a strategic role in patient-clinician communication: based on the user's preferences and health status, it generates a list of personalized questions for the user to ask at their next medical appointment.

This dual-layer functionality—event scheduling plus personalized question generation—makes the calendar module a bridge between algorithmic insight and patient empowerment, encouraging informed conversations with healthcare professionals while ensuring timely follow-up.

IV. Results

The development and evaluation of BreastCareAI's components yielded significant insights into the effectiveness of multimodal AI for breast cancer guidance and support. Our segmentation pipeline evolution demonstrated clear performance improvements across iterations, with the initial Basic U-Net implementation on the CBIS-DDSM dataset achieving only a 0.04 Dice coefficient. The model exhibited significant issues including overfitting, unstable validation metrics, and poor lesion delineation. Transitioning to the BUSI dataset and implementing a Genetic Algorithm-optimized Attention U-Net architecture substantially improved performance, reaching a Dice coefficient of 0.562. This model achieved 40 percent faster training time compared to manual hyperparameter tuning approaches, with more stable convergence patterns. The final manually configured Attention U-Net with the BUSI dataset delivered the best performance with a Dice coefficient of 0.728, showing smoother convergence and superior lesion boundary detection. The Attention U-Net's effectiveness was further validated through Grad-CAM visualization, which demonstrated the model's focus on relevant lesion boundaries and internal features. This visualization approach provided critical explainability insights, showing how attention mechanisms direct the model's focus to diagnostically relevant image regions. When compared to the basic U-Net architecture, the attention-based model showed substantially improved feature localization. To contextualize the performance of our model, we compared the final Attention U-Net architecture against results from other published segmentation methods on the same BUSI dataset. For instance, AAU-Net, which incorporates adaptive attention mechanisms, achieved a Dice coefficient of 0.730 on BUSI [15], while DAU-Net, using dual attention strategies, reported a Dice of 0.700 [23]. Our model reached a Dice of 0.728, showing competitive accuracy while prioritizing architectural simplicity and reproducibility. These results suggest that our model performs at the level of current state-of-the-art methods, confirming its suitability for real-world deployment scenarios. Our Fuzzy Weighted Rule-Based System for BI-RADS classification translated the segmentation outputs into clinically meaningful categories based on lesion morphology and acoustic characteristics. This approach replicated the radiologist reasoning process through formalized fuzzy logic rules evaluating parameters such as lesion shape, margins, orientation, echo pattern, and posterior acoustic features. The radiologist-inspired visual explanation system successfully integrated the outputs of the different AI modules to provide clinically coherent summaries. By taking inputs from the ultrasound image, segmentation mask, probability map, Grad-CAM heatmap, and morphological features, the system generated natural language reports that characterized lesions following standard reporting conventions. For the conversational AI component, we evaluated several language models for final response generation, comparing performance across medical accuracy, empathy, and appropriate detail level. Our initial attempt at LoRA fine-tuning on the Microsoft Phi-2 model faced significant challenges, including dataset limitations leading to overfitting and excessive computational demands. Consequently, we shifted to a prompt engineering and RAG-based approach using the llama3:8b model enhanced with a specialized system prompt tailored for breast cancer information delivery. ...using the llama3:8b model enhanced with a specialized system prompt tailored for breast cancer information delivery. To assess the real-world usability of the conversational module, we conducted a formative evaluation with five participants (three breast cancer survivors and two medical students). Each completed tasks such as querying BI-RADS results, treatment side effects, and self-care advice, followed by a short questionnaire.

Overall, the system received highly positive feedback: all users rated it as easy to use, 80% reported improved understanding of clinical terms, and all would recommend it to others. These findings suggest that BreastCareAI combines technical robustness with a supportive and accessible user experience (see Appendix I).

V. Conclusions

BreastCareAI represents a significant advancement in AI-assisted breast cancer support systems through several key innovations. We demonstrated that a unified, multimodal AI approach combining conversational interfaces, medical image analysis, and practical health management tools can effectively address the multifaceted needs of breast cancer patients throughout their care journey. The conversational component, powered by retrieval-augmented generation and specialized prompting techniques, provides evidence-based responses that maintain both accuracy and conversational fluency. The progression from basic U-Net (Dice = 0.04) to attention-enhanced architectures (final Dice = 0.728) for breast ultrasound segmentation illustrates the critical importance of attention mechanisms and dataset selection in medical image analysis. The combination of the BUSI dataset with an Attention U-Net architecture proved particularly effective, enabling accurate lesion delineation even in challenging ultrasound images with variable appearance and noise characteristics. Our explainability approach, combining Grad-CAM visualizations with radiologist-inspired natural language explanations, demonstrates how AI systems can bridge the interpretability gap between clinical professionals and patients. This transparency is essential for building trust in AI-driven healthcare tools, particularly for sensitive applications like breast cancer diagnosis and monitoring. The development of BreastCareAI has revealed several important implications for AI integration in breast cancer management. While much AI research focuses solely on diagnosis, our findings highlight the importance of systems that provide continuous support across the entire patient journey, from screening through survivorship. The integration of complementary components—conversational AI, image analysis, medication reminders, and calendar management—creates a more comprehensive support system than any single technology could provide. Despite promising results, several limitations and opportunities for future work remain. The BUSI dataset, though valuable, includes a limited variety of ultrasound equipment and clinical settings. Future work should validate performance across more diverse imaging conditions and patient populations. While BreastCareAI functions as a standalone system, deeper integration with electronic health records and clinical decision support systems would enhance its practical utility. Future research directions should include expanding the segmentation capability to other imaging modalities, particularly mammography and breast MRI, to create a comprehensive multi-modality breast imaging assessment system. BreastCareAI demonstrates the potential of AI to transform breast cancer care from a fragmented, often overwhelming experience into a more integrated, understandable, and supportive journey. By combining advanced AI techniques with human-centered design principles, we've created a system that not only delivers accurate technical outputs but also addresses the educational needs of patients. The strong performance of our attention-based segmentation models and RAG-enhanced conversational AI validates our technical approach. As AI continues to evolve in healthcare, we believe that multimodal, explainable systems like BreastCareAI represent a promising direction for technology that truly serves patient needs. The breast cancer journey remains challenging, but with thoughtfully designed AI assistance, patients can be better equipped with the knowledge, understanding, and support they need to navigate their care with greater confidence and agency.

VI. Appendices

Appendix A
System Architecture Diagram

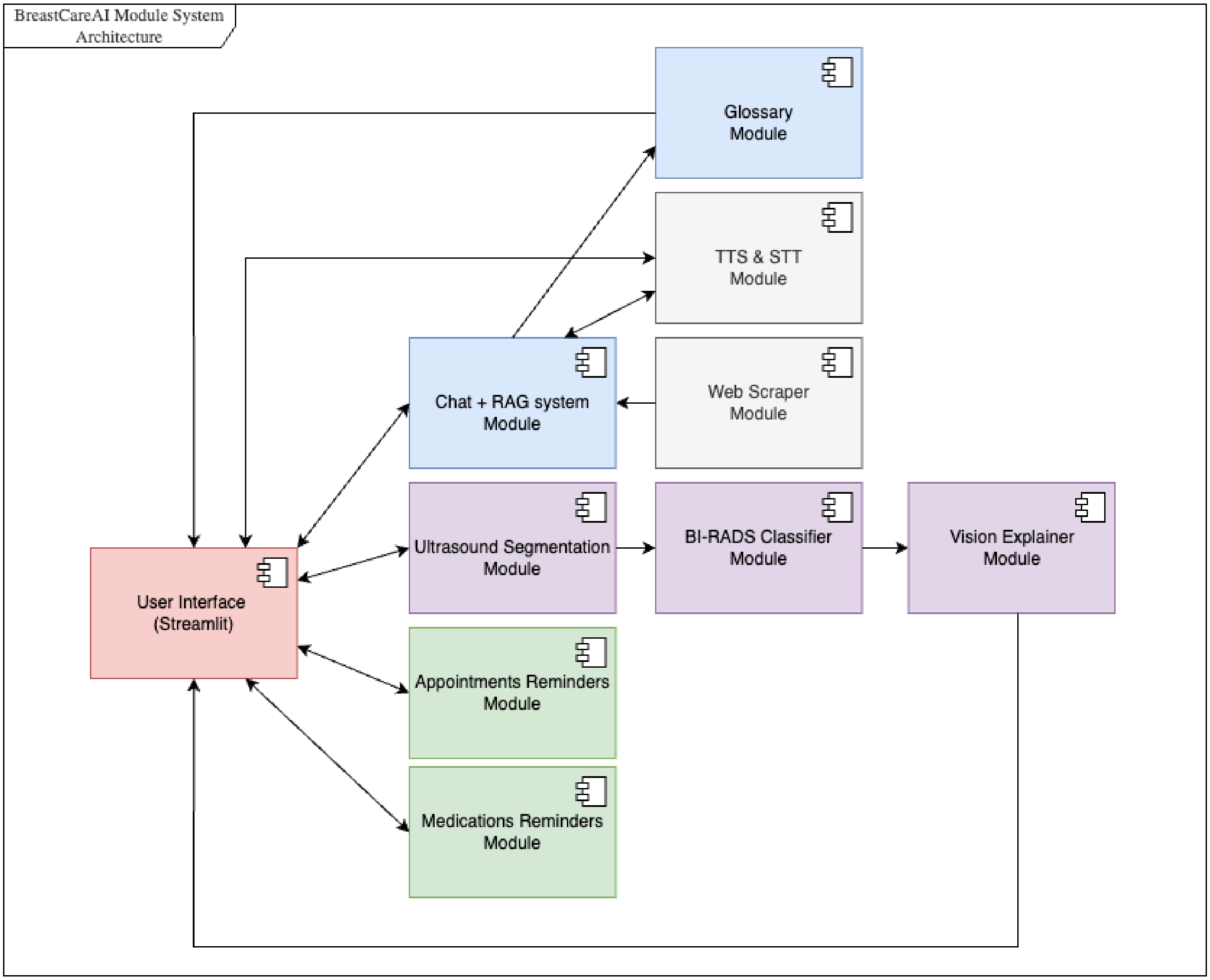

Figure A.1: *System architecture of BreastCareAI showing the modular integration of specialized components.*

Appendix B
Model Evaluation Questions and Scores

TABLE I: Evaluation scores for LLM responses to 30 questions across accuracy, empathy, and appropriateness dimensions.

Question	**breast-cancer-llama3**	**LLaMA3-8B**	**phi2-breast**
What are the first signs of breast cancer?	0.90	0.80	0.10
How is breast cancer diagnosed?	0.90	0.70	0.10
What is a mammogram and does it hurt?	0.90	0.85	0.05
Can I get breast cancer if it doesn't run in my family?	0.90	0.90	0.00
How do I interpret my BI-RADS score?	0.85	0.80	0.05
What does a BI-RADS 3 mean?	0.90	0.85	0.1
How often should I do breast self-exams?	0.90	0.85	0.2
Is breast cancer always hereditary?	0.85	0.80	0.05
What are the treatment options for stage 1 breast cancer?	0.70	0.70	0.10
Will I lose my hair during chemotherapy?	0.85	0.65	0.20
What should I eat during treatment?	0.85	0.80	0.00
Can I still have children after treatment?	0.85	0.75	0.10
What is the risk of recurrence after treatment?	0.80	0.80	0.00
How long will my treatment last?	0.90	0.85	0.10
Can stress worsen breast cancer?	0.85	0.80	0.20
How should I talk to my children about my diagnosis?	0.90	0.80	0.00
What emotional support options are available?	0.85	0.85	0.15
Can I keep working during treatment?	0.85	0.75	0.15
Is it normal to feel anxious before results?	0.90	0.65	0.10
How can I manage side effects of chemotherapy?	0.90	0.85	0.00
Does breast cancer always require surgery?	0.85	0.80	0.10
What is hormone therapy in breast cancer?	0.85	0.85	0.10
Can men get breast cancer too?	0.90	0.85	0.00
Should I freeze my eggs before chemo?	0.90	0.80	0.00
How can I prepare for a mastectomy?	0.85	0.75	0.00
Are second opinions important?	0.90	0.80	0.30
Can I exercise during treatment?	0.85	0.85	0.10
Is breast cancer always a lump?	0.80	0.80	0.40
What does triple-negative mean?	0.90	0.55	0.00
Will my breasts look the same after surgery?	0.85	0.80	0.05

Appendix C
Segmentation Architecture Comparison

TABLE II: Comparison of Segmentation Architectures Used in BreastCareAI

Architecture	**Dataset**	**Dice Score**	**Training Time**	**Key Advantage**	**Description**
Basic U-Net (ours)	CBIS-DDSM	0.040	∼4h	Baseline approach	Classic encoder-decoder, no attention
GA Attention U-Net (ours)	BUSI	0.562	∼4h	Faster training	Evolved via genetic algorithm, reduced size
Standard Attention U-Net (ours)	BUSI	**0.728**	∼6h	Superior accuracy	Adds attention gates for focused segmentation
AAU-Net [15]	BUSI	0.730	–	Adaptive attention	Dynamically weighted attention blocks
DAU-Net [23]	BUSI	0.700	–	Dual attention	Combines spatial and channel attention

Appendix D
Training Performance of U-Net on CBIS-DDSM

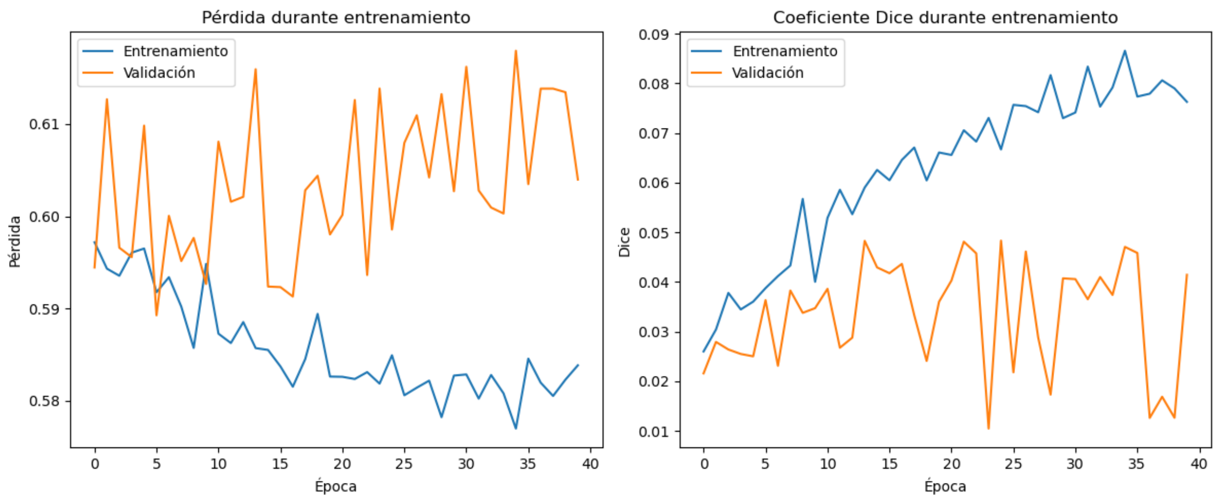

Fig. 1: Training and validation loss (left) and Dice coefficient (right) for the basic U-Net model trained on the CBIS-DDSM dataset.

Appendix E
Training Performance of GA Attention U-Net on BUSI

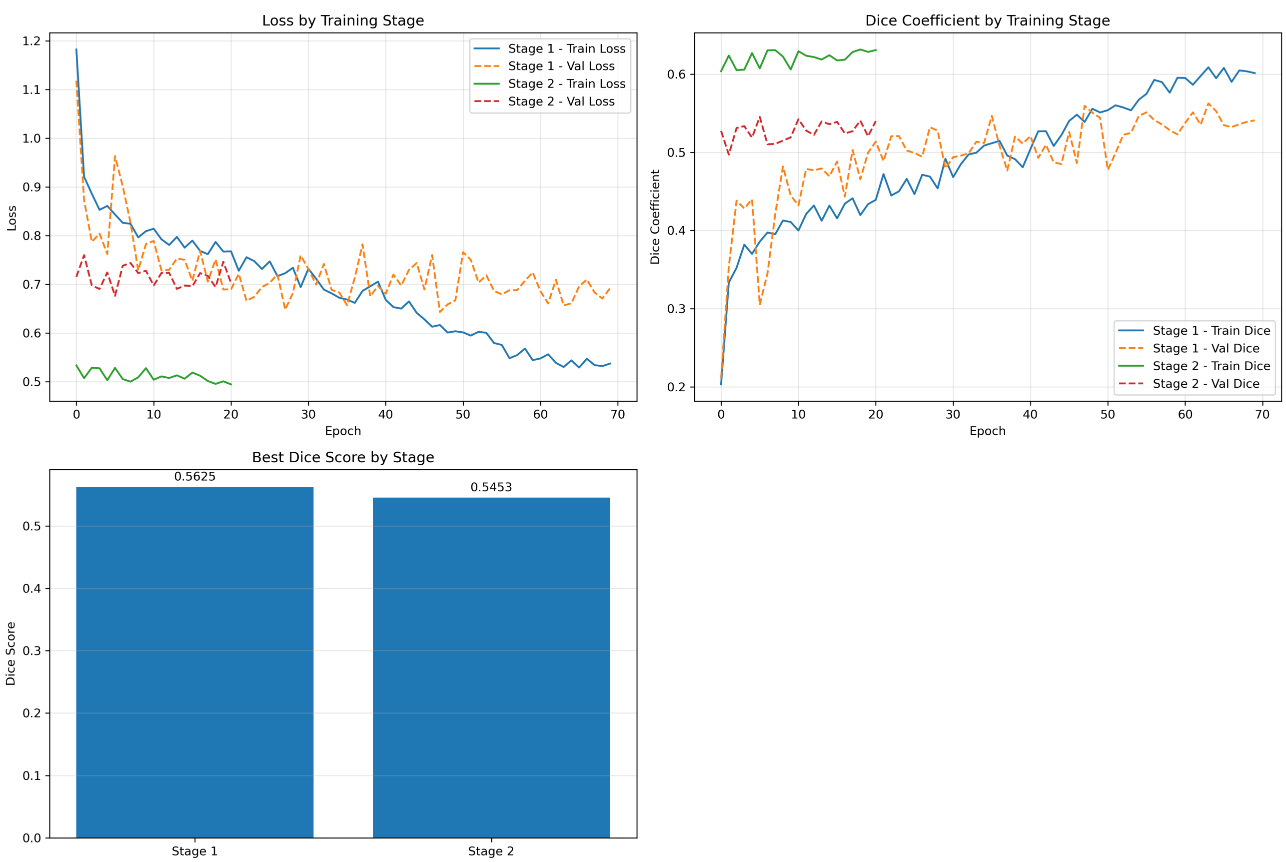

Fig. 2: Training loss and Dice score curves across both training stages for the GA-Optimized Attention U-Net.

Appendix F
Dice score distribution of GA Attention U-Net on BUSI

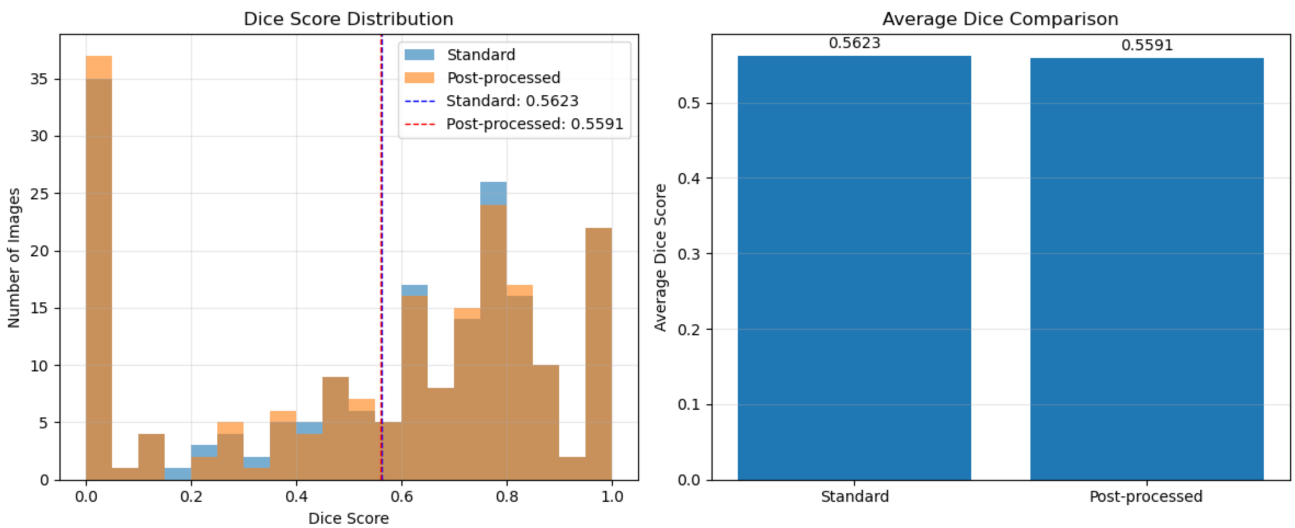

Fig. 3: Dice score distribution and average comparison between standard predictions and post-processed masks.

Appendix G
Training and validation curves for the final Attention U-Net model

Fig. 4: Training and validation loss and Dice curves for the final Attention U-Net model trained on the BUSI dataset.

APPENDIX H
VISUAL EXAMPLE OF THE SEGMENTATION OUTPUT

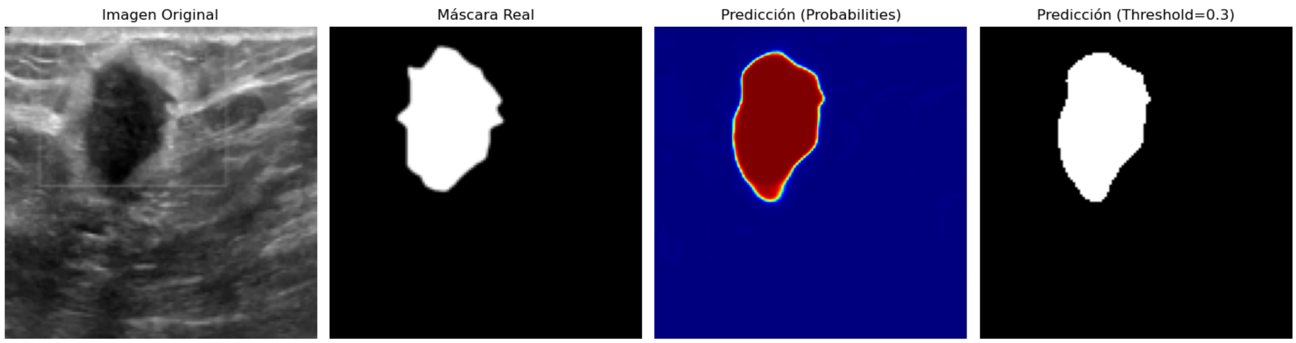

Fig. 5: Visual example of the segmentation output: original image, ground truth, predicted probability map, and binary prediction after applying a 0.3 threshold.

APPENDIX I
USER EVALUATION OF THE CONVERSATIONAL MODULE

TABLE III: Individual results from the formative evaluation of BreastCareAI's conversational interface. Scores reflect subjective user assessments following system interaction.

ID	Profile	Ease of Use	Clarity of Explanations	Trust in System	Qualitative Comment
U1	Survivor	Yes	Yes	Yes	"Very helpful and reassuring"
U2	Survivor	Yes	Yes	Yes	"Clear and understandable info"
U3	Survivor	Yes	No	Yes	"Some terms still too technical"
U4	Med student	Yes	Yes	Yes	"Surprisingly informative"
U5	Med student	Yes	Yes	Yes	"Easy to navigate and well-structured"

REFERENCES

[1] World Health Organization, “Breast cancer fact sheet,” 2024, accessed: 2025-05-06. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[2] American Cancer Society, “Understanding your mammogram report,” 2024, accessed: 2025-05-06. [Online]. Available: https://www.cancer.org/cancer/types/breast-cancer/screening-tests-and-early-detection/mammograms/understanding-your-mammogram-report.html

[3] B. Chaix, J.-E. Bibault, A. Pienkowski, G. Delamon, A. Guillemassé, P. Nectoux, and B. Brouard, “When chatbots meet patients: One-year prospective study of conversations between patients with breast cancer and a chatbot,” *JMIR Cancer*, vol. 5, no. 1, p. e12856, 2019. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6521209/

[4] M. AI, “Llama 3.2 11b vision,” https://huggingface.co/meta-llama/Llama-3.2-11B-Vision, 2024, accessed: 2025-05-07.

[5] A. Grattafiori, A. Dubey, ..., and Z. Ma, “The llama 3 herd of models,” 2024. [Online]. Available: https://arxiv.org/abs/2407.21783

[6] H. Liu, C. Li, Q. Wu, and Y. J. Lee, “Visual instruction tuning,” 2023. [Online]. Available: https://arxiv.org/abs/2304.08485

[7] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust speech recognition via large-scale weak supervision,” *arXiv preprint arXiv:2212.04356*, 2022. [Online]. Available: https://arxiv.org/abs/2212.04356

[8] E. S. Burnside, E. A. Sickles, L. W. Bassett, D. L. Rubin, C. H. Lee, D. M. Ikeda, E. B. Mendelson, P. A. Wilcox, P. F. Butler, and C. J. D’Orsi, “The acr bi-rads experience: Learning from history,” *Journal of the American College of Radiology*, vol. 6, no. 12, pp. 851–860, 2009. [Online]. Available: https://doi.org/10.1016/j.jacr.2009.07.023

[9] C. Armbruster, M. Knaub, E. Farin-Glattacker, and R. von der Warth, “Predictors of adherence to cancer-related mhealth apps in cancer patients undergoing oncological or follow-up treatment—a scoping review,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 20, 2022. [Online]. Available: https://www.mdpi.com/1660-4601/19/20/13689

[10] G. Goumas, T. I. Dardavesis, K. Syrigos, N. Syrigos, and E. Simou, “Chatbots in cancer applications, advantages and disadvantages: All that glitters is not gold,” *Journal of Personalized Medicine*, vol. 14, no. 8, p. 877, 2024. [Online]. Available: https://doi.org/10.3390/jpm14080877

[11] B. Chaix, A. Guillemassé, P. Nectoux, G. Delamon, and B. Brouard, “Vik: A chatbot to support patients with chronic diseases,” *Health*, vol. 12, no. 7, pp. 783–793, 2020. [Online]. Available: https://www.scirp.org/journal/paperinformation.aspx?paperid=101598

[12] J.-E. Bibault, B. Chaix, A. Guillemassé, S. Cousin, A. Escande, M. Perrin, A. Pienkowski, G. Delamon, P. Nectoux, and B. Brouard, “A chatbot versus physicians to provide information for patients with breast cancer: Blind, randomized controlled noninferiority trial,” *Journal of Medical Internet Research*, vol. 21, no. 11, p. e15787, 2019, pMID: 31774408. [Online]. Available: https://doi.org/10.2196/15787

[13] J. Miao, C. Thongprayoon, S. Suppadungsuk, O. A. Garcia Valencia, and W. Cheungpasitporn, “Integrating retrieval-augmented generation with large language models in nephrology: Advancing practical applications,” *Medicina (Kaunas, Lithuania)*, vol. 60, no. 3, p. 445, 2024. [Online]. Available: https://doi.org/10.3390/medicina60030445

[14] P. Pramanik, A. Roy, E. Cuevas, M. Perez-Cisneros, and R. Sarkar, “Dau-net: Dual attention-aided u-net for segmenting tumor in breast ultrasound images,” *PLOS ONE*, vol. 19, no. 5, p. e0303670, 2024. [Online]. Available: https://doi.org/10.1371/journal.pone.0303670

[15] G. Chen, L. Li, Y. Dai, J. Zhang, and M. H. Yap, “Aau-net: An adaptive attention u-net for breast lesions segmentation in ultrasound images,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 5, pp. 1289–1300, May 2023. [Online]. Available: https://doi.org/10.1109/TMI.2022.3226268

[16] B. Zhang, A. Vakanski, and M. Xian, “Bi-rads-net: An explainable multitask learning approach for cancer diagnosis in breast ultrasound images,” in *IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2021. [Online]. Available: https://doi.org/10.1109/mlsp52302.2021.9596314

[17] A. Sulaiman, V. Anand, S. Gupta, A. Rajab, H. Alshahrani, M. S. Al Reshan, A. Shaikh, and M. Hamdi, “Attention based unet model for breast cancer segmentation using busi dataset,” *Scientific Reports*, vol. 14, no. 1, p. 72712, 2024.

[18] N. S. Punn and S. Agarwal, “Rca-iunet: A residual cross-spatial attention-guided inception u-net model for tumor segmentation in breast ultrasound imaging,” *Machine Vision and Applications*, vol. 33, no. 2, p. 27, 2022.

[19] W. Al-Dhabyani, M. Gomaa, R. Khaled, and A. Fahmy, “Dataset of breast ultrasound images,” *Data in Brief*, vol. 28, p. 104863, 2020.

[20] S. Suara, A. Jha, P. Sinha, and A. A. Sekh, “Is grad-cam explainable in medical images?” *arXiv preprint arXiv:2307.10506*, 2023. [Online]. Available: https://arxiv.org/abs/2307.10506

[21] M. Talaat, A. Salem, R. Aly, E. H. Houssein, and H. Zaki, “Grad-cam enabled breast cancer classification with a 3d inception-resnet v2: Empowering radiologists with explainable insights,” *Diagnostics*, vol. 13, no. 7, p. 1241, 2023.

[22] A. Dost Muhammad, A. Khan, A. Rehman, and et al., “Unveiling the black box: A systematic review of explainable artificial intelligence in medical image analysis,” *Computational and Structural Biotechnology Journal*, vol. 22, pp. 2226–2243, 2024.

[23] S. Pramanik, N. Singh, and N. Dey, “Dau-net: Dual attention u-net for breast ultrasound image segmentation,” *Biomedical Signal Processing and Control*, vol. 86, p. 105374, 2024. [Online]. Available: https://doi.org/10.1016/j.bspc.2023.105374