Cloud-Based Diagnostic Support for Early Breast Cancer Detection

AI-powered triage system enabling clinics to detect malignancy instantly

Group 12

Jad Zoghaib, Miguel de Faria, Ana Petrescu, Jan Philipp Gnau

The Clinical Challenge

Access Gap

Many clinics lack oncologists for diagnosis

Capability Mismatch

Staff collect samples but cannot assess malignancy

Critical Delays

Missed cases and delayed treatment

Why Existing Tools Fall Short

Cost Barrier

Too expensive for low-resource settings

Complexity

Too complex for non-specialists

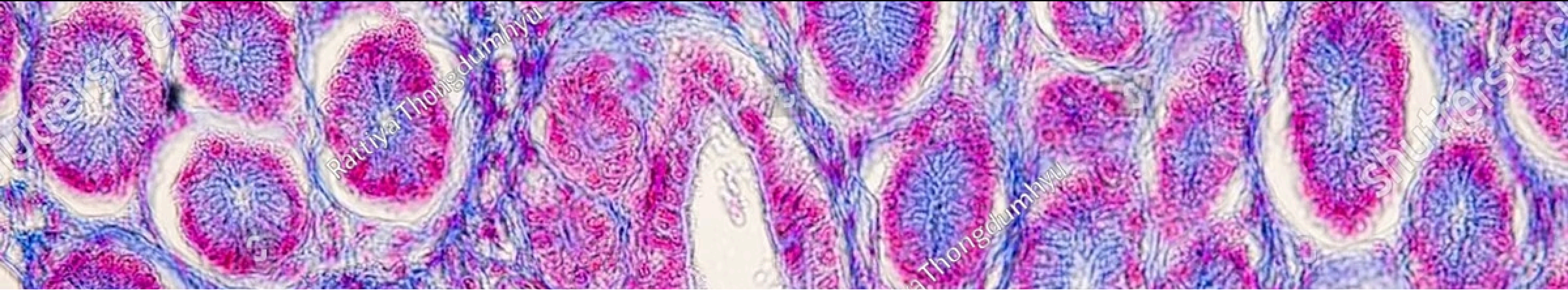
Speed

Manual review is too slow

Scalability

Not scalable in resource-limited environments

The Gap: Need for simple, low-cost tool that analyzes cellular features instantly



Our Solution

01

SVM Classifier

Trained on 30 cellular features

02

Instant Prediction

Malignant/benign diagnosis in real-time

03

Cloud-Powered

Scalable via serverless AWS components

04

Clinical Support

Reliable, interpretable results for decision-making

Why Cloud Architecture?

Real-Time Processing

AWS Lambda enables instant analysis

Zero Infrastructure

No local hardware or ML setup required

Auto-Scaling

Handles varying clinical workloads automatically

Seamless Integration

Works with existing medical systems

Secure Storage

DynamoDB ensures traceability and data protection

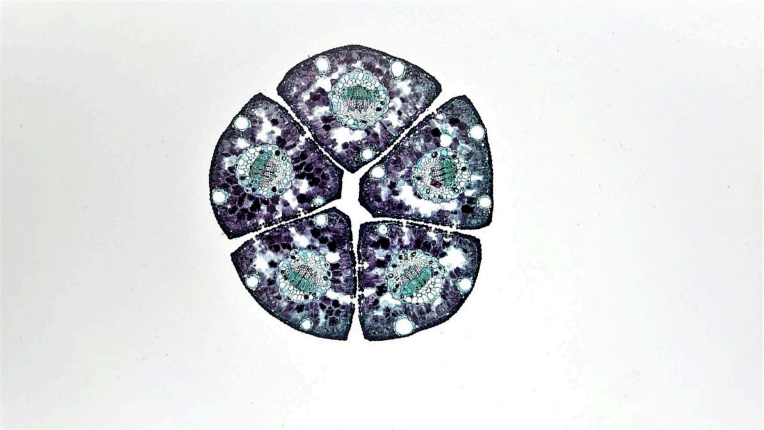
> Enables early breast cancer triage and accelerates life-saving interventions

Training Data Foundation

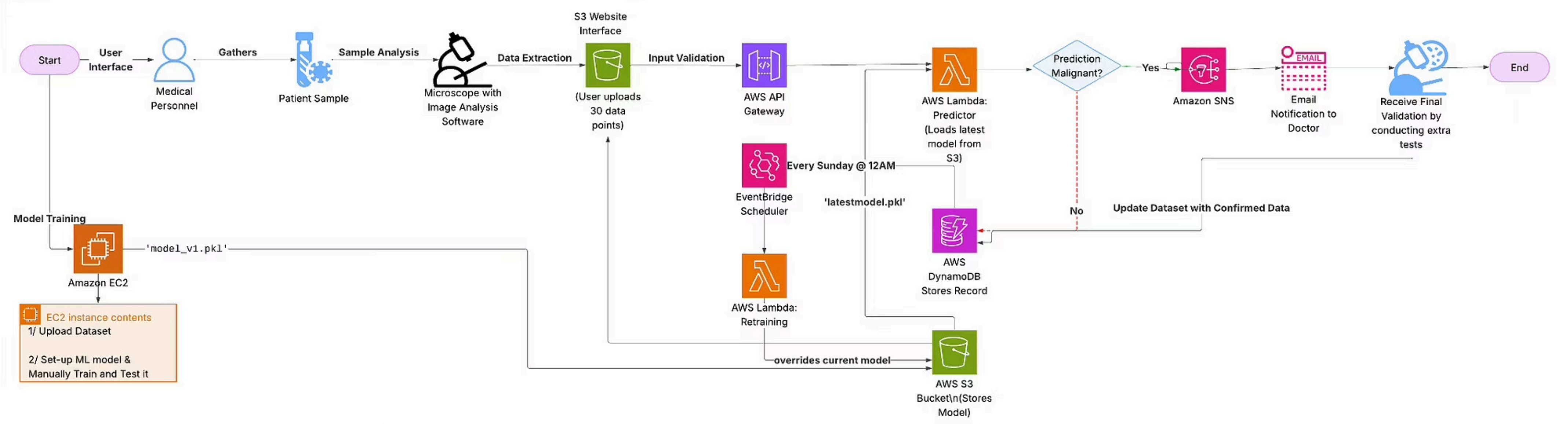
Source	Wisconsin Diagnostic Breast Cancer (WDBC) dataset
Collection Method	Digitized fine needle aspirate images
Data Points	569 instances
Features	30 real-valued vectors
Target	M = Malignant, B = Benign

Key Features Analyzed

- Radius
- Texture
- Perimeter
- Area
- Smoothness



AWS Architecture



Key Architectural Decisions

AWS Lambda

- Fully serverless—no infrastructure management
- Cost-efficient, scales instantly
- Flexible model versioning from S3

DynamoDB

- Flexible schema for clinical data
- Fast record storage
- Easy retraining data access

Unlimited Model Storage

S3 provides unlimited model storage.

Parallel Execution

Stateless Lambda allows for parallel execution.

Separate Retraining Pipeline

Dedicated pipeline for model retraining.

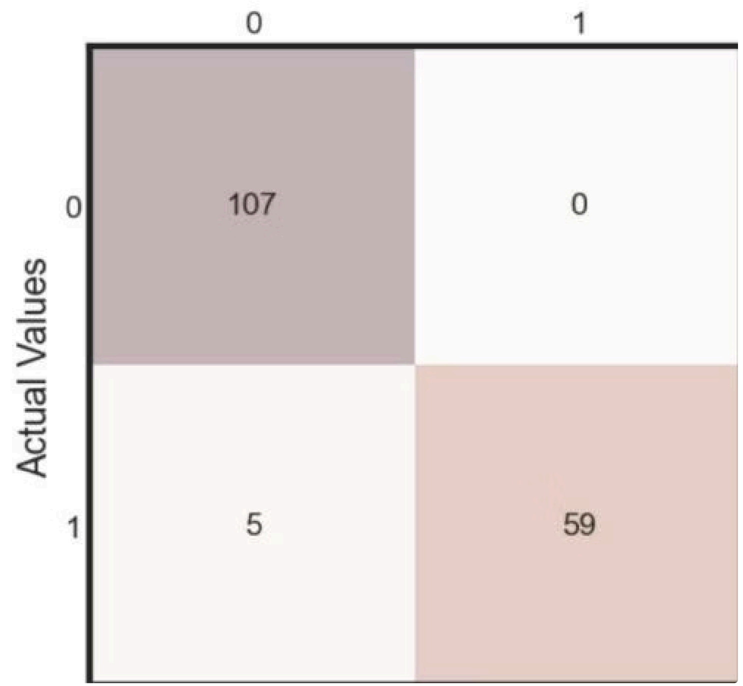
Mobile/API Ready

Plug-in ready for mobile and API integration.

Leveraged AWS Services

PHASE	COMPONENT	PURPOSE & FUNCTION	TECHNICAL DETAILS / CONFIGURATION
Phase 1: Model Development	**Data Source**	Wisconsin Breast Cancer Dataset used as the baseline.	**File:** clean-data.csv (Golden Copy)
	Amazon S3	Stores the source of truth for training data.	**Bucket Name:** breast-cancer-cleaneddata
	Training Script *train.py*	Downloads data, trains Logistic Regression (scikit-learn), evaluates accuracy, and serializes (pickles) the model.	**Library:** scikit-learn **Output:** Serialized model artifact
	Lambda Layer	Deployment package to handle AWS Lambda size limits.	**File:** sklearn_light.zip **Contents:** Stripped-down scikit-learn & numpy
Phase 2: Real-Time Prediction	**DynamoDB**	Operational database storing every prediction request.	**Table:** Patient_Entries **Key:** id (Partition Key) **Attributes:** features, prediction, doctor_resolution (Default: Pending), is_exported
	Lambda Function 1 Predictor	1. Receive data via API Gateway. 2. Load model & predict. 3. **SNS Alert:** If 'Malignant', email doctor. 4. Save to DynamoDB.	**Trigger:** HTTP POST via API Gateway **Permissions:** DynamoDBFullAccess, SNSFullAccess, BasicExecutionRole
	EC2 Frontend	Virtual Server hosting the web application interface for doctors.	**OS:** Linux (Ubuntu/Amazon Linux) **Software:** Web Server (Apache/Nginx) **File:** index.html
Phase 3: Continuous Learning	**Lambda Function 2** Sync Agent	Moves "Confirmed" data from DynamoDB to S3 to expand the training set.	**Logic:** Scans for resolved cases, flips labels if "False Positive", downloads latest CSV from S3, appends new rows, uploads as new version. **Permissions:** S3FullAccess, DynamoDBFullAccess
	Amazon EventBridge	Scheduler that triggers the automation loop.	**Type:** Schedule Rule (Cron Job) **Schedule:** cron(0 0 ? * SUN *) (Every Sunday at Midnight)

Model Performance & Impact



	precision	recall	f1-score	support
0	0.96	1.00	0.98	107
1	1.00	0.92	0.96	64
accuracy			0.97	171
macro avg	0.98	0.96	0.97	171
weighted avg	0.97	0.97	0.97	171

Evaluation Metrics

- **Recall/Sensitivity:** Minimizes false negatives
- **Precision & F1:** Balanced to avoid false alarms
- Stable across train/test splits

Clinical Impact: Rapid triage in resource-limited environments, prioritizes high-risk patients, reduces diagnostic delays

Live Demo

http://breast-cancer-frontend-team12.s3-website.eu-central-1.amazonaws.com/

Limitations

Manual Data Entry

True diagnosis labels must be entered manually into the system

Domain Shift Risk

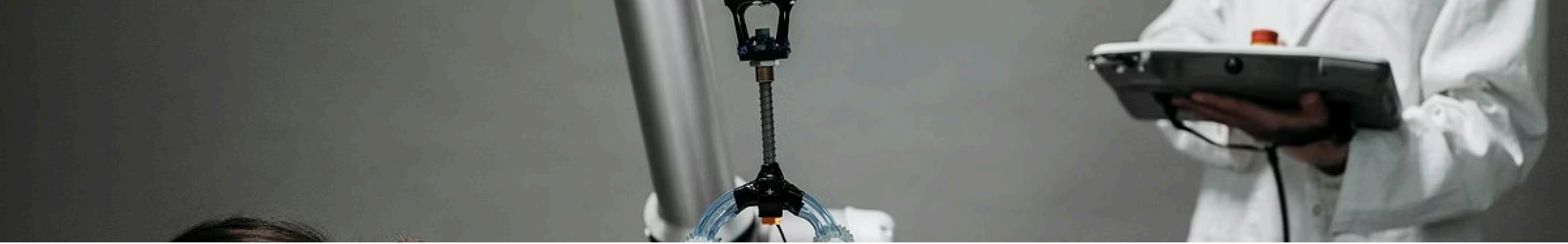
Real clinical data may vary from the WDBC dataset

Model Simplicity

Simple SVM model; performance depends on feature scaling

Feature-Based Only

No image-based inference (only numerical features supported)



Future Directions

Image-Based Models

Integrate CNNs via SageMaker for direct image inference

Enhanced Retraining

Move to SageMaker Training Jobs for larger datasets

Mobile & Offline

Expand to mobile-first, offline data capture

Multi-Class Support

Richer tumor classification with clinical metadata

Conclusion

Our Solution

SVM, instant prediction, scalable

Clinical Impact

Clinical Impact

Enables early triage and faster interventions

Why the Cloud?

Real-time, no local hardware, secure