

## Módulo RESTART

### Sistema de Apoio ao Diagnóstico de Câncer de Mama (OncoVision AI)

**Integrantes:** Gabriel de Aragão A. Oliveira; Weriton Luis Petreca;

#### 1. Identificação do Problema

**Descrição do Problema:** O câncer de mama é a neoplasia maligna mais incidente na população feminina brasileira (excluindo o câncer de pele não melanoma). Segundo a Cartilha Câncer de Mama: vamos falar sobre isso? (10<sup>a</sup> ed., 2025) do Instituto Nacional de Câncer (INCA), estimam-se 73.610 novos casos para o ano de 2025. Apesar de ser curável se detectado precocemente, a doença ainda apresenta alta mortalidade devido a diagnósticos tardios e à dificuldade de acesso a exames especializados em tempo hábil. Muitas mulheres dependem exclusivamente das Unidades Básicas de Saúde (UBS), que frequentemente carecem de ferramentas tecnológicas para auxiliar na triagem rápida e priorização de casos suspeitos, atrasando o encaminhamento para tratamento especializado.

#### Stakeholders (Partes Interessadas):

- **Pacientes:** Mulheres, especialmente na faixa etária de risco (50 a 74 anos), e aquelas em regiões com escassez de médicos especialistas.
- **Profissionais da Atenção Básica:** Médicos generalistas e enfermeiros das UBS que necessitam de suporte à decisão clínica para encaminhamentos assertivos.
- **Gestores de Saúde Pública:** Interessados na otimização de recursos, redução de filas e diminuição dos custos associados a tratamentos de câncer em estágios avançados.
- **Hospitais de Referência:** Instituições que receberão os casos já triados, otimizando o fluxo de atendimento oncológico.

**Justificativa:** A democratização do diagnóstico é urgente e estratégica. O projeto *OncoVision AI* propõe o uso de computação em nuvem (AWS) e Inteligência Artificial para levar capacidade analítica avançada a regiões remotas. A literatura científica recente demonstra que algoritmos de aprendizado de máquina supervisionado, como *AdaBoost* e *Random Forest*, podem atingir precisão

diagnóstica superior a 75% na predição de sobrevida e risco. Ao alinhar essa tecnologia com as diretrizes do INCA – que reforçam a importância da detecção precoce para tratamentos menos agressivos e maior chance de cura – a solução oferece uma "segunda opinião" automatizada. O sistema processará imagens e dados clínicos (considerando fatores de risco como idade e histórico familiar) para classificar a urgência, apoiando o profissional de saúde na tomada de decisão rápida.

## 2. Levantamento de Requisitos

### Requisitos Funcionais (RF):

- **RF01** – Cadastro e Upload: O sistema deve permitir que o profissional de saúde cadastre os dados clínicos do paciente (baseados nos fatores de risco do INCA) e realize o upload de imagens de exames para a nuvem AWS.
- **RF02** – Análise Preditiva: O sistema deve processar os dados utilizando um modelo de *Machine Learning* treinado (AdaBoost/Random Forest) para classificar o risco de malignidade.
- **RF03** – Visualização de Resultados: O sistema deve apresentar um *dashboard* simplificado com o score de risco e sugestão de conduta (ex: encaminhamento prioritário).
- **RF04** – Histórico Seguro: O sistema deve armazenar o histórico de análises, permitindo acompanhamento longitudinal do paciente.

### Requisitos Não Funcionais (RNF):

- **RNF01** – Acessibilidade: O sistema deve ser acessível via navegador web, leve e otimizado para conexões de internet instáveis (comuns em áreas remotas).
- **RNF02** – Segurança e Privacidade: O sistema deve garantir a criptografia de dados em trânsito e repouso, em conformidade com a LGPD e normas de dados sensíveis de saúde.
- **RNF03** – Escalabilidade: A arquitetura deve utilizar serviços *serverless* da AWS para suportar picos de acesso sem degradação de performance.
- **RNF04** – Confiabilidade do Modelo: O algoritmo deve manter métricas de acurácia alinhadas com o estado da arte (aprox. 78% conforme literatura base ).

### **MVP (Produto Mínimo Viável):**

1. Módulo de autenticação para profissionais de saúde.
2. Interface de upload de dados clínicos essenciais (Idade, Tamanho Nódulo, Histórico) e imagens.
3. API de processamento na nuvem retornando a classificação de risco baseada no modelo preliminar.

### **3. Planejamento Ágil**

#### **Backlog de Tarefas:**

##### **1. Pesquisa & Dados:**

- o Análise exploratória do dataset METABRIC.
- o Mapeamento de variáveis compatíveis com os fatores de risco brasileiros (INCA 2025: idade >50, histórico familiar).

##### **2. Infraestrutura AWS:**

- o Configuração de IAM (permissões) e Bucket S3 para armazenamento seguro de exames.

##### **3. Desenvolvimento Backend:**

- o Setup do projeto Java com Spring Boot e Gradle.
- o Criação da API REST para recebimento de dados.

##### **4. Desenvolvimento Frontend:**

- o Prototipagem de interface focada em usabilidade para UBS.

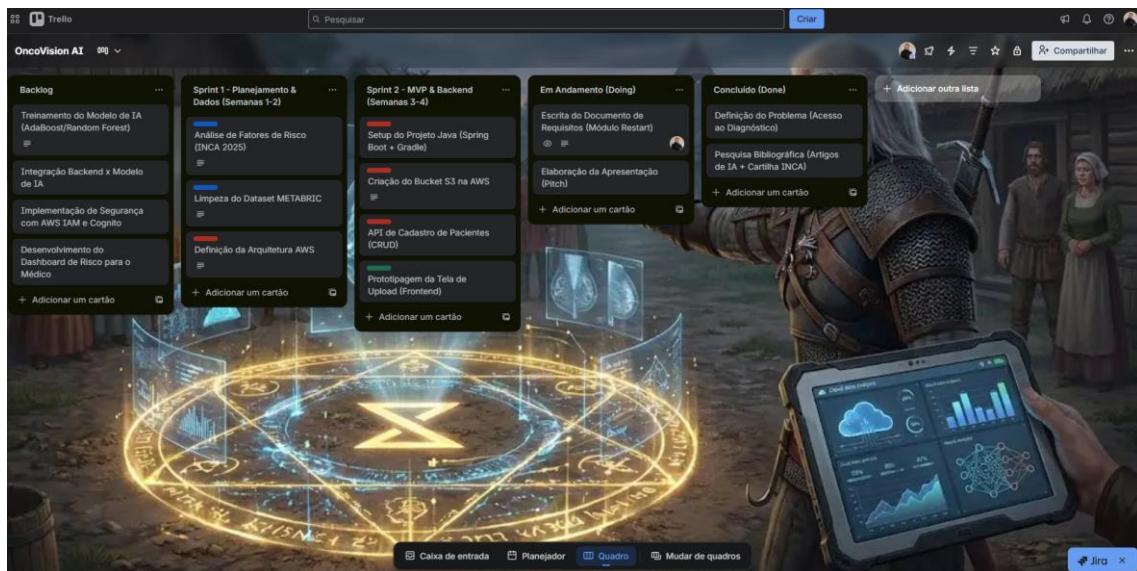
##### **5. Inteligência Artificial (Módulo Seguinte):**

- o Treinamento e validação dos modelos AdaBoost e Random Forest.

## Sprints:

- **Sprint 1 (Setup & Discovery):** Configuração do ambiente de desenvolvimento, Repositório Git, Definição da Arquitetura AWS e Estudo dos dados do INCA/METABRIC.
- **Sprint 2 (Core Backend):** Implementação da API de cadastro de pacientes e integração inicial com AWS S3.
- **Sprint 3 (MVP Interface):** Desenvolvimento da tela de upload e conexão com o Backend para fluxo ponta-a-ponta (sem IA ativa ainda).

## Quadro (Board):



<https://trello.com/invite/b/69221545a1f696fb4924410f/ATTI971f032c7e7ab6bab60b04e9ec15ab4B3D6743F/oncovision-ai>

## Módulo de Extensão em Inteligência Artificial

### 1. Componentes e Serviços

Nesta etapa, redefinimos o arsenal tecnológico utilizado para viabilizar o **Oncovision AI**. A solução evoluiu para uma arquitetura **Assíncrona Orientada a Eventos**, garantindo desacoplamento total, alta resiliência para processamento de arquivos pesados (DICOM) e conformidade rigorosa com normas de saúde (LGPD/HIPAA).

## A. Front-end & Autenticação

- **AWS Amplify Hosting (com CloudFront integrado):** Responsável pela entrega da aplicação Web (SPA) com baixa latência global e cacheamento de conteúdo estático via CDN.
- **Amazon Cognito:** O guardião da identidade. Gerencia a autenticação dos médicos com obrigatoriedade de **MFA (Múltiplo Fator de Autenticação)**, fornecendo tokens JWT para acesso seguro.

## B. Computação & Orquestração (Serverless)

- **Amazon API Gateway (Regional):** A porta de entrada da API REST. Gerencia o tráfego, aplica *throttling* (limite de requisições) e valida tokens do Cognito.
- **AWS Lambda 1 (API Handler - Java 21 SnapStart):** Recebe metadados, valida regras de negócio e gera **Pre-signed URLs** para upload seguro direto ao S3. Utiliza *SnapStart* para performance imediata (sem *cold start*).
- **AWS Lambda 2 (Dispatcher - Python):** Acionada por eventos do EventBridge, prepara o payload e submete a requisição para a fila de inferência do SageMaker.
- **AWS Lambda 3 (Result Processor - Python):** Consome notificações do SNS, processa o resultado JSON da IA e atualiza o diagnóstico final no DynamoDB.

## C. Inteligência Artificial (Inference)

- **Amazon SageMaker (Asynchronous Inference Endpoint):** Endpoint configurado para processamento assíncrono de grandes payloads (imagens médicas). Possui **Auto-Scaling para Zero**, desligando a infraestrutura (custo zero) quando não há exames na fila.

## D. Armazenamento & Dados

- **Amazon S3 (Simple Storage Service):**
  - *Bucket Raw-Input:* Armazena imagens DICOM criptografadas.
  - *Bucket Inference-Output:* Armazena os JSONs de resultado da IA.
  - *Bucket Access-Logs (Novo):* Repositório centralizado e imutável para logs de auditoria e acesso.
- **Amazon DynamoDB:** Banco NoSQL para controle de estado do exame e metadados do paciente.

## E. Integração & Mensageria

- **Amazon EventBridge:** Monitora o bucket S3 e dispara o fluxo de análise (Lambda 2) assim que uma imagem é carregada.
- **Amazon SNS (Simple Notification Service):** Canal nativo de notificação do SageMaker. Avisa a Lambda 3 imediatamente após o sucesso ou falha da inferência.

## F. Segurança, Governança & Observabilidade (Well-Architected)

- **AWS IAM (Identity & Access Management):** Implementa o princípio do **Privilégio Mínimo**. Cada Lambda possui uma *Role* exclusiva, permitindo acesso apenas aos recursos estritamente necessários.
  - **AWS WAF (Web Application Firewall):** Acoplado ao API Gateway, bloqueia ataques comuns (SQL Injection, XSS) e aplica **Geo-Blocking** (permitindo acesso apenas de IPs do Brasil).
  - **AWS Shield Standard:** Proteção nativa e contínua contra ataques DDoS na camada de infraestrutura (camadas 3 e 4).
  - **AWS CloudTrail:** Auditoria de conformidade. Registra **toda** chamada de API feita na conta (quem acessou o quê e quando), enviando os trilhas para o bucket de logs.
  - **AWS Config:** Monitoramento de configuração. Vigia se os recursos estão em *compliance* (ex: alerta se alguém criar um Bucket S3 sem criptografia).
  - **AWS X-Ray:** Rastreamento distribuído (*Distributed Tracing*). Permite visualizar a latência de ponta a ponta, desde o clique do médico até a resposta da IA, facilitando a detecção de gargalos.
  - **Amazon GuardDuty:** Detecção inteligente de ameaças. Monitora logs de rede e DNS em busca de comportamentos maliciosos (ex: instâncias minerando criptomoedas ou exfiltração de dados).
  - **VPC Endpoints (PrivateLink):** Canais privados que conectam as Lambdas aos serviços AWS (S3, DynamoDB, SageMaker) sem passar pela internet pública.
-

## 2. Arquitetura do Projeto

A arquitetura do **OncoVision AI** foi reestruturada para seguir uma abordagem **Assíncrona Orientada a Eventos (Event-Driven)**. Este modelo desacopla a ingestão de dados do processamento pesado de IA, garantindo que a aplicação permaneça responsiva e economicamente viável, mesmo ao processar arquivos de imagem médica de grande porte (DICOM).

O fluxo de dados e a interação entre os componentes seguem a sequência detalhada abaixo:

### 1. Ingestão Segura e Validação (Camada Síncrona)

- **Interação:** O médico envia os dados clínicos do paciente (anamnese) através do Front-end (Amplify).
- **Processamento:** O **Amazon API Gateway** intercepta a requisição, protegida pelo **AWS WAF** (contra ataques web) e validada pelo **Amazon Cognito** (garantindo que apenas médicos autenticados com MFA acessem).
- **Execução:** A requisição é encaminhada para a **Lambda 1 (API Handler - Java 21 SnapStart)**. Esta função:
  1. Valida os dados clínicos e grava o registro inicial no **Amazon DynamoDB** com status PENDING.
  2. Gera uma **S3 Presigned URL** de curta duração (ex: 5 minutos), permitindo que o navegador do médico faça o upload seguro da imagem diretamente para o bucket Raw-Input.

### 2. Upload e Disparo de Eventos (Desacoplamento)

- **Ação:** O upload da imagem ocorre diretamente do navegador para o **Amazon S3**, sem passar pelos servidores de aplicação (padrão *Presigned URL*), eliminando gargalos de memória na API.
- **Gatilho:** A conclusão do upload no bucket Raw-Input emite um evento de sistema que é capturado pelo **Amazon EventBridge**.
- **Roteamento:** O EventBridge aplica uma regra de filtragem (ex: apenas arquivos .dcm ou .jpg) e aciona a **Lambda 2 (Dispatcher - Python)**.

## 6. Inferência de Inteligência Artificial (Camada Assíncrona)

- **Orquestração:** A Lambda 2 não processa a imagem; ela apenas invoca o **Amazon SageMaker Asynchronous Endpoint**, passando a localização da imagem no S3.
- **Processamento Inteligente:**
  - O SageMaker coloca a requisição em uma fila interna gerenciada.
  - **Auto-Scaling (Scale-to-Zero):** Se o sistema estiver ocioso, o SageMaker provisiona automaticamente a infraestrutura de GPU. Se não houver exames na fila, a infraestrutura é desligada para zerar custos.
  - O modelo de IA (AdaBoost/CNN) analisa a imagem, calcula o risco de malignidade e salva o resultado (JSON) no bucket Inference-Output.

## 7. Conclusão e Persistência

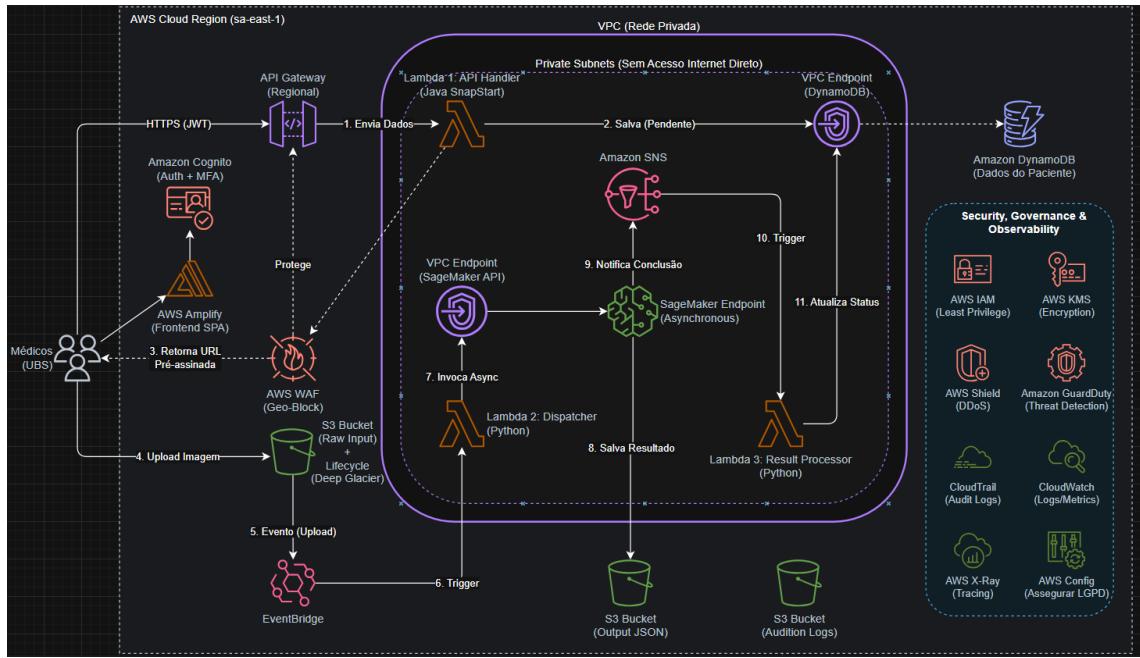
- **Notificação:** Ao finalizar a análise, o SageMaker publica automaticamente uma mensagem em um tópico do **Amazon SNS**.
- **Atualização:** A **Lambda 3 (Result Processor – Python)** é ativada por essa notificação SNS. Ela lê o resultado da inferência e atualiza o registro do paciente no **Amazon DynamoDB** para o status COMPLETED, anexando o score de risco.

## 8. Governança e Segurança Transversal

Todo o tráfego sensível entre as funções Lambda, S3, DynamoDB e SageMaker ocorre exclusivamente dentro da rede da AWS através de **VPC Endpoints (PrivateLink)**, garantindo que dados de saúde nunca trafeguem pela internet pública. O **AWS CloudTrail** registra todas as chamadas de API para auditoria, e o **Amazon GuardDuty** monitora continuamente a rede em busca de comportamentos anômalos.

O diagrama da arquitetura OncoVision AI está disposta a seguir e pode ser melhor analisada no link:

<https://edn-project.s3.us-east-1.amazonaws.com/OncoVisionAI.drawio.png>



### 3. Próximos Passos (Evolução Futura)

Para garantir a evolução sustentável do **Oncovision AI** e sua prontidão para um ambiente de produção em escala nacional, os seguintes passos estratégicos foram mapeados:

#### A. Automação de Infraestrutura (IaC e CI/CD)

- **Pipeline de Deploy:** Implementar pipelines de CI/CD utilizando **AWS CodePipeline** e **AWS CodeBuild**. Isso permitirá que qualquer alteração no código das Lambdas ou na configuração da infraestrutura (definida via AWS SAM ou Terraform) seja testada e implantada automaticamente, reduzindo o erro humano.
- **MLOps (SageMaker Pipelines):** Automatizar o ciclo de re-treinamento do modelo. Assim que novos diagnósticos confirmados por biópsia forem inseridos no sistema, um pipeline deve ser acionado para re-treinar o modelo, garantindo que a IA se torne mais precisa ao longo do tempo (Mitigação de *Model Drift*).

#### B. Otimização da Experiência do Usuário (Real-Time)

- **WebSockets via AWS AppSync:** Substituir o mecanismo atual de *polling* (onde o frontend pergunta periodicamente "já está pronto?") por uma conexão **GraphQL Subscription** via AWS AppSync. Isso permitirá que

o médico receba o resultado na tela instantaneamente assim que a análise for concluída, melhorando a percepção de performance e economizando bateria/dados em dispositivos móveis.

### C. Analytics e Saúde Populacional

- **Data Lake de Saúde:** Configurar o **Amazon Athena** para consultar diretamente os dados anonimizados no S3 e DynamoDB. Isso permitirá a criação de dashboards no **Amazon QuickSight** para gestores de saúde pública, visualizando mapas de calor de incidência de risco por região, apoiando a alocação estratégica de mamógrafos e especialistas.