VASTAVIK: Deep Fake Medical Detection System

CPG NO : 10

A pioneering AI-driven system designed to detect manipulated medical images and safeguard diagnostic integrity in healthcare environments.

By: Yash Saxena, Danveer, Nikhil Garg, Preetmannat, Arman

Mentors: Dr. Vibha Jain,
Dr. Komal Bharti

The Challenge: Safeguarding Medical Imaging Integrity

The Problem

- AI-generated deepfakes can manipulate critical medical scans
- Risks include misdiagnoses and fraudulent insurance claims
- Erodes trust in healthcare AI systems

Our Scope

- Detect manipulation in X-ray, CT, MRI, and Ultrasound images
- Ensure integrity across diagnostic workflows and patient records
- Develop a scalable, explainable AI detection system

Project Objectives

1

Develop AI Detection System

Successfully developed an AI-driven system capable of identifying deepfake alterations in medical images.

2

Enhance Medical Imaging Security

Implemented robust security measures to protect the authenticity and integrity of diagnostic data.

3

Evaluate Across Diverse Datasets

Rigorously evaluated the system's performance on various medical imaging datasets.

4

User-Friendly Verification Tool

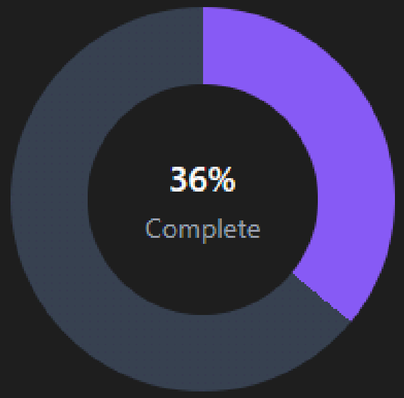
Designing an intuitive interface for radiologists to easily verify image authenticity.

5

Ensure Scalability and Deployment

Optimizing the system for seamless integration into existing hospital PACS/EHR systems.

Overall Progress Distribution



Completed: 2 (0.4%) Pending: 3 (0.6%)

✅ Completed Objectives (2/5)

- Develop AI Detection System
- User-friendly Verification Tool

⏳ Pending Objectives (3/5)

- Enhance Medical Imaging Security
- Evaluate Across Diverse Datasets
- Scalability & Deployment Readiness

No of Project Objectives Approved

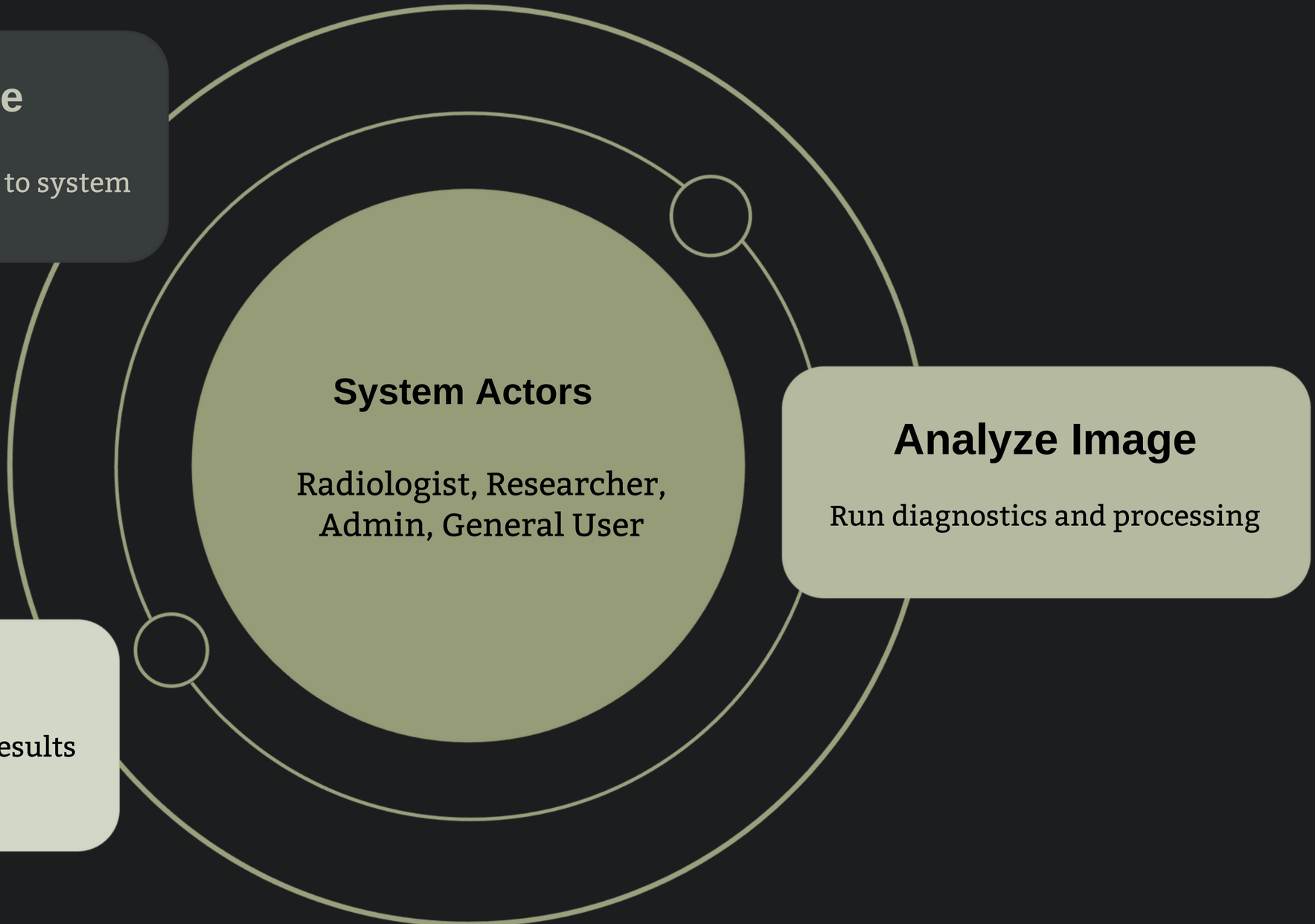
Pie Chart Distribution

Project Analysis & Design

Understanding User Interactions

Upload Image

Submit medical images to system

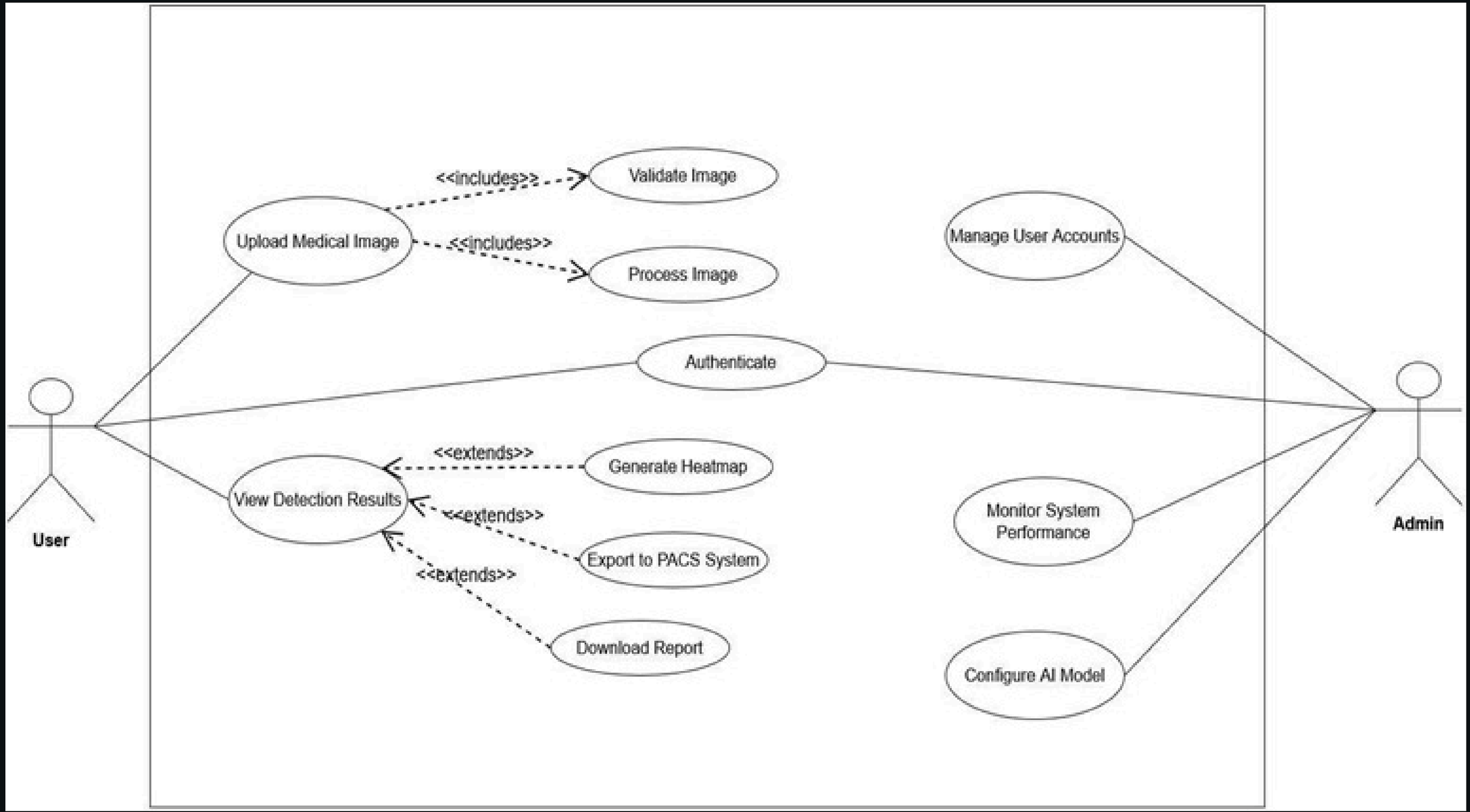


Analyze Image

Run diagnostics and processing

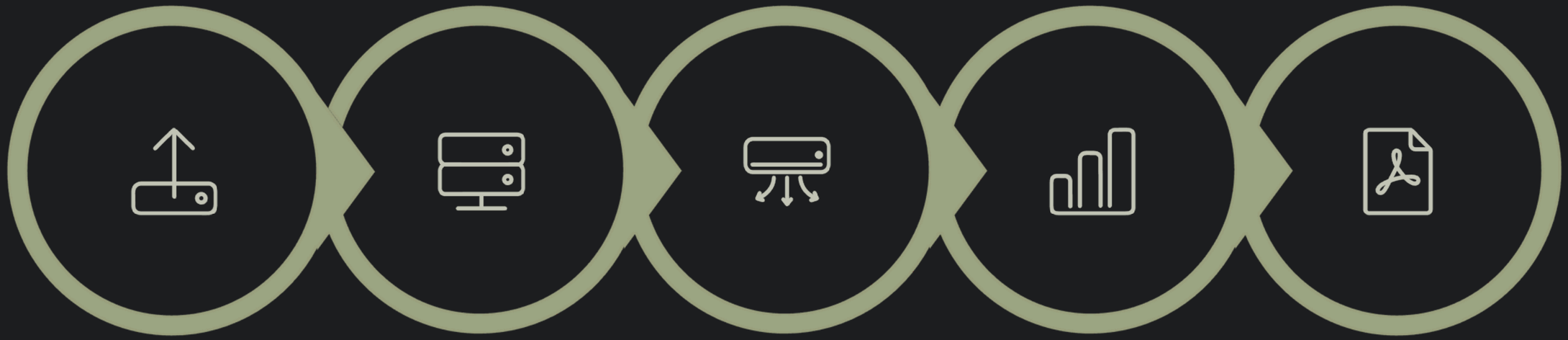
Generate Report

Create and export analysis results


Validate Image
<<includes>>
Upload Medical Image
<<includes>>
Process Image
Manage User Accounts
Authenticate
User
<<extends>>
Generate Heatmap
View Detection Results
<<extends>>
Export to PACS System
<<extends>>
Download Report
Monitor System Performance
Admin
Configure AI Model

Project Analysis & Design

Streamlined Detection Workflow



The Sequence Diagram details the seamless flow of image processing, from user upload to final report generation. This ensures efficient and timely analysis, crucial for clinical environments where rapid, accurate insights are paramount. Each step is designed for optimal performance and data integrity.

Literature Review: Deepfakes in Medical Imaging

Key studies on the emerging threat of deepfakes in medical imaging:

Authors	Study Focus	Key Contributions	Limitations
Mirsky & Lee (2021)	Analysis of adversarial attacks on CT and MRI scans	Showed how fake medical images could deceive radiologists and AI systems	Primarily theoretical, lacked large-scale real-world testing
Chen et al. (2020)	GAN-based synthesis of medical images	Demonstrated realistic synthetic data useful for training models	Ethical concerns about data misuse; authenticity hard to verify
Yu et al. (2021)	Deepfake detection in medical imagery	Achieved >90% detection accuracy in controlled datasets	Performance drops when tested on unseen datasets

These studies highlight the growing concern and need for robust detection systems like VASTAVIK.

Detailed Design

Tools & Architecture

Key Tools

- **Google Colab:** For training and dataset preprocessing.
- **FastAPI + Ultralytics YOLOv8:** Powering the backend inference.
- **React.js, Tailwind, Lucide Icons:** Crafting the intuitive frontend UI.
- **ReportLab:** For generating comprehensive PDF reports.

Data Design

Dataset split into **train, validation, and test** sets, comprising real and fake medical images. YOLO-compatible .txt label files ensure accurate model training.

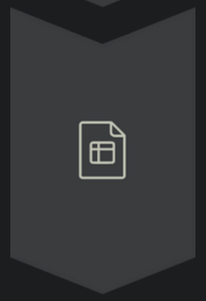
Detailed Architecture Flow

Frontend Interaction

User uploads image; API call initiated.

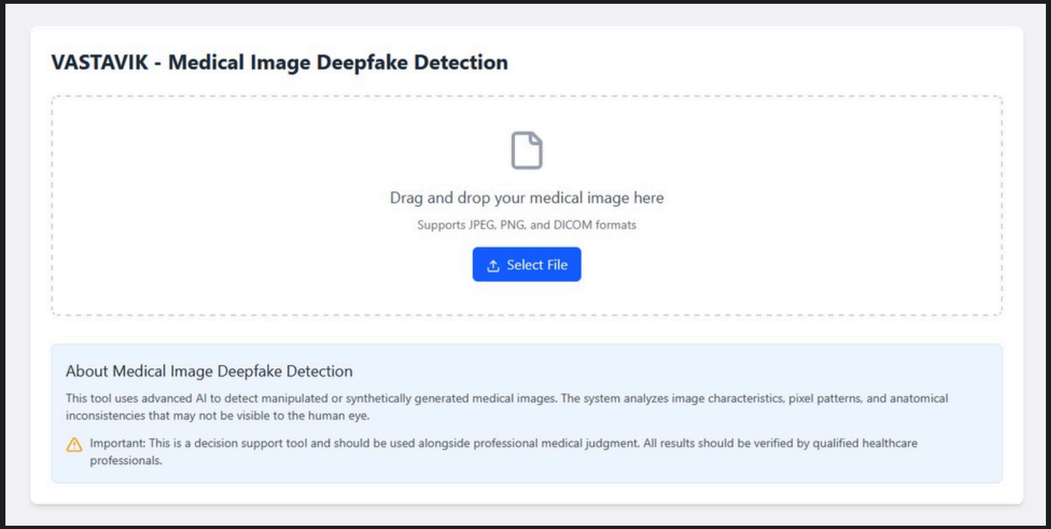
Backend Processing

Image received, YOLO inference performed, anomalies detected.



Report Engine

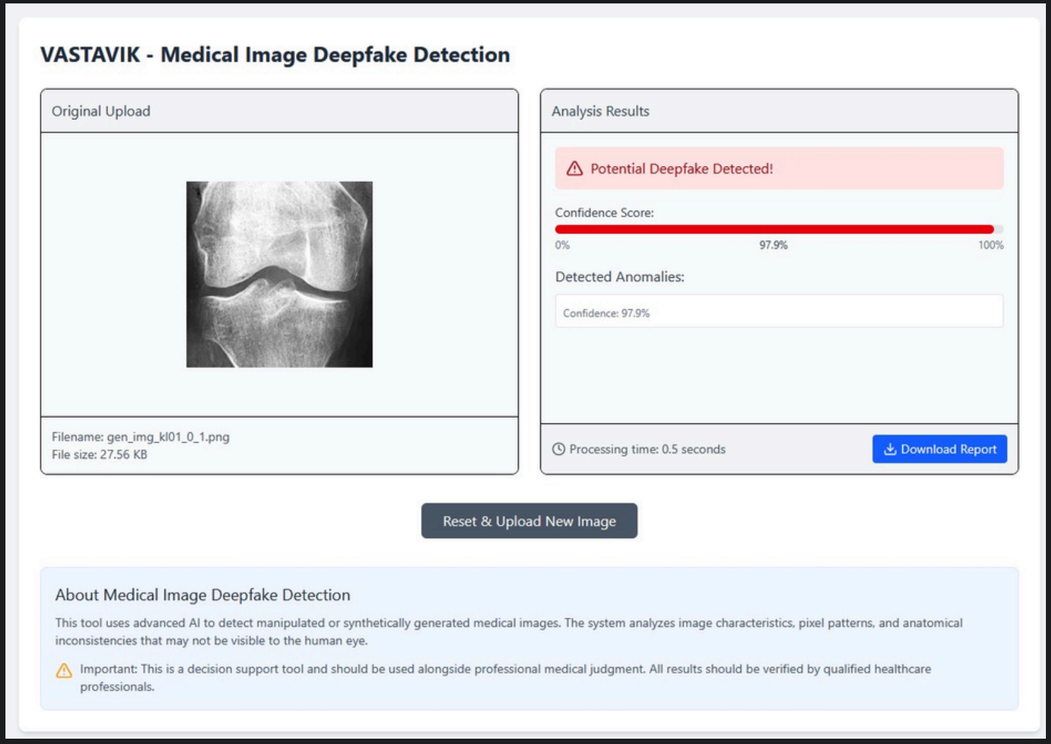
PDF generation including predictions, anomalies, and annotated image.

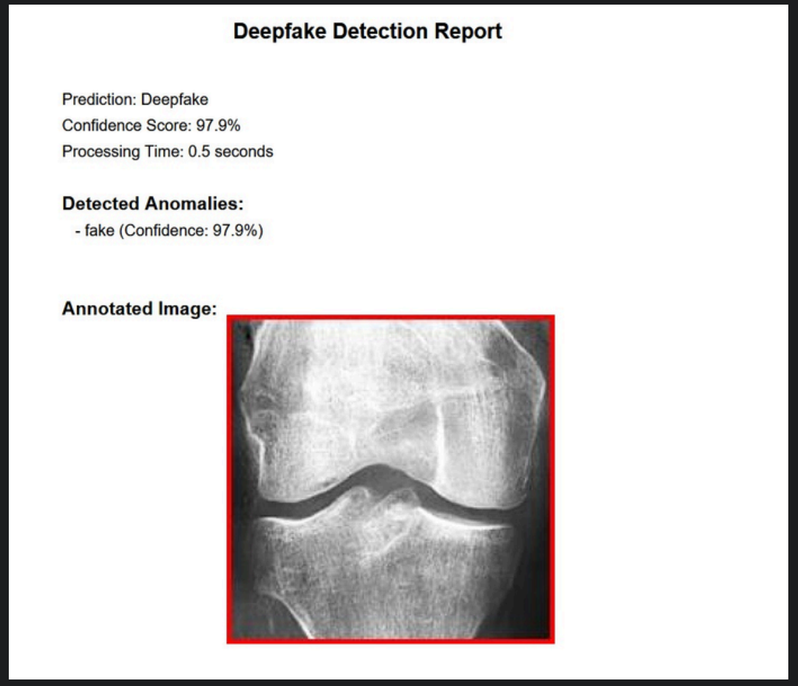


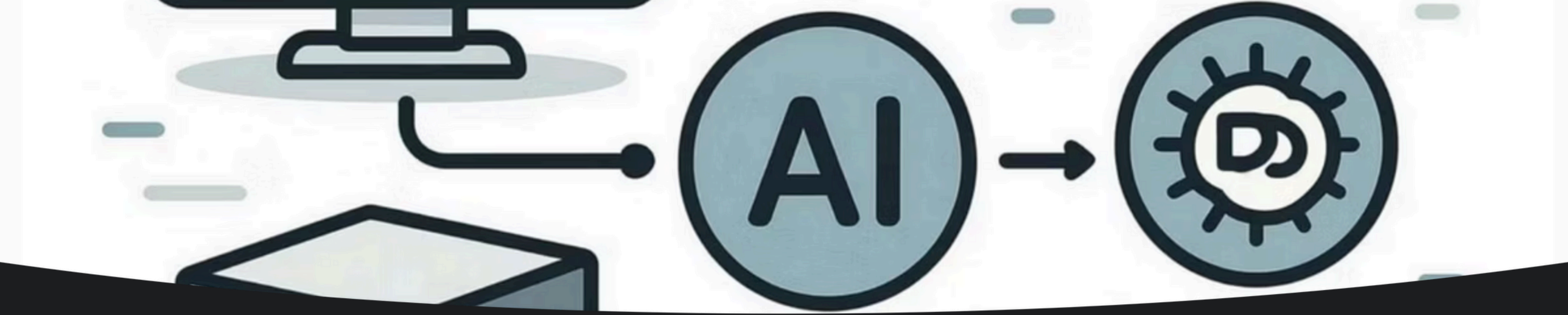
Detailed Design

Intuitive User Interface

The UI is designed for clarity and ease of use, featuring a clean dashboard with drag-and-drop upload functionality. Key elements include a confidence score visualization, immediate deepfake alerts, and a prominent report download button. An informational section also guides users on responsible AI use.







System Architecture: A Robust Foundation

Frontend

Intuitive user interaction via React.js, focusing on seamless workflow for medical professionals.

Backend

High-performance Node.js/Flask pipeline for AI inference and data processing.

AI Models

YOLOv5/v8 for precise tampered region detection and ResNet152 fine-tuned for detailed CT scan analysis.

Enhanced transparency with Grad-CAM heatmaps and YOLO bounding boxes to highlight suspicious areas.

Detailed Architecture: From Upload to Diagnosis

Image Upload

User uploads medical images (DICOM/JPEG/PNG formats).

Preprocessing

Automated resizing, normalization, and augmentation of image data.

Model Inference

Deepfake detection via YOLO and CNN models.

Anomaly Highlight

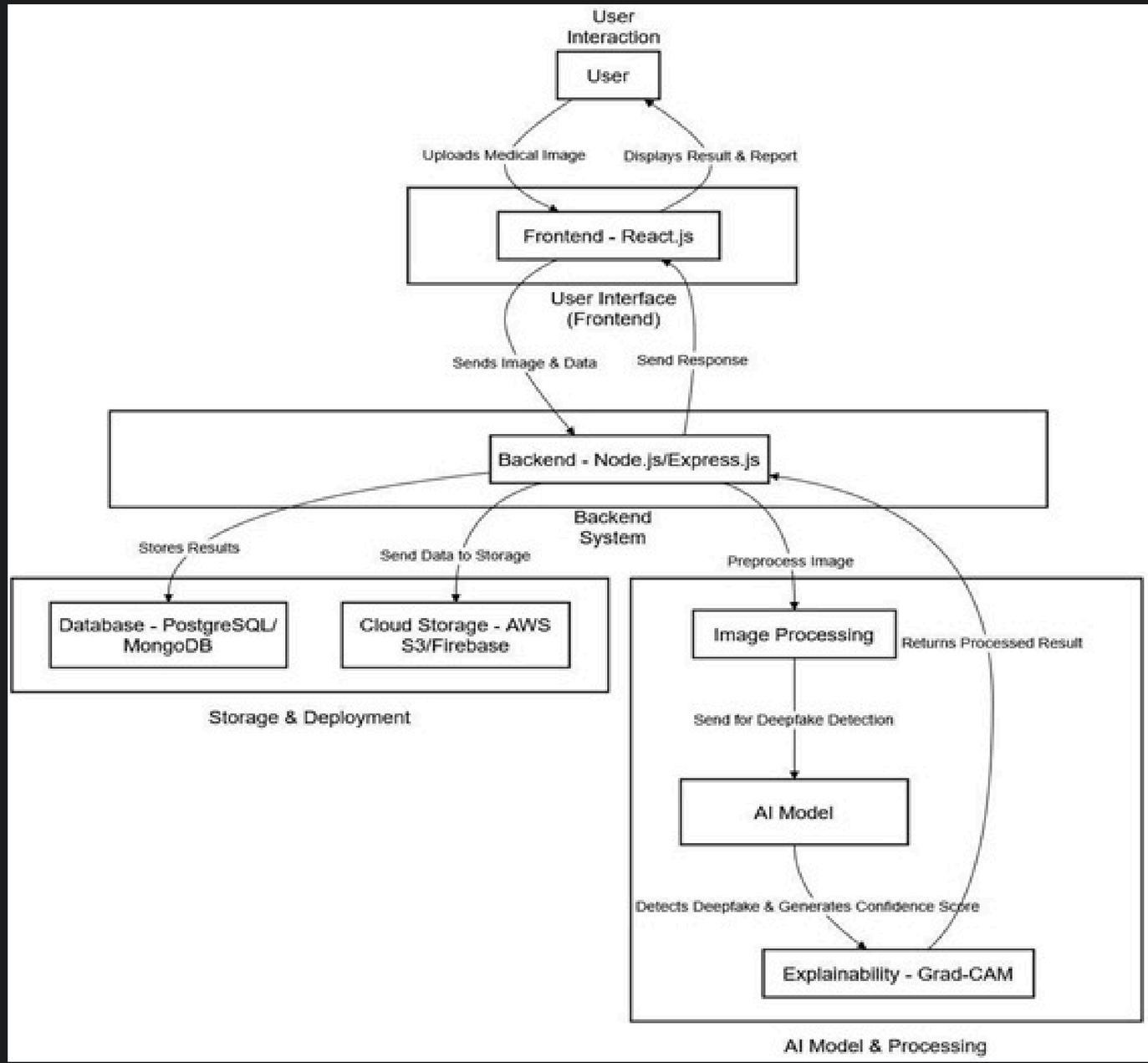
Grad-CAM heatmaps and bounding boxes pinpoint tampered areas.

Report Generation

System generates confidence scores and detailed reports.

Radiologist Review

Results displayed for expert review and diagnostic decision-making.



Project Outcomes

Introducing VASTAVIK: Our Solution

VASTAVIK is a web-based AI platform developed to provide real-time deepfake detection for radiological scans. It offers a secure, professional-grade solution designed to enhance trust and accuracy in medical imaging diagnostics.

- **Real-time Analysis:** Instantly classifies scans as real or fake.
- **Professional Reports:** Generates secure PDF reports with detailed annotations.
- **Web-Based Platform:** Accessible and user-friendly for clinical review.

VASTAVIK's Impact: Key Project Outcomes

Exceptional Accuracy

~80% accuracy on Lung CT datasets with CNN/YOLO and 100% detection accuracy on Knee X-ray datasets using YOLO.

Transparent AI

Visual heatmaps and bounding boxes ensure explainable AI outputs for radiologist confidence.

Real-time Performance

Average detection speed of ~5 seconds per scan, suitable for hospital workflows.

Scalability Ready

Designed for cloud deployment, supporting high volumes of scans in clinical settings.

Project Outcomes

Working Prototype & Core Technologies

Backend: FastAPI & YOLOv8

Powered by FastAPI for efficient API serving and Ultralytics YOLOv8 for deepfake inference on medical datasets.

Frontend: React + Tailwind

Intuitive user interface with drag-and-drop uploads, real-time visualization, and seamless report downloads.

Mathematical Models

Utilizes YOLO loss functions for binary classification, evaluated by Accuracy, Precision, Recall, and mAP.

Performance Metrics: A Closer Look

Knee X-ray Dataset

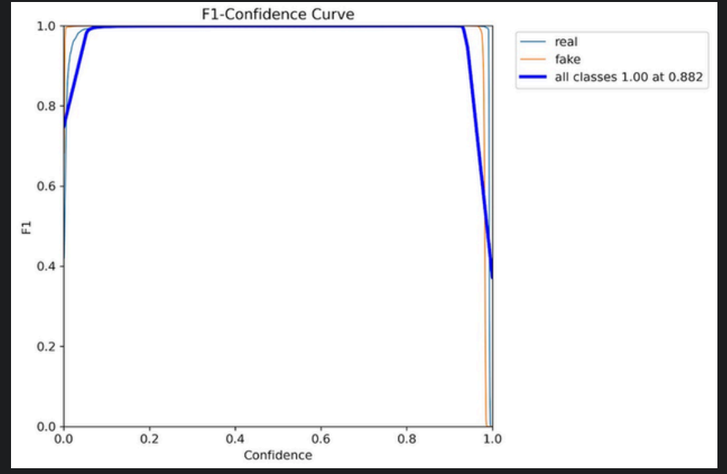
- 100% Accuracy
- Recall of 1.0
- Perfect detection of deepfakes

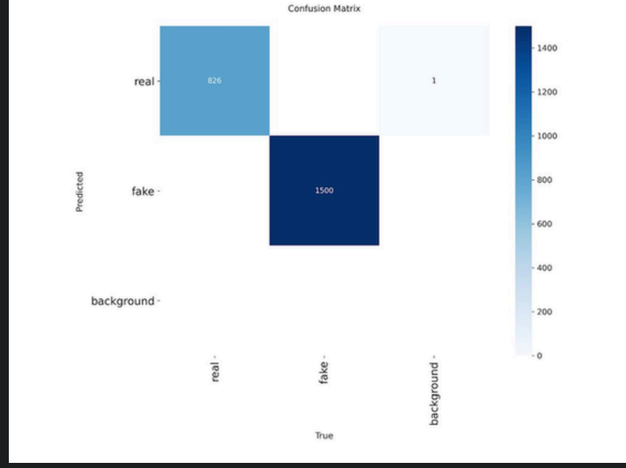
Lung CT Dataset

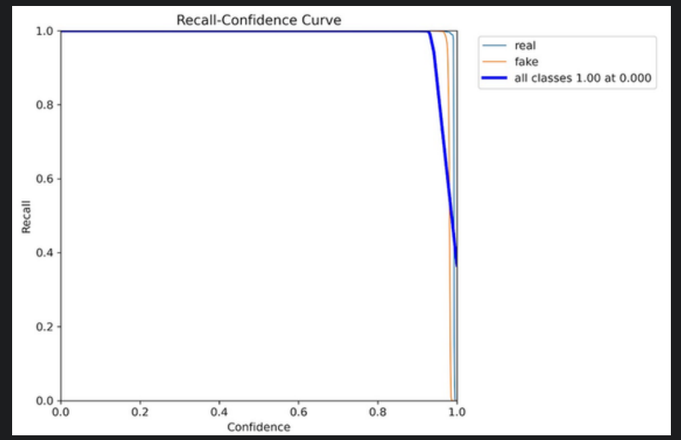
- 80% Accuracy
- Robust detection in complex 3D imaging

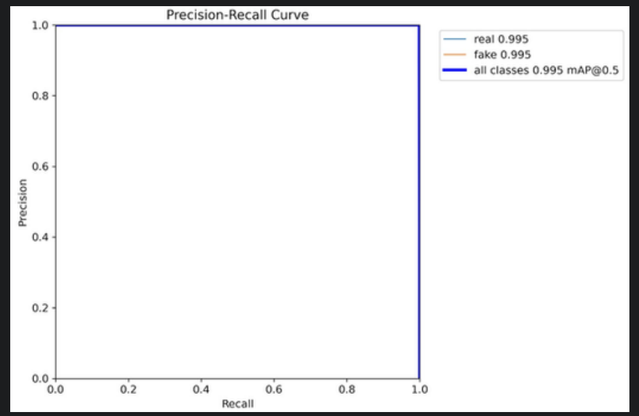
Operational Metrics

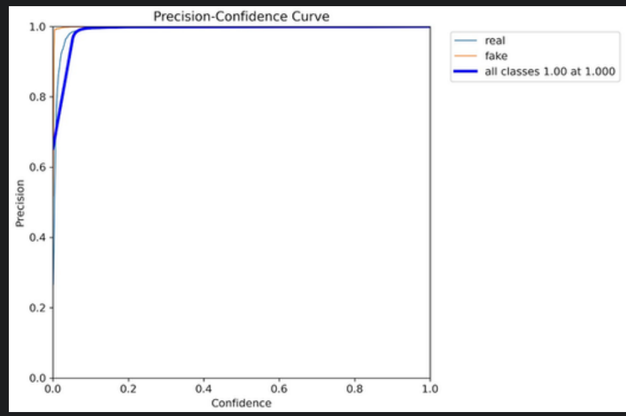
- 5 seconds per scan detection speed
- Fully integrated Grad-CAM visualizations
- Suitable for clinical environments











Project Team:
Our Dedicated Contributors

Nikhil Garg
Integrated the backend + made ppt

Danveer
worked on yolo model implementation and creatued UI

Preetmannat
made Report + backend integration

Yash Saxena
worked on lung ct scan data using cnn and yolo + made ppt

Arman
worked on lung ct scan data using cnn + made report

Mentors : Dr. Vibha Jain
Dr. Komal Bharti

Future Horizons: Expanding VASTAVIK's Reach

Modality Expansion

Extend YOLO-based detection to MRI and Ultrasound imaging.

Dataset Augmentation

Leverage GAN-based techniques to significantly expand training datasets.

Multi-class Detection

Instead of binary (real/fake), classify types of manipulations (insertion, removal, texture tampering)

Explainability & Trust

- Add visual heatmaps showing why the system thinks an area is fake.
- Helps doctors verify results.

Our roadmap includes building a dedicated radiologist dashboard and establishing processes for ongoing error analysis and model retraining to adapt to evolving deepfake techniques.

THANK YOU