# VISIONFUSION

## MULTI-MODAL ANALYSIS TOOL

Overview:

- Purpose: A platform for extracting actionable data and insights from images.
- Key Uses: TACO dataset analysis, glioblastoma early diagnosis, simple image text extraction.

# ANALYSIS CATEGORIES & MODES

Categories of Insight

1. TACO dataset Analysis: Identify and assess gemstones.
2. Glioblastoma Early Diagnosis: Assist in detecting brain tumors.
3. Simple Image Text Extraction: Extract text from various image types.

Analysis Modes

1. Detection: Identifies and outlines objects with confidence levels.
2. Segmentation: Isolates objects from the background.
3. Full Analysis: Combines detection and segmentation for comprehensive results.

# HOW IT WORKS

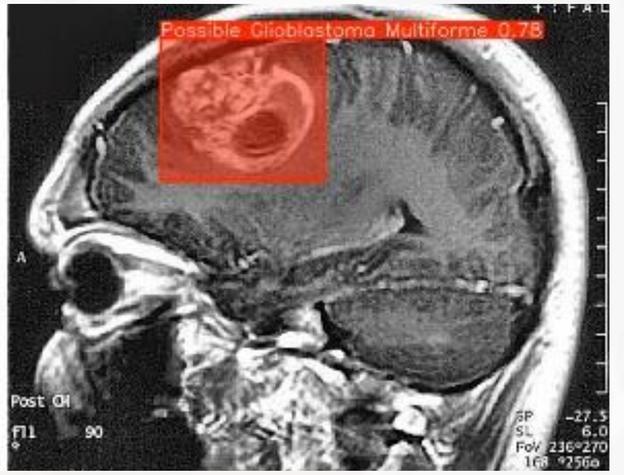
Steps to Analyze Your Images

1. **Image Upload:** Drag and drop your image into the tool.
2. **Select Analysis:** Choose the category and mode (Detection, Segmentation, or Full Analysis).
3. **Pipeline for Custom Training:** The tool allows users to train their own datasets, offering flexible detection and segmentation results tailored to specific needs.
4. **Analyze:** Click "Analyze" to start processing.
5. **Results:**
    - Visuals: Detection plots, outlined objects, segmented images.
    - Data: Downloadable CSV with object details, folder of segmented objects.

# WHAT CAN THE TOOL DETECT?

What Can the Tool Detect?

1. Text from Images
   - Extracts text from documents, signs, handwritten notes, and artwork using advanced Optical Character Recognition (OCR) technology.
   - Tool Used: Tesseract OCR for accurate and fast text extraction.
2. Brain Tumors (Glioblastoma Detection)
   - Assists in identifying early signs of glioblastoma from medical imaging (e.g., MRI scans).
   - Aids medical professionals in early detection, potentially leading to quicker diagnosis and treatment.
3. TACO
   - Detects and classifies different types of waste in images.
   - Useful for environmental and waste management analysis, enabling efficient waste identification and processing.




# VISION & MISSION

## BEHIND THE SCENES

- Image Processing: Preprocessing, normalization.
- Detection & Segmentation: YOLO model for fast and efficient analysis.
- Text Extraction: Tesseract OCR for accurate text recognition.
- AI Summarization: LangChain and Gemini AI for insightful summaries.

## CHALLENGES & SOLUTIONS

- Initial Hurdle: DETR’s long training times.
- Adaptation: Shift to YOLO for faster training and real-time performance, with selective use of DETR where beneficial.

# FUTURE DIRECTIONS

Future Possibilities

- Advanced Summarization: Richer, more detailed summaries with LLM models.
- Expanded Capabilities: Color analysis, shape detection, and more.
- Database Integration: Secure storage and organization of results.

Get Involved

- Explore the project repository, contribute improvements.
- Visuals: Repository links, contact information.

# CONTRIBUTIONS

## Anubhav Mazumder (22051145)

- Project Lead: Directed conceptualization, development, and strategic oversight.
- Model Optimization: Led the integration and tuning of YOLO and DETR models to balance performance and accuracy.
- UI/UX Designer: Designed an intuitive interface to streamline user interaction.

## Debjit Mandal (22051069)

- Text Extraction: Spearheaded the implementation of the OCR engine for accurate text recognition.
- Summarization Feature: Developed the summarization system, using language models to create insightful reports.
- Documentation: Managed project documentation and

# THANK YOU

FOR YOUR ATTENTION