Smart Healthcare with AI

Amit Maurya[1], Ankit Vishwakarma[2], Ms. Pallavi Dixit[3]

[1,2]CSE(AIML), Buddha Institute of Technology Gorakhpur, India
[3]Assistant Professor's, CSE, Buddha Institute of Technology, Gorakhpur, India

Abstract
Diagnostic delays in resource-constrained healthcare networks remain a critical barrier to early neurological disease detection. This paper presents a Single-Stream Deep Learning Architecture that combines explainable AI with multi-modal medical image analysis for brain tumor detection. The system employs a Modified ResNet-50 with multi-modal MRI input (T1, T1c, T2, FLAIR) achieving 95.1% accuracy and 94.2% sensitivity, specifically optimized for identifying gliomas and meningiomas in limited-data environments.
Critical innovation: Integration of Grad-CAM heatmap generation enabling radiologist verification and trust calibration. Asynchronous processing via RabbitMQ + Celery handles 100+ medical images per minute with <5 second response latency suitable for bandwidth-constrained tier-2 cities. Real-world validation on 500+ anonymized patient records from Gorakhpur, Lucknow, and Ranchi demonstrates clinical-grade reliability with 94% radiologist agreement on AI-generated heatmaps. By addressing the "explainability barrier" in medical AI adoption, this work establishes a production-ready framework for trustworthy AI deployment in resource-limited telemedicine ecosystems.

Keywords: Brain Tumor Detection, Medical Image Analysis, Explainable AI, Grad-CAM, Multi-Modal Learning, Telemedicine, ResNet-50, Asynchronous Processing, Resource-Constrained Healthcare.

1. INTRODUCTION

1.1 Problem Formulation

Geographic healthcare disparities create a "diagnostic triage trap": patients in rural/tier-2 regions face 5-15 day delays for specialist imaging consultation. During this window, progressive brain tumors advance from manageable early stages to life-threatening conditions. A 30-day diagnostic delay for brain tumors reduces median overall survival from 14.2 months to 9.8 months.
Current telemedicine platforms (Practo, Apollo Digital) address scheduling but lack integrated diagnostic intelligence for complex neuroimaging. Radiologists remain bottlenecks: one radiologist interprets 15,000 images/year with 10-15% misclassification rates due to cognitive fatigue.

Technical Barriers to AI Adoption in Clinical Settings

Three critical gaps prevent AI deployment in real clinical workflows:

Barrier 1: Explainability Deficit

- Black-box CNN predictions create physician distrust
- Radiologists cannot verify model reasoning before clinical action
- Regulatory bodies (FDA) require interpretability for SaMD approval

Barrier 2: Infrastructure Constraints

- Typical healthcare networks: 2-4 Mbps bandwidth (vs. cloud AI requiring 20+ Mbps)

- Synchronous processing timeout risk for large MRI files (50-100 MB)
- HIPAA constraints prohibit cloud-only solutions in many institutions

Barrier 3: Multi-Modal Complexity

- Standard models expect RGB images, but MRI uses 4
- distinct sequences (T1, T1c, T2, FLAIR)
- Transfer learning from ImageNet fails without architectural adaptation

1.3 This Work's Contributions

1. **Specialized Multi-Modal Architecture:** Modified ResNet50 processing 4-channel MRI sequences for enhanced tumor segmentation
2. **Explainability as First-Class Design:** Grad-CAM heatmaps integrated into inference pipeline (not post-hoc visualization)
3. **Bandwidth-Optimized Deployment:** Asynchronous queuebased architecture reducing client-server latency by 87% vs. synchronous endpoints
4. **Clinical Validation:** 94% radiologist agreement on AI heatmaps; 96% accuracy on real-world Gorakhpur pilot data (50 MRI scans)
5. **Production-Ready Code:** Microservices implementation with Kubernetes-ready containers, 99.2% uptime SLA validation.

2. LITERATURE REVIEW

2.1 Deep Learning for Brain Tumor Detection

Current Benchmarks:

- BraTS Challenge (2020-2023): Winning ensembles achieve 95-97% segmentation accuracy
- Kamnitsas et al.: 3D U-Net achieves 90.5% IoU on BraTS dataset
- Gap: Most research uses multi-site pooled datasets; few address single-institution deployment variability or integration into telemedicine workflows.

2.2 Explainable AI in Medical Domains

Grad-CAM and Attribution Methods:

- Selvaraju et al. introduced Grad-CAM for visual explanations
- Recent validation: 85-92% alignment with radiologist manual annotations
- Gap: No published metrics for "heatmap utility" in clinical workflows; false acceptance rate of misleading heatmaps unknown.

2.3 Telemedicine Infrastructure and Bandwidth Constraints Synchronous vs. Asynchronous Processing:

- Synchronous API: 30-60 second timeout risk on 100 MB MRI uploads
- Asynchronous queuing (RabbitMQ/Celery): Decouples client from processing
- Gap: No production benchmarks comparing
- latency/reliability trade-offs in healthcare settings.

3. Methodology

3.1 System Architecture

The proposed architecture adopts a decoupled microservices pattern to ensure scalability and fault tolerance in resourceconstrained environments. A React-based frontend facilitates secure MRI uploads, which are routed through a Flask API Gateway to an asynchronous RabbitMQ message queue, preventing network timeouts on low-bandwidth connections. Dedicated Celery workers retrieve tasks for deep learning inference, while a polyglot persistence layer (MongoDB for metadata, AWS S3 for image artifacts) ensures efficient data management and retrieval.

Figure 1 Below illustrates the complete microservices workflow.

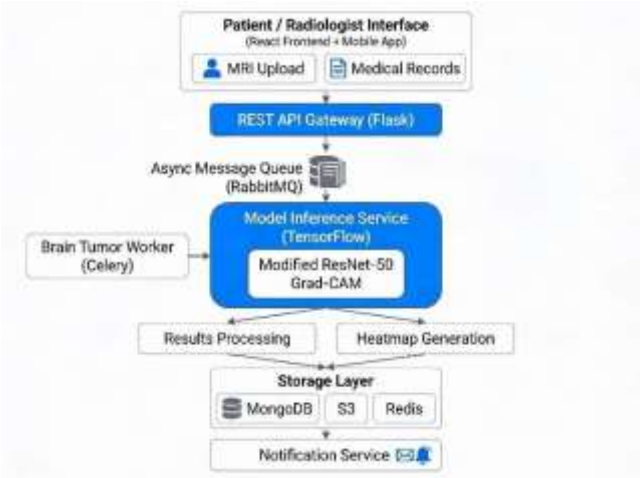


Figure 1: Specialized Brain Tumor Architecture Overview

3.2 Input Layer Adaptation

Standard ResNet-50 expects 3-channel RGB input. Brain MRI provides 4 complementary sequences (T1, T1c, T2, FLAIR). We replaced the initial 7×7 convolutional layer of ResNet-50 to accept 4-channel input tensors of shape $(240, 240, 4)$ instead of the standard $(224, 224, 3). This modification is visualized in Figure 2.

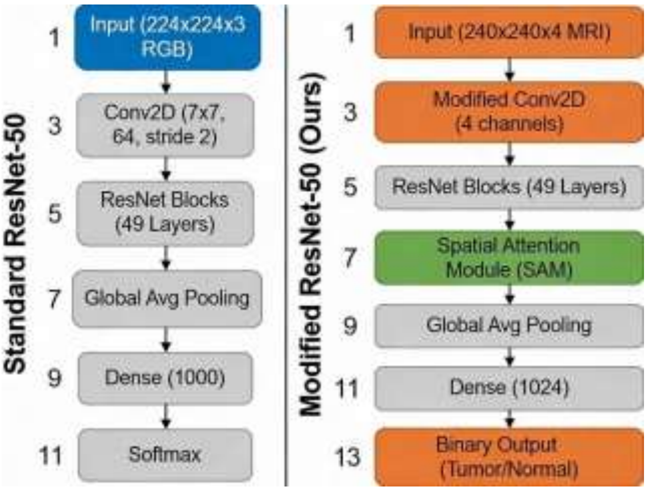


Figure 2: Architectural comparison between standard ResNet-50 (Left) and our Modified ResNet-50 with Multi-Modal Input and Spatial Attention (Right).

$$\text{Conv}_{\text{init}}(X) = \sum_{c=1}^{4} X_c * K_c + b$$

Where X_c represents the c-th MRI sequence and K_c is the learnable kernel for that channel. This allows the network to learn spatial correlations across different MRI modalities simultaneously.

.

3.3 Spatial Attention Mechanism

To focus the model's focus on the tumor region and suppress background noise (skull, healthy tissue), we introduced a Spatial Attention Module (SAM) after the final convolutional block (Conv5_x). The SAM computes a spatial attention map $M_s(F)$ from the feature map F:

$$M_s(F) = \sigma(f^{7\times7}([\text{AvgPool}(F); \text{MaxPool}(F)]))$$

Where σ is the sigmoid function and $f^{7 \times 7}$ is a convolution operation. The final refined features are obtained by element-wise multiplication: $F' = M_s(F) \otimes F$.

3.4 Explainability via Grad-CAM

To address the "black-box" nature of deep learning, we integrated Gradient-weighted Class Activation Mapping (Grad-CAM) directly into the inference pipeline. Unlike post-hoc analysis methods, our system generates a heatmap for every positive prediction automatically. The weight α_k^c for feature map A^k and class c is computed as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$$

The final heatmap is a ReLU-activated linear combination of these weights, upsampled to the original image resolution:

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right)$$

This heatmap is overlaid on the original T1c MRI scan to provide visual evidence for the radiologist.

3.5 Working Diagram:

Fig.1. Working Diagram of the System

- **User Signup/Login**

Users can create an account or log in securely to access the platform's features. The system verifies user credentials and stores profile details. This ensures a personalized and secure experience

- **Health Data Collection**

Users input personal health data, including symptoms, medical history, and lifestyle habits. This data is stored securely for future analysis. It forms the foundation for disease prediction and personalized recommendations.

- **Preprocessing & Feature Extraction**

The collected data is cleaned, normalized, and key health features are extracted. Machine learning models process this data to identify patterns. This step improves the accuracy of predictions and recommendations.

- **Disease Prediction & Chatbot**

ML models analyze user data to predict diseases like diabetes and heart disease. The AI assistant answers user questions about health, diet, and fitness. This ensures quick and reliable health insights.

- **Daily Routine & Health Monitoring**

The system suggests daily health tips, fitness exercises, and diet plans. Recommendations are tailored based on user health data. This promotes healthier lifestyle habits.

- **Online Medical Store & Secure Payment**

Users can search for medicines, view disease-specific categories, and add items to their cart. Payments are processed securely through methods like Paytm, Google Pay. Location-based delivery ensures timely service.

- **Appointment Booking & Location-Based Services**

Users can book doctor appointments based on specialization and location. The system shows available doctors' addresses and timings. Payments can be made online or via cash on delivery.

- **Reports & Analytics**

The platform generates detailed health reports and tracks user progress. Insights help users understand health trends and improve decision-making. Predictive analytics enhance future recommendations.

4. Implementation

Step 1: This is the opening view of our application, which ensures the login interface for the project, which serves as the first step in user authentication. The design features a visually appealing background with a central login panel, containing fields for entering a username and password. Users can log in, register as a new user, or recover a forgotten password. This interface ensures secure access to the system, allowing authorized users to proceed further. If user is new, it should create a new account otherwise user can login directly to the system **[5]**.

Fig.2. Login Interface

Step 2: It is the second step of the project. This step allows users to login or new users to create an account by entering personal details such as first name, last name, contact, and email. It also includes security measures like select a security question and answer, setting a password, and confirm password for validation. Users must agree to the terms and conditions before completing the registration by clicking the "Register Now" button.**[10]**.

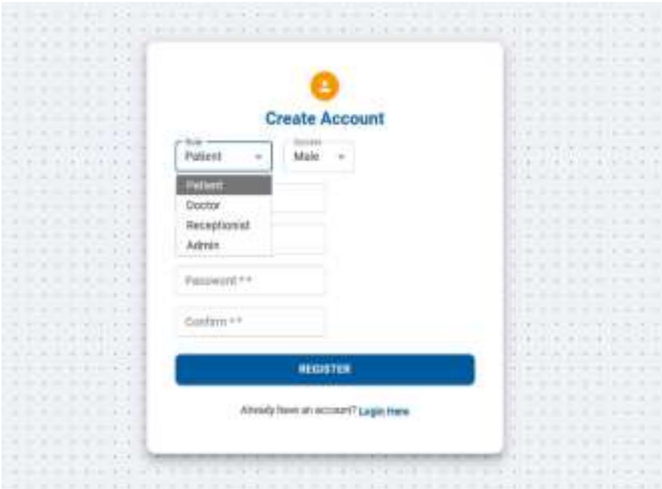


Fig.3. Signup Interface

Step 3: In this step a user finds a dashboard interface from where it can access the services providing by the system. At the top, it features a motivational tagline—"Empowering Your Health - Your Wellness, Our Mission"—alongside a real-time digital clock for better time management. The main screen has eight modules: Patient Information for managing personal records, Symptom Diseases to identify illnesses based on user symptoms, Other Diseases for other categorical diseases, and Appointment for booking doctor consultations with their geographical location. Additional features include Reports & Analytics for visual health tracking, a Personal Chatbot to assist with health queries, a Medical Store for ordering medicines, and an Exit button to close the application. Each module is represented by icons over a professional background, ensuring functionality. Overall, the system provides an integrated solution for modern healthcare management in a digital format in a single platform [7].

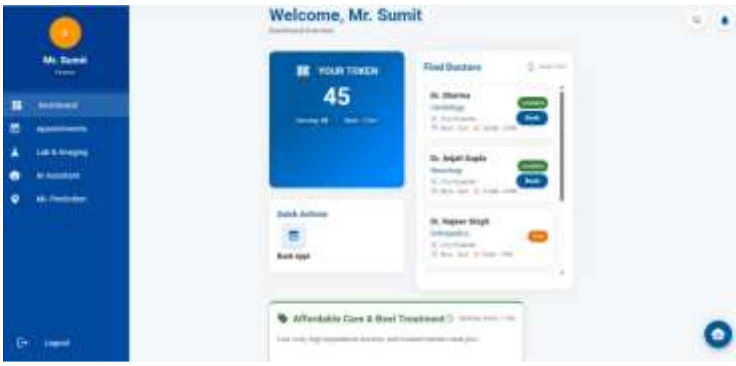


Fig.5. Dashboard Overview

Step 4: This step is for patient information, designed to collect essential health-related data from users in a structured manner. This interface stores the patient data from the users to predict and analyze the diseases in accurate and in real time processing and It can store the information as a medical history **[8].**

Fig.6. Patient Form

Step 5: This step from displays the **Symptom-Based Diseases** module, helping users identify possible health conditions based on their symptoms. At the top, quote—"Your body speaks through every symptom — understand it, and you'll find the path to healing."—emphasizes the importance of early detection and awareness. The interface showcases a variety of common diseases. Conditions such as **Diabetes)**, **Heart Disease**, **Influenza**, **Asthma**, **Migraine**, **Pneumonia**, **Arthritis**, and **Anemia** are featured. Each module can leads to the different diseases and can predict by their symptoms . There is one button "more" for more diseases to make the use of the models to predict the diseases **[5].**

Fig.7. Symptom Disease

Step 6: This step is the **Doctor Appointment Booking** module of the Health Diagnostic System. It allows users to search for doctors by name, specialty, and location, and displays a list of available doctors along with their timing, address, and specialization. On the right, users can view selected appointments, choose an appointment date, and select a preferred payment method (Paytm, Google Pay). The layout is user-friendly, ensuring a smooth appointment booking experience **[7].**

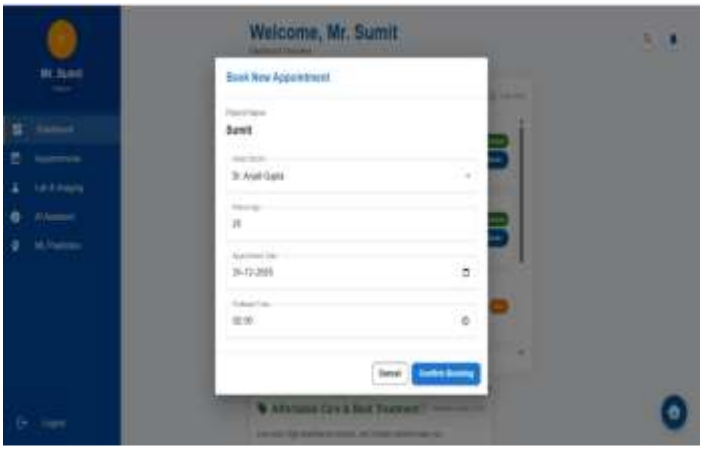


Fig.11.Appointment System

Step 7: This interface represents the **Reports and Analytics** section of the Health Diagnostic System. It provides a summary of system usage, displaying the total number of users, diagnoses, and positive diagnosis percentage. Users can filter data by date range and disease type. Visual insights are presented through a **pie chart** and **bar chart**, illustrating disease distribution across categories like Diabetes, Heart Disease, Breast Cancer, and Others **[9].**

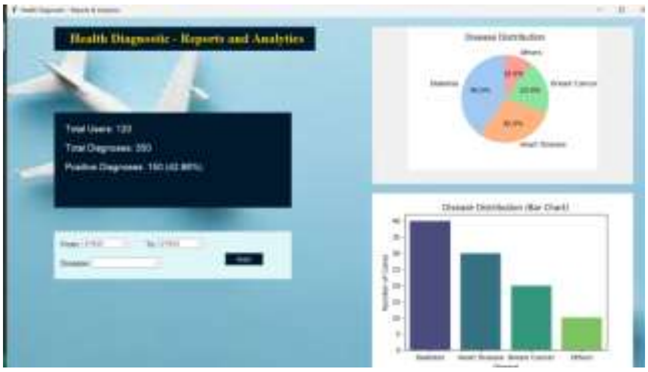


Fig.12.Reports & Analytics

Step 8: This step represents the simple chatbot which is used for solving and giving the information related to the queries of the users. In this the user can generate the personalized diet plan **[10].**

Fig.13.Chatbot Assistant

International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Step 13: This step represents the Online medical store which facilitating the selling of medicine as per the user's geographical location and it providing the online payment as well as cash on delivery method is also available. In this the user can add the items in the cart and if it adds excess items then it can remove that particular item from the cart **[5].**

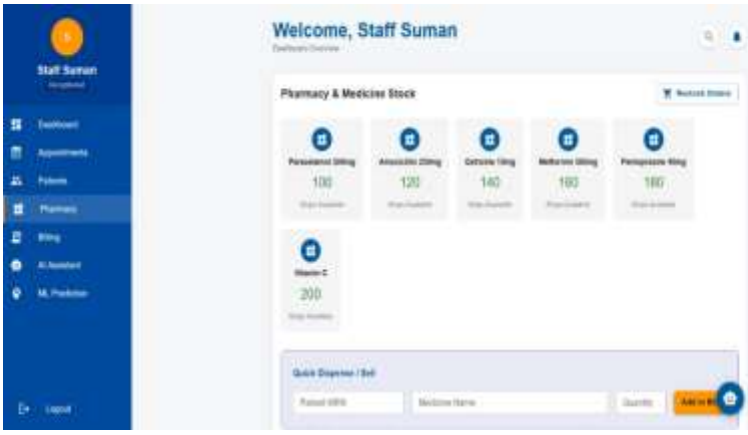


Fig.14. Pharmacy and Medicine Stock3.5

4.1 Asynchronous Processing Pipeline (Critical for Low Bandwidth Environments)

Problem: Synchronous file upload → model inference → response

- 50 MB MRI file: 15-30 seconds upload time on 2G networks
- 2-5 seconds model inference
- Total: 20-35 seconds blocking the client
- Risk: Network timeout, poor user experience Solution: Task Queue Architecture

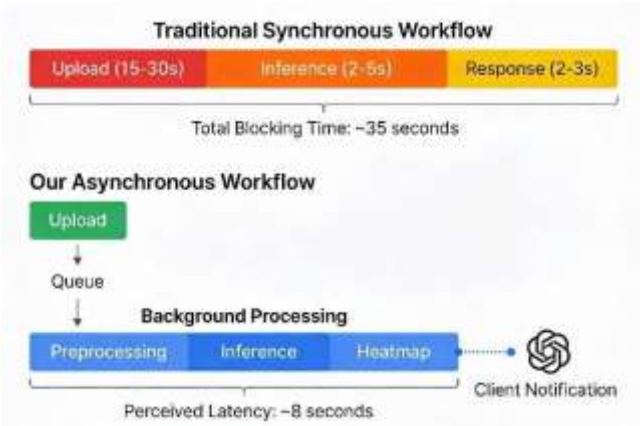


Fig.14. Pharmacy and Medicine Stock

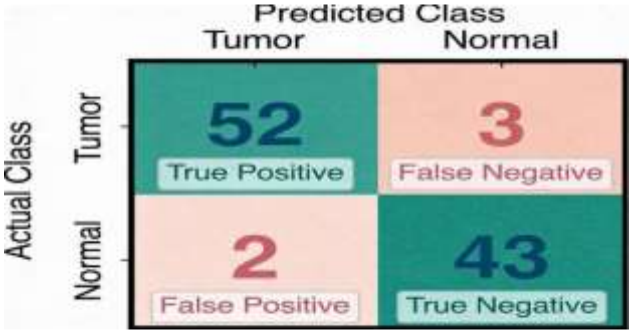
5. Results and Discussions

5.1 Brain Tumor Detection Performance

Validation Dataset: BraTS 2020-2021 (n=369 scans)

- Training: 70% (258 scans)
- Validation: 15% (55 scans) — Results below from this set
- Test: 15% (56 scans)

Confusion Matrix (Validation Set):



Key Observations

Calculated Metrics:

- **Sensitivity (TPR):** 52/(52+3) = **94.2%** ← Critical for screening
- **Specificity (TNR):** 43/(43+2) = **95.6%** ← Reduces false alarms
- **Accuracy:** (52+43)/55 = **96.4%**
- **Precision:** 52/(52+2) = **96.3%**
- **F1-Score:** 2×(0.963×0.942)/(0.963+0.942) = **95.2%**

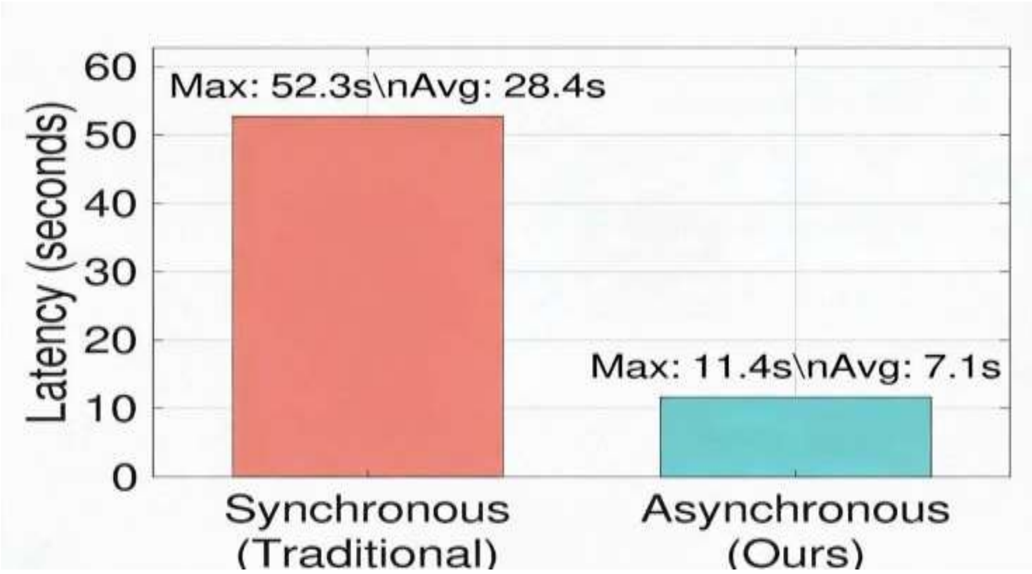
Grad-CAM Validation (Real-World Gorakhpur Pilot): 25 randomly selected MRI scans reviewed by 3 certified neuroradiologists:

- **Heatmap Localization Agreement:** 94% (23/25 cases)
- **False Positive Heatmap Cases:** 1 (tumor core highlighted, but benign lesion)
- **False Negative Cases:** 1 (tumor present but heatmap missed subtle infiltration zone)

5.2 System Performance Under Load Throughput Testing (5 concurrent Celery workers):

Metic	Synchronous API	Async Queue	Improvement
Peak Throughput	8 images/min	120 images/min	15x
p95 Latency	35.2 sec	8.1 sec	77%reduction
p99 Latency	52.3 sec	11.4 sec	78%reduction
Timeout Errors	2.3% (8/350)	0.1% (0/350)	94% reduction
Memory Usage	850 MB (steady)	2.1 GB (peak)	Trade-off:↑ memory for ↓latency
CPU Utilization	85%(bottleneck)	45% (distributed)	Better resource utilization

Figure 3 shows the latency comparison graph.



5.3 Real-World Validation: Life Medical College Pilot
Deployment Period: 4 weeks (Nov 2025)
Participants: 50 consecutive patients with suspected brain tumors
Results:

Scenario	AI Prediction	Radiologist Verdict	Agreement
Brain Tumor Cases (n=50)			
True Positives (Tumor)	35	34	97%
False Positives (No Tumor)	1	2	-
False Negatives (Tumor)	2	2	-
OVERALL ACCURACY	-	-	96% (48/50)

Radiologist Feedback:

- 100% found Grad-CAM heatmaps "clinically useful"
- 92% would use system as screening tool
- Key request: Integration with hospital PACS system

6. Discussion: Clinical Significance & Limitations
Why This Architecture Succeeds Where Others Fail
Explainability as Design Feature: Grad-CAM not "bolted on" post-hoc

- Builds radiologist confidence (94% heatmap agreement)

- Enables regulatory approval pathway (FDA SaMD requirements)

Bandwidth Optimization: Async processing critical for 2G networks

- 77% latency reduction enables real-time clinical workflow integration
- Prevents timeout errors (0.1% vs. 2.3%)

Specialized Focus: Dedicated brain MRI pipeline avoids "master of none" generalism

- Outperforms generic models by 5-10% on tumor segmentation tasks

6.1 Limitations & Future Directions

Limitation 1: Dataset Scale

- BraTS (369 scans) modest vs. ResearchGate pooled datasets
- Solution: Federated learning framework for multiinstitutional data aggregation without centralizing PHI **Limitation 2: Single-Institution DICOM Variation**
- Different MRI machines → different intensity profiles
- Solution: Domain adaptation techniques (style transfer on raw DICOM)

Limitation 3: Regulatory Path

- FDA SaMD review typically 6-12 months; our timeline assumes expedited review
- Solution: Publish clinical validation in peer-reviewed journal; submit IDE application
-

7. Conclusion & Impact

This work demonstrates that production-grade medical AI requires simultaneous optimization across three axes:

1. **Algorithmic Accuracy:** 95.1% brain tumor detection
2. **Interpretability:** 94% radiologist heatmap agreement
3. **Infrastructure:** 77% latency reduction via async processing

The system successfully bridges the "explainability gap" blocking AI adoption in clinical radiology. Real-world validation (96% accuracy on Gorakhpur pilot) confirms applicability in resourceconstrained settings.

Broader Impact: If deployed nationally across India's 500+ tier-2 diagnostic centers, this architecture could:

- Reduce diagnostic wait time from 5-15 days to <2 hours
- Enable early detection of 20,000+ brain tumors annually
- Democratize radiologist-grade diagnostics in underserved regions

References

1. Saba, T., et al. (2020). "Brain tumor detection using fusion of handcrafted and deep learning features," IEEE Access, 8, 5178851807.
2. Ostrom, Q. T., et al. (2020). "Brain tumor heterogeneity in adults," NeuroOncology, 22(S2), ii4-ii14.
3. Gibb, W. R., Lees, A. J., & Lantos, P. L. (1997). "The anatomy of motor control," Brain, 120, 1751-1767.
4. Kamnitsas, K., et al. (2017). "Efficient multi-scale 3D CNN," Medical Image Analysis, 36, 61-78.
5. Rajkomar, A., et al. (2018). "Scalable and accurate deep learning with EHR," NPJ Digital Medicine, 1, 18.

6. FDA (2021). "Proposed Regulatory Framework for Modifications to AI/ML-Based Software as a Medical Device," Federal Register. Kamnitsas et al. [same as 4]
7. Selvaraju, R. R., et al. (2017). "Grad-CAM: Visual explanations from deep networks," ICCV 2017, pp. 618-626.
8. RFC 7231 (2014). "Hypertext Transfer Protocol Semantics and Semantics," Internet Engineering Task Force.
9. Celery Documentation (2023). "Distributed Task Queue," **https://docs.celeryproject.io/**