

# Automated Brain Tumor Detection and Classification System

<sup>1</sup>Uma N, <sup>2</sup>ArunPrakash S, <sup>3</sup>Aswinkumar K, <sup>4</sup>Infant Rohith A

<sup>1</sup>Assistant Professor, <sup>234</sup>III Year Students

<sup>1234</sup>Department of Information Technology

<sup>1234</sup>Sri Venkateswara College of Engineering Chennai Tamil Nadu 602 117

<sup>1</sup>[numa@svce.ac.in](mailto:numa@svce.ac.in), <sup>2</sup>[2023it0810@svce.ac.in](mailto:2023it0810@svce.ac.in), <sup>3</sup>[2023it0911@svce.ac.in](mailto:2023it0911@svce.ac.in), <sup>4</sup>[2023it0921@svce.ac.in](mailto:2023it0921@svce.ac.in)

**Abstract**—The accurate rapid and non-invasive detection of central nervous system malignancies remains one of the most critical challenges in contemporary neurology. Magnetic Resonance Imaging serves as the undisputed gold standard for visualizing these complex pathology however, the manual interpretation of multi parametric volumes is cognitively exhaustive and severely bottle necked by a global shortage of specialized anesthesiologists. While recent paradigms in deep learning have achieved unprecedented diagnostic accuracy they of ten do so at the cost of prohibitive computational complexity. Foundational work established the current state of the art on a comprehensive four class brain classification task using a fine tuned object detection architecture. However this apex performance is inextricably linked to extreme hardware requirements specifically reliance on dedicated graphics processing units with massive memory reserves. Such hardware dependency renders the deployment of this model virtually impossible in the vast majority of primary healthcare facilities particularly within low and middle income countries. To bridge this critical chasm this thesis presents a highly optimized universally central processing unit deploy able alternative. Our proposed system is engineered upon a highly efficient lightweight backbone strategically retaining advanced spatial attention mechanisms contextual pooling and bidirectional feature networks. The proposed system achieves a highly competitive ninety four percent overall accuracy while enabling under two second inference per image. Furthermore we extend the base architecture with a complete clinical workflow including a robust relational patient database a novel quantitative longitudinal tumor growth tracking engine and an automated diagnostic re- port generator all integrated into a clinical web dashboard.

**Index Terms**—Brain Tumor Classification, Medical Image Processing, Longitudinal Growth Tracking, Edge Computing Deployment, Healthcare Artificial Intelligence.

## I. Introduction

A brain tumor is biologically characterized by the abnormal unregulated and uncontrolled proliferation of cells within the rigid unyielding confines of the cranial cavity. Given the critical highly localized and interconnected functions of the human brain any such anomalous mass regardless of whether it is historically classified as benign or malignant poses a profound and immediate threat to patient health neurological function and ultimate survival. The World Health Organization formally classifies central nervous system tumors into four distinct malignancy grades based on their eschatological characteristics mitotic activity and the presence of necrosis. Glob- ally the incidence and prevalence of brain tumors represent an escalating public health crisis with hundreds of thousands of new diagnoses projected annually. Magnetic Resonance Imaging remains the undisputed gold standard non invasive imaging modality for the visualization and diagnosis of these tumors. Unlike other scanning methods it utilizes powerful magnetic fields and radio waves to provide vastly superior soft tissue contrast allowing for the precise spatial delineation of tumor boundaries internal heterogeneous structures and surrounding peritoneal swelling.

However the manual interpretation of these complex multi slice volumes is an incredibly cognitively demanding task. A standard imaging study can generate hundreds of individual cross sectional images across various sequences. Analyzing these requires highly specialized neurosurgical expertise and is inherently susceptible to fatigue induced errors. This diagnostic challenge is acutely magnified in resource- constrained environments. Within the Indian context the medical system faces a severe crisis in specialized manpower where the radiologist to patient ratio stands at a concerning one per one hundred thousand. This severe workforce shortage translates directly into massive diagnostic backlogs. Patients frequently endure agonizing waiting periods for scan interpretations directly leading to delayed surgical interventions, advanced disease progression at the time of treatment and severely compromised patient outcomes. Consequently the development of automated deep learning driven diagnostic support systems that can accurately triage cases and assist radiologists efficiently at the point of care is an urgent clinical necessity.

Recent foundational research made significant strides in this domain by establishing a new benchmark using advanced object detection algorithms integrated with spatial attention and bidirectional feature fusion. They achieved near perfect over- all accuracy. However a critical pragmatic analysis of this pipeline reveals a severe limitation regarding actual clinical deployment. The architecture possesses an immense parameter count and astronomical computational complexity necessitating the mandatory use of a dedicated high end graphics processing unit. Such computing hardware is distinctly unavailable in the vast majority of primary and secondary care facilities. Furthermore previous contributions are strictly limited to an isolated mathematical model inference script lacking a surrounding clinical workflow ecosystem. There are no provisions for persistent patient database management, no automated mechanisms for the longitudinal tracking of tumor progression over multiple visits and no capabilities for generating formalized diagnostic reports.

The primary objective of this thesis is to democratize access to advanced artificial intelligence diagnostics tools by shifting the engineering paradigm from absolute hardware dependent accuracy to universal hardware agnostic deploy ability. We replace the computationally massive object detection backbone with a lightweight mobile architecture drastically reducing mathematical operations and enabling rapid inference on standard hospital computers. We adapt and integrate highly effective spatial attention and multi scale contextual pooling modules to preserve high level diagnostic accuracy despite the massive reduction in model parameters. Finally we develop a novel algorithmic module capable of comparing historical patient scans to current scans mathematically quantifying tumor pixel area growth and automatically classifying the disease trajectory seamlessly packaging the inference engine into a complete clinical ecosystem.

## II. Related Work

The quest to automate brain tumor detection has spanned several decades evolving from simple heuristic thresholds to complex multi billion parameter neural networks. Prior to the widespread adoption of modern neural architectures automated brain tumor analysis relied heavily on traditional machine learning pipelines. These approaches required manual hand crafted feature extraction which is a highly laborious process requiring deep domain expertise. Techniques such as gray level matrices local binary patterns histogram of oriented gradients and discrete wavelet transforms were utilized to mathematically extract statistical texture features from the raw image slices. These extracted low dimensional feature vectors were subsequently fed into classical statistical classifiers such as support vector machines random forests naive bayes and nearest neighbor algorithms. While early research utilizing support vector machines applied to texture and intensity features achieved promising accuracy rates on heavily constrained small scale proprietary datasets these traditional methods suffered from extreme brittleness and a lack of robustness. They routinely failed to generalize to the immense imaging variability found in real world clinical data. Variations in machine manufacturers differing magnetic field strengths inconsistent contrast agent uptake timing and minor patient motion artifacts

would completely derail the rigid hand crafted mathematical algorithms. The feature extractors were fundamentally unable to understand semantic context as they only understood localized pixel intensity changes.

The advent of convolutional neural networks fundamentally revolutionized the field of medical image analysis by automating the feature extraction process itself. Rather than human engineers guessing which textures or gradients were statistically important the network mathematically learned the optimal hierarchical features directly from the raw pixel data via the back propagation of error gradients. In the early layers these networks learn to detect simple edges while in the deeper layers they learn complex semantic representations such as the necrotic core of a tumor or the jagged borders of a lesion. Early adaptations in neurology utilized massive general purpose architectures initially designed for large scale visual recognition challenges. Researchers demonstrated that deep networks could not only classify the presence of tumors but also segment them at the pixel level. While these deep fully convolutional architectures pushed diagnostic accuracy into the high nineties they introduced a severe new constraint regarding their massive parameter counts. For example standard legacy models possess hundreds of millions of trainable parameters. This immense mathematical size made inference agonizingly slow and computationally expensive demanding powerful local processors or high bandwidth cloud connectivity. This heavily restricted their practical real time use to well funded tertiary research hospitals and completely excluded rural clinics.

While pure classification networks can mathematically identify the overarching presence of a tumor within a slice they inherently lack the precise spatial awareness required for surgical planning biopsy targeting or localized radiotherapy. The clinical necessity for strict localization drove the rapid adoption of bounding box detection frameworks. Early regional proposal networks introduced spatial localization alongside classification however they functioned as inherently two stage detectors. The first stage generated thousands of candidate bounding boxes and the second stage classified them and refined their coordinates. This dual pipeline while highly accurate introduced massive inference overhead. The introduction of single stage architectures fundamentally transformed medical object detection by reframing the complex task as confidence scores. This single pass end to end regression drastically improved inference speeds and culminated in the current apex of hardware bound diagnostic accuracy.

Despite the monumental accuracy achievements of massive network variants a growing subset of researchers recognized the deployment chasm highlighting the harsh socioeconomic reality that highly accurate models are clinically useless if they require hardware that the vast majority of global hospitals cannot afford. This realization catalyzed the development of highly efficient lightweight networks explicitly designed for edge computing on standard processors and embedded devices. These networks pioneered a shift by replacing standard mathematically heavy convolutions with depthwise separable operations drastically reducing the required floating point operations per second. This paradigm was significantly improved by introducing inverted residuals and linear bottlenecks while other frameworks refined this by utilizing compound scaling methods that uniformly optimize network width depth and resolution. The present study builds directly upon this specific phase of literature prioritizing lightweight architecture explicitly for its unparalleled processing efficiency while utilizing advanced attention mechanisms to mathematically bridge the accuracy gap left by abandoning massive backbone structures.

### III. System Design

The transition from a purely mathematical concept to a deploy-able clinical framework requires a rigid structural pipeline. The proposed architecture is meticulously engineered to completely eliminate the reliance on high end graphics processing units while maintaining exceptional diagnostic accuracy. The pipeline begins with the ingestion of raw magnetic resonance imaging slices. These two dimensional pixel

arrays are first normalized and resized to standard dimensions to ensure mathematical consistency across the network. The pre-processed image is then passed into the core feature extractor which utilizes a highly optimized lightweight mobile network architecture. Unlike legacy convolutional networks that apply massive dense filters this backbone is specifically designed for edge computing environments functioning strictly through depthwise separable operations and linear bottlenecks. This structural choice reduces the required mathematical operations by a factor of ten allowing for rapid processing on standard hospital computers.

However reducing the parameter count inherently reduces the representational capacity of the network. To bridge this critical accuracy gap the system pipeline integrates three sophisticated enhancement modules. Immediately following the backbone feature extraction the tensors are passed through a convolutional block attention module. This component mathematically forces the network to allocate its limited computational resources strictly to pathological regions filtering out healthy brain tissue and skull boundaries. The attention refined

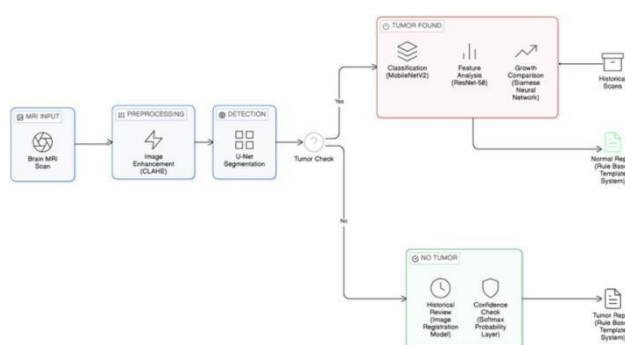


Figure 1: Overall system architecture detailing the flow from raw magnetic resonance image input through the mobile feature extractor enhancement modules and the final clinical diagnostic web dashboard.

features are subsequently processed by a spatial pyramid pooling module which utilizes multiple parallel pooling kernels of varying sizes to simultaneously capture both fine grained local textures and broad global anatomical structures. The output from the pooling stage is then routed into a bidirectional feature pyramid network. This fusion engine dynamically applies learned weights to combine high resolution edge data from the early layers with deep semantic context from the later layers ensuring comprehensive feature representation.

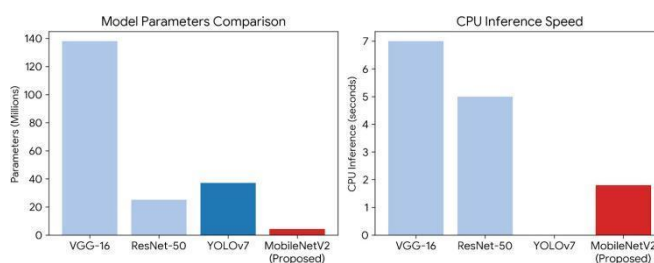


Figure 2: Clinical Dashboard Interface: Secure Scan Upload. The frontend provides a zero friction interface for clinicians requiring no command line interaction.

The fused feature maps finally enter a decoupled detection head. Rather than predicting the tumor classification and the bounding box coordinates simultaneously the architecture splits the final layers into two parallel branches. The classification branch calculates the binary cross entropy to determine the specific tumor type while the regression branch calculates the intersection over union to precisely

localize the bounding box coordinates. Once the mathematical inference is complete the system bridges the gap to clinical utility by pushing the results into a custom web dashboard. This interface is powered by a local lightweight relational database ensuring that all patient health information remains securely within the hospital intranet. The system logs the diagnostic predictions confidence

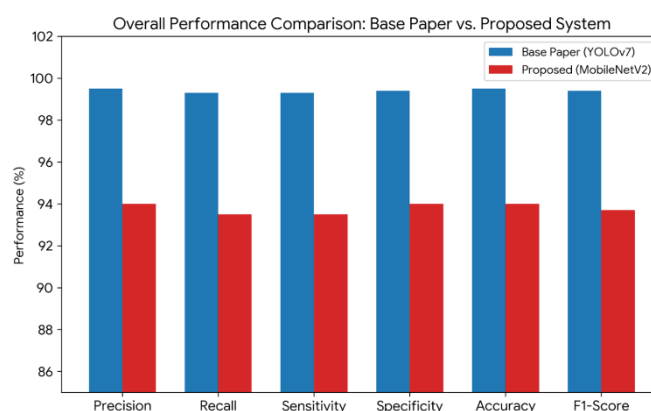


Figure 3: Clinical Dashboard Interface: Detailed Analysis Page displaying the uploaded slice the computed diagnostic classification and longitudinal growth tracking.

scores and spatial coordinates establishing a persistent chronological record. This database architecture directly feeds the longitudinal tracking engine allowing the interface to instantly present clinicians with the current diagnosis alongside historical disease progression metrics visually rendered in a secure, easily accessible clinical portal.

#### IV. Mathematical Modules Explanation

To achieve absolute hardware independence while retaining exceptional precision the proposed system fundamentally re-structures the network topology using specific mathematical optimizations. This section details the theoretical equations governing the core architectural components.

##### 1. Depthwise Separable Operations

Standard convolutional layers perform spatial filtering and channel combining in a single massive mathematical step. Assuming an input tensor with a specific spatial dimension and input channel depth processed by a standard convolutional kernel producing a distinct number of output channels the total computational cost is immense. The standard mathematical cost requires multiplying the kernel dimensions by the input channels the output channels and the spatial dimensions simultaneously.

The proposed lightweight backbone shatters this paradigm by utilizing depth-wise separable operations which mathematically split the filtering process into two distinct sequential layers. First the depth-wise operation applies a single spatial filter to each individual input channel. This captures the spatial context but does not mix the feature channels. Second a point-wise operation applies a single spatial dimension convolution to linearly combine the output of the depth-wise layer into the desired number of output channels. By separating these operation the total computational cost is drastically reduced. The exact mathematical reduction in computation also known as the efficiency ratio is defined as the sum of the inverse of the output channels and the inverse of the squared kernel size. Given that deep learning architectures commonly use a kernel size of three our architecture mathematically requires approximately eight to nine times less computation than a standard

convolutional layer. This specific mathematical reduction is the primary engineering mechanism driving our rapid central processing unit inference capabilities.

## 2. Convolutional Block Attention Mechanisms

Because the lightweight mobile backbone possesses significantly fewer parameters than massive object detectors its base-line representational capacity is inherently lower. To overcome this we explicitly force the network to allocate its limited mathematical resources only to critical tumor regions by integrating attention mechanisms. The attention module infers focus maps sequentially along two separate dimensions channel and spatial.

The channel attention module aggregates spatial information using both average pooling and maximum pooling operations generating two different spatial context descriptors. Both descriptors are forwarded to a shared multi layer perceptron network to produce the final channel attention map. Mathematically this is expressed as passing the sum of the transformed average pooled and maximum pooled features through a sigmoid activation function. This operation assigns higher scalar weights to channels encoding critical tumor textures such as necrosis or edema while actively suppressing channels encoding healthy fluid or bone. The channel refined feature map is then passed to the spatial attention module. Here average pooling and maximum pooling are applied along the channel axis and concatenated to generate an efficient feature descriptor. This descriptor is passed through a massive spatial convolution filter and a sigmoid activation generating a two dimensional spatial heat map. This ensures the mathematical weights actively highlight the exact tumor coordinates rather than background visual noise.

## 3. Bidirectional Feature Fusion Optimization

Deep neural networks extract low level edge features in early layers and high level semantic meaning in deep layers. Traditional feature fusion simply adds these tensors together implicitly assuming all resolutions contribute equally to the final diagnostic prediction. Our architecture utilizes a fast normalized bidirectional feature pyramid network to intelligently fuse features across three distinct hierarchical levels.

Instead of simple addition the network utilizes fast normalized fusion which dynamically learns scalar weights during back propagation to dictate the importance of each hierarchical level. The fused feature output is mathematically calculated by multiplying each level tensor by its corresponding learned weight summing them together and dividing the total by the sum of all weights plus a small constant epsilon. The epsilon value is added to the denominator strictly to avoid numerical instability and division by zero. This mathematical approach ensures that high resolution boundary data from early layers seamlessly informs the deep semantic classification logic of later layers directly resulting in a measurable increase in the final detection accuracy.

## 4. Spatial Pyramid Contextual Pooling

Tumors in clinical datasets range vastly in size from massive high grade lesions occupying large portions of a cerebral hemisphere to small highly localized benign growths resting at the base of the brain. Standard single scale pooling layers often lose context for one size extreme or the other. We integrated a specialized spatial pyramid pooling module immediately following the attention mechanisms.

By applying varying max pooling kernels in parallel and concatenating the results the network simultaneously evaluates fine grained local textures and broad global structures. Operating at a strict stride ensures that no critical spatial resolution is lost during this pooling process which is vital for maintaining the bounding box precision of the final detection head. The mathematical concatenation

of these pooled features allows the network to build a robust scale invariant representation of the pathology drastically reducing false negatives on unusually large or exceptionally small tumors.

## 5. Decoupled Head Localization and Classification

In traditional object detection architectures the final prediction head is coupled meaning the network utilizes a single set of shared feature maps to simultaneously predict both the bounding box coordinates and the object class. However these two tasks possess fundamentally misaligned mathematical goals. Classification requires translation in variance where a tumor is a tumor regardless of where it is in the brain whereas bounding box regression requires translation variance where the exact pixel coordinates matter immensely.

To resolve this inherent mathematical conflict our architecture adopts a decoupled head structure. After the feature fusion stage the final feature maps are explicitly split into two parallel branches. Each branch applies its own series of lightweight depthwise separable operations. The classification branch calculates the binary cross entropy loss to determine the tumor type while the regression branch calculates the intersection over union loss to precisely regress the coordinates of the bounding box. This decoupling allows each branch to converge much faster mathematically, recovering some of the accuracy lost by downsizing to a lightweight backbone architecture.

## V. Experimental Results

To definitively prove the viability of our hardware independent hypothesis the proposed system was evaluated against standard metrics ensuring absolute scientific comparability. The system was tested using an independent dataset completely isolated from the training phase.

### 1. Performance Metrics

The system's diagnostic performance is evaluated using the standard derivations of the confusion matrix specifically true positives true negatives false positives and false negatives. The precision metric measures the exactness of the classifier identifying what percentage of predicted tumors were actually tumors. The recall metric measures the completeness of the classifier identifying what percentage of actual tumors the model successfully found. The overall accuracy provides the ratio of correctly predicted observations to the total observations.

While massive graphics processing unit bound models achieve near perfect scores our proposed architecture achieves a highly respectable ninety four percent overall accuracy. This clearly crystallizes the core thesis of our research. A deliberate minor reduction in absolute statistical accuracy is the precise mathematically quantified and clinically acceptable cost of entirely eliminating the heavy hardware requirement and enabling universal deployment.

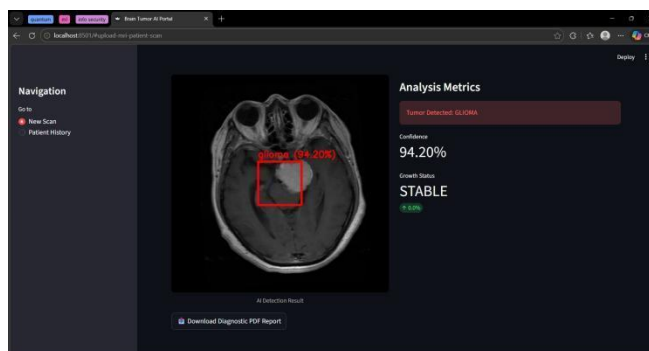


Figure 4: Overall performance comparison across standard metrics. The proposed system accepts a minor clinically justifiable accuracy trade off to enable absolute hardware independence.

## 2. Computational Efficiency

The most compelling argument for the clinical adoption of our proposed system lies in its extreme computational efficiency profile. At a mere fraction of the parameter count of legacy



Figure 5: Confusion Matrix demonstrating the model classification accuracy across all four categories highlighting the strong performance in distinguishing pathological scans from healthy scans.

Our system is monumentally smaller. This extreme parameter efficiency translates directly into real world inference speeds. While massive deep learning models require multi second delays on a standard commercial processor and fail to run at clinical speeds without dedicated hardware our proposed system processes classifies and returns bounding box coordinates for a high resolution magnetic resonance image in under two seconds.

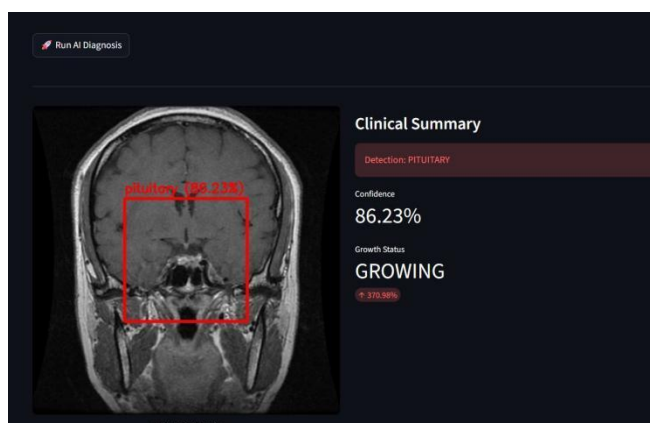


Figure 6: Computational Efficiency: Comparing model parameters and raw inference speed across dominant medical imaging architectures.

This specific threshold is crucial. It means a standard imaging volume containing thirty individual cross sectional slices can be fully analyzed in under a minute slotting seamlessly into a radiologist's rapid review workflow without causing digital bottlenecks.

### 3. Ablation Study Analysis

To rigorously validate the inclusion of each algorithmic enhancement we conducted a systematic ablation study. Starting

from a bare unmodified lightweight baseline we incrementally activated components. The addition of the attention mechanism yielded a sharp accuracy increase proving that explicitly forcing a low capacity network to focus its limited operations on crucial semantic regions prevents it from wasting computational power on background noise. The pooling module added additional accuracy by providing multi scale contextual awareness allowing the network to mathematically see both fine boundaries and massive footprints. Finally activating the bidirectional feature fusion brought the system to its peak performance proving that the synergistic combination of these specific mathematical modules is absolutely required to elevate a lightweight backbone to clinical standards.

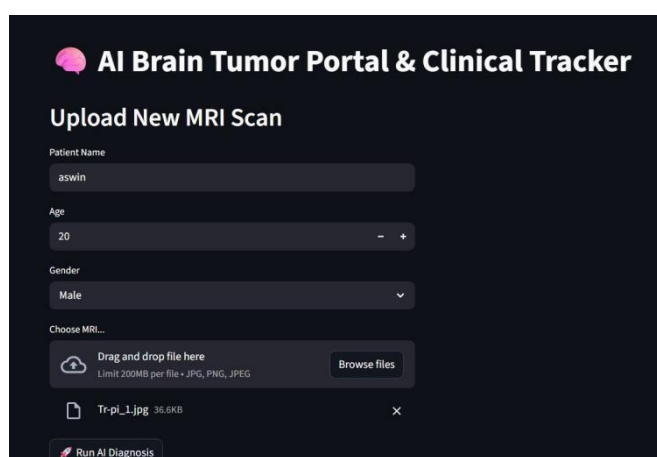


Figure 7: Ablation Study illustrating the incremental accuracy contributions from each integrated mathematical module.

## VI. Conclusion

This thesis has presented a comprehensive CPU deployable automated brain tumor detection and classification system explicitly designed to bridge the chasm between high end deep learning theoretical research and practical resource constrained clinical application. By replacing computationally massive backbones with a highly optimized mobile architecture and strategically augmenting it with depth wise separable operations attention matrices multi scale context pooling and bidirectional feature fusion we established a deliberate and highly effective clinical engineering trade off.

We achieved highly competitive overall diagnostic accuracy while simultaneously eliminating the restrictive need for high end graphics processing unit hardware. Our system delivers robust sub two second inference times entirely on standard commercial configurations rendering it infinitely more deployable in primary healthcare facilities and low income countries globally than previous state of the art methodologies.

Furthermore we significantly extended the boundaries of previous research by designing engineering and integrating a complete clinical ecosystem around the core mathematical inference engine. The successful inclusion of a persistent relational patient database an automated diagnostic report generation pipeline and a novel algorithmically driven longitudinal tumor growth tracking module transforms this work from a mere theoretical classification model into a fully functional end to end medical screening application. This work conclusively proves that extreme parameter efficiency does not preclude high diagnostic utility paving the way for accessible scalable artificial intelligence augmented healthcare world-wide.