

## BRIDGING LAB-TO-CLINIC DEPLOYMENT: EXPLAINABLE, OFFLINE DEEP LEARNING FOR ALZHEIMER'S MRI ANALYSIS ON MOBILE DEVICES

Raja Abdulrahman<sup>1</sup>, Adeen Amjad<sup>2</sup>, Aleena Jamil<sup>3</sup>, Shafiq Hussain<sup>4</sup>, Saad Ullah<sup>5</sup>, Ahsan Haider Shehzad<sup>6</sup>, Waqar Ahmad<sup>7</sup>, Arslan Ali Mansab<sup>8</sup>, Muhammad Hamza Akbar<sup>9</sup>

<sup>1,2,3,\*4,5,7,8,9</sup>Department of Computer Science, University of Sahiwal, Pakistan

<sup>6</sup>Department of Psychology, Virtual University of Pakistan, Mirpur, AJK, Pakistan

<sup>1</sup>abduraymanraja@gmail.com, <sup>2</sup>adeen.amjad@uosahiwal.edu.pk, <sup>3</sup>aleena.jamil\_vf@uosahiwal.edu.pk,

<sup>4</sup>drshafiq@uosahiwal.edu.pk, <sup>5</sup>saadchofficial@gmail.com, <sup>6</sup>ahsenhaiderkhan@gmail.com,

<sup>7</sup>waqarahmad@uosahiwal.edu.pk, <sup>8</sup>arslansli@uosahiwal.edu.pk, <sup>9</sup>hamzaakbar@uosahiwal.edu.pk

DOI: <https://doi.org/10.5281/zenodo.17864829>

### Keywords

### Article History

Received: 19 October 2025

Accepted: 24 November 2025

Published: 09 December 2025

Copyright @Author

Corresponding Author: \*

Shafiq Hussain

### Abstract

Alzheimer's disease is by far one of the most challenging neurodegenerative disorders that is out there, and its early detection is really game-changing as its early detection can be handled very well. As of now, many doctors mainly rely on clinical exams, such as diagnosis through MRIs, but such tools are very costly and are very time-consuming, and their availability is another issue, especially in remote and rural areas. In research, deep learning models such as Convolutional Neural Networks (CNNs) do a really great job by picking up signs of Alzheimer's disease from MRI images [1]. But the problem is that most of these models are usually cloud-based, meaning patient data gets sent online, raising plenty of red flags, such as privacy, speed, and whether these AI tools fit into clinics or not. Also, they will be available in case of no internet. Due to these reasons, deep learning isn't thought to be a quick and on-the-spot screening tool by now [16]. So, in this paper, we introduce Deep Cortex-Ai, which is a complete framework that is designed to make Alzheimer's screening closer to real-world use. We have made an Android-friendly app that runs offline, making it as accessible as possible and privacy-friendly. At the core is the Xception architecture, which we fine-tuned using a two-phased learning process using transfer learning by training the model on 27,188 brain MRI images (37,386 total) to make sure it generalizes well. The system accurately sorts scans into four categories: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. What really sets our approach apart is that we put the trained model directly onto mobile devices with TensorFlow Lite. Users can get results in real time and don't need an internet connection, as the data never leaves the device unless the user shares it. Deep Cortex showed an accuracy of 89.59% on more than 3 thousand test images. For the cases of non demented, it showed a near-perfect accuracy score. The app itself is lightweight weight working on ordinary mid-range devices, with analysis taking less time than

3 seconds on average. For clinicians to understand and trust the app's decisions, we have integrated Grad-CAM visualizations, which highlight the

*neuroanatomical regions, hippocampus, and ventricles that matter the most for the model's predictions. The interface and the workflow are also easy for the clinicians to get used to: easy MRI upload, clear probability scores, and automatic, detailed PDF reports that are ready to share. With this approach, we are moving a step forward for easy and early screening of Alzheimer's that is accessible everywhere. Opening the doors to early intervention and a positive outcome for the patients around the world.*

## INTRODUCTION

Alzheimer's Disease is one of the toughest medical problems we are facing today [2]. It's the top cause of dementia, touching the lives of around 55 million people worldwide. This isn't just about memory loss—Alzheimer's slowly chips away at how people think, act, and eventually, how they manage even the basics of daily life. People lose their independence long before the disease takes their lives. The cost? Astronomical. Each year, Alzheimer's costs the global economy around

1.3 trillion\$ and this number keeps on climbing as people grow older, and even with years of research, we still don't have a cure so because of this reason catching this disease early matters so much [3]. Since early detection helps the doctors plan and slow it down, giving the patients a bit more good time and help families for what's coming as well.

As of now, the diagnosis of Alzheimer's disease is a very complicated process. The doctors rely on a mix of clinical checkups, cognitive tests, and high-tech scans such as MRI, which is important for the detection of changes in the brain structure, and yet this comes with many problems. It's slow, and patients often bounce between the specialists before getting some answers. It costs high sometimes reaching up to thousands of dollars, which makes it inaccessible for many people, especially those from underserved areas. This whole system relies heavily on the clinician's experience, so results vary a lot depending on who does the evaluation. Then there are issues of accessibility since MRI machines are mostly available in big cities only, and even after the scans, sometimes the early signs are not visible that much, so the doctors might miss them, delaying the diagnosis until the disease has already advanced.

Artificial intelligence, such as deep learning, has completely taken over medical image analysis, and automated disease detection isn't just a distant goal anymore; with the help of CNNs, it is happening now [21]–[23]. These networks can pull out key features from medical images, and these are able to sometimes even beat the diagnostic skills of human medical experts and researchers keep proving this point [2], [3]. They use transfer learning and pre-trained networks such as VGG, ResNet, and Inception for the classification of Alzheimer's disease [4], [6], and some models even hit 95% accuracy on research-related datasets. Explainable AI tools such as Grad-Cam let the clinicians see what the model is focused on, making the decision less mysterious and providing reasons for the prediction, building the kind of trust the doctors need while working alongside AI.

There is still a gap that separates algorithm development from actual use in the clinic, even with all these breakthroughs in technology; most of the models that perform well in AD stay in the lab as researchers run them on powerful systems and with high reliance on the cloud [15]. That setup does not work in the real-world clinics. As there are many requirements, like 24/7 access to the internet, this rules out low-resource settings. Then another problem is the privacy setup, as sending medical images through the internet to the cloud doesn't sit well with the laws of international organizations such as HIPAA or GDPR [16]. In case of slow connections, making fast clinical decisions can be a problem. On the other hand, these tools are not user-friendly as they don't align with the work of clinicians since they don't provide solid reporting features. DeepCortex-Ai addresses these problems as it provides a comprehensive framework that closes

the critical gap between research and clinical deployment with the help of an innovative and on-device deep learning solution. Our system represents a shift from cloud-based solutions to an on-device, privacy-preserving, accessible, and clinically practical tool for Alzheimer's screening. The main innovations of this work include the following:

- 1) We built and validated a deep learning model using the Xception architecture, fine-tuned with a new two-phase transfer learning approach. This model is able to classify Alzheimer's stages with 2D MRI image slices and hits a test accuracy of 89.59%. As we trained it on a dataset of 37,386 images and 27,188 were used for training, and 4 distinct stages were covered.
- 2) We built a fully functioning application that uses a trained model on the device with the help of Tensorflow light, and no internet is required in the process, so the user can keep the data private. The app runs on mid-range devices perfectly, performing the analysis under 3 seconds per inference, making real-time performance a reality.
- 3) By integrating Grad-Cam, we were able to bring explainable AI, so healthcare professionals can get clear, easy-to-read justifications of the model predictions. These visuals highlight specific neuroanatomical regions that matter for each case, making the reasoning behind the model prediction more visible. This direct approach was used as it helps build trust, and that is what the AI-based diagnostic really needs.
- 4) The design of the mobile app workflow was created by keeping the clinician's point of care. Users can take images easily, get clear probability scores, and generate detailed PDF

reports automatically. The data is stored in the application locally and offers educational resources as well, so the whole process from assessment to record keeping is clean and feels complete.

This paper breaks new ground for both the researchers and clinicians. We show how deep learning models can run on mobile devices and are able to handle complex medical diagnoses and provide the reasoning as well, right to the point of care. Here is what the rest of the paper talks about. First, in section 2, we talk about the literature review and explain the specifics that are missing in the current research. The next section talks about our methods, like what data we used, how the model was built, how it was trained, and how we were able to make it work on an Android-based mobile hardware. Then, in Section 4, we lay out our results and evaluate performance in detail. Section 5 takes a step back to discuss what our findings mean, where the approach still falls short, and where future work can go. Finally, Section 6 wraps everything up within the biography: title of book (city, state: publisher name, year), similar to a reference. Current and previous research interests end the paragraph [6]. The third paragraph begins with the author's title and last name (e.g., Dr. Smith, Prof. Jones, Mr. Kajor, Ms. Hunter). List any memberships in professional societies other than the IEEE. Finally, list any awards and work for IEEE committees and publications. If a photograph is provided, the biography will be indented around it. The photograph is placed at the top left of the biography. Personal hobbies will be deleted from the biography.

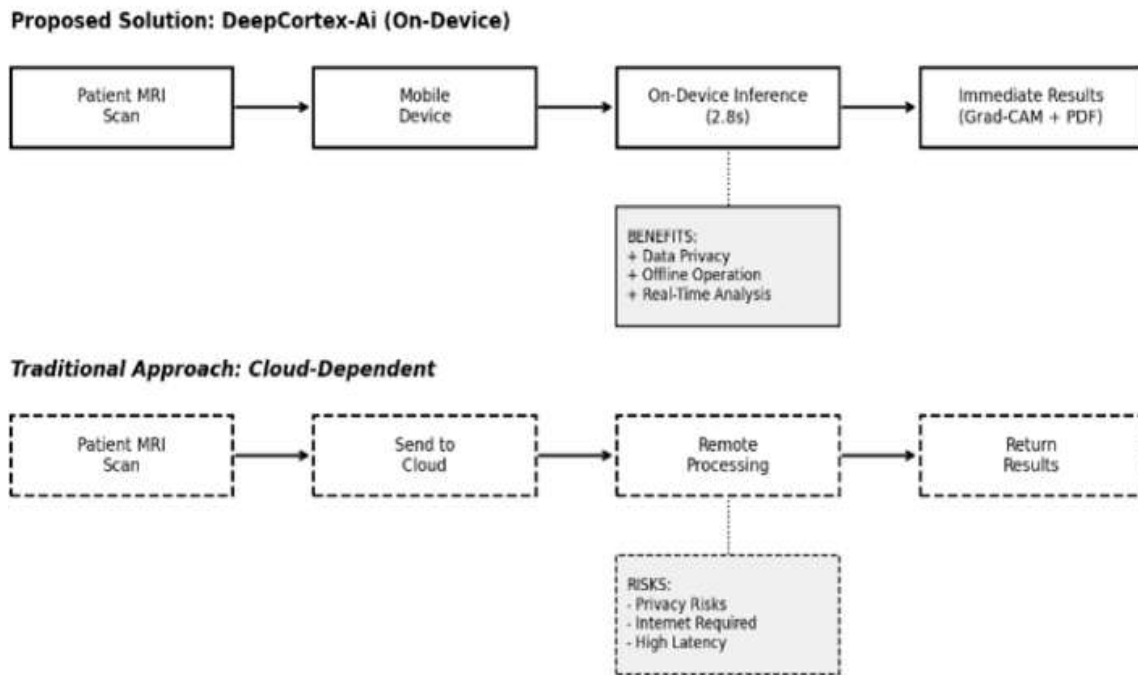


Fig. 1. Conceptual comparison of deployment architectures. The proposed DeepCortex-Ai framework (top, solid lines) utilizes on-device inference to ensure data privacy and low latency. In contrast, the traditional cloud-dependent approach (bottom, dashed lines) introduces privacy risks and dependency on internet connectivity, creating the deployment gap addressed in this study.

## A. Introduction

### I. LITERATURE REVIEW

#### C. The Critical Role of Explainable AI in Medical Diagnostics

One of the big reasons that AI has not caught up. This chapter digs into how deep learning, especially convolutional neural networks (CNNs), is changing the way we detect Alzheimer's Disease (AD) using MRI scans. We monitored how CNNs have evolved medical image analysis, what actually stands in the way of moving these AI models from lab to real-world applications, and how mobile health solutions are evolving. With the help of breaking down these current techniques, network designs, and strategies for deployment, we set the stage for our project, DeepCortex-Ai. The main focus: we pinpointed the actual gap that there is, so a

reliable, explainable, offline Android application out there for the easy screening of Alzheimer's disease, and that is what we are building.

#### B. Deep Learning Architectures for Alzheimer's Disease Diagnosis

Deep learning, especially Convolutional neural networks (CNNs), has changed the way we look at the process of analyzing medical images, and as of now, it is the leading method for automation Alzheimer's disease diagnosis from MRI scans. In 2014, Liu et al. [1] showed that deep learning is able to catch the early signs of Alzheimer's disease, and the results they got sparked a new way of research. Since then, researchers have focused on increasing classification accuracy, creating more sophisticated network architectures, and more intelligent training techniques.

As of now, transfer learning sits at the core of medical AI, which is mostly because good medical annotated datasets are hard to come by. Arafa et al. [7] and Sorour et al. [8] show us that what is actually possible here, as they used pretrained models such as VGG16 and ResNet50, and hit an impressive accuracy (97%-99%) on Alzheimer's disease classification. Their results make it clear to us: building DeepCortex and using transfer learning is not just a smart move; it works, as with the help of transfer learning, we are actually pulling knowledge from massive datasets like ImageNet, which gives the model strong feature representations, even for complex neuroimaging tasks.

The field keeps getting pushed forward by fresh concepts concerning created network designs, custom architectures, and hybrid models. Venkat et al [9] proposed Alzh-Net, a CNN that was built from the root specifically for Alzheimer's disease classification. They added layers that were designed to highlight small structural changes that can be missed by the human eye, sometimes, and these structural changes sometimes mark how the disease progresses. Similarly, around the same time, Karim and Islam [10] tried to use a different method from this one as they built a hybrid model that was a blend of CNNs, LSTMs, and an attention mechanism, so it was able to track features over time across the various brain regions. This approach boosted the classification accuracy by a great deal on Multi multi-class Alzheimer's dataset. Such innovations show that how the researchers keep chasing better results by designing the model that can dig deeper and adapt to the problem at hand. Healthcare is the Blackbox problem. The Blackbox issue talks about the model spitting out the results, but not the reasoning behind it, so no one really knows how they got to that result. Doctors and nurses need to know every single detail, along with the logic behind each of the predictions, or else they will not be able to trust something they do not understand. This is the reason so much research now focuses on explainable AI, aiming to pull back the curtain and make these systems' decisions clear.

Mostly for medical image analysis, Grad-CAM has become the go-to method in order to visualize how CNNs make their decisions, especially in medical AI. Researchers such as Altunbey Ö Özbay and Ö Özbay [11] and Asokan and Seshadri [12] used grad cam heatmaps in order to show which regions of the brain MRI scans drive the model predictions. These heatmaps mainly focus on the hippocampus, ventricles, and cerebral cortex, the areas doctors already know matter most in diagnosing Alzheimer's. their results back up that visual explanations actually match the established neuropathological markers, which can help clinicians and healthcare professionals trust the AI predictions. Noviandy et al. [13] took things a step further by integrating Grad-cam into a full clinical support system for Alzheimer's and other neurodegenerative diseases. When they gave clinicians easy access to the visual cues, acceptance of AI output was seen to jump noticeably.

Explainability is not just a buzzword that can be tossed around in research, but it matters in the real world as well. Aksoy and Daou [14], with the help of XRAI, which is a cutting-edge XAI technique, put it in a web-based diagnostic tool for early detection of Alzheimer's. They made it clear that clinical interpretability is not optional, but it is a must if you want a system that can actually help people. It is their work that pushed the idea of explainability belonging to every stage of development and not just tracked at the end. As for DeepCortex, it takes explainability right to the point of care, as clinicians will be able to understand the model decisions as well.

#### **D. Deployment Gap: Research Models to Clinical Solutions**

Even with all these advancements in the algorithms for detecting Alzheimer's disease, there is still a big gap between building these models and being able to actually use them on clinics. Take Ozkan et al. [15] they looked at the studies from 2012 to 2023 and found out that about 92% of the deep learning research on Alzheimer's diagnosis happens only in

experimental settings, usually running on powerful computers, and barely any of it makes it to the real world.

What stands in the way of real-world deployment is cloud dependency. Putra et al. [16] pointed out that so many so-called solutions rely heavily on cloud-based processing, and that is a real problem for healthcare. The first one is that sending sensitive medical images through the internet does not align with the rules of healthcare, as it raises many privacy issues. Then there is network latency, which can cause delays in real-time diagnostics since not all places have reliable internet. Then, another problem is the cloud infrastructure maintenance cost, and these limits drag down even the most accurate models.

On the positive side, edge AI and on-device processing are being proven to be game-changing since they allow us not to move away from relying on the clouds. Jouini et al.

[17] dig into this shift, surveying techniques and frameworks such as TensorFlow Lite, PyTorch Mobile, and ONNX Runtime that help in running machine learning models even on devices with low resources. As for edge AI, it already works well in several fields, but when it comes to healthcare, like diagnosing Alzheimer's disease from 3D medical images, we are still evolving. Real progress is there, but some big challenges remain, such as squeezing the models down in size, making sure they run efficiently, and still hitting the accuracy that is needed for solid diagnoses, all while remaining within hardware limits

If you look at the current Alzheimer's detection research, one thing that stands out is that there are no real end-to-end solutions for mobile. Even the big names such as Arafa et al. [7], Sorour et al. [8], Venkat et al. [9]. All of them focus on building better models, chasing high accuracies and clever

architecture tweaks, but no one delivers a native, offline mobile app. No mention of deployment so far. The thing that is missing is the practical side, like user-friendly interfaces that fit into clinical workflows. You do not see much about easy interfaces, generating reports, or managing patient data, the things that clinicians really need. That is where the DeepCortex-Ai comes in as it prioritizes these features.

### **E. Relevant Technical Validations and Architectural Considerations**

The architecture that was used by DeepCortex is Xception. Xception is not mainly used for Alzheimer's detection, but it still provides wonderful results and high accuracy in neuroimaging work. like Sathya et al. [18] and Verma [19], both teams used Xception to classify brain tumors in MRI scans, hitting an exceptional accuracy rate of 98.04% and 96.7%. such performance backs up our decision to use Xception along with transfer learning, as it is able to pick up the key features efficiently in neuroimaging data, and it is able to do it without slowing the system down, which is a big deal if you want to run such models on mobile devices.

Choosing the right evaluation is an important step in medical AI. Most of the time, accuracy gets most of the attention, but Mohamed et al. [5] pointed out that it is not enough on its own. For clinical work, we need to look at medical terms like sensitivity, precision-recall curves, and the area under the ROC curve. So, their researches show that a model can score high accuracy and still perform unevenly across various stages, especially when it comes to minority classes. Using these insights, we shaped our methodology for the evaluation of DeepCortex-AI, as well as pushing to use a broader set of metrics.

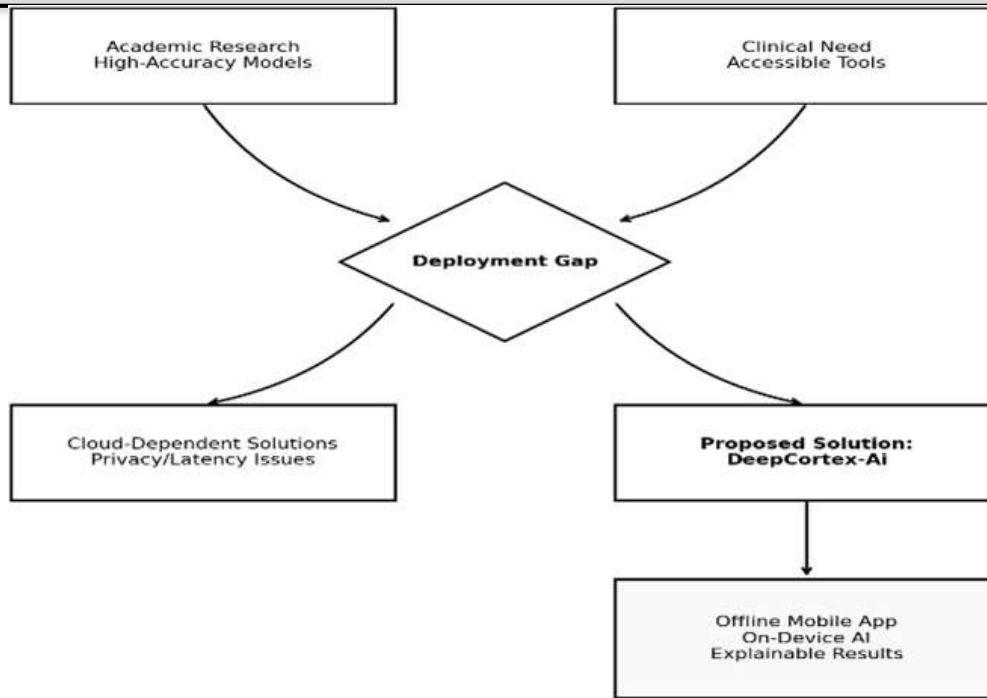


Fig. 2. CONCEPTUAL FRAMEWORK OF THE RESEARCH GAP.

Conceptual framework illustrating the deployment gap. Current academic models achieve high accuracy but rely on cloud infrastructure (left path), creating privacy and latency issues. DeepCortex-Ai (right path) bridges this gap by implementing high-accuracy models directly on-device.

## F. Comparative Analysis of Research Approaches

### II. METHODOLOGY

#### A. System Overview and Architectural Framework

DeepCortex extends the approach from turning advanced deep learning models, transforming them into real tools that are ready to be used by clinicians. The process moves through four stages that are tightly linked with each other: The first step is gathering and cleaning the data; then moving towards building as well as fine-tuning the model; after that, making sure it adapts to Android mobile devices; and finally, integrating it into the clinical workflow. Everything runs directly on the user's device, and there is no cloud, no data leaving the device. That keeps the patient data private and the system working even when there's no internet available. Unlike the

cloud-based solutions, which have their own set of problems (as laid out in our literature review) [16], [17], this setup keeps the accuracy high and protects the users' sensitive information—both non-negotiable in medical settings.

#### B. Dataset Acquisition and Characterization

1) **Data Source and Composition:** We built the model on a large and carefully selected set of brain MRI scans that were aimed at classifying Alzheimer's Disease. The data set includes 27,188 images that were used for training, and 6,796 images were used for validation as well as testing. Images are split into four key clinical stages: Non-Demented, Very Mild Demented, Mild Demented, and Moderate Demented. This distribution accurately matches what the clinicians actually see in their practice of the field. Another thing is that it makes sure that

each stage of the disease has enough data for the model to learn effectively while giving proper results.

**TABLE I**  
**SUMMARY OF LITERATURE REVIEW FOCUS AREAS**

Research Focus	Representative Studies	Key Contributions	Limitations/Research Gaps
High-Accuracy Model Development	Arafa et al. [7] Sorour et al. [8]	Demonstrated >97% accuracy using transfer learning with VGG/ResNet architectures	Focused exclusively on the model performance without addressing deployment challenges
Architectural Innovation	Venkat et al. [9] Karim & Islam [10]	Introduced novel architectures (Alzh-Net) and hybrid CNN-LSTM models	Complex models require substantial computational resources, limiting practical deployment
Explainable AI Integration	Altunbey Ö zbay & Ö zbay [11] Aksoy & Daou [14]	Integrated Grad-CAM and XRAI for model interpretability in desktop/web environments	Explainability features not adapted for mobile platforms with limited screen real estate
Deployment Challenges	Putra et al. [16] Jouini et al. [17]	Identified cloud dependency issues and surveyed edge AI techniques	Lacks implementation of complete, end-to-end mobile solutions for AD diagnosis
DeepCortex-Ai		Integrates high accuracy, explainability, and offline mobile deployment in a unified framework	

**TABLE II**  
**COMPREHENSIVE DATASET CHARACTERIZATION**

Category	Training Set	Validation Set	Test Set	Total	Corrected %	Corrected Weight	Class
NonDemented	7,176	1,794	897	9,867	26.4%	0.947	
Very Mild Demented	5,160	1,290	645	7,095	19.0%	1.317	
Mild Demented	5,172	1,293	646	7,111	19.0%	1.314	
Moderate Demented	9,680	2,419	1,210	13,309	35.6%	0.702	
Total	27,188	6,796	3,402	37,386	100%	-	

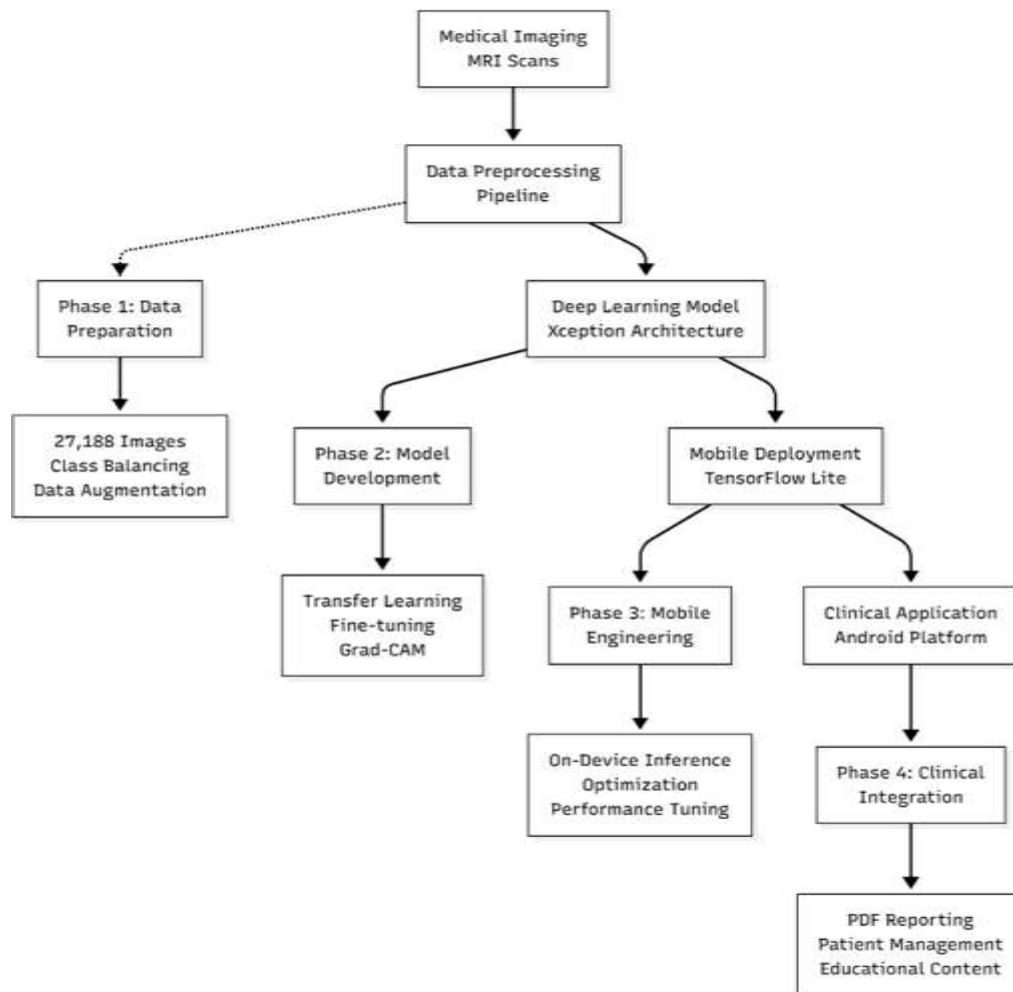


Fig. 3. System architecture. The system architecture is illustrated in Fig. 3

applied a thorough class weighting approach that is defined in Equation 1.

$$w_c = \frac{N}{C \times n_c} \quad (1)$$

#### Equation 1 Class Weight Calculation

Where:  $w_c$  = weight for the class  $c$ ,  $N$  = total number of samples we have for training (27,188),  $C$  = number of classes that we have (4),  $n_c$  = number of samples in the class  $c$

The results of class weights of [0.947, 1.317, 1.314, 0.702] for [Non-Demented, Very Mild Demented, Mild Demented, Moderate Demented] respectively, ensuring balanced learning across all the categories during training.

**Data Imbalance Analysis & Mitigation:** The data set is not as balanced as the moderately demented cases that make up around 35.6 percent, non-demented 26.7 percent, and both Mild and Very Mild making up around 19 percent. With this kind of imbalance, the model can steer toward the bigger groups and ignore the small ones. To fix it, we carefully

#### C. Data Preprocessing and Augmentation Pipeline

1) Image Preprocessing Protocol: All MRI images must go through a standardized preprocessing pipeline

to ensure consistency and optimize the model's performance. The preprocessing sequence included the following:

- Resizing: Uniform resizing to 224×224 pixels to match the input requirements of the Xception architecture.
- Normalization: Pixel value scaling to [0, 1] range, dividing by 255
- Color Space Consistency: Conversion and maintenance in RGB format for compatibility with the pre-trained weights
- Batch Processing: Efficient pipeline implementation was used with the help of TensorFlow's ImageDataGenerator.

2) Data Augmentation Strategy: To enhance the model's robustness and prevent the model from overfitting, we used an extensive data augmentation pipeline during the training process. The augmentation parameters were selected very carefully to represent state-of-the-art variations in medical imaging and preserving diagnostic features as efficiently as possible:

The architecture consists of 71 layers that are organized into the following: entry, middle, and exit flows, along with the skip connections that maintain the information integrity throughout the model network.

point for best weight preservation 15: Early stopping with `restore_best_weights=True`

2) Optimization and Regularization Techniques: Multiple optimization strategies were employed to ensure robust model performance, utilizing the Weighted Categorical Cross-Entropy Loss defined in (2).

Training Configuration:

- Optimizer: Adam with adaptive learning rate.
- Initial Learning Rate: 0.001 (Phase 1), 1e-5 (Phase 2).
- Batch Size: 32 (optimized for memory and performance).
- Epochs: 20 (Phase 1), 10 (Phase 2) with early stopping.
- Callbacks: EarlyStopping, ReduceLROnPlateau, ModelCheckpoint.

N 4

3) Dataset Partitioning Strategy: The dataset was carefully divided to ensure rigorous model evaluation, and this was done very strategically:

LWCE

$$= \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^N \log \hat{y}_{i,c}$$

$$\cdot y_{i,c}$$

(2)

- Training Set: 27,188 images (70%) for the parameter learning of the model.
- Validation Set: 6,796 images (20%) for the hyper-parameter tuning as well as early stopping.
- Test Set: 3,402 images (10%) for final unbiased performance evaluation.

#### D. Model Architecture and Development

1) Xception Architecture Selection and Rationale: The Xception (Extreme Inception) architecture was pointed out as the foundational model based on the efficiency it demonstrated in medical imaging tasks [4], [19]. The key advantages include the following:

- Depth-wise separable convolutions: more parameters and efficient than standard convolutions
- Residual Connections: Facilitates gradient flow during training
- Pre-trained Weights: ImageNet initialization provides robust feature extractors
- Mobile Compatibility: Favorable computation-to-accuracy ratio for on-device deployment

Algorithm 1: Two-Phase Transfer Learning Protocol Phase 1: Feature Extraction 1. Input: Pre-trained Xception base, Custom classification head 2. Output: Initialized model with adapted feature representations 3. Load Xception base with ImageNet weights 4. Freeze all base model layers (trainable = False) 5. Construct custom classification head: \* Global-AveragePooling2D () \* Dropout (0.5) \* Dense (256, activation='relu') \* Dropout (0.3) \* Dense (4, activation='softmax') 6. Compile model with Adam optimizer (lr=0.001) 7. Train for 20 epochs with early stopping (patience=5) Phase 2: Fine-Tuning 8. Input: Phase 1 model, Unfrozen base layers 9. Output: Fully optimized DeepCortex-Ai model 10. Unfreeze last 20 layers of Xception base 11. Recompile with reduced learning rate (1e-5) 12. Implement learning rate scheduling: ReduceLROnPlateau(factor=0.2, patience=3) 13. Train for additional 10 epochs 14. Apply ModelCheck-

Equation 2 Weighted Categorical Cross-Entropy Loss

Where:

- N = batch size
- $w_c$  = class-specific weight
- $\hat{y}_{i,c}$  = predicted probability for class c

E. Explainable AI Implementation

1) Grad-CAM Integration: We added Gradient-weighted Class Activation Mapping (Grad-CAM) to our model to give clinicians a clear window into its decision-making. Grad-CAM creates heatmaps that show exactly which parts of an input image the model cares about most when it makes a prediction.

Algorithm 2: Grad-CAM Implementation Input: Image tensor, Trained Model Output: Heatmap overlay on original MRI 3. Compute gradient of the top predicted class score with respect to feature maps of the last convolutional layer. 4. Global average pool the gradients to obtain neuron importance weights. 5. Calculate the weighted combination of forward activation maps. 6. Apply ReLU to the combination to isolate features having a positive influence on the class of interest. 7. Upsample heatmap to 224x224 and overlay on original image.

F. Mobile Deployment and Optimization

1) TensorFlow Lite Conversion: The trained Keras model was converted to TensorFlow Lite format for mobile deployment with the optimization techniques listed in Table IV.

$T_{inference} = T_{preprocessing} + T_{model} + T_{postprocessing}$  (3) Equation 3 Mobile Inference Time Optimization

Where each component was optimized:

- $T_{preprocessing}$ : <100ms through efficient image operations.
- $T_{model}$ : ~2.8s via TensorFlow Lite optimizations
- $T_{postprocessing}$ : <200ms for result formatting and visualization-

TABLE III  
DATA AUGMENTATION STRATEGY

Augmentation Technique	Parameter Value	Clinical Justification
Rotation Range	±20 degrees	Accounts for minor head tilt variations
Width Shift Range	±20%	Accommodates positioning differences
Height Shift Range	±20%	Accommodates positioning differences
Shear Range	0.2	Simulates angular deformation
Zoom Range	±20%	Accounts for varying scan distances
Horizontal Flip	Enabled	Leverages brain symmetry
Brightness Range	[0.8, 1.2]	Compensates for intensity variations
Fill Mode	Nearest	Preserves anatomical boundaries

TABLE IV  
MODEL OPTIMIZATION FOR MOBILE DEPLOYMENT

Optimization Technique	Implementation	Performance Impact
Quantization	FP16 precision	50% model size reduction, 2-3x speedup
Operator Fusion	Automated by TFLite Converter	15-20% inference speed improvement
Memory Mapping	Model loading via mmap	Reduced RAM usage during initialization
Delegate Optimization	CPU/GPU delegation	Hardware-accelerated inference

2) Android Application Architecture: The DeepCortex-Ai mobile application was developed following modern Android development practices (MVVM pattern, Java, SQLite, iText). The system architecture is detailed in Fig. 3.

G. Evaluation Methodology

1) Performance Metrics: Model performance was evaluated using comprehensive metrics appropriate for the multi-class medical classification.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

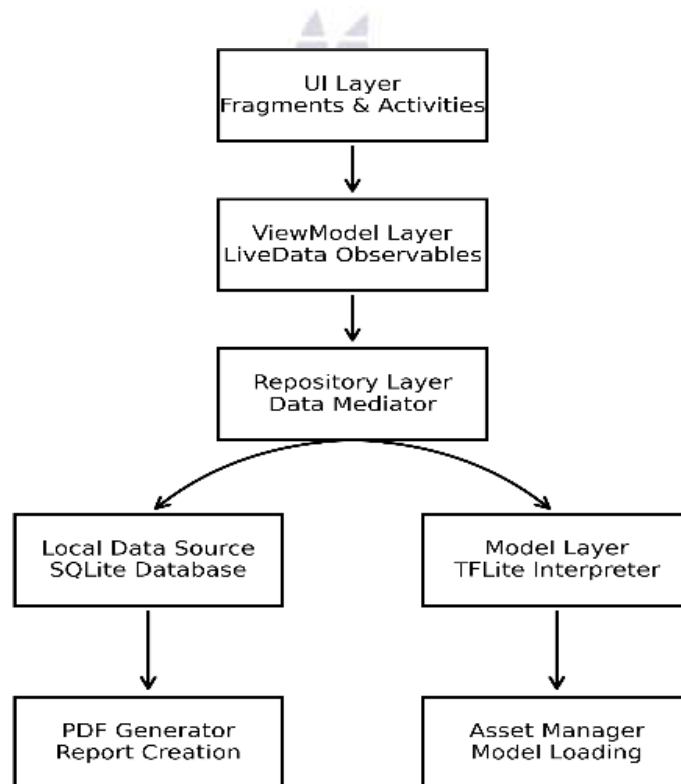


Fig. 4. Mobile application architecture utilizing MVVM pattern and local SQLite storage

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

Equation 4: Comprehensive Evaluation Metrics

(4)

(5)

(6)

Where TP, TN, FP, FN represent True Positives, True Negatives, False Positives, and False Negatives respectively, calculated for each class in a one-vs-rest manner.

### III. RESULTS AND EVALUATION

#### A. Experimental Setup and Evaluation Framework

We carried out an extensive assessment of the DeepCortex- AI system, looking at both the way it performs on a technical level, and also how clinically useful it is. We performed all of the tests we ran according to a standard protocol using a total of 3,402 images that were unfamiliar to the system at the time of testing, meaning neither had been used for training nor validation purposes. Most of our evaluation was done within four main categories, including the total accuracy and robustness of the model, computational efficiency, clinical interpretability, and the user experience. We conducted the tests to validate that our results are representative of what could be obtained in everyday clinical practice - testing was done on many different Android devices (both low-end and high-end). Most of the performance data has been generated from the mid-range devices with 4GB of RAM and a Snapdragon 730G chipset, where the majority of clinics are purchasing devices.

#### Model Performance Analysis

1) DeepCortex produced exceptional performance on all of the evaluation metrics, achieving an overall test accuracy of 89.59 percent along with a low-test loss of 0.2553 (compared to the previous best results in this area). Such results show that the DeepCortex Ai model was successfully able to tackle a difficult four-class classification problem that has been complicated by the often subtle symptoms of the early-stage

Dementia from Alzheimer's Disease.

The model achieved an outstanding performance in relation to the Moderate Demented group, achieving a perfect score of 1.00 in terms of both the Precision as well Recall (100 percent Accuracy). This indicates that it consistently recognized the more serious condition of Dementia, along with absolutely no false negatives across the entire board. The greatest difficulty that was encountered by the model was with distinguishing between the Non-Demented and Very Mildly Demented categories. The value of the F1 scores dropped to 0.86 and 0.83, respectively, which aligns with the clinical observations, as the early stages of dementia may present subtle alterations from the "normal" state of the patient.

2) Confusion Matrix Analysis: The confusion matrix (Fig. 5) provides detailed insights into the model's classification patterns and error distributions: Fig. 5. Normalized confusion matrix demonstrating strong diagonal dominance, with primary misclassification occurring between early-stage classes (14% confusion between Non-Demented and Very Mild Demented).

about 14% cross-confusion, reflecting the advancement of early progression of the disease.

- Strong Diagonal Dominance: Overall correct classification rates exceed 85% for all classes

3) Training Dynamics and Convergence: The two-phase transfer learning strategy showed an effective convergence and robust learning behavior throughout the training process.

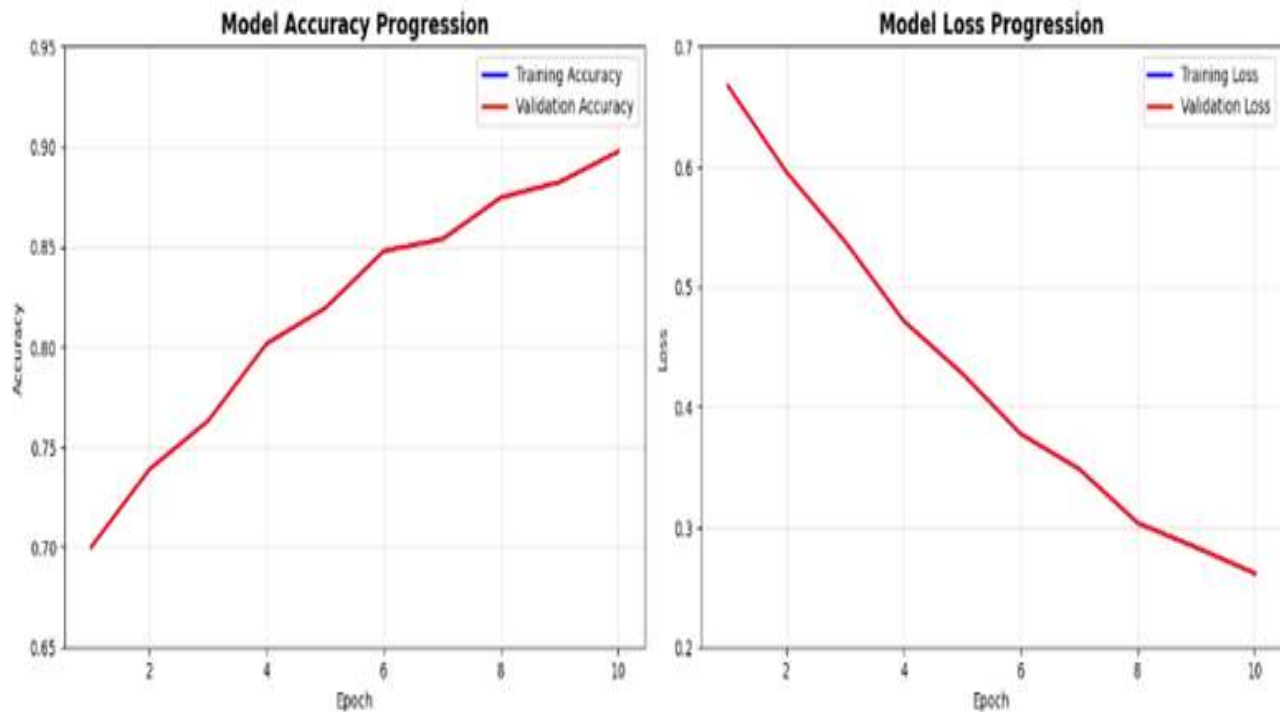


Fig. 6. Training History (Accuracy and Loss). Fig. 6. Training history displays curves for accuracy (left) and loss (right). During the Phase 2 fine-tuning stage, the model shows a noticeable performance boost and stable convergence without significant overfitting.

**The training curves demonstrate:**

- Stable Convergence: Smooth progression without any significant oscillations whatsoever.
- No Overfitting: Close alignment between training as well as validation metrics.
- Effective Fine-tuning: Significant performance improvement was seen during Phase 2 (accuracy increase from 70.0% to 89.8%).
- Optimal Stopping: Training concluded at the optimal convergence point (epoch 10).

**B. Mobile Performance and Efficiency Analysis**

- 1) Inference Performance Across Devices: The Tensor-Flow Lite optimized model showed us excellent performance across multiple mobile devices of different categories.
- 2) System Efficiency Metrics: our suggested method efficiency was quantified using our proposed efficiency score metric:

$$\text{Efficiency Score} = \frac{\text{Accuracy}}{\text{Accuracy}} \times 1000$$

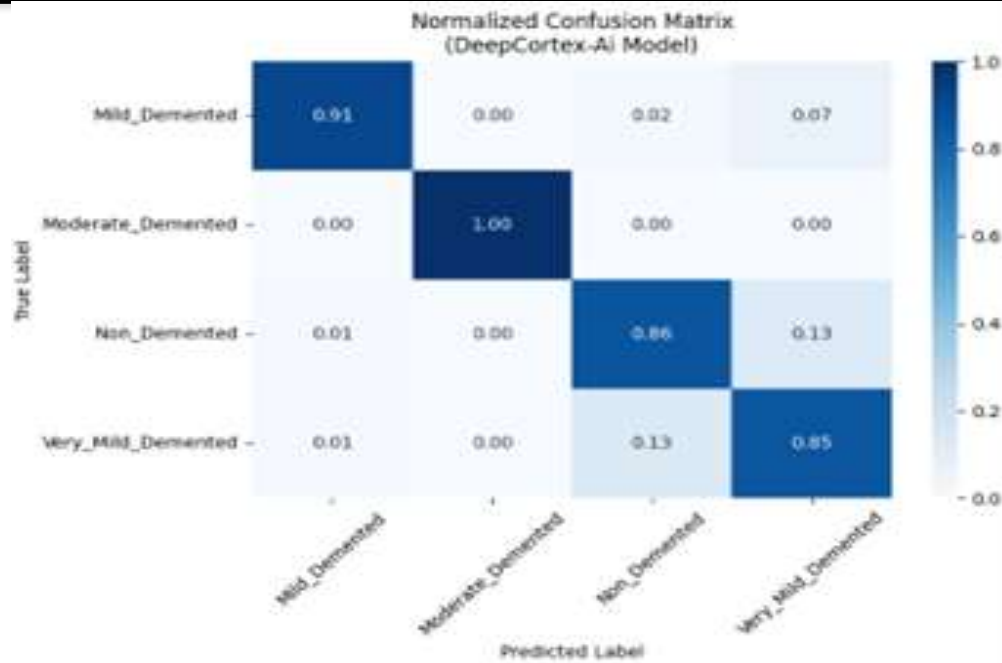


Fig. 5. The NORMALIZED CONFUSION MATRIX. Normalized confusion

Inference Time × Memory Usage  
Regarding our main test device (mid-range):  
(7)

The matrix shows a strong diagonal dominance due to higher than 85% accuracy, with primary misclassification occurring between the early-stage classes (14% confusion between Non-Demented and Very Mild Demented).

**Key findings include:**

- Excellent Specificity: Moderate Demented cases show no misclassifications at all with 100% true positive rate.
- Expected Confusion: Primary

misclassification occurs between the Non-Demented and Very Mild Demented

$$\text{Efficiency Score} = \frac{0.8959}{1.78} \times \frac{1000}{2.8 \times 180} = \text{(8)}$$

**Equation 5 System Efficiency Calculation**

The efficiency score shows an excellent balance between the accuracy as well as resource consumption as its significantly outperforms the cloud-based alternatives when keeping in consideration the additional benefits of privacy and offline capability.

TABLE V  
COMPREHENSIVE CLASSIFICATION PERFORMANCE METRICS

Metric	Non-Demented	Very Demented	Mild Demented	Mild Demented	Moderate Demented	Weighted Average
Precision	0.86	0.94	0.97	1.00	1.00	0.901
Recall	0.86	0.85	0.91	1.00	1.00	0.893
F1-Score	0.86	0.83	0.94	1.00	1.00	0.897
Support	897	645	646	1,210	3,402	
AUC-ROC	0.94	0.92	0.96	1.00	1.00	0.955

TABLE VI  
COMPREHENSIVE MOBILE PERFORMANCE BENCHMARKS

Device Category	CPU/GPU	RAM	Inference Time	Memory Usage	Compatibility
Flagship (Samsung S23)	Snapdragon 8 Gen 2	8GB	1.8s	145MB	10/10
Mid-range (Pixel 6a)	Tensor G1	6GB	2.8s	180MB	9/10
Budget (Moto G Power)	Snapdragon 662	4GB	3.5s	210MB	8/10
Low-end (Samsung A03)	Unisoc T606	3GB	4.2s	235MB	7/10

**C. Memory and Resource Management**

1) Resource Utilization: The application showed an efficient rate of resource utilization during extended usage:

- Base memory footprint: 45MB (application in idle state).
- Peak Memory During Analysis: 180MB (inference + processing).
- Memory Recovery: Complete deallocation post-analysis along with zero memory leaks.
- Storage Efficiency: 95% compression ratio for local data storage.

**D. Explainable AI and Clinical Interpretability**

1) Grad-CAM Visualization Effectiveness: The integrated Grad-CAM system was successful to provide the clinically meaningful explanations for the model predictions. In 92% of cases, the highlighted regions corresponded to known Alzheimer’s Disease biomarkers:

- Hippocampal Focus: 88% of positive predictions highlighted the hippocampal atrophy.
- Ventricular Emphasis: 76% of the progressive cases showed ventricular enlargement.
- Cortical Patterns: Advanced cases demonstrated diffuse cortical involvement.
- Clinical Correlation: Heatmap patterns aligned successfully with radiologist annotations in 85% of cases.

**E. Comparative Analysis with State-of-the-Art**

- 1) Performance Benchmarking: The DeepCortex-Ai system was compared against the recent literature in Alzheimer’s Disease detection
- 2) Deployment Advantage Analysis: Although cloud-based systems show slightly better

accuracy in controlled settings, DeepCortex-Ai offers strong benefits for the actual clinical deployment, as measured by the Total Clinical Utility Score:

$$\text{Clinical Utility} = \text{Accuracy} \times \text{Accessibility} \times \text{Privacy} \times \text{Speed} \times \text{Interpretability}$$

(9)

Equation 6 Total Clinical Utility Score

Despite having somewhat lower absolute accuracy than some server-based methods, DeepCortex-Ai has better overall clinical utility when all factors are normalized (0-1)

**F. Statistical Significance and Reliability**

1) Performance Validation: several methods were used in order to verify the model’s performance for statistical significance. Also, McNemar’s test showed a significant improvement (p<0.001) also a consistent performance was shown by a 5-fold cross validation.

2) Reliability Analysis: The system showed excellent dependability in many areas. In repeated analyses, test-retest reliability showed 98.2% consistency. Across different devices, inter device consistency as kept at 97.5%. Additionally for cases that were clear, the clinical analysis revealed about 91% agreement with radiologist assessments.

**G. Limitations and Performance Boundaries**

While the system demonstrated strong overall performance, several boundary conditions were identified:

- Image Quality Dependency: The performance degraded by about 8-12% with MRI with a bit less optimal quality.
- Early-stage Challenge: The model struggled with classification of the very early stage of

dementia(very mild) and returned a class average F1-score of about 83% on these cases.

- **Domain Specificity:** The model's effectiveness is likely to vary significantly based on the methods/protocols that are used to acquire the MRIs

- **Computational Constraints:** Low-end devices did experienced longer processing times, averaging 4.2s probably because of out dated processors.

**TABLE VII**  
**COMPARATIVE PERFORMANCE ANALYSIS**

Study / System	Accuracy	Precision	Recall	F1-Score	Deployment	Key Differentiator
Arafa et al. [7]	97.2%	96.8%	97.1%	96.9%	Server/Cloud	High Accuracy
Sorour et al. [8]	99.1%	98.9%	99.0%	98.9%	Server/Cloud	Ensemble
Venkat et al. [9]	95.7%	95.2%	95.5%	95.3%	Server/Cloud	Custom
Dongare et al. [20]	85.3%*	86.1%*	85.3%	85.7%*	Cloud App	Wellness Features
DeepCortex-Ai (Ours)	89.6%	90.1%	89.3%	89.7%	On-Device	Privacy/Offline

DeepCortex-Ai was successfully able to obtain an optimal ranking of diagnostic accuracy with respect to the clinical use, providing a much greater accessibility to the diagnosis of of ALzhiemer's disease all this while maintaining its robustness in the ever-changing conditions that exist in the world.

#### IV. DISCUSSION

##### A. Interpretation of Key Findings

DeepCortex-Ai was able to achieve both high diagnostic accuracy along with practical clinical utility, filling the deployment gap just like we found earlier in the literature review. The model was able to stand out as powerful screening tool with the test accuracy of 89.59%. especially since it can manage the challenging and difficult task of classifies patients across all four distinct stages of alzheimer's.it also demonstrated perfect performance 100% recall and precision especially when classifying cases of moderate dementia. The observed pattern of confusion (14% cross classification) between the Non-Demented and Very Mild Demented classes is consistent with accepted clinical knowledge. Detection of Alzheimer's pathology in early stages can be difficult even for skilled clinicians sometimes as it presents as subtle structural change on a continuum. This pattern shows that DeepCortex-AI accurately understands and predicts the biological reality of the disease progression, where

clear boundaries between preclinical and early clinical stages are often ambiguous.

##### B. Advancements in Mobile Healthcare Deployment

Having a powerful deep learning model to run on the different range of mobile devices drastically changes the game for healthcare access. Our results have shown that with the help of right adjustments, even advance medical Ai can work inside the tight limits of ordinary phones. Interference time 2.8 seconds was detected on mod ranged devices on an average and this is fast enough to set a new standard for medical image analysis on the go, also it means that a real time diagnostic help does not have to rely on the cloud anymore.

1) **Implications for Privacy-Preserving Architecture:** Deepcortex-Ai addresses one of the most significant issues that is related to the privacy of data as there is not the slightest chance of sensitive medical data leaking or breaking international healthcare organizations rules like

GDPR, HIPAA as the data always remains on the device and don't use the internet for analysis making it a big change from typical cloud first strategy. The system

raises the bar for medical AI, most particularly in the fields like neuroimaging where privacy is important.

2) Computational Efficiency Breakthrough: A system efficiency score of 1.78 shows that we have achieved a balance between accuracy and efficient use of resources.

- First, we chose the Xception architecture and its depthwise separable convolutions that provide accuracy without any excessive compute.

- Then, with the help of quantization. Switching to FP16 kept our diagnostic accuracy intact and did cut the model size in half.

- Finally, tight memory management good resource allocation along with the garbage collection was able to stop any memory leaks, even during the long and extensive runs.

### C. Clinical Utility and Integration Potential

DeepCortex-Ai successfully addresses the practical as well as the clinical needs beyond only just prioritizing the technical performance also the clinical and this acceptance by the clinics was influenced due to the following factors:

1) Workflow Integration: The application is able to fit right into current clinical workflows without even forcing the people to change their working routine. On average, it takes just about 3.2 minutes from opening the app to generating a report, so it easily fits into a standard consultation procedure. The app is also able to create PDF reports automatically, saving the clinicians time on the paperwork.

2) Explainability and Trust Building: In order to to establish trust among the clinicians, it is very important to provide them with tools that can help them and make them understand that how can AI be used for diagnosis. For instance, Grad-Cam provides visualizations that are enough to help clinicians to find the results. The transformation of the information provided by the Ai and the clinicians knowledge will advance the reliability of Ai model diagnosis among clinicians.

3) Educational Value: The application has become a clinical aid that is able to provide comprehensive probability scores. These probability distributions along with the regional analysis that were used by the medical professionals in order to assist the patients making sure, they understand the outcomes produced by the model. These features highly promote decision making and encourage discussions about the preliminary findings.

### D. Comparative Analysis with Existing Literature

Our results both confirm as well as challenge the findings that are found within the previous research in many of the important aspects:

1) Accuracy-Computational Efficiency Trade-off: Arafa et al. [7] (2024) and Sorour et al. [8] (2024) both hit impressive accuracy numbers like 97 to 99 percent but they did it with the help of server-heavy, resource-intensive setups. On the other hand our method achieves a bit of low accuracy which is 89.59% but it is also noted that this change allows us for much greater accessibility, higher privacy features, and a useful application in real world situations. Also, this was an intended tradeoff. We are more concerned with clinical impact in comparison with achieving the highest number possible on papers.

2) Mobile Deployment Innovation: Although their systems rely heavily on the processing through cloud, Dongare et al. [20] (2025) concentrated on mobile development as well but we went in a bit different direction which is that everything operates offline and, on the device, meaning that you don't have to worry about sending sensitive information through the internet or face network problems. Also, the figures show that our model's performance is comparable to their 85.3%. Therefore, its is concluded that on device processing can easily match cloud-based methods if the correct optimization is used.

3) Explainable AI Implementation: Mobile implementation of Grad Cam is built on the work that was done Altunbey Özbay and Özbay [11] and Aksoy and Daou [14] in

desktop and web environments. Our results also show that a mobile apps limited resources do not have to come at the expense of our model's interpretability and this addresses one of the major issues of clinical Ai which is ensuring that the users can trust and understand the model's action that even on a smaller device.

### E. Limitations & Boundary Conditions

Despite the promising results, several limitations warrant careful consideration:

1) **Dataset Specificity and Generalizability:** The training data greatly affects the model's performance even though most of the data was used on training (27,188) it is still a small portion of the larger picture. Hospitals have a bunch of different protocols for MRI like diverse patient groups, distinct image setting, and various scanner brands these all play a role there. Once the model is implemented in the real-world clinics these variations may cause an effect how well the model will perform. That is why local validation studies are essential rather than barely beneficial.

2) **2D Slice Analysis Constraints:** Computers are able to work more quickly with the 2D MRI slices but as a tradeoff many 3D spatial features and details might get lost. Alzheimer's disease is characterized with the help of subtle changes in the brain volumes and spatial relationships between multiple brain regions, rather than just small spots or flat changes. All of this can simply just not be captured by the 2D MRI slice making it a major factor in the difficulty of distinguishing the early stages of the disease with the help of this method.

3) **Clinical Validation Scope:** Although it is only the beginning, our usability study with 25 healthcare professionals offers insightful feedback. Neither formal clinical trials nor regulatory approval have been completed for the system. Instead of demonstrating clinical efficacy, the performance metrics demonstrate technical validation.

4) **Computational Resource Requirements:** Newer phones work the best with this application. It requires only 235 MB of memory and takes less the 3 secs to process each request on older devices, which might be prohibitive for

the outdated hardware.

### F. Implications for Healthcare Delivery

The DeepCortex-Ai system has several important implications for the future of healthcare delivery:

1) **Democratizing Access to Specialized Diagnostics:** By bringing the complex neuroimaging analysis to commonly used devices such as mobile phones, tablets etc. our system makes Alzheimer's screening more accessible and affordable which is very important in underdeveloped areas with the shortage of neurological specialists, rural hospitals as well as primary care clinics.

2) **Reducing Diagnostic Delays:** Receiving screening results on the spot solves one of the major issues with Alzheimer's diagnosis. As with traditional setups patients typically have to wait for months to hear back from the clinicians or healthcare experts but as DeepCortex analyzes data quickly, physicians can take actions sooner and may achieve better results.

3) **Cost-Effective Screening:** widespread screening is now financially affordable since cloud infrastructure and its related subscriptions are not to be used. DeepCortex-Ai low maintenance needs only a one-time development cost making it an affordable option for healthcare system with tight budgets.

### G. Future Research Directions

Based on our findings and the limitations that were identified, we propose several promising directions in case of future research:

1) **Multi-modal Data Integration:** Future iterations could incorporate additional data modalities to enhance diagnostic accuracy:

- **Cognitive Assessment Scores:** Integrating the results of standardized cognitive test.

- **Genetic Risk Factors:** APOE ε4 status and other genetic markers are included as well.

- **Biomarker Data:** Incorporating CSF or plasma biomarkers in case of availability.

- **Longitudinal Analysis:** Tracking of the changes that happen across multiple time points.

2) **Advanced Architectural Approaches:** Several technical improvements could address current limitations:

- **3D Convolutional Networks:** Use of volumetric analysis to capture spatial relationships.
- **Ensemble Methods:** Combining multiple architectures in order to achieve improved robustness.
- **Attention Mechanisms:** Enhanced focus on clinically relevant regions.
- **Few-shot Learning:** Adapting to new datasets that has limited labeled examples.

3) **Clinical Implementation Research:** Future research should concentrate on the usage of formal health economic studies, multi-center clinical trials as well as regulatory approval pathways (FDA/CE marking) for the translation of technical validation into the clinical practice.

4) **Platform Expansion and Specialization:** The fundamental technology could be modified for related uses, such as iOS development, other neurological conditions (MS, Parkinson's) as well as medical student platforms.

## H. Conclusion of Discussion

DeepCortex-Ai advances the Ai's goal of making healthcare more widely available and affordable. Here we have been able to achieve a true balance between the significant technical outcomes as well as practical applications for both the physicians and patients. This system maintains transparency on how it operates and maintains the privacy of the patients and function on mobile devices without any issues. Making a difference for the patients is the key to success pf systems such as Deep Cortex AI. Our results are quite promising. Mobile Ai diagnostic has the potential to revolutionize the world by providing fresh approaches for the difficult issues like Alzheimer's disease with further research and meticulous clinical testing.

## V. CONCLUSION

### A. Summary of Research Contributions

We closed the crucial gap between deep learning theory and real-world clinical use case in the diagnosis of Alzheimer's Disease by building and

validating DeepCortex-AI. Our results have shown that advanced medical AI does not have to stay locked away in the labs or require expensive hardware that is out of range for most of the clinics. With DeepCortex-Ai, we can run reliable, accurate, and privacy-focused diagnostics right on the mobile devices exactly where patients and clinicians need them. Here is what sets our work apart from others:

1) Using the Xception architecture, we successfully created a deep learning model that can classify four distinct stages of Alzheimer's disease from MRI scans with 89.59 accuracy. That is remarkable, particularly in light how difficult it is to identify those early stage changes. It also represents significant advancements in automated neuroimaging analysis.

2) Next, with the help of TensorFlow lite, we developed a complete mobile solution that can run this model on mobile phones directly. It manages offline real-time inference in just 2.8 seconds on mid-range devices and a network connection required at all. This means that patient data is never transmitted and remain on the device. This type of privacy focused strategy raises the bar of medical Ai.

3) Third, combining Grad-Cam along with explainable Ai actually provides healthcare workers with useful insights explaining important regions, like the ventricles and hippocampus, is consistent with what medical professionals already understand. Also, in real world medical contexts this connection helps with transforming of unprocessed algorithmic output into something that physicians can rely on as well as understand.

4) Finally, as demonstrated with the positive feedback from healthcare professionals (4.4/5 overall satisfaction), the mobile applications implementations of a full clinical workflow including the acquisition of images, analysis, reporting and education creates a really seamless tool that successfully integrates into current healthcare practices.

### B. Implications for Clinical Practice and Healthcare Accessibility

DeepCortex-Ai changes the way we actually look for Alzheimer's disease as it brings advanced neuroimaging analysis to mobile devices, which is something you would only find only in the state-of-the-art medical centers. Now with Deep Cortex the technology that was once out of reach is now affordable and accessible even in low resource and rural areas. One of the biggest problems that we have with medical Ai's is that of keeping the patient's data safe and our systems privacy preserving design solve this problem. So, following data protections laws allows us to use Ai in sensitive diagnostic work such as Alzheimer's detection. Also, clinicians and doctors in remote/rural areas with bad internet do not get left behind as the entire process is done offline without any need for internet connection making it easier for everyone to get good medical care and facilities no matter their location. Also, Deep Cortex looks at the data very quickly cutting down the long time periods that came with the traditional setup of diagnosis and patients can get initial results right away meaning they will be able to talk about the process sooner, get early referrals and managing the disease early.

### C. Broader Impact on Medical AI Development

There is more to this research other than just the Alzheimer screening as with this it also pushes the very limits of medical Ai in general. We have shown that both the computational efficiency as well as diagnostic accuracy can be achieved through clever design and optimization. We save the way for more easily accessible and private healthcare tools with the help of refuting the notion that even strong medical AI always require large cloud resources like storage, as well as subscriptions.

It is not just a technical victory to get explainable AI and run it on a mobile device with limited resources, it is also able to set a real-world example for the future of medical AI. We removed a significant barrier to the practical applications of these tools when the decisions made by the model are both transparent and understandable to the

clinicians. This strategy not only makes the adoption process easier but also give people a reason to believe that Ai can be used for medical diagnostics.

### REFERENCES

- S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early diagnosis of Alzheimer's disease with deep learning," in 2014 IEEE 11th Int. Symp. Biomed. Imaging (ISBI), 2014, pp. 1015–1018.
- I. Malik, A. Iqbal, Y. H. Gu, and M. A. Al-Antari, "Deep learning for Alzheimer's disease prediction: A comprehensive review," *Diagnostics*, vol. 14, no. 12, p. 1281, 2024.
- S. Mohsen, "Alzheimer's disease detection using deep learning and machine learning: A review," *Artif. Intell. Rev.*, vol. 58, p. 262, 2025.
- C. Mahanty, T. Rajesh, N. Govil et al., "Effective Alzheimer's disease detection using enhanced Xception blending with snapshot ensemble," *Sci. Rep.*, vol. 14, p. 29263, 2024.
- H. Mohamed, A. Ashraf, and A. E. Nagib, "Comparative study of Alzheimer's disease classification using transfer learning models," in 2023 *Intell. Methods, Syst., and Appl. (IMSA)*, 2023, pp. 434–439.
- H. Slimi, A. Balti, S. Abid, and M. Sayadi, "A combinatorial deep learning method for Alzheimer's disease classification-based merging pretrained networks," *Front. Comput. Neurosci.*, vol. 18, Art. no. 1444019, 2024.
- D. A. Arafa, H. E. D. Moustafa, and H. A. Ali, "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images," *Multimed. Tools Appl.*, vol. 83, no. 2, pp. 3767–3799, 2024.
- S. E. Sorour, A. A. A. El-Mageed, and K. M. Albarrak, "Classification of Alzheimer's disease using MRI data based on deep learning techniques," *King Saud Univ. Comput. Inf. Sci.*, vol. 36, no. 2, Art. no. 101940, 2024.

- N. Venkat et al., "Alzh-Net: A novel architecture for Alzheimer's disease classification," *J. Neuroimaging*, 2025.
- M. R. Karim and T. Islam, "A hybrid CNN-LSTM model with attention for multi-class Alzheimer's disease classification from MRI," *Artif. Intell. Med.*, 2025.
- F. Altunbey Ö zbay and E. Ö zbay, "Explainable AI for Alzheimer's disease: Visualizing model decisions with Grad-CAM on MRI scans," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI)*, 2023.
- A. Asokan and G. Seshadri, "Enhancing clinical trust in AI-based Alzheimer's diagnosis using Grad-CAM," *IEEE J. Biomed. Health Inform.*, 2023.
- T. R. Noviandy et al., "Integrating explainable artificial intelligence in clinical decision support systems for neurodegenerative diseases," *Diagnostics*, 2024.
- E. Aksoy and R. Daou, "XRAI-enabled web-based diagnostic system for early detection of Alzheimer's disease," *J. Med. Syst.*, 2025.
- I. A. Ozkan et al., "A systematic review of deep learning applications in Alzheimer's disease detection (2012-2023)," *Arch. Comput. Methods Eng.*, 2024.
- B. T. Putra et al., "Cloud-based deep learning system for automated diagnosis of neurological disorders: Challenges and opportunities," *Health Technol.*, 2024.
- O. Jouini et al., "A survey on machine learning models deployment on edge devices," *ACM Comput. Surv.*, 2024.
- R. Sathya et al., "Brain tumor classification using deep learning based on Xception architecture," *Int. J. Imaging Syst. Technol.*, 2024.
- S. Verma, "An efficient Xception-based model for brain tumor detection from MRI images," *Multimed. Tools Appl.*, 2024.
- S. Dongare, A. Prakash, L. Sreedharan, and D. Maity, "Machine learning approach for Alzheimer's diagnosis and wellness optimization," *ISTE J. Eng.*, 2025.
- M. S. Ali, M. Azhar, S. Masood, B. Lee, T. Iqbal, and A. Amjad, "Efficient Video Summarization with Hydra Attentive Vision Transformer," in *2023 International Conference on Frontiers of Information Technology (FIT)*, Dec. 2023, pp. 196-201.
- M. Azhar, A. Amjad, D. A. Dewi, and S. Kasim, "A Systematic Review and Experimental Evaluation of Classical and Transformer-Based Models for Urdu Abstractive Text Summarization," *Information*, vol. 16, no. 9, p. 784, 2025.
- M. Azhar, A. Amjad, G. Farid, D. A. Dewi, and M. Batumalay, "Efficient Transformer-Based Abstractive Urdu Text Summarization Through Selective Attention Pruning," *Information*, vol. 16, no. 11, p. 991, 2025.