Deep Learning & Generative AI in Healthcare

Session 04

Convolutional Filters and Feature Detectors

Motivation for CNNs

Fully Connected Networks Problem:

- $10^3 \times 10^3$ color image = 3×10^6 pixels
- 1,000 hidden units → 3×10^9 weights in first layer alone!
- Must learn invariances from data (huge datasets needed)

CNN Solution:

Incorporate inductive biases about image structure → dramatically reduce parameters and improve generalization

Key Concepts

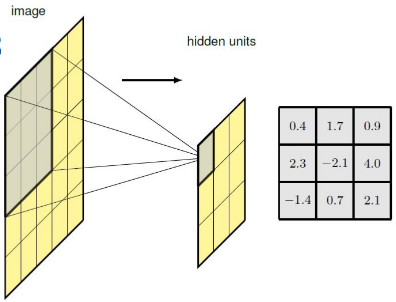
Hierarchy: Natural hierarchical structure (edges → shapes → objects)

Locality: Features detected from local patches

Equivariance: Translation equivariance (shift input → shift output)

Invariance: Small translations don't affect classification

Receptive Field &



Receptive Field: Small rectangular patch from image that a unit receives as input

Example: 3×3 patch visualized as kernel/filter

Feature Detector Output:

z = ReLU(w^T x + w_0)

Maximum response when image patch matches kernel (up to scaling)

CNN Architectures: ImageNet and VGG16

ImageNet Challenge

Dataset:

- 14 million natural images
- ~22,000 categories (hand-labeled)
- Challenge: 1,000 non-overlapping categories
- 1.28M training, 50K validation, 100K test

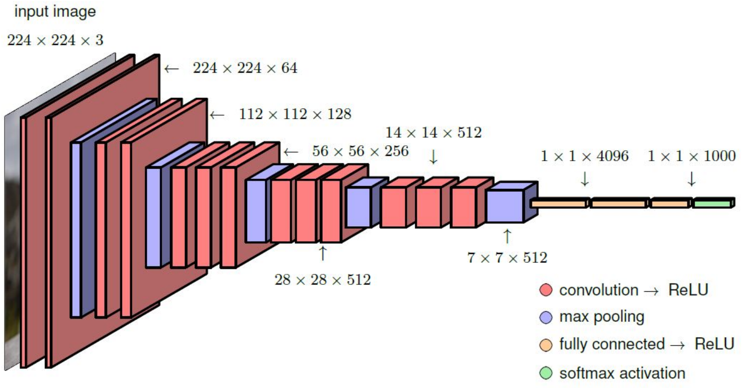
Evaluation Metrics:

- Top-1 error: true class at rank 1
- Top-5 error: true class in top 5 predictions
- Random guessing: 99.9% error

Historical Performance

Model	Year	Top-5 Error
Early results	~2010	~25.5%
AlexNet	2012	15.3%
Later advances	~2017	~3%
Human-level	-	~5%

VGG-16 Architecture



Design Principles (Simonyan & Zisserman, 2014):

- Input: 224×224×3 pixels
- All conv filters: 3×3, stride 1, same padding, ReLU
- All pooling: 2×2, stride 2 (down-samples by 4×)
- Channels: 64→64→128→256→512 (doubling with down-sampling)
- Final: 3 fully connected (4096, 4096, 1000 units)
- ~138M parameters (103M in first FC layer!)

Introduction to Medical Image Analysis

Why Medical Image Analysis?

Essential for: Precise diagnosis, treatment planning, non-invasive monitoring, anomaly detection, and improving patient outcomes

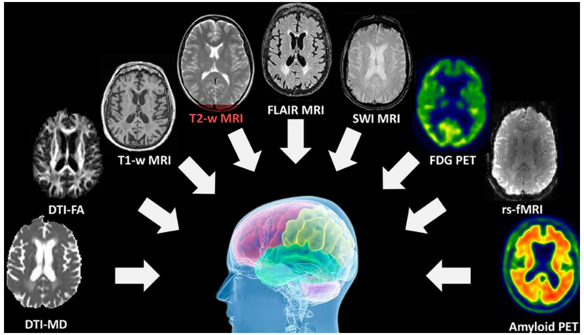
Key Achievement: Deep learning achieves up to 95% accuracy in disease detection, outperforming traditional methods by 10–15%

Traditional vs Deep Learning

Traditional Methods	Deep Learning
Manual effort required	Automated processing
Subjective interpretation	Consistent results
Limited scalability	Highly scalable
Hand-crafted features	Automatic feature learning

Medical Image Analysis Pipeline

1. **Image Acquisition:** MRI, CT, X-ray, Ultrasound
2. **Pre-processing:** Noise removal, normalization, filtering
3. **Feature Extraction:** Edges, textures, shapes (automatic in DL)
4. **Model Training:** Learn patterns from annotated datasets
5. **Classification:** Disease detection, severity grading, prognosis



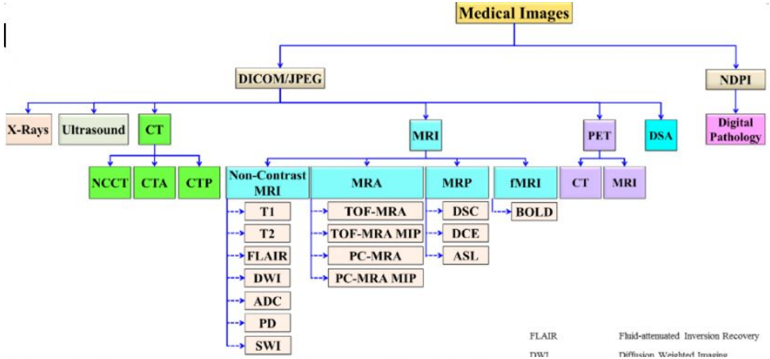
Medical Imaging Modalities Overview

File Format Categories
DICOM/JPEG: X-Ray, Ultrasound, CT, MRI, PET, DSA | NDPI: Digital Pathology

Modality	Subtypes	Applications
CT	NCCT, CTA, CTP	Anatomy, vessels, perfusion
MRI	T1, T2, FLAIR, DWI, ADC, PD, SWI	Soft tissue, brain lesions
MRA	TOF-MRA, PC-MRA, MIP variants	Vascular imaging
MRP	DSC, DCE, ASL	Tissue perfusion
fMRI	BOLD	Brain activity mapping
PET	PET, PET-CT, PET-MRI	Metabolic imaging

CT Abbreviations
NCCT: Non-contrast CT | CTA: CT Angiography | CTP: CT Perfusion

MRI Sequences
FLAIR: Fluid-attenuated Inversion Recovery | DWI: Diffusion Weighted | ADC: Apparent Diffusion Coefficient | SWI: Susceptibility Weighted



MR Angiography (MRA)
TOF-MRA: Time-of-Flight | PC-MRA: Phase-contrast | MIP: Maximum Intensity Projection

Perfusion Imaging (MRP)
DSC: Dynamic Susceptibility Contrast | DCE: Dynamic Contrast Enhanced | ASL: Arterial Spin Labeling | BOLD: Blood-oxygen-level Dependent

Deep Learning Implications
- Different modalities require modality-specific preprocessing
- Multi-sequence MRI enables richer feature extraction
- Transfer learning across modalities remains challenging
- Digital pathology (NDPI) uses whole-slide imaging at gigapixel scale

Medical Imaging Modalities

Primary Imaging Techniques

Modality	Technology	Primary Applications
MRI	Magnetic fields + radio waves	Brain, spinal cord, soft tissues
CT	X-rays (cross-sectional)	Chest, abdomen, fractures
X-ray	Ionizing radiation	Bones, lungs, dental
Ultrasound	High-frequency sound waves	Fetal monitoring, soft tissues

Classification Applications by Domain

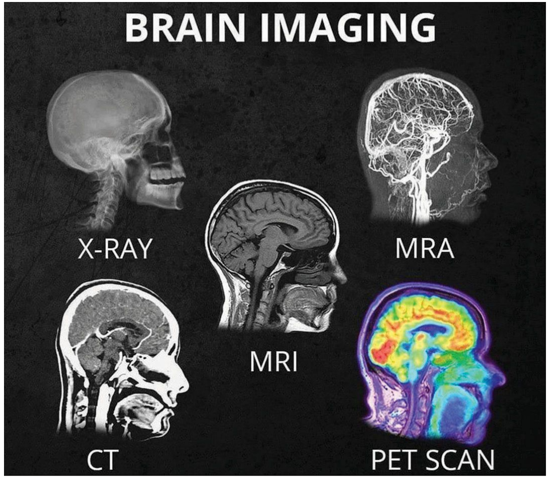
Cancer Classification: Breast, lung, brain, liver tumors via mammography, MRI, CT

Neurological: Alzheimer's, stroke, multiple sclerosis classification via brain MRI/CT

Cardiovascular: Coronary artery disease risk stratification via echocardiography, CT angiography

Pulmonary: Pneumonia, tuberculosis, COPD classification via chest X-rays and CT

Ophthalmology: Diabetic retinopathy severity grading via retinal imaging



Deep Learning Architectures For Classification

Core Architectures for Predictive Modeling

Architecture	Strength	Classification Task
CNN	Spatial feature extraction	Disease detection, grading
LSTM/RNN	Sequential dependencies	Progression prediction
Transformers	Global context	Multi-class classification
Ensemble	Combined predictions	Robust classification

Supervised Learning for Classification

Binary Classification: Healthy vs diseased, benign vs malignant

Multi-class Classification: Disease subtypes, severity grades (Stage I–IV)

Multi-label Classification: Multiple conditions from single image

Key Breakthrough Factors

The deep learning breakthrough in the 21st century was fueled by three key factors:

- Exponential growth in computational power (GPUs)
- Availability of large-scale labeled datasets
- Significant algorithmic innovations

DL vs Traditional ML

Traditional ML: Decision Trees, SVM, Random Forest require manually engineered features from domain experts

Deep Learning: Automatically learns hierarchical features directly from raw data, eliminating manual feature engineering

Convolutional Neural Networks

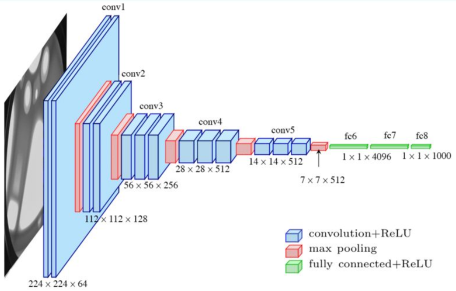
CNN Building Blocks

Convolutional Layer: Applies learnable filters to extract local features (edges, textures, patterns). Each filter produces a feature map.

Activation (ReLU): Introduces non-linearity by setting negative values to zero, enabling complex pattern learning

Pooling Layer: Reduces spatial dimensions via max/average pooling, preventing overfitting while preserving features

Fully Connected Layer: Integrates learned features for final classification output



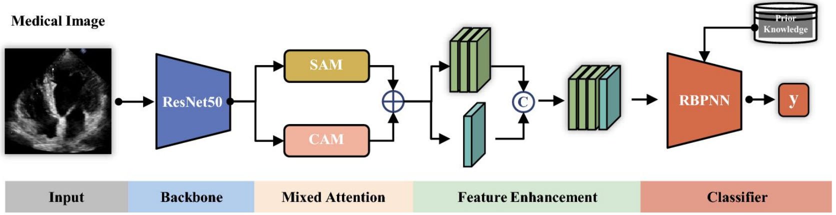
Why CNNs for Classification?

Advantages:

- Hierarchical feature learning (edges → shapes → objects)
- Translation invariance for robust classification
- State-of-the-art on image classification benchmarks
- Efficient for single-modality imaging (MRI, CT, X-ray)

Classification Output

Softmax Layer: Converts final layer outputs to probability distribution across classes



CNN Classification Applications in Medical Imaging

Disease Detection & Classification

Binary Classification:

- Chest X-rays → Pneumonia (present/absent)
- Mammograms → Breast cancer (benign/malignant)
- Skin images → Melanoma (cancerous/non-cancerous)

Multi-class Classification:

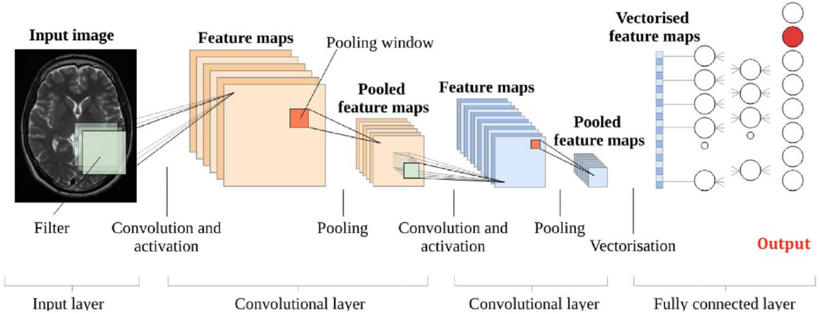
- Brain MRI → Tumor type (glioma, meningioma, pituitary)
- Retinal images → DR severity (Grade 0–4)
- Lung CT → Nodule classification (benign, malignant, indeterminate)

Popular CNN Architectures

Architecture	Key Feature
VGG-16/19	Simple, uniform structure
ResNet	Skip connections, very deep
Inception	Multi-scale feature extraction
DenseNet	Dense connectivity

Classification Performance

Application	Architecture	Accuracy
Skin cancer	Ensemble DCNN	99.53%
Brain tumor	CNN + GradCAM	98%
Lung cancer	CNN + UQ	94%
Breast tumor	ResNet + FL	96%
COVID-19	Ensemble CNN	89%



Transfer Learning for Medical Image Classification

Why Transfer Learning?

Challenge: Medical imaging datasets are often small and expensive to annotate compared to natural image datasets

Solution: Leverage pre-trained models (trained on ImageNet) and fine-tune for medical classification tasks

Transfer Learning Strategies

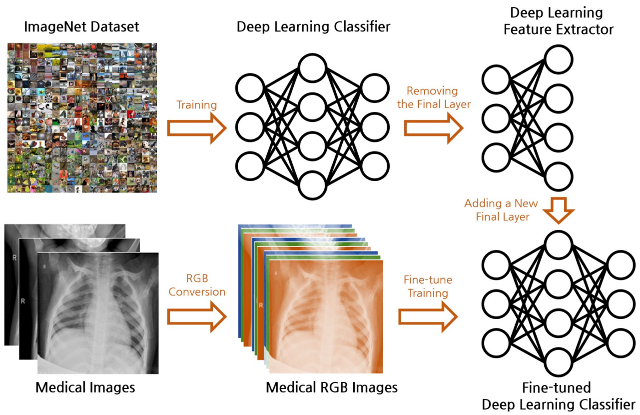
Feature Extraction: Freeze pre-trained layers, train only classifier head

Fine-tuning: Unfreeze some/all layers, train with lower learning rate

Domain Adaptation: Adapt model to handle distribution shift between natural and medical images

Pre-trained Models for Medical Imaging

Model	Parameters	Medical Application
VGG-16	138M	Retinal, chest X-ray
ResNet-50	25M	Tumor classification
InceptionV3	24M	Skin lesion, pathology
DenseNet-121	8M	Chest X-ray, CT
EfficientNet	5-66M	Multi-modal imaging



Ensemble Methods for Robust Classification

Why Ensemble Learning?

Principle: Combine multiple models to achieve better predictive performance than any single model

Benefits:

- Reduced variance and overfitting
- Improved generalization
- Robust to individual model errors
- Better uncertainty estimation

Ensemble Strategies

Model Averaging: Average predictions from multiple models (VGG, ResNet, DenseNet)

Voting: Majority vote for final classification decision

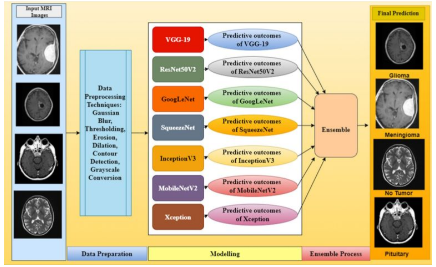
Stacking: Meta-learner combines base model outputs

Medical Imaging Examples

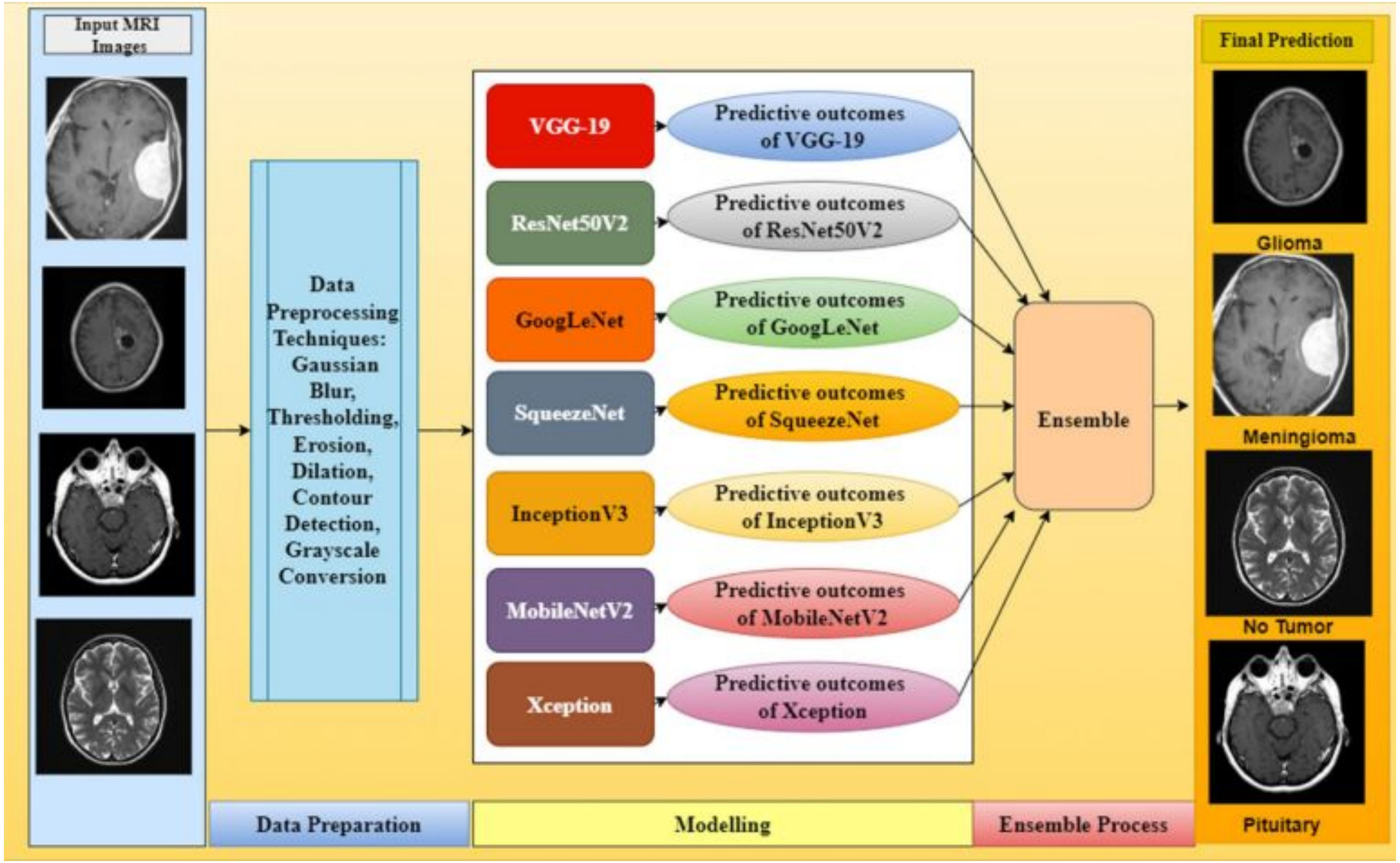
Application	Ensemble	Performance
Skin cancer	3 DCNNs	99.53% acc
COVID-19	ResNet + Inception + DenseNet	89% acc, 0.998 AUC
Tumor seg.	VGG19 + ResNet50 + DenseNet121	98.1% acc
Lung cancer	ResNet + DenseNet + Inception	F1: 0.845

Uncertainty Quantification

Monte Carlo Dropout: Multiple forward passes with dropout for uncertainty estimation in classification



Ensemble Methods for Robust Classification



Classification Performance Metrics

Primary Metrics

Accuracy: (TP + TN) / Total
Overall correctness of classification

Sensitivity (Recall): TP / (TP + FN)
Ability to detect positive cases — critical for disease screening

Specificity: TN / (TN + FP)
Ability to identify negative cases correctly

Precision: TP / (TP + FP)
Proportion of positive predictions that are correct

F1-Score: 2 × (Precision × Recall) / (Precision + Recall)
Harmonic mean balancing precision and recall

Advanced Metrics

AUC-ROC: Area under ROC curve measuring discrimination ability across thresholds

Cohen's Kappa: Agreement measure accounting for chance — important for multi-rater consistency

Benchmark Performance

Study	Metric	Value
Skin cancer (Chanda 2024)	Accuracy	99.53%
Brain tumor (Alshuhail 2024)	ROC-AUC	99–100%
Prostate cancer (Li 2024)	AUC	0.83
Pancreatic cancer (Hong 2024)	AUC	0.998
COVID-19 (Chatterjee 2024)	F1-Score	0.875

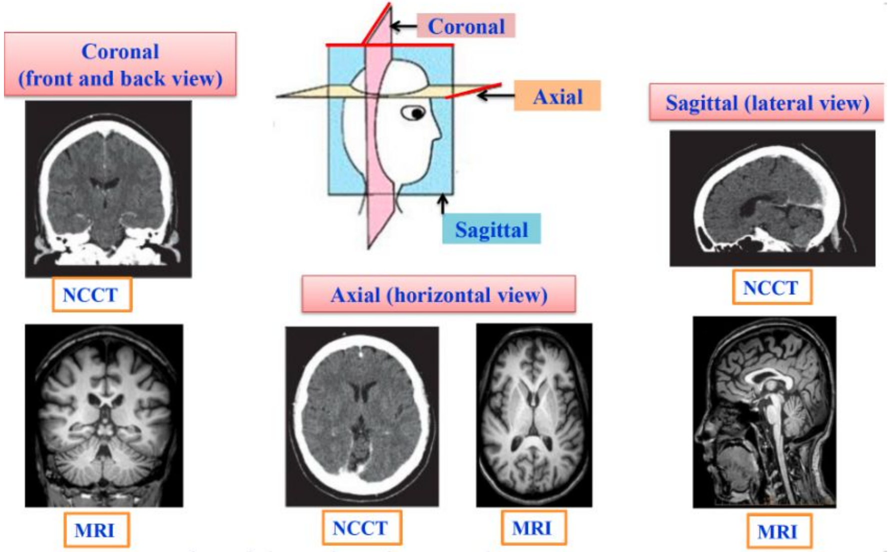
Anatomical Imaging Planes

Plane	Orientation	View
Axial	Horizontal slice	Top-down (superior/inferior)
Coronal	Vertical front-to-back	Front/back (anterior/posterior)
Sagittal	Vertical left-to-right	Side view (lateral)

Axial (Transverse)
Divides body into superior and inferior portions; most common plane for brain CT/MRI scans

Coronal (Frontal)
Divides body into front and back; useful for assessing bilateral structures

Sagittal
Divides body into left and right; midsagittal passes through midline



Clinical Relevance
Deep learning models must handle all three planes; multi-planar reconstruction enables 3D analysis from 2D acquisitions

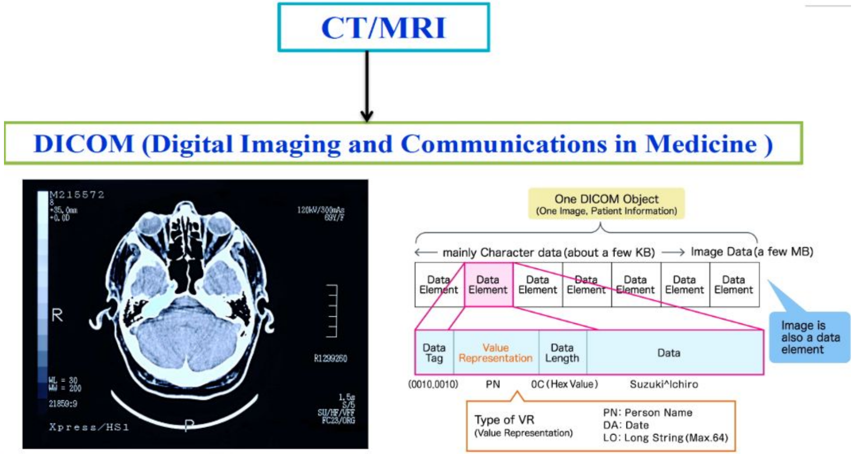
Medical Imaging File Formats

DICOM (Digital Imaging and Communications in Medicine)
International standard for medical image storage and transmission; combines image data with rich metadata

Feature	DICOM	NIfTI
Extension	.dcm	.nii, .nii.gz
Primary Use	Clinical workflow	Research / analysis
Metadata	Extensive (patient, device)	Minimal (spatial info)
Structure	One file per slice	Single 3D/4D volume
Privacy	Requires de-identification	Inherently anonymous

DICOM Data Elements
Each element contains: Data Tag (group, element) + Value Representation (VR) + Data Length + Data value

Common VR Types
PN: Person Name | DA: Date | LO: Long String (max 64) | UI: Unique Identifier | IS: Integer String



NIfTI (Neuroimaging Informatics Technology Initiative)
Designed for neuroimaging research; stores entire 3D/4D volumes in a single file with spatial transformation matrices

Deep Learning Considerations

- DICOM → NIfTI conversion common for preprocessing
- Libraries: pydicom, nibabel, SimpleITK
- Image data typically stored as last element (largest)
- Pixel spacing and orientation critical for model training

Recent Advances in Medical Image Classification

Innovative Classification Architectures

Technique	Application	Performance
Deep Forest Model	Melanoma classification	88.5% accuracy
3D Efficient CapsNet	Prostate cancer risk	AUC: 0.83
MIST Transformer	Histopathology subtype	93.6% accuracy
EfficientNet-b3	Pancreatic cancer	96.7% accuracy
VisionDeep-AI	Fundus classification	81.5% accuracy

Semi-Supervised Classification

SPLAL: Similarity-based pseudo-labeling for classification with limited labeled data

COSST: Comprehensive supervision and self-training for multi-organ classification

Explainable Classification

Grad-CAM: Visual explanations for CNN classification decisions
Application: Brain tumor classification with interpretability

RXD Framework: Explanation supervision improving nodule classification (24–80% AUC improvement)

Federated Learning for Classification

Privacy-Preserving Classification:

- Train across institutions without sharing data
- ResNet + FL for breast tumor classification
- DCFL for pulmonary nodule detection