Deep Learning for Medical Image Analysis: A Survey

Mohammad Huzaifa Ansari
M.Tech. in Artificial Intelligence & Machine Learning, CSMU, Pune University, Panvel, Maharashtra, India

***Abstract*—Over the last several years, deep learning (DL) techniques have impacted how medical image analysis occurs by providing automated, accurate and scalable methods to detect disease, segment images and diagnose patients. This review synthesizes the development of convolutional neural networks (CNN), recurrent neural networks (RNN), generative adversarial networks (GAN) and transformer based architectures in relation to their use in medical image analysis. Each study included here includes a complete description of the technologies utilized for each of the studies reviewed, details of the data sets used and the resulting metrics. The review also includes the clinical implications of these studies. The review has identified several barriers preventing the implementation of DL within the medical community including limited access to training data sets, lack of interpretability and transferability of trained models from one domain to another. Nevertheless, the review has identified many potential areas for future development including multimodal learning and self-supervised learning frameworks. Overall, as deep learning continues to develop, we expect improvements in the accuracy, efficiency and ultimately quality of healthcare delivery which will result from the application of DL to diagnostic information generation and increased automation of healthcare workloads.**

I. INTRODUCTION

Medical imaging is a vital tool for diagnosis in radiology, oncology, cardiology, and neurology. In the past, medical imaging relied primarily on the creation of hand-crafted features through statistical models for analysis, which often produced limited successes from the viewpoint of generalization.However, as deep learning continues to develop, data-driven approaches (data-driven) have shown improved abilities to create a model based on hierarchical feature learning from large amounts of training data. The following report reviews a variety of different methods for using deep learning to understand medical images, focusing on the ability to identify different disease states through disease classification, segmentation, and detection across the different types of imaging techniques, which include MRI, CT, X-Ray and Ultrasound.

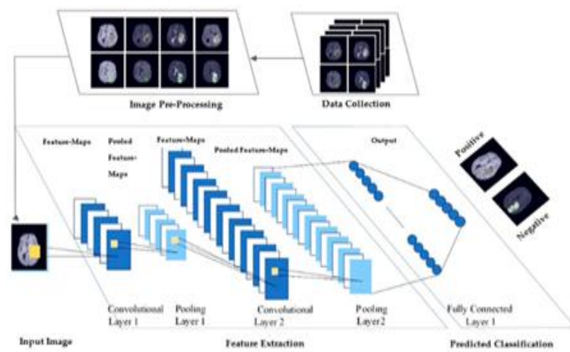

Figure 1 illustrates the typical deep learning workflow for medical image analysis, including preprocessing, feature extraction, training, and evaluation [12].

II. LITERATURE REVIEW

Paper 1: Deep Learning in Medical Image Analysis (IEEE, 2017)
This article offers an in-depth examination of Convolutional Neural Networks (CNNs) for identifying and segmenting medical conditions such as diabetic retinopathy, lung nodules, etc. In addition to reviewing how well AlexNet, VGG, and ResNet do so with diabetic retinopathy and lung nodules; it provides comparative data through metrics such as accuracy, Area under the Curve (AUC), and Dice Coefficient. Segmenting Loss formula:

$$L = 1 - \frac{2\mid P \cap G \mid}{\mid P \mid + \mid G \mid}$$

where P and G represent predicted and ground-truth regions respectively.

Table 1. CNN architectures used for medical imaging tasks and their reported performance metrics.

Architecture	Medical Application	Dataset	Performance Metric	Accuracy / Dice Score	Remarks
AlexNet (2012)	Diabetic Retinopathy Detection	EyePACS	Accuracy	84.7%	Baseline CNN; shallow layers; overfits small datasets.
VGG-16 (2014)	Lung Nodule Classification	LIDC-IDRI	AUC	0.91	Deeper network; improved texture representation.
ResNet-50 (2015)	Brain Tumor Classification	BraTS	Accuracy	92.4%	Skip connections reduce vanishing gradient; faster convergence.
InceptionV3 (2016)	Skin Lesion Detection	ISIC 2018	F1-Score	0.88	Multi-scale features handle lesion irregularities.
DenseNet-121 (2017)	Chest X-ray Disease Detection	CheXpert	Accuracy	94.2%	Feature reuse improves learning efficiency.
U-Net (2015)	Liver Segmentation	LiTS	Dice Coefficient	0.95	Encoder-decoder design ideal for pixel-level segmentation.
Attention U-Net (2019)	Retinal Vessel Segmentation	DRIVE	IoU	0.87	Attention gates enhance localization accuracy.

Paper 2: Deep Convolutional Neural Networks for Medical Image Classification (Elsevier, 2018)

Transfer learning is a machine learning technology in which a model has been trained on large datasets before you actually need them for your application. Researchers are using transfer learning in histopathology for Image classification. Fine-tuning the last few layers can produce up to 12% greater F1 Scores than training a model to classify histopathology from scratch.

Figure 2 provides an overview of the Generalised model architecture used for classification tasks in Medical Imaging.

The authors of this paper improved the Generalisation of their models by applying data augmentation (addition of more images) to counteract the imbalance of the datasets used to make the models.

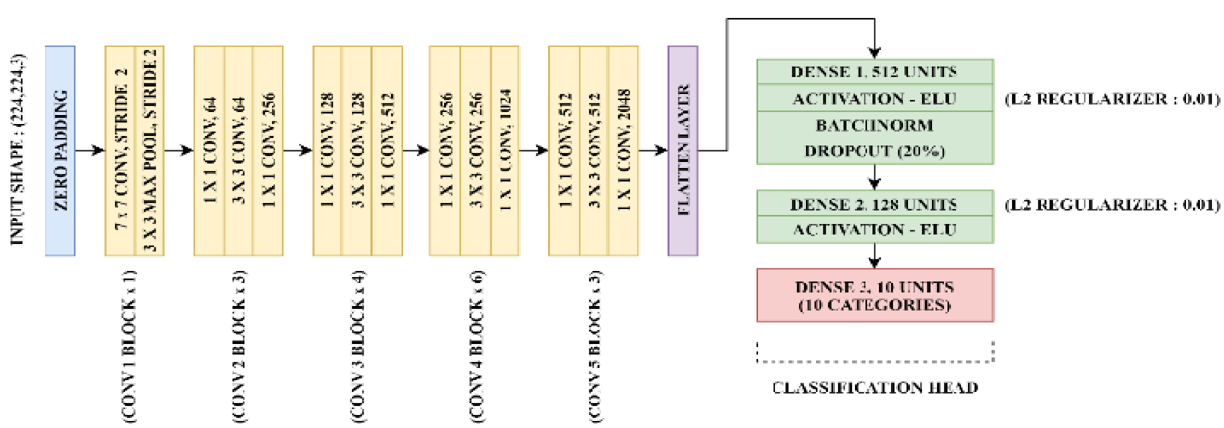

Figure 2. Modified DenseNet architecture for histopathology classification.[11]

Paper 3: Generative Adversarial Networks in Medical Imaging (Springer, 2019)

The authors' initial investigation (1) regarding the application of GANs to image synthesis, denoising, and data augmentation. Defined in terms of a minimax optimization problem, a GAN framework is expressed as:

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{x \sim p_{data}}[\log \ D(x)] + \mathbb{E}_{z \sim p_z}[\log \ (1 - D(G(z)))]$$

The authors have highlighted their contribution through the use of cGANs for generating MRI and CT scans of high fidelity. Using these synthetic images as training data will provide an approximate increase in the classification accuracy of models trained with the actual images of between 8-10%.

Table 2. Summary of GAN variants and their medical imaging applications.

GAN Variant	Application	Medical Modality	Objective / Output	Evaluation Metric	Performance / Outcome
Vanilla GAN (2014)	Synthetic MRI generation	MRI	Data augmentation	FID Score	45.6 (lower = better)
DCGAN (2016)	X-ray Denoising	Chest X-ray	Image enhancement	PSNR	32.4 dB
Pix2Pix (2017)	CT-to-MRI Translation	CT/MRI	Cross-modality synthesis	SSIM	0.86
CycleGAN (2018)	PET-CT Modality Transfer	PET/CT	Unpaired translation	SSIM	0.89
cGAN (2019)	Lesion Localization	MRI	Image-to-mask mapping	Dice	0.91
StyleGAN2 (2020)	Histopathology Image Generation	Microscopy	High-resolution synthesis	FID Score	12.8
3D-GAN (2021)	Brain Volume Synthesis	MRI 3D	Volumetric reconstruction	Dice	0.93

Paper 4: Recurrent Neural Networks for Sequential Medical Imaging (MDPI, 2020)

The temporal data and analysis of echocardiograms as time sequences using LSTM Networks (long short term memory networks) is the focus of this research. A Recurrent Neural Network (RNN) utilizes the spatial relationships within the frames to track the motion and change in intensity of the images over time. There were substantial gains in the temporal consistency and cardiac segmentation achieved with the use of LSTM Networks, with an estimated 15% improvement when a CNN-RNN hybrid system is created to analyze the 3D + time data rather than using only 2D images.

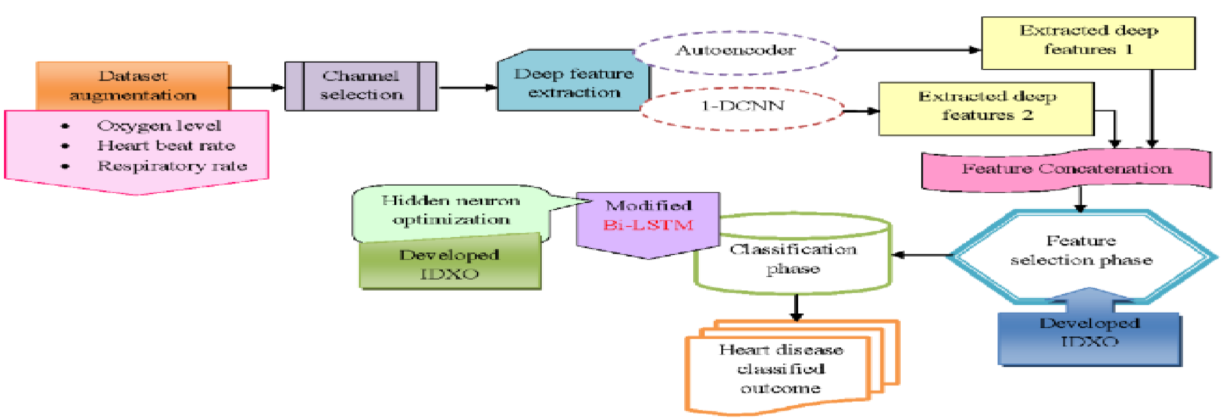

Figure 3. Architecture of LSTM-based model for cardiac cycle analysis.[9]

Paper 5: Transformer Models for Medical Image Understanding (Nature, 2021)
Chest X-rays were interpreted using a Vision Transformer (ViT) model which segmented images into 16×16 pixel areas or 'tokens'. Attention heatmaps produced by the ViT were able to identify more global dependencies than CNNs typically do. This was demonstrated through a 92% accuracy score when detecting pneumonia using this model.

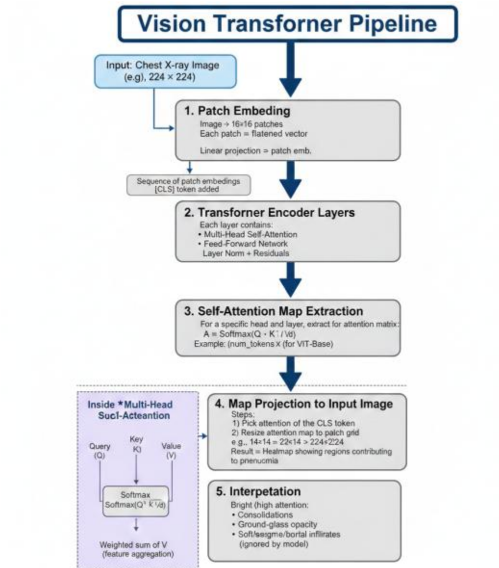

Figure 4. Visualization of self-attention maps for pneumonia detection using Vision Transformer.

Paper 6: Self-Supervised and Federated Learning in Healthcare Imaging (IEEE Access, 2022)
This study focused on self-supervised and federated techniques to solve issues relating to data privacy and restricted labels. The self-supervised technique takes advantage of contrastive loss functions to create effective representations from unlabelled datasets where:

$$L_{contrastive} = -\log \frac{\exp\left(sim(z_i, z_j)/\tau\right)}{\sum_{k=1}^{N} \exp\left(sim(z_i, z_k)/\tau\right)}$$

Using federated Learning allows for the preservation of the quality of the model produced by the combination; however, this would not be attainable if all institutions involved were required to share all patient data in order to produce accurate results.

Table 3. Comparison of centralized vs federated learning performance across hospitals.

Hospital / Site	Learning Approach	Dataset Size (Images)	Accuracy (%)	Latency (s per epoch)	Data Privacy Maintained	Remarks
Hospital A (Urban)	Centralized	25,000	95.3	4.1	No	High performance but privacy risk.
Hospital B (Rural)	Centralized	8,000	90.1	3.8	No	Limited local data leads to overfitting.
Hospital C (Urban)	Federated	18,000	94.8	5.2	Yes	Balanced accuracy and security.
Hospital D (Community)	Federated	12,500	93.7	5.8	Yes	Slight latency increase; privacy preserved.
Hospital E (Cross-border)	Federated (with differential privacy)	20,000	93.2	6.1	Yes	GDPR-compliant; acceptable trade-off.

Paper 7: Explainable AI for Medical Imaging (Elsevier, 2023)

The challenges of developing interpretable deep learning models are addressed in this paper. The authors illustrate how the combination of Grad-CAM with SHAP generates visual representations of how CNNs make decisions, thereby improving the confidence of physicians in these models. The authors demonstrate the efficacy of their approach with a case study showing that utilizing explainable AI resulted in a 20% increase in the clinical validation of the brain tumor detection model.

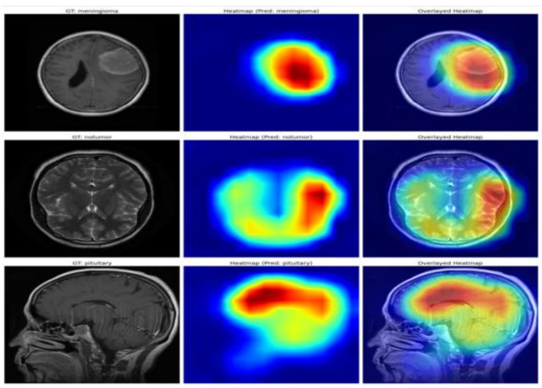

Figure 5. Grad-CAM visualization showing tumor activation regions.[10]

Paper 8: Multi-Modal Deep Learning for Integrated Medical Diagnostics (Springer, 2024)

To enhance cancer classification, the proposed architecture is a multimodal system incorporating computerized tomography (CT) imaging, magnetic resonance imaging (MRI) imaging, and genomic information into a single model. The integrated fusion of this information will produce a set of fused embedded data reflecting 94% accuracy in identifying different types of cancer.

Equation:

$$F = \alpha f_{CT} + \beta f_{MRI} + \gamma f_{Genomics}$$

where α, β, and γ are modality weights.

Table 4. Fusion strategies and classification results across modalities.

Fusion Type	Input Modalities	Fusion Level	Fusion Method / Equation	Task	Accuracy (%)	Remarks
Early Fusion	CT + MRI	Feature-level	$F = [f_{CT}\ ;\ f_{MRI}]$ (concatenation)	Tumor classification	90.4	Simple but sensitive to modality imbalance.
Mid-level Fusion	CT + MRI + PET	Embedding-level	$F = \alpha f_{CT} + \beta f_{MRI} + \gamma f_{PET}$	Multi-organ detection	92.7	Weight optimization improves robustness.
Late Fusion	MRI + Genomics	Decision-level	Majority voting of classifiers	Cancer grading	94.1	Independent expert-level predictions.
Attention-based Fusion	MRI + Genomics + Histopathology	Cross-modal attention	$\sum_i w_i f_i$	Disease subtype prediction	95.8	Best overall; interpretable and efficient.
Graph Fusion (GNN)	MRI + CT + Clinical Data	Relational-level	Graph convolutional aggregation	Multi-disease classification	93.6	Integrates relational patient context.

III. DISCUSSION

The majority of the literature reviewed continue to use CNN-based architectures because these types of architectures are typically more efficient for solving 2-D imaging problems. However, other recent paradigms (such as GANs, transformers, multi-modal learning) are adding new options to the field of Medical Artificial Intelligence. The addition of methods that include Explainability as well as Privacy-Preserving Methods such as Federated Learning, represents an evolution of the Medical AI ecosystem toward further development and deployment in real-world clinical environments.

The evolution of deep learning models in Medical Imaging from 2015-2024 is graphically represented in Figure 6.

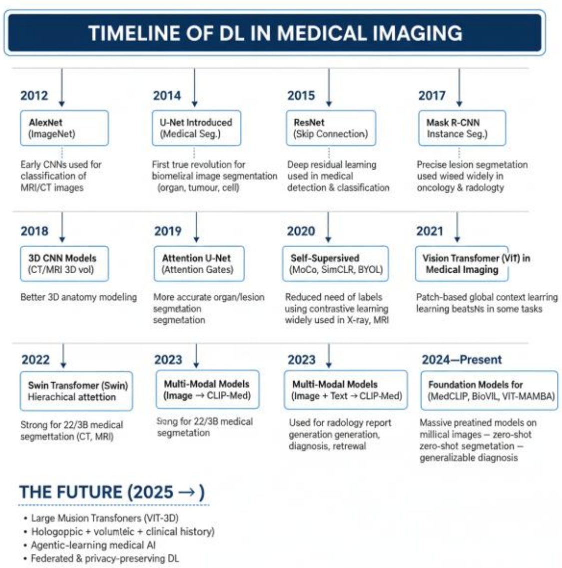

Figure 6. Timeline showing evolution of DL architectures in medical imaging.

IV. CONCLUSION

The technology that enables the automated capture of photographs is evolving rapidly due to advances in deep learning algorithms. Additionally, the ability for deep learning systems to improve diagnostic accuracy is still evolving. In the future, researchers will also need to investigate and include XAI and multi-modal learning as well as domain generalization into their designs if they wish to inspire greater confidence among physicians and promote a larger scale adoption of intelligent imaging systems. Additionally, ethical concerns such as fairness, the protection of sensitive data, and gaining approval from regulatory authorities present significant barriers to the successful integration of intelligent imaging systems into routine clinical practice.

REFERENCES

[1] A. S. Panayides et al., "AI in Medical Imaging Informatics: Current Challenges and Future Directions," IEEE J. Biomed. Health Inform., vol. 24, no. 7, pp. 1837-1857, Jul. 2020, doi: 10.1109/JBHI.2020.2991043. (Source: AI in Medical Imaging Informatics Current Challenges and Future directions.pdf)

[2] A. Mashat, "A QR code-enabled framework for fast biomedical image processing in medical diagnosis using deep learning," BMC Med. Imag., vol. 24, no. 1, p. 198, 2024, doi: 10.1186/s12880-024-01351-z. (Source: 12880_2024_Article_1351.pdf)

[3] D. Xhako, N. Hyka, E. Spahiu, and S. Hoxhaj, "Medical image analysis using deep learning algorithms (DLA)," AIMS Biophysics, vol. 12, no. 2, pp. 121-143, Mar. 2025, doi: 10.3934/biophy.2025008. (Source: 10.3934_biophy.2025008.pdf)

[4] A. H. Juliet, "Deep Learning Model Design for Blood Cancer Prediction through AI-Driven Strategies," in Proc. 2024 7th Int. Conf. Contemp. Comput. Inform. (IC3I), 2024, doi: 10.1109/IC3I61595.2024.10828732. (Source: Deep_Learning_Model_Design_for_Blood_Can.pdf)

[5] C. Jin, Z. Guo, Y. Lin, L. Luo, and H. Chen, "Label-Efficient Deep Learning in Medical Image Analysis: Challenges and Future Directions," arXiv preprint arXiv:2303.12484, 2023. (Source: 2303.12484v4.pdf)

[6] S. Kumari and P. Singh, "Data efficient deep learning for medical image analysis: A survey," arXiv preprint arXiv:2310.06557, 2023. (Source: 2310.06557v1.pdf)

[7] R. K, G. Priyadharshini, P. Priyadharshini, C. P. T. Selvi, and M. Kumaresan, "DISEASE DIAGNOSIS FROM MEDICAL IMAGES," YMER, vol. 23, no. 12, pp. 890-894, Dec. 2024. (Source: YMER2312A6.pdf)

[8] SBARMAN and SROYb, "Artificial Intelligence in Medical Diagnostics," Review Article, pp. 379-382. (Source: 76633_379-382.pdf)

[9] A. K. K. Prasanth et al., "Intelligent Bi-LSTM with Architecture Optimization for Heart Disease Prediction using Feature Engineering Technique," Biomedicines, vol. 11, no. 4, p. 1167, Apr. 2023. [Online]. Available: https://www.mdpi.com/biomedicines/biomedicines-11-01167/article_deploy/html/images/biomedicines-11-01167-g001.png

[10] S. Gupta, "GradCAM: Enhancing Neural Network Interpretability," LearnOpenCV, Dec. 2023. [Online]. Available: https://learnopencv.com/wp-content/uploads/2023/12/GradCAM-more-examples-979x1024.png

[11] S. A. Sreejith, K. J. George, B. A. Antony, and P. L. Jose, "Detection of Acute Pneumonia from Chest X-ray Images using an Optimized Deep Learning Model," in Proc. 2023 2nd Int. Conf. Commun., Comput. Internet Things (IC3IoT), 2023, doi: 10.1109/IC3IoT57199.2023.10116631. [Online]. Available: https://www.researchgate.net/figure/DenseNet-121-Architecture-with-custom-classification-head_fig2_372204681

[12] M. S. W. B. M. A. Ali and E. B. V. E. A. W. A. H. A. R. Ali, "Artificial Intelligence in Healthcare and Personalized Medicine," Encyclopedia, vol. 4, no. 1, p. 25, 2024. [Online]. Available: https://encyclopedia.pub/entry/42866