International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

A Review on Medical Image Classification with Multi-Modal Diffusion Network through Clinical Data Fusion and Imaging.

K Saiprasad Goud[1], Morarjee Kolla[2]

[1,2]*Department of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Hyderabad, Telangana, India*

Abstract— **The analysis of medical images is a crucial process in the contemporary healthcare system as it allows the examination of diseases based on such imaging techniques as X-rays, computed tomography (CT), magnetic resonance imaging (MRI), and dermoscopic images. In spite of the fact that deep learning has made a tremendous step forward in the medical image classification, a lot of current methods are based only on the visual data and do not use important clinical data like the history of the patient, the symptoms, and the diagnostic notes. This weakness lowers their usefulness in practice clinical processes where decision-making is multimodal by nature. This analysis will explore the current developments in the field of multi-modal diffusion-based medical image classification with specific attention on methods that utilize medical images and clinical text. We examine two-granularity feature extraction techniques to acquire global anatomical features and local pathological features, clinical text representation based on domain-specific language models, like ClinicalBERT, and cross-attention patterns to achieve successful image-text fusion. In the review, the heterologous diffusion strategies that boost robustness in the presence of noisy and uncertain scenarios are also discussed. Through a synthesis of the findings of current research findings, this article demonstrates the advantages, challenges, and open research recommendations of multi-modal diffusion frameworks, their potential to enhance contextual insight and clinical significance in medical image classification.**

Keywords—**Clinical BERT, Computed Tomography, Magnetic Resonance Imaging, Multi-Modal Diffusion-based Medical Image Classification, X-rays**

I. INTRODUCTION

Medical imaging is the most prominent aspect of the contemporary healthcare system as it allows imaging anatomical structures and pathological conditions in a non-invasive manner using various modalities, including X-ray, CT, MRI, ultrasound, PET, and OCT. These imaging technologies produce vast amounts of multi-dimensional data which are vital in the diagnosis of diseases, planning of treatment and monitoring of the patients. But, the process of medical image interpretation is time-consuming, subjective, and vulnerable to inter-observer variability, especially in those medical conditions with subtle abnormalities or early diseases.

One solution that is being brought to help overcome these shortcomings is artificial intelligence, specifically deep learning. Convolutional neural networks have been shown to be very successful in medical image classification including cancer detection, disease grading and anatomical structure classification. This has been facilitated by more powerful computers, the existence of annotated datasets and constant enhancements of neural network designs. Medical image classification is a field that has been experiencing a major shift between 2022 and 2026 with the development of more advanced models that are not based on CNN.

ViTs have received attention because of their capability to address the long-range spatial features with self-attention architectures, which can help to understand the global context better. Simultaneously, the self-supervised learning has emerged as a central approach to reducing the lack of labeled medical data by learning powerful representations on large-scale unlabeled data. A further improvement in performance has been made by the use of hybrid CNN-Transformer architectures, that use local feature extraction with global context modeling, and multimodal learning models that use multiple imaging modalities or clinical data that have demonstrated promising increases in diagnostic accuracy. Although these developments have been made, a number of problems persist to hamper clinical implementation. These are limited access to annotated datasets, errors of imbalance in classes, institutional domain shift, excessive computational complexity and lack of interpretability of the models.

In addition, the absence of standardized evaluation procedures and multi-center validation make it difficult to compare the suggested methods and make realistic generalization. Since recent developments in medical image classification are evolving rapidly and are still faced with a number of difficulties, it is necessary to review the latest advances in this field. The review is a systematic examination of highly influential publications published within the past two years, 2022-2026, and emphasizes the important trends in architecture, learning paradigms, practice evaluation, and open research problems, and are aimed at informing future research on the development of reliable and clinically implementable medical image classification systems.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

II. Related Work

In recent years, medical image classification has received a wide range of research interest owing to the development of the deep learning models and the growing accessibility of medical imaging data. This part outlines literature that made an impact and that was widely referenced over the period of 2022-2026 and classified thematically based on their major methodological contributions.

A. Standardized Evaluation and Benchmark Data.

Benchmark datasets have an important role in making research in medical image classification repeatable and per comparable. In the study by Yang et al., MedMNIST v2 is a curated set of 18 biomedical image datasets across a variety of modalities, such as X-ray, CT, MRI, OCT, and pathology [1]. The benchmark offers single preprocessing pipelines, uniform train-test partitions, and standardized assessment guidelines of a binary, multi-class, ordinal, and multi-label classification undertaking. These features have enabled MedMNIST v2 to become a popular tool of comparing algorithms and validating a method.

B. Vision Transformers in Medical imaging.

ViTs are a substitute to convolutional architectures, which means that long-range spatial dependencies are modeled instead of relying on self-attention mechanisms. The original ViT framework was proposed by Dosovitskiy et al. [2], and it processes images in the form of sequence of patch embeddings. This design has been modified to medical imaging to replicate global contextual relationships. Manzari et al. [3] introduced an architecture named MedViT, a transformer model that is specialized to classify medical images and has integrated domain-based strategies of patch embedding and regularization that can be effectively used with small amounts of training. In a study of transformer-based techniques in medical imaging, Azad et al. [4] have offered an in-depth survey of the techniques with clear effectiveness in whole-slide histopathology and multi-organ CT analysis although they have high computational requirements. Li et al. [5] also examined the transformer scalability and robustness, highlighting the issues with data efficiency and interpretability. Bao et al. [6] suggested hierarchical transformer structures of gigapixel histopathology images and illustrated that the advantages of multi-scale attention come at the price of more memory.

C. Self-Supervised learning strategies.

SSL has become a recent approach to lessen the reliance of big labeled medical datasets.

Huang et al. [7] conducted a review of the currently used means of SSL in medical imaging and stated that contrastive learning approaches always yield higher results compared to their predictive and generative counterparts in all possible modalities. Li et al. [8] proposed a momentum contrast-based learning approach to learning the medical image representation, which demonstrated a high level of performance in the low-label condition. The article by Azizi et al. [9], investigated large scale pretraining in SSL with large datasets of medical images showing better transferability of features to downstream tasks. Wang et al. [10] compared the most popular methods of representative SSL, such as SimCLR, MoCo, BYOL, and SwAV, and offered a viable understanding of augmentation and optimization tactics. Investigating the power imbalance and architectural differences in the context of the workflow of the system of the HTML-based authentication, Zhang et al. [11] provided the guidelines to deploy it efficiently in the context of the medical image classification.

D. Convolutional Neural Network Architectures.

Although transformers have increased in popularity, CNNs are still popular because they are computationally efficient and include powerful inductive biases. The residual networks by He et al. [12] allow deep model training with shortcut connections and dense connectivity with feature reuse, respectively. Zhou et al. [14] showed poorly supervised CNN models in the classification of chest X-rays through the use of class activation maps. Zhang et al. [15] have included both spatial and channel attention mechanisms in order to enhance disease localization. Kermany et al. [16] emphasized the power of transfer learning on natural image in classification of retinal OCT.

E. CNN-Transformer Architectures Hybrids.

Hybrid CNN-Transformer models are models which combine both local feature extraction and global context modeling. Chen et al. [17] introduced sequential hybrid models of multi organ disease classification that incurred less computational cost than pure transformer models. Ma et al. [18] created cross-attention fusion networks, which explicitly represent interactions between convolutional and transformer representations. Li et al. [19] proposed adaptive dual-branch models that trade-off between local and global characteristics and show stable better outcomes at various medical imaging problems.

F. Fine-Tuning and Transfer Learning.

Transfer learning continues to play a crucial role in the classification of medical images with only a few annotations.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

Shin et al. [20] demonstrated that ImageNet-trained CNNs are superior to models trained on barebones at the low-data regimes. Tajbakhsh et al. [21] also examined the strategies of fine-tuning, which highlights on the need to have both layer-wise adaptation and control of learning rate to achieve optimal results.

G. Data Fusion and Multimodal Learning.

Multimodal methods of learning combine imaging data with clinical data to enhance the context of diagnosis. Liu et al. [22] combined radiology with electronic health records through attention-based mechanisms, and showed better diagnostic accuracy. Patel et al. [23] suggested models of image-text transformer which learn visual and text images together, especially in the classification of rare diseases.

H. Uncertainty Quantification and Predictable Prediction.

Medical AI systems must be deployed with doubts being estimated in order to be reliable. To measure predictive uncertainty, Ghesu et al. [24] used Bayesian deep learning accompanied by Monte Carlo dropout. Kendall and Gal [25] identified two types of uncertainty, aleatoric and epistemic, and demonstrated that well-calibrated estimates of uncertainty can be used to improve the clinical decision support.

I. Summary and Research Gaps.

The literature review has shown that medical image classification has made great advances in terms of architectural innovation, self-supervised learning, and multimodal fusion. Nonetheless, there are still some issues such as a lack of multicentre validation, interpretability, a standardised evaluation protocol, and a discrepancy between benchmark performance and the clinical application in the real world. This is due to these limitations that encourage more studies on strong, understandable and clinically consistent medical image classification systems.

III. MIMIC-CHEST X-RAY(CXR)-DATASET

MIMIC-CXR Dataset is a large-scale, publicly-accessible set of de-identified images of chest X-rays and radiology free-text reports, which was created to facilitate the study of automated analysis of chest radiographs. The dataset will consist of both frontal and side views of the chest X-ray of patients admitted to intensive care unit (ICU) under real-life clinical settings and is subject to variability of patient positioning, X-ray machine, acquisition protocols, and presentation of the disease.

One of the most important features of MIMIC-CXR is the presence of paired radiology reports that allow not only image-based classification but also multimodal learning based on image-text fusion. The data base comprises the broad scope of thoracic outcomes, such as lung diseases, cardiovascular issues, and medical aids. MIMIC-CXR with its scale, diversity, and clinical realism is popular in the classification of chest diseases, multimodal medical AI studies, and benchmark deep learning models and it studies issues related to class imbalance, label noise, and domain variability. Fig 1 illustrates sample of chest X-ray images of the MIMIC-CXR dataset with frontal and lateral views taken in a clinical real-world setting. The images indicate differences in positioning of patients, quality of the image and presence of medical apparatus, which are common issues in automated analysis of chest X-rays.

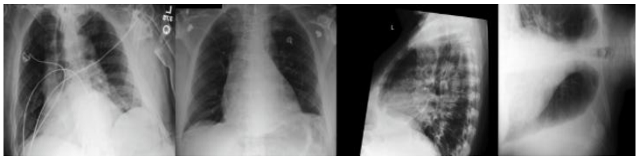

Fig 1:Samples of chest X-ray images of the MIMIC-CXR dataset

IV. TAXONOMY OF DEEP LEARNING MEDICAL IMAGE CLASSIFICATION METHODS

Medical image classification techniques based on deep learning have developed quickly, owing to the development of neural network designs and learning algorithms. The modern methods can be divided into five broad paradigms namely, convolutional neural network based models, transformer based models, hybrid CNN-Transformer structures, self-supervised learning systems, and multimodal fusion networks.

A. Convolutional Neural Network-Based Models

CNNs are the most popular models used in the medical image classification because they are computationally more efficient as well as have the inductive biases towards the extraction of spatial features. CNNs acquire hierarchical feature representations by using stacked convolutional layers using local receptive fields.

1. Residual Networks (ResNet): He et al. introduced residual connections that allow deep networks to be trained by allowing gradient flow. Modifications of ResNet have become widely used as the backbone in most medical imaging tasks especially when trained on large-scale natural image datasets and fine-tuned on medical data.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

2. Dense Networks (DenseNet): DenseNet architectures encourage the re-use of features both by dense connectivity, which minimises redundancy of parameters, and by enhancing gradient propagation. These qualities render DenseNet especially useful in cases when there is a paucity of labelling information.

3. Attention-Guided CNNs: CNNs have attention mechanisms that allow the models to pay attention to areas that are diagnostically relevant. The weakly supervised structures with the help of class activation maps do not require pixel-level annotations, and they can be used to localize disease, but are susceptible to size and noise in a dataset.

4. Transfer Learning: Large natural image datasets have been shown to be useful in transfer learning of medical imaging particularly when only limited data is available. The general visual patterns are recorded in lower network layers, whereas the high-level layers are adjusted to domain-specific features. Limitations: The CNNs mainly depend on local context, and it cannot model long-range interactions and global anatomical interactions. They can also need a lot of labeled data and close architectural fine-tuning to different resolutions of images.

B. The Transformer-Based Architectures

The architecture that is of interest to this study is transformer-based architectures. Transformer models are based on self-attention mechanisms that learn to accept relationships between the global context of the whole image by learning to handle it as a sequence of patch embeddings.

1. Vision Transformer: Vision Transformers break down images into regular sized patches and process them with transformer encoders. It is a design which allows holistic image interpretation and is modified to medical imaging tasks that need to be aware of global context.

2. Medical-Specific Transformers: A number of variants of transformers have been designed with adaptations to a medical domain including multi-scale patch embeddings, domain-rich regularization, and hierarchical processing, especially on images with high resolution histopathology.

3. Comparative Characteristics: Transformers typically exhibit more of the ability to model long-range dependencies than CNNs, but require more training data and more computing power. Limitations: Transformer implementation is constrained by high computational complexities, high number of parameters, and decreased data efficiency in resource-constrained clinical setting.

C. CNN-Transformer Hybrid Structures

Hybrid architecture Hybrid architecture is the combination of CNNs and transformers to exploit the complementary properties of the two paradigms.

*1. Sequential Hybrid Models:*Local feature extraction is done through CNN layers and global context modeling is performed through transformer modules. This method makes the computation much cheaper and maintains information about the surrounding environment.

*2. Parallel Dual-Stream Models:*CNN and transformer branches are independent of each other and take inputs simultaneously, and feature fusion happens at the later stages to integrate between local and global representations.

*3. Cross-Attention Fusion:*The cross-attention mechanisms explicitly learn the interaction between CNN-derived and transformer-derived features, which allows one to selectively combine supplementary information.

*4. Performance Characteristics: :*Architecture Hybrid models tend to be able to have trade-offs between accuracy and efficiency of a balanced nature, although implementation costs are harder, due to architectural complexity and design decisions.

D. Learning Systems that are Self-Supervised

Self-supervised learning (SSL) solves the problem of data scarcity by training feature representations during unlabeled medical images with pretext tasks.

1.Comparison and Contrast Learning Models: Contrastive learners are trained to represent images in an invariant way by maximizing consensus across dissimilar augmented representations of the same image, and enhance downstream low-label-regiming performance.

2. Large-Scale Pretraining: The scale and diversity of datasets improve the benefits of SSL, allowing cross-modal representations in imaging to be used.

*3.Implementation Insights:*Dominance of domain specific augmentations, momentum based optimization, and cautious use of fine-tuning methods is stressed in the domain of empirical studies.

4. Limitations: SSL takes considerable computation to pretrain and performance improvements depend on modality and clinical tasks.

E. Multi Modal Fusion Network

Multimodal methods are imaging methods that combine more than one imaging or integrates imaging data with clinical data to develop diagnostic context.

1.Image-Image Fusion: The complementary imaging modalities can be combined to give richer representations especially when there is fusion of features in the middle stages in the network.

2.Image-Text and Image-Clinical Data Fusion Image-Text: Fusion of image and text Image-Clinical Data: Fusion of image and clinical data Combining radiological images with clinical text or electronic health records allow the context-based decision-making process particularly on complicated diagnostic activities.

F. Comparative Analysis

The CNN-based models are also feasible in limited resource and small size datasets. Transformer architecture is good in modeling global dependencies with high costs in computation. Hybrid structures are able to balance efficiency and contextual modeling, whereas self-supervised learning is successful in alleviating the lack of annotations. Multimodal methods provide a better diagnostic context where extra support data is reliable. Regardless of this development, there are still problems pertinent to interpretability, generalization, standardized assessment, and clinical validation.

Table I shows comparative description of key deep learning paradigms in medical image classification. It outlines the representative models and their main strengths, weaknesses and common clinical uses. This comparison serves to point out the trade-offs in computational efficiency, ability to model the context, and data requirements with a view to providing insight into which approach is more appropriate in particular medical imaging situations.

Table I.
Deep Learning Comparative study on medical image classification.

Category	Representative Models	Key Strengths	Key Limitations	Typical Applications
CNN	ResNet, DenseNet	Efficient, stable, well-understood	Limited global context	Screening, resource-limited settings
Transformer	ViT, MedViT	Global dependency modeling	High computational cost	Complex multi-organ analysis
Hybrid	CNN–Transformer	Balanced accuracy and efficiency	Architectural complexity	General-purpose classification
SSL	MoCo, SimCLR	Reduced annotation dependency	High pretraining cost	Low-label scenarios
Multimodal	Image–Text, Image–EHR	Context-aware diagnosis	Data alignment challenges	Complex clinical decision support

V. CHALLENGES AND LIMITATIONS

Although there are some significant improvements in deep learning-mediated medical image classification, there are still a range of challenges that impede the implementation of this approach on a massive clinical scale. One of the significant drawbacks is the lack of quality annotated medical information since an expert labeling is expensive, time-consuming, and prone to inter-observer differences.

The privacy control also constrains the cross-institutional data sharing thus restricting the diversity of datasets. Moreover, medical data is frequently characterized by extreme class imbalance and demographic bias, which may decrease sensitivity to rare diseases and increase the question of fairness. Scanner, acquisition protocol, and patient population heterogeneity in datasets leads to domain shift that reduces their performance when models are used in other clinical environments.

The other issue is the reliability and deployment of models. Most deep learning applications are not interpretable enough, which lowers their clinician trust and complicates governmental approval. Computational complexity, in specific case of transformer based architecture, limits its use in resource limited environments. The overfitting of benchmark datasets is a common practice, as it underscores the disparity between the research outcomes and the practice in clinical reliability. Also, there is a lack of uniformity in evaluation protocols and multi-center validation which discourages equitable comparison between studies. The actual implementation into clinical practice, as well as ethical, legal, and regulatory concerns, remain a major impediment towards regular clinical implementation.

VI. Conclusion

This review has critically synthesized recent progress in the field of deep learning-based medical image classification, including innovations in the field of 2022-2026. The work has demonstrated the strengths and weaknesses of the current methodologies by examining convolutional neural networks, transformer-based architectures, hybrid models, self-supervised learning methods and multimodal models. In the analysis, it can be stated that there is a definite tendency towards hybrid and multimodal systems which utilize local and global representations and introduces the clinical context.

The self-supervised learning has demonstrated significant potential in resolving the problem of data scarcity, and uncertainty-conscious models can assist in enhancing reliability. Nevertheless, issues concerning heterogeneity of datasets, interpretability, generalization, consistency of evaluation and regulatory compliance still hamper universal clinical use. In general, despite the significant advancement in technical terms of medical image classification, much more work is needed to overcome the gap between research standards and clinical practice. This is the reference that can be used by scholars and practitioners who wish to come up with reliable and efficient, and reliable medical image classification systems that meet clinical requirements.

REFERENCES

[1] Yang, Jiancheng, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister and Bingbing Ni. Medmnist v2 large-scale lightweight benchmark on 2 and 3D biomedical image classification. Scientific Data 10, no. 1 (2023): 41.

[2] Dosovitskiy, Alexey. Image recognition at scale: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020).

[3] Manzari, Omid Nejati, Hamid Ahmadabadi, Hossein Kashiani, Shahriar B. Shokouhi and Ahmad Ayatollahi. MedViT: a generalized medical image classification vision transformer. Computers in biology and medicine 157 (2023): 106791.

[4] Azad, Reza, Amirhossein Kazerouni, Moein Heidari, Ehsan Khodapanah Aghdam, Amirali Molaei, Yiwei Jia, Abin Jose, Rijo Roy as well as Dorit Merhof. M. I. Vision transformers Advances in medical image analysis. Medical Image Analysis 91 (2024): 103000.

[5] Li, Jun, Junyu Chen, Yucheng Tang, Ce Wang, Bennett A. Landman and S. Kevin Zhou. Transforming medical imaging using Transformers? Comparison of the main properties and the present development, as well as the prospects. Medical image analysis 85 (2023): 102762.

[6] Yu, Jinze, Shuo Li, Luxin Tan, Haoyi Zhou, Zhongwu Li and Jianxin Li. HI ViT: Hierarchical attention-based Transformer on multi-scale entire slide histopathological image classification. Expert Systems with Applications 277 (2025): 127164.

[7] Huang, Shih-Cheng, Anuj Pareek, Malte Jensen, Matthew P. Lungren, Serena Yeung, and Akshay S. Chaudhari. Self-supervised learning of medical images: a review and implementation recommendations. NPJ Digital Medicine 6, no. 1 (2023): 74.

[8] Ghesu, Florin C., Bogdan Georgescu, Awais Mansoor, Youngjin Yoo, Dominik Neumann, Pragneshkumar Patel, Reddappagari Suryanarayana Vishwanath et al. "Contrastive self-supervised learning using 100 million medical images with optional supervision. Journal of Medical Imaging 9, no. 6 (2022): no. 064503-064503.

[9] Azizi, Shekoofeh, Basil Mustafa, Fiona Ryan, Zachary Beaver, Jan Freyberg, Jonathan Deaton, Aaron Loh et al. "Big self-supervised models promote medical image classification." Proceedings of the IEEE/CVF international conference on computer vision, pp. 3478-3488. 2021.

[10] Wang, Wei-Chien, Euijoon Ahn, Dagan Feng and Jinman Kim. A survey of predictive and contrastive self-supervised medical image learning. Machine Intelligence Research 20:483-513, vol. 20, no. 4.

[11] Chen, Liang, Paul Bentley, Kensaku Mori, Kazunari Misawa, Michitaka Fujiwara and Daniel Rueckert. Self-supervised learning of medical image analysis by image context restoration. Medical image analysis 58 (2019): 101539.

[12] Kaiming, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep residual learning, image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[13] Ouyang, Xi, Srikrishna Karanam, Ziyan Wu, Terrence Chen, Jiayu Huo, Xiang Sean Zhou, Qian Wang, and Jie-Zhi Cheng. "Learning hierarchical attention to weakly-supervised chest X-ray abnormality localization and diagnosis. IEEE transactions on medical imaging 40, no. 10 (2020): 2698-2710.

[14] Huang, Gao, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. Procedings of the IEEE conference on computer vision and pattern recognition, pp. 4700-4708. 2017.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 02, February 2026)

[15] Guan, Qingji and Yaping Huang, Zhun Zhong, Zhedong Zheng, Liang Zheng and Yi Yang. "Attention guided convolutional neural network Thorax disease classification. Pattern Recognition Letters 131 (2020): 38-45.

[16] Kermany, Daniel S., Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L. Baxter, Alex McKeown et al. "Deep learning to identify medical diagnoses and treatable diseases using images only.' cell 172, no. 5 (2018): 1122-1131.

[17] Wu, Xin, Yue Feng, Hong Xu, Zhuosheng Lin, Tao Chen, Shengke Li, Shihan Qiu, Qichao Liu, Yuangang Ma and Shuangsheng Zhang. CTransCNN: Transformer+CNN in multilabel medical image classification. Knowledge-Based Systems 281 (2023): 111030.

[18] Zheng, Jianwei, Hao Liu, Yuchao Feng, Jinshan Xu and Liang Zhao. Crossover: Cross-attention and cross-scale fusion network medical image segmentation. Computer Methods and Programs in Biomedicine 229 (2023): 107307.

[19] [19]Chowdary, G. Jignesh and Vivek Kanhangad. A dual-branch network to diagnose the thorax diseases based on the X-rays of the chest. IEEE Journal of Biomedical and Health Informatics 26, 12 version, 6081-6092.

[20] Shin, Hoo-Chang, Holger R. Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura and Ronald M. Summers. Deep convolutional neural networks: computer-aided detection. CNN architectures, dataset characteristics and transfer learning Deep convolutional neural networks. IEEE transactions on medical imaging 35, vol. 5, no. 5 (2016): 1285-1298.

[21] Tajbakhsh, Nima, Jae Y. Shin, Suryakanth R. Gurudu, R. Todd Hurst, Christopher B. Kendall, Michael B. Gotway and Jianming Liang. Convolutional neural networks to medical image analysis: Full training or fine tuning? The article is found in IEEE transactions on medical imaging 35, no. 5 (2016): 1299-1312).

[22] Guo, Zhe, Xiang Li, Heng Huang, Ning Guo and Quanzheng Li. Deep learning-based multimodal medical grade image segmentation. IEEE transactions on radiations and plasma medical sciences 3, no. 2 (2019): 162-169.

[23] Nowak, Sebastian, David Biesner, Y. C. Layer, Maike Theis, Helen Schneider, W. Block, Benjamin Wulff, U. I. Attenberger, Rafet Sifa and A. M. Sprinkart. " Structuring of free-text radiology report databases using transformers. European radiology 33, no. 6 (2023): 4228-4236.

[24] Ghesu, Florin C., Bogdan Georgescu, Awais Mansoor, Youngjin Yoo, Eli Gibson, R. S. Vishwanath, Abishek Balachandran et al. "Quantifying and leveraging predictive uncertainty to medical images assessment." Medical Image Analysis 68 (2021): 101855.

[25] Kendall, Alex, and Yarin Gal. What do we require in bayesian deep learning of computer vision? Recent progress in neural information processing systems 30 (2017).