

# *The Application of ML in Medical Image Analysis Research*

Ze Zheng Yang<sup>1,a,\*</sup>

<sup>1</sup>*Xi'an Jiaotong-Liverpool University, School of Advanced Technology, Suzhou, 215000, China*

*a. Zezheng.Yang23@student.xjtlu.edu.cn*

*\*corresponding author*

**Abstract:** The development of Machine Learning (ML) and Deep Learning (DL) has brought revolutionary changes to medical image analysis and provided a new solution for diagnosing and treating all categories of disease. This paper is going to review the applications of ML and DL techniques in the field of medical image analysis such as X-rays, CT scans, MRI, and DEXA, in particular, their role in the diagnosis of orthopedic diseases. With the progress of the AI's processing power and the universality of open-source platforms, ML's algorithms have developed rapidly, it has significantly boosted the classification, segmentation and lesion detection of the medical image. However, there also exist challenges like data quality, model interpretability, and class imbalances. The paper will summarize and analyze the present situation of ML and DL in medical images and emphasize their advantages, limitations, and future development potential. Studying recent literature and experiment reports clarifies that solving these limitations is crucial for the clinical adoption and usefulness of ML and DL. Finally, this paper calls for more research to improve the accuracy, efficiency, and safety of using ML and DL in medical images, enabling these technologies to truly enhance patient outcomes.

**Keywords:** AI, ML, Medical Imaging, Image Classification, Medical Imaging

## **1. Introduction**

Over the past few decades, the combination of advanced imaging technologies and computational methods has led to significant advances in the field of medical image analysis. Medical imaging techniques including computed tomography (CT), magnetic resonance imaging (MRI), and X-ray are quite necessary for identifying a wide range of illnesses, from cancer to orthopedic issues[1]. The ability to visualize internal structures non-invasively has transformed clinical practice, which makes early detection and intervention available [2].

In the last few years, the application of Artificial Intelligence (AI), specifically ML and DL, has obtained outstanding results in medical imaging. Traditionally, analyzing medical images is a task for humans and it needs a lot of expertise, but now, AI technologies are capable of automating image analysis tasks. The introduction of ML algorithms has promoted the classification, and detection of diseases by using large datasets of medical images [3]. For example, a subset of deep learning called convolutional neural networks (CNNs), has shown impressive accuracy in detecting anomalies in medical images, it can almost reach the performance levels equivalent to that of qualified radiologists. [3]

Although ML and DL have demonstrated their promising capabilities in enhancing the accuracy of diagnosis, several difficulties still exist. For instance, noisy or incomplete datasets are questions related to data quality, which can significantly affect the performance of AI models [3]. Moreover, it will be very difficult for the doctors and the patients to understand how the decisions are made due to the "black box" nature of many DL algorithms. And it causes concerns about the in-model interpretability, and also the problems of safety and trust.[4]. Additionally, the imbalance in classes within medical datasets can hinder the training of effective models, leading to biased predictions [5].

This review's purpose is to summarize the past literature of the use of ML and DL in medical image analyses. Especially focusing on recent papers and experiments on the progress, limitations and future development directions of ML and DL. Synthesizing the key findings from existing studies provides insights that can guide future research efforts and facilitate the integration of AI technologies into clinical practice.

## **2. ML and DL in Medical Image Analysis**

### **2.1. Evolution of Medical Imaging Technologies**

The field of medical imaging has been hugely influenced by many of these imaging techniques, each of which offers a unique perspective on understanding the human body. X-rays and CT scans are traditional techniques that have played an important role in the diagnosis of various diseases, including orthopedic conditions[1]. The integration of these computational methods into ML marks a new era in medical imaging [3]. Through the development of ML algorithms, the ability to process and interpret large amounts of image data has been enhanced, thereby improving the accuracy of diagnosis and the development of personalized treatment plans [2].

### **2.2. ML and DL Applications in Medical Imaging**

ML is a subset of AI, it is now used in a wide range of medical imaging, and it is particularly used for image classification and segmentation. These applications benefit from the increased computing power and the availability of large data sets [6].

#### **2.2.1. Image Classification**

ML algorithms, especially DL models like CNNs, have been instrumental in classifying medical images. For instance, CNNs have shown promising results in detecting tumors in MRI scans and differentiating between normal and abnormal tissues [7].

#### **2.2.2. Image Segmentation**

DL models have revolutionized the process of image segmentation, which is crucial for precise diagnosis and treatment planning. These models can delineate organs and lesions with high accuracy, thus aiding in surgeries and radiotherapy planning [8].

#### **2.2.3. Lesion Detection**

The application of AI in detecting lesions has improved the sensitivity and specificity of medical imaging. Early detection of abnormalities can lead to more effective treatment outcomes, highlighting the importance of ML in enhancing diagnostic accuracy [9].

### **2.3. Key Challenges in Medical Image Analysis**

Significant progress has indeed been made, but there are still several challenges in the domain of medical image analysis: The first thing is the data quality issues. Inconsistencies in standards, insufficient datasets, and overall data quality concerns hold the development and training of robust ML models back.[10]. Secondly, model interpretability can still be a big challenge. The complexity of DL models often leads to the lack of transparency, which means it will be difficult to understand how their decisions are made, it is a crucial problem for clinical practice [11]. The last issue is the imbalance of data. Medical imaging datasets often suffer from uneven variety, where certain conditions are underrepresented. This will cause biased models that perform poorly in diagnosing rare conditions.

## **3. ML and DL Models for Medical Imaging**

### **3.1. Traditional ML Models**

In the evolution of medical image analysis, traditional ML models are the foundations. Among them, support vector machines (SVMs), random forests, and decision trees deserve special attention because of their effectiveness in various image processing tasks.

#### **3.1.1.SVM**

SVM's principle is to classify data by identifying the best hyperplane in a high-dimensional space that separates different classes. They are very useful for medical imaging tasks, such as tumor classification and X-ray abnormality detection.[3]. SVMs are suitable for medical applications when the circumstances are complex and diverse due to their ability to handling high-dimensional data and its stability.

#### **3.1.2.Random Forests**

Random forests are a type of machine learning classifier consisting of several decision trees whose output classes are based on different classes generated by each tree. In this learning approach, several decision trees are created during training and their prediction models are derived to perform classification tasks. Random forests can aggregate the results of different decision trees, which means it can enhance the diagnostic accuracy. That is why it has been used in medical imaging, and it can also increase the reliability of the model and reduce variability[12]. These random trees have proved to be very helpful in areas such as diagnosing different tumor types based on imaging parameters and predicting treatment outcomes[13].

#### **3.1.3.Decision Trees**

Decision trees represent judgments and their possible outcomes visually and are therefore obvious and easy to understand. Decision trees are frequently used in medical diagnosis to categorize photos according to different attributes including texture and shape. However, clustering techniques like random forests are frequently used since judgment trees are prone to overfitting issues, particularly in complicated datasets[5].

### 3.2. DL Models

In medical image analysis, DL has become an effective technique, particularly when combined with CNNs. CNNs are very useful for image classification because they have the ability to automatically and adaptively learn hierarchies of spatial features from images.

#### 3.2.1. CNNs

CNNs have revolutionized the field of medical imaging by enabling automated feature extraction and classification. CNNs have been proven to be very powerful in several medical imaging applications, for instance, they can identify lung nodules in chest X-rays and classify skin lesions in dermoscopy images [6]. They are composed of multiple layers, such as convolution layers, cluster layers and fully coupled layers. They work together to identify different patterns in images [8]. Diagnostic efficiency and accuracy have increased dramatically as a result of their capacity to learn from massive datasets [7].

#### 3.2.2. Transfer Learning

This method leverages pre-trained models on large datasets, allowing for the adaptation of these models to specific medical imaging tasks with relatively smaller datasets. In medical imaging, where it is difficult to obtain annotated data, transfer learning is useful. By trimming models including VGGNet or ResNet, researchers can achieve excellent performance in tasks such as tumour detection and organ segmentation without the need for large numbers of training data. [14].

### 3.3. Evaluation Metrics and Performance

The results of ML and DL models in medical imaging are evaluated using various types of standards that include recall, accuracy and precision.

#### 3.3.1. Performance Evaluation Criteria

Examples of metrics commonly used are accuracy, recall, precision, recall, F1 score, and area under the curve (AUC). These metrics show the system's ability to recognize irregularities and classify images[15]. For example, accuracy evaluates the model's ability to detect all relevant situations, while precision calculates the ratio of actual positive detections to all positive predictions[16].

#### 3.3.2. Current Performance Benchmarks

Recent studies have brought attention to the differences in performance between DL models and standard ML models. CNNs, for example, have steadily outperformed traditional methods in applications showing higher accuracy and lower false-positive rates [9]. Because the choice of model is highly dependent on the details and processing requirements of the dataset, the context of the application must be considered in detail [17].

## 4. Performance and Outcome Metrics

### 4.1. Metrics to Measure Performance

Different models have different metrics and parameters that offer accuracy and relevance to the model for performance and validation purposes. This further enables the comparison of diverse models and provides clinical validation.

#### 4.1.1. Accuracy

This is one of the most basic metrics. It defines the ratio of the number of true results (both true positives and true negatives) to the number of cases studied. However, it can be misleading when there are a number of stratified imbalances such that one category significantly outnumbers the other [15].

#### 4.1.2. Accuracy

Accuracy is the ratio of the number of times the model predicted positive to the number of times this prediction was correct. It is the ratio of true positives to all positives [16]. As such, high accuracy means less FP for the model.

#### 4.1.3. Recall (Sensitivity)

Recall numbers of relevant cases out of the total cases the model found to be relevant. It is defined as the true positive results divided by true positives plus false negatives [18]. A model with high recall recovers many relevant cases and few false negatives.

#### 4.1.4. F1-Score

The F1-Score is the harmonic average of precision and recall, it gives a better balance between precision and recall. This is very effective in cases that an uneven class distribution [19].

#### 4.1.5. AUC

AUC represents the performance of a model in distinguishing between classes. The ROC curve is the generative of the threshold [20], in which the true positive rate is measured across various thresholds.

### 4.2. Existing Performance Benchmarks

Comparative studies have been used to set performance benchmarks for medical image analysis models. These benchmarks are necessary to benchmark against the state-of-the-art and to direct future research. Recent comparison studies have reported results showing that, especially CNNs, DL models outperform traditional ML models in image classification and segmentation [7]. A work demonstrating that CNNs were more accurate and sensitive than traditional ML algorithms in classifying lung nodules [9, 12] provided an example. Specifically, DL models have produced accurate results comparable to human - level diagnosis and, in some cases, even superior in mammogram classification[6]. Moreover, in targeted medical domains such as mammogram classification, DL models have been empirically demonstrated to have diagnostic performance on par or substantially better than healthcare professionals[6]. In addition, they have accomplishing to make consensus inter-observer variability a central key for standardizing diagnostic and guaranteeing medical evaluations.

## 5. Conclusion

The use of ML and AI in medical image analysis has a great impact on the diagnosis as well as the interpretation of images', and brings forth tremendous change in these processes. This paper focuses on how these technologies, such as MRI, CT, and X-ray, have positively changed the treatment of orthopedic and other medical issues.

The tasks of image classification, segmentation, and lesion identification have shown advances with the integration of ML and DL approaches in medical imaging. Numerous researches have shown

that CNN-based DL models are more sensitive and accurate in diagnosing conditions than standard ML models. These developments may enable radiologists to focus on more complex problems and provide patients with the tools to recognize impending problems earlier.

Although advancements in ML and AI have been made, not all challenges are overcome. For instance, poor labeling and imbalance between classes are problems related to insufficient data that negatively impact how well a model can work. In addition, a lot of these deep learning models are viewed as black boxes, and clinicians find it hard to provide a high trust to AI models due to a lack of transparency in the decision output. These obstacles must be solved if AI technologies are wanted to be used in clinical settings. Achieving the pre-set targets would be in the best interests of all stakeholders. However, special emphasis should be placed on collaboration in the area of regulatory measures for AI governance, addressing bias resulting from missing data, and protecting data through federated learning. Solving these problems on a global scale will lead to revolutionary changes in the accuracy of diagnosis and resulting treatment of patients.

Lack of empirical research and insufficient references pose significant limitations to this study. It would be more focused on enhanced model explainability, stronger performance, lower bias, and higher quality data for AI models. In addition, further increases in prediction power can be obtained from AI models by integrating a variety of data such as images and electronic medical records. Multi-disciplinary collaborative approaches between researchers, clinicians, and regulatory agencies are important in determining the optimal approaches, protocols, and limits for AI use in medical imaging.

## References

- [1] S. Hussain et al., "Modern Diagnostic Imaging Technique Applications and Risk Factors in the Medical Field: A Review," *Biomed Res. Int.*, vol. 2022, 2022, doi: 10.1155/2022/5164970.
- [2] S. Vahdati et al., "A Guideline for Open-Source Tools to Make Medical Imaging Data Ready for AI Applications: A Society of Imaging Informatics in Medicine (SIIM) Survey," *J. Imaging Informatics Med.*, vol. 37, no. 5, pp. 2015–2024, 2024, doi: 10.1007/s10278-024-01083-0.
- [3] M. Avanzo, J. Stancanello, G. Pirrone, A. Drigo, and A. Retico, "The Evolution of AI in Medical Imaging : From Computer Science to Machine and DL," pp. 1–23, 2024.
- [4] E. Neri et al., "Explainable AI in radiology: a white paper of the Italian Society of Medical and Interventional Radiology," *Radiol. Medica*, vol. 128, no. 6, pp. 755–764, 2023, doi: 10.1007/s11547-023-01634-5.
- [5] S. K. Zhou et al., "A Review of DL in Medical Imaging: Imaging Traits, Technology Trends, Case Studies with Progress Highlights, and Future Promises," *Proc. IEEE*, vol. 109, no. 5, pp. 820–838, 2021, doi: 10.1109/JPROC.2021.3054390.
- [6] A. Esteve et al., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017, doi: 10.1038/nature21056.
- [7] E. Bejnordi et al., "A survey on DL in medical image analysis Doi link to publisher : <https://doi.org/10.1016/j.media.2017.07.005> Version of the following full text : Publisher ' s version Published under the terms of article 25fa of the Dutch copyright act . Pleas," no. 2017, pp. 60–88, 2024.
- [8] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, pp. 16591–16603, 2021, doi: 10.1109/ACCESS.2021.3053408.
- [9] N. Antropova, B. Q. Huynh, and M. L. Giger, "A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets," *Med. Phys.*, vol. 44, no. 10, pp. 5162–5171, 2017, doi: 10.1002/mp.12453.
- [10] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. June 2002, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [11] Rudin, C. Stop explaining black box ML models for high stakes decisions and use interpretable models instead. *Nat Mach Intell* 1, 206–215, 2019, doi:10.1038/s42256-019-0048-x
- [12] Z. Jin, J. Shang, Q. Zhu, C. Ling, W. Xie, and B. Qiang, "RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12343 LNCS, pp. 503–515, 2020, doi: 10.1007/978-3-030-62008-0\_35.
- [13] M. L. Giger, H. P. Chan, and J. Boone, "Anniversary paper: History and status of CAD and quantitative image analysis: The role of Medical Physics and AAPM," *Med. Phys.*, vol. 35, no. 12, pp. 5799–5820, 2008, doi: 10.1118/1.3013555.

- [14] P. Kora et al., "Transfer learning techniques for medical image analysis: A review," *Biocybern. Biomed. Eng.*, vol. 42, no. 1, pp. 79–107, 2022, doi: 10.1016/j.bbe.2021.11.004.
- [15] Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Informa Process Manag.* 45,427–37,2009,doi: 10.1016/j.ipm.2009.03.002
- [16] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," *ACM Int. Conf. Proceeding Ser.*, vol. 148, no. June 2006, pp. 233–240, 2006, doi: 10.1145/1143844.1143874.
- [17] Y. Lecun, Y. Bengio, and G. Hinton, "DL," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015, doi: 10.1038/nature14539.
- [18] Zweig, M.H.; Campbell, G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine, *Clinical Chemistry, Volume 39, Issue 4, 1 April 1993, Pages 561–577*, doi:10.1093/clinchem/39.4.561
- [19] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," pp. 37–63, 2020, [Online]. Available: <http://arxiv.org/abs/2010.16061>
- [20] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, 2006, doi: 10.1016/j.patrec.2005.10.010.