Library Progress International
Vol.44 No. 3, July-December 2024:
P. 25941-25957

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Deep Learning Models for Medical Image Classification: A Comprehensive Review

Janet Lobo[1], Dr N Anandakrishnan[2]

1. Research Scholar,
PG & Research Department of Computer Science,
Providence College for Women Coonoor (Autonomous)
sr.janetlobo@providencecnr.org

2. Associate Professoor, Department of Computer Science,
Nilgiri College of Arts and Science,
Konnachal Post, Thaloor, The Nilgiris 643239
anandpjn@gmail.com

How to cite this article: Janet Lobo, Dr N Anandakrishnan (2024) Deep Learning Models for Medical Image Classification: A Comprehensive Review. *Library Progress International,* 44 (3), 25941-25957

Abstract
The field of medical informatics involves the study of integrating imaging and biomedical record data. Medical image data consists of pixels representing different parts of a physical object. Analyzing this data requires expertise in value analysis and disease diagnosis. Medical image classification is crucial for Computer-Aided Diagnosis (CAD) and improving healthcare services. It involves analyzing pixel data to categorize medical images and identify affected areas. This is challenging due to the high dimensionality and complex structures of medical images. Experts are needed to interpret image features and verify classification results. The key goal is to maximize categorization accuracy for precise disease diagnosis. In recent decades, Artificial Intelligence (AI) models, including Machine Learning (ML) and Deep Learning (DL) algorithms, have been developed for medical image classification. Traditional ML relies on hand-engineered features, while DL models automatically extract discriminative features at multiple levels of abstraction. However, challenges such as limited training data, class imbalance, and inter-class similarities hinder the learning of salient visual characteristics. This study provides a thorough examination of the most advanced DL models used for classifying medical images. Furthermore, it evaluates the advantages and constraints of various models on a range of medical image datasets. The review provides valuable insights on how to optimize medical picture categorization and offers guidance for future improvements in Computer-Aided Diagnosis (CAD) to improve precision healthcare.

Keywords—Medical image classification, Computer-aided diagnosis, Artificial intelligence, Machine learning, Deep learning

I. INTRODUCTION

Medical Informatics combines Information Technology (IT) and healthcare, integrating biomedical records and imaging data. Biomedical records contain patient medical test results while imaging data is formed by pixels representing physical objects [1]. Medical image analysis mainly aims to pinpoint the exact areas of the body impacted by disorders. Improving clinical treatment relies heavily on developing automatic diagnosis algorithms using image data [2]. Because medical picture categorization is still a difficulty, this study gives a comprehensive overview of the methods currently used to solve this problem.

1.1 Basics of Medical Imaging

Undetectable waves, such as electromagnetic radiation, sound waves, and magnetic fields, are essential in medical imaging [3]. Comprehending medical imaging studies requires a thorough understanding of the many sorts of waves. Typically, these waves originate from a source located on one side of the body, traverse through the body (including the specific area of interest), and ultimately reach a detector positioned on the opposite side. The waves are absorbed by the body's tissues to varied extents, and the detector generates an image that displays the

"shadows" of distinct biological tissues. Previous medical imaging techniques, like radiography, utilized photodetector plates that necessitated film processing prior to image visualization [4]. Nevertheless, contemporary medical imaging technology enables immediate image acquisition and presentation on digital monitors [5].

1.2 Applications of Medical Imaging

Medical imaging is predominantly utilized for diagnostic purposes; however, it also has significant applications in other areas.:

- Medical imaging is a rapid method to diagnose the cause of a patient's disease, including bone fractures, cysts, tumors, and anomalies [6].
- Monitoring disease progression: Imaging techniques like contrast-enhanced CT or MRI can help track the stage and progression of diseases, such as cancer or Parkinson's disease [7].
- Treatment planning: Medical imaging facilitates surgical planning by offering precise data regarding the dimensions and positioning of lesions [8].
- Evaluating treatment effectiveness: Imaging is used to assess the success of treatments, such as monitoring tumor size in cancer patients or ensuring proper alignment of bones and implants during surgery [9].
- Age determination: Ultrasonography and radiographs, can be used to determine fetal age, maternal gestational age, and patient age for legal purposes [10].

1.3 Medical Imaging Modalities

Methods including mammography, Computed Tomography (CT), and Magnetic Resonance Imaging (MRI) provide medical pictures from biomedical instruments. Nuclear medicine, optical techniques, ionizing radiation, and magnetic resonance imaging are all part of this category of imaging modalities. The structure and organ tissue of the human body react differently to each modality. Among the most popular forms of medical imaging are:

- **X-rays:** This is one of the most ancient and extensively employed methods for medical imaging. X-rays employ ionizing radiation to provide images of skeletal structures and certain types of pliable tissues [11]. They are often used to look at bones for fractures, dental issues, chest images, etc.
- **CT:** CT use computer technology and X-rays to generate detailed cross-sectional images, sometimes known as "slices," of the human body.. It provides more detail than regular X-rays and can show both bone and soft tissues [12]. CT is used to diagnose tumors, internal organs, blood clots, fractures, etc.
- **MRI:** MRI utilizes powerful magnetic fields as well as radio waves to produce intricate three-dimensional anatomy pictures without the need of ionizing radiation [13]. It is good for soft tissue visualization and imaging of organs, muscles, ligaments, cartilage, etc. MRI can assess spinal cord injuries, brain abnormalities, torn ligaments, etc.
- **Ultrasound:** Obtains images of the inside of the body by use of high-frequency sound waves. Breasts, abdomen, hearts, and pregnancies are common areas of examination. No ionizing radiation, portable, and non-invasive [14].
- **PET (Positron Emission Tomography):** The purpose of PET is to reveal biochemical and metabolic processes by means of radioactive tracers. By analyzing changes in tissue function, it can identify illnesses at an early stage [15]. It often combined with CT and used in cancer, heart disease, brain disorders, etc.
- **Mammography:** It is a medical imaging technique that uses X-rays to provide detailed pictures of the breast's internal anatomy. It is an effective tool for detecting breast cancer in its early stages when the tumor is too small to be felt or detected by other methods [16].

1.4 Fundamentals of Medical Image Classification

As depicted in Figure 6, there are multiple stages to medical picture classification [17]. Each stage is described in detail in this section.

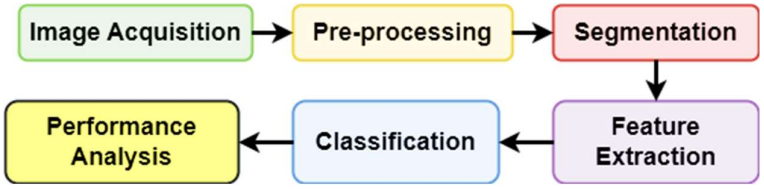

Figure 1. Basic Steps in Medical Image Classification

1.4.1 Image Acquisition

The initial pivotal stage in medical image categorization involves obtaining raw imaging data, which encompasses primary information regarding the interior components of the body. Digital Radiography (DR) and CT detect different physical quantities, whereas PET detects different physical quantities, MRI detects different physical quantities, and ultrasonography detects different physical quantities based on acoustic echoes [18]. Acquiring data involves detecting and converting a physical amount into an electrical signal, preconditioning the signal, and digitizing, regardless of the modality. This process is represented in a general block diagram as shown in Figure 2, which is applicable to most medical imaging modalities.

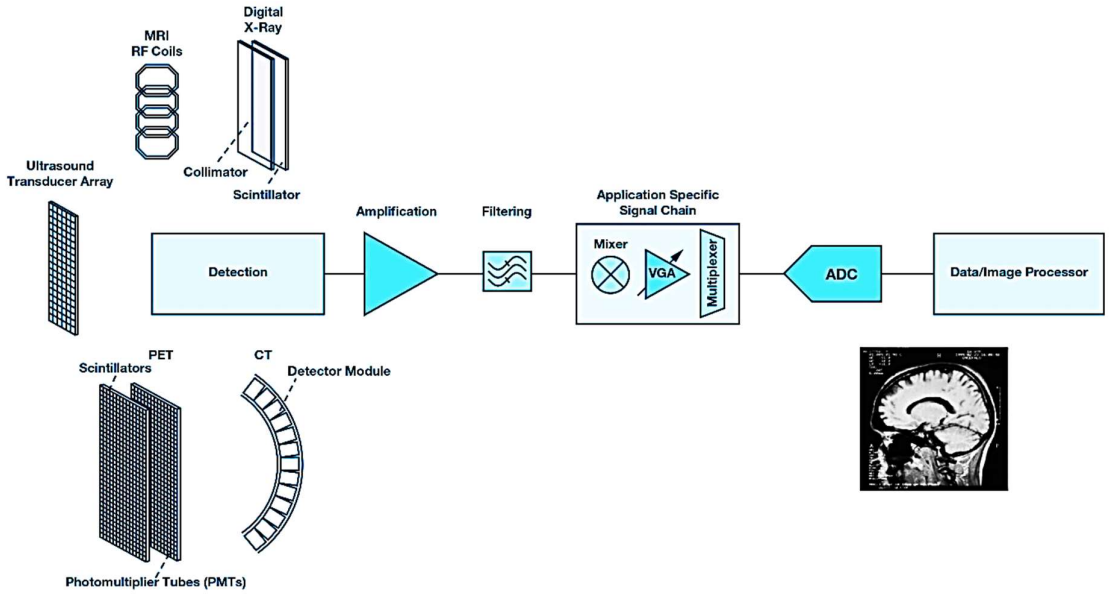

Figure 2. General Block Diagram of Medical Image Acquisition Practice

1.4.2 Pre-Processing

In order to get medical images ready for analysis, particularly for classification tasks, pre-processing is essential [19]. Reducing image acquisition artifacts and ensuring picture consistency throughout a dataset are among the main objectives of medical image pre-processing. This involves techniques to enhance image quality, reduce noise, standardize features, and extract relevant information. Common pre-processing steps include:

- **Background Removal:** Image segmentation is the process of isolating the main subject of interest from the surroundings in an image to enhance the effectiveness and precision of classification. An instance of this is skull stripping (see Figure 3), a method that removes the head and its environs from MRI brain scans [20]. A common step in this procedure is to use morphological procedures to create a mask of the Region of Interest(ROI).

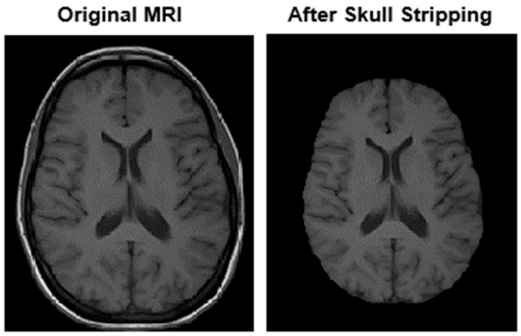

Figure 3. Example of Background Removal

- **Denoising:** Noise can have an impact on medical imaging modalities, leading to unpredictable variations in image intensity. In order to diminish noise, pictures might undergo filtering in both the spatial and frequency domains. Speckle is a type of multiplicative noise that can affect coherent imaging modalities such as ultrasound images. It occurs when the transmitted waveform and its echoes interfere with one another. Figure 3 shows how the speckle filter (specklefilt) function reduces speckle in a picture using a Speckle-Reducing Anisotropic Diffusion (SRAD) method [21].

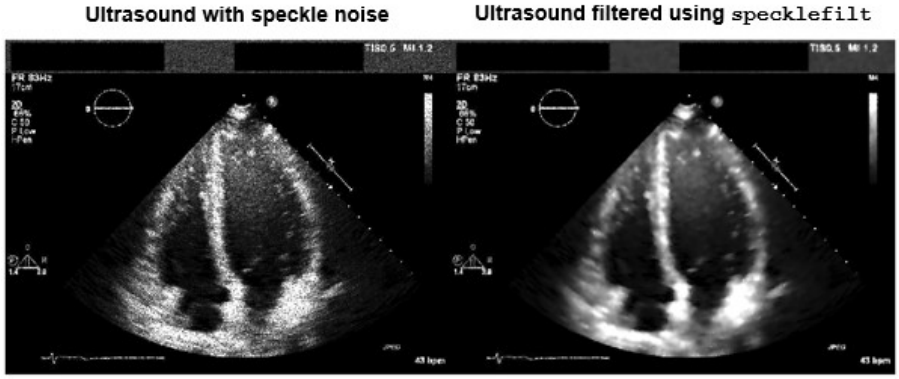

Figure 3. Example of Denoising

- **Resampling:** This technique is used to change the size of specific image pixels or voxels without changing their location in the patient's coordinate system (Refer to Figure 4). This aids in maintaining uniform image resolution throughout a collection that comprises images obtained from various scanners [22].

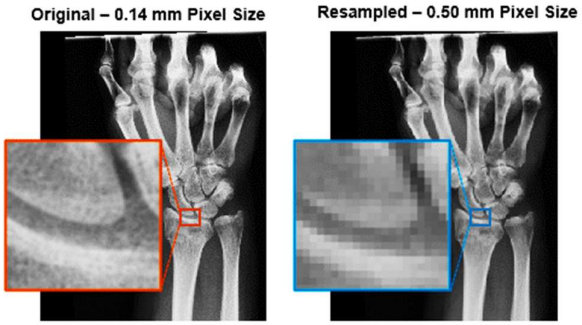

Figure 4. Example of Resampling

- **Registration:** Aligning two- or three-dimensional medical images inside a dataset is known as image registration [23]. Its goal is to achieve consistent spatial alignment in images captured from many patients, or from the same patient obtained at different periods using multiple scanners or other imaging modalities (Refer to Figure 5).

Figure 5. Example of Image Registration

- **Intensity Normalization:** The procedure uniformness the intensity value range for all images in a dataset [24]. There are usually two processes to accomplish this: first, reducing the intensity range to a more manageable size; and second, adjusting the intensity range to match the data type of the image, whether it's digital or binary. One way to make the images more consistent is to use the image's minimum and maximum values to rescale the intensity values. Another tactic is to make sure that all of the photographs use the same intensity range of values. The use of intensity windowing in CT scans achieves this by limiting the intensity values (in Hounsfield Units, or HU) to a range suitable for the tissue being studied. Using the HU range of -1400 to 2400 for bone segmentation and the -1200 to 600 range for lung segmentation from a chest CT are the two main approaches. Figure 6 shows a chest CT slice recorded with and without intensity windowing, as well as with and without a "lung" and "bone" intensity window, respectively.

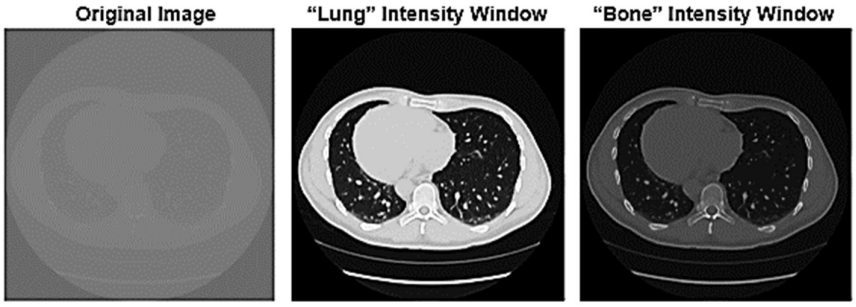

Figure 6. Example of Image Intensity Normalization

1.4.3 Segmentation

Medical image segmentation involves extracting ROIs from 3D medical image data, such as MRI or CT scans, to identify specific areas of the anatomy for various purposes [25]. Recent advancements in AI software have made this task more efficient, allowing for precise analysis of anatomical data by isolating necessary areas, removing unwanted details, and generating segmented masks for further analysis by clinicians [26]. However, segmenting medical images automatically is challenging due to their complex nature and various factors affecting the output of segmentation algorithms. Various techniques are used, tailored to specific imaging modalities, anatomical structures, and clinical goals. Some commonly used techniques include:

- **Thresholding:** It is a simple technique that classifies pixels or voxels in an image as foreground or background based on a fixed intensity threshold [27]. It is particularly useful for segmenting images with clear intensity differences between ROIs and background, such as in binary or grayscale images (Refer to Figure 7). However, selecting the right threshold values can be challenging and can impact performance when artifacts are present.

Figure 7. Example of Thresholding-Based Medical Image Segmentation

- **Region Growing:** It is a segmentation method that starts from seed points or regions and adds neighboring pixels with similar characteristics to the segmented region (See Figure 8). It is effective for segmenting homogeneous structures in medical images but can be sensitive to noise and intensity variations [28]. It includes region merging, region splitting, and split and merge methods. It has limitations such as under-segmentation and over-segmentation.

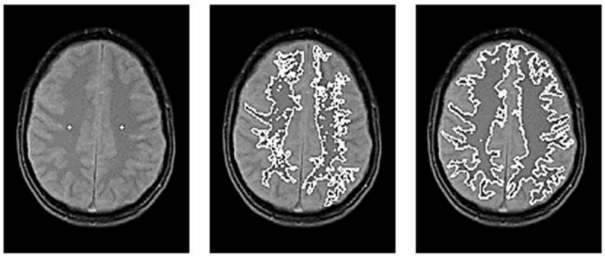

Figure 8. Example of Region Growing-Based Medical Image Segmentation

- **Edge-Based:** Edge detection is the process of identifying boundaries between different structures in an image based on changes in intensity or gradient (Refer to Figure 9). Techniques like the Canny edge detector can be combined with region-based approaches for better segmentation accuracy [29]. However, edge detection methods may be affected by noise and can produce fake or weak edges, which can impact segmentation results. Thus, it is often necessary to use edge detection in combination with region-based techniques for complete segmentation.

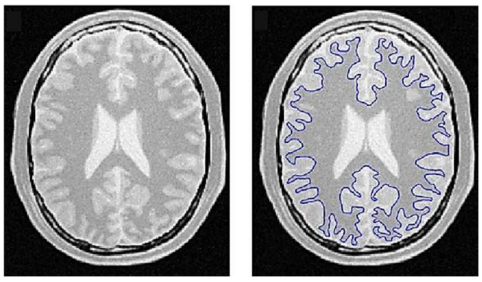

Figure 9. Example of Edge-Based Medical Image Segmentation

- **Active-Contour (Snakes):** Figure 10 depicts active contours, which are flexible systems that reduce an energy function using picture characteristics and limitations on contour smoothness. They are effective for segmenting structures with irregular boundaries, such as tumors or blood vessels, as they can adapt to complex object shapes [30].

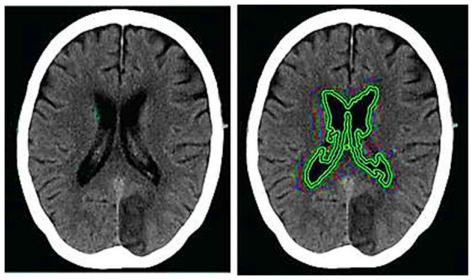

Figure 10. Example of Active Contour-Based Medical Image Segmentation

- **Atlas-Based:** It involves registering a pre-segmented atlas or template image to a target image and transferring the corresponding segmentations (See Figure 11). It uses anatomical priors and spatial information to achieve accurate segmentation, especially in cases with significant inter-subject variability [31]. However, it may struggle with complex structures that have variable shapes, sizes, and properties, and building the database requires expert knowledge.

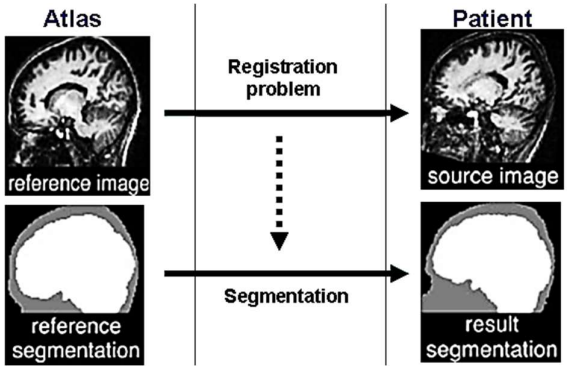

Figure 11. Example of Atlas-Based Segmentation

1.4.4 Feature Extraction

It involves identifying and quantifying important information from images for classification. Common techniques and features used in medical image classification include:

- **Intensity-Based Features:**
  1. Histogram-based features: Statistical measures derived from the intensity histogram of the image, including mean intensity, standard deviation, skewness, and kurtosis [32].
  2. Gradient-based features: Measures of local intensity gradients or edges, including gradient magnitude, gradient orientation, and edge density [33].
- **Shape-Based Features:**
  1. Geometric features: Quantitative shape characteristics such as area, perimeter, compactness, eccentricity, and circularity [34].
  2. Region-based features: Measures based on the spatial distribution of pixels within segmented regions, including centroid coordinates, moments, and Euler numbers [35].
  3. Skeletonization features: Features extracted from the skeleton or medial axis representation of segmented objects, including branch points, endpoints, and branch lengths [36].
- **Texture-Based Features:**
  1. Texture features: Statistical measures describing spatial variations in intensity, such as Gray-Level Co-Occurrence Matrix (GLCM) features, Gray-Level Run Length Matrix (GLRLM) features, and Gray-Level Size Zone Matrix (GLSZM) features [37].
  2. Filter-based texture analysis: Computing texture features using filter banks or convolutional kernels, such as Gabor filters, wavelet transforms, or Local Binary Patterns (LBP) [38].
- **Frequency Domain Features:**
  1. Fourier transform features: Extracting frequency-domain information from images using Fourier or Discrete Cosine Transform (DCT), including frequency components, power spectra, and spectral entropy [39].

2. Wavelet transform features: Decomposing images into different frequency bands using wavelet transforms and computing features from wavelet coefficients, such as energy, entropy, and skewness [40].

- **Local Descriptors [41]:**
  1. SIFT (Scale-Invariant Feature Transform) is used for detecting and describing local features that are invariant to scale, rotation, and illumination changes, making it useful for matching and registration tasks.
  2. SURF (Speeded-Up Robust Features) is similar to SIFT but computationally faster, making it suitable for real-time applications in medical imaging.
  3. ORB (Oriented FAST and Rotated BRIEF) is a combination of FAST keypoint detection and BRIEF descriptor, offering robustness and efficiency for feature extraction.
- **Deep Learning-Based Features:**
  1. Feature Extraction from Pre-trained CNNs: Feature extraction from VGG, ResNet, or Inception, or other pre-trained CNNs, through the use of global pooling or intermediary layers [42].
  2. Autoencoder-based Features: Generating compact representations of image data using autoencoders or Variational Auto-Encoders (VAEs) and utilizing the learned features for downstream tasks [43].

1.4.5 Classification

The process of automatically classifying medical photographs into predetermined groups according to their visual content is known as medical image classification. This is important for disease diagnosis, prognosis, treatment planning, and patient management. Traditional classification methods include [44]:

- Support Vector Machines (SVM): This supervised ML model excels in two-group classification issues and regression analysis in general. To categorize input data, SVMs use decision functions; they are nonparametric classifiers.
- K-Nearest Neighbors (K-NN): Data regression and categorization using the number of k-neighbors are two of its primary uses.
- Naive Bayes: It applies Bayes' hypothesis with strong independent assumptions and is used for medical image classification.
- Decision Trees: It is a decision support tool that works with discrete-valued parameters and aims to create a small decision tree. DT Ensembles, such as Bagging (Random Forest (RF)) and Boosting (Gradient Boosting DT (GBDT) and Extreme Gradient Boosting (XGBoost)), are also used for ensemble classification.

DL is a type of ML that allows computers to understand the world through a hierarchy of ideas. This allows for complex ideas to be learned by building them from simpler ones. Neural Networks (NNs) based on the human brain and neurons are used as classifiers in classifying medical images. Some common DL techniques include [45]:

- Deep Neural Network (DNN): A multi-layer Artificial NN (ANN) connects the input and output nodes. It is trainable in the same way as other ML algorithms and has neurons, synapses, weights, biases, and functions.
- CNN: It assigns significance to different aspects/objects in an input image and can distinguish between them with minimal pre-processing.

This review examines the application of DL models in classifying medical images. It surveys recent studies using DL for medical image classification, highlighting benefits and limitations. It also aims to identify research gaps and suggest future directions to improve classification accuracy across various medical imaging modalities. Section II reviews DL-based medical image classification models, while Section III summarizes the analyses and suggests potential improvements in this field.

II. SURVEY ON DEEP LEARNING-BASED MEDICAL IMAGE CLASSIFICATION

The advent of DL techniques has sparked interest in medical image categorization. These models have shown promise in accurately classifying medical images. This section reviews recent research on advanced techniques for medical image classification. Huang et al. [46] created a Hybrid Network (HybridNet) by combining a modified Principal Component Analysis Network (PCANet) with a simplified DenseNet. The modified PCANet has two stages that produce feature maps by convoluting inputs with learned kernels. The simplified DenseNet uses the feature maps from the PCANet as inputs and employs dense shortcut connections for accurate medical

image classification. In order to identify medical images, Ashraf et al. [47] presented a new way of image representation that makes use of a DL algorithm. On the DNN's final three layers, they adjusted a pre-trained deep CNN.

Fuzzy Discriminative Sparse Representation (FDSR) is a novel approach to medical picture classification that was proposed by Ghasemi et al. [48]. To improve both the similarities and differences between intra-class representations and between classes, this strategy uses fuzzy words. Dictionary atoms were also learned using a flexible fuzzy dictionary learning method. Diagonal Bilinear Interpolated Deep Residual Network (DBI-DRSN) was developed by Assad and Kiczales [49] and is a novel approach to feature extraction and biological image classification. The DRSN uses the extracted features for picture classification, while the DBI uses an interpolation function to improve information and content features.

The Adversarial Lesion Enhancement Neural Network (ALENN) model was created for medical picture categorization by Zhang & Hu [50]. A structure-based adversarial inpainting mechanism and a lesion data-fusion based classification module made up the two stages of the framework. The initial step was to locate the area of interest in the photos that represented the lesion; the subsequent step was to merge that area and classify the results using the improved data.

A network model for the detection of colonic polyps in colonoscopy images was developed by Wang et al. [51]. In an effort to overcome the difficulty of colonic polyps' delicate surface texture, they developed a Channel Information Interaction Perception (CIIP) module. The IIP-Net, or Information Interaction Perception Network, was built on top of this module. Using three different classifier structures—Fully Connected (FC), Global Average Pooling FC (GAP-FC), and Convolution GAP (C-GAP)—the network was able to improve classification accuracy while decreasing computational cost. Using Deep Tree Training (DTT), Yang et al. [52] presented a CNN branch selection algorithm with two stages. To solve the problem of disappearing gradients and reduce computing load, DTT hierarchically trains a succession of networks built from CNN's hidden layers. An ideal classifier was then constructed by merging the CNN branches according to diversity and accuracy standards.

In order to facilitate finer granularity reasoning and enhance information extraction from medical images, An et al. [53] presented a Multiscale Convolutional Neural Network (MCNN) model with a visual attention mechanism. Machine convolutional neural networks (MCNNs) enhance training strategies and medical image classification tasks by automatically extracting high-level discriminative appearance elements from the source image and using a Mahalanobis distance optimization model for the loss function. In order to classify medical images using Ensemble Learning (EL), Abd Elaziz et al. [54] created an IoMT model that is 6G enabled. To enhance classification accuracy and precision, the model employs a blend of MobileNet and DenseNet architecture for feature extraction, and a Honey Badger Algorithm (HBA) based on Levy Flight (LFHBA) for feature selection.

An AI-based Fusion Model (AIFM) was presented for biological picture classification by Mansour et al. [55]. As a preliminary step in the processing, it employs Gaussian filtering to eliminate noise and improve contrast. The feature extraction approach is fusion-based, meaning it uses both deep features from Inceptionv4 and SIFT-based handmade features. To further improve the classification performance of a deep Support Vector Machine (SVM), a whale optimization was employed.

A highly sparse descendent network strategy was created by Zhu et al. [56] using an evolutionary synthesis scheme, which is an evolution-based collective learning approach. In inference, these networks can serve as foundational networks for ensemble learning. Using this approach as a foundation, the Medical Image Classification using Ensemble Bio-inspired Evolutionary DenseNets (MEEDNets) model was created. MEEDNets is composed of several evolutionary DenseNet-121s generated by evolution.

A novel model for medical image classification using several areas was created by Ashwath et al. [57] and is called TS-CNN. The goal of developing this model was to make it easier to understand and use for classification tasks. Medical photos featuring dispersed and irregularly shaped lesions worked especially well with it. There are three parts to this model: one that learns patterns from the input image on a global scale, one that focuses on specific regions and ignores the rest so that the local branch can learn, and a third that makes use of information obtained from the global and local divides to classify the input image.

In order to create a convolutional neural network (CNN) model for medical image classification, Ghosh et al. [58] created a Two-Phase Evolutionary model known as TPEvo-CNN. First, the best CNN architecture number of layers was found using Differential Evolution (DE). During the second stage, the Genetic Algorithm (GA) adjusts the hyperparameters of the constructed CNN layer. The hyper-parameter search space was explored using GA's crossover and mutation processes, and a candidate hyper-parameter for the next generation was chosen using an

elitism selection technique.

The CTransCNN hybrid DL model was created by Wu et al. [59]. It combines Transformer and CNN and consists of three primary modules: MMAEF, MBR, and IIM. A multi-label multi-head attention enhanced feature module and a multi-branch residual module round out the model. For medical image classification using multiple labels, MMAEF investigates implicit label correlations, MBR optimizes the model, and IIM improves feature transmission and increases branch-to-branch nonlinearity.

The ICNN-Ensemble, created by Musaev et al. [60], is an enhanced CNN ensemble that uses RHRIC and SMDE, or an organized model dropout ensemble. It use the original RGB images in conjunction with RHRIC-processed image channels to obtain a deeper understanding of image channels and access additional residual feature connections. In the ICNN-Ensemble model, SMDE chooses ensemble members according to changes in the Accurate Prediction Field (APF). The model can be taught with larger batches of images by performing ensembling during the test set prediction, which makes the most efficient use of the GPU.

Table 1 summarizes the advantages, disadvantages, and performance of the medical image classification models discussed above.

Table 1. Comparative Study of DL-Based Medical Image Classification Models

Ref. No.	Models	Merits	Demerits	Dataset	Performance
[46]	HybridNet	The model had fewer parameters, resulting in lower computational cost and preventing overfitting.	The accuracy was not satisfactory.	Digital Database for Screening Mammography (DDSM), osteosarcoma histology images, and the Mammographic Image Analysis Society (MIAS) dataset	**DDSM Dataset:** Sensitivity=0.862; Specificity=0.787; Accuracy=0.83; Area Under Curve (AUC)=0.897 **Osteosarcoma Histology Image Dataset:** Sensitivity=0.872; Specificity=0.936; Accuracy=0.872; AUC=0.965 **MIAS Dataset:** Sensitivity=0.867; Specificity=0.933; Accuracy=0.867; AUC=0.938
[47]	Deep CNN	It achieved the highest accuracy in classifying medical images.	The execution time and computational cost were high for large-scale datasets.	Multiple open-access datasets of medical images for various human body organs	Accuracy=97.73%; Execution time=682min 10s
[48]	FDSR	It achieved maximum classification accuracy by capturing representative and discriminative patterns of all classes.	The usage of fuzzy discriminative terms and fuzzy learning with dictionary atoms contributed to the high computational complexity.	Repository of Molecular Brain Neoplasia Data (REMBRANDT), The Cancer Genome Atlas Low Grade Glioma (TCGA-LGG), and MIAS dataset	**REMBRANDT:** Accuracy=99.383%; Sensitivity=98.144%; Specificity=99.629% **TCGA-LGG:** Accuracy=98.599%; Sensitivity=95.82%; Specificity=99.162% **MIAS:** Accuracy=99.176%; Sensitivity=97.826%; Specificity=99.5%
[49]	DBI-DRSN	Efficient and reliable.	Accuracy was not efficient for large-scale	Kaggle dataset containing Optical Coherence	Accuracy=90.45%

			medical image datasets.	Tomography (OCT) images	
[50]	ALENN	It improved lesion positioning efficiency with coarse-grained labels.	Time complexity of the sliding window was relatively high.	MURA dataset containing a musculoskeletal X-ray dataset	Accuracy=85.59%; Precision=81.11%; Sensitivity=93.19%; F1 score=86.73%
[51]	IIP-Net-GAP-FC	It achieved the highest level of accuracy and specificity.	Collecting and pre-processing the colonoscopy image dataset was challenging.	Colonic polyp image dataset	Accuracy=99.59%; Precision=99.4%; Sensitivity=99.4%; Specificity=99.7%; F1 score=99.4%
[52]	DTT using GoogLeNet	It achieved high specificity and feasibility.	Complex model with excessive hyperparameters and inappropriate settings can result in poor performance.	Breast Histopathology Images (BHI) dataset and Chest X-ray dataset	**BHI Dataset:** Accuracy=88.8%; Sensitivity=88.56%; Specificity=93.16%; F1 score=86.33% **Chest X-ray Dataset:** Accuracy=86.7%; Sensitivity=89.34%; Specificity=95.3%; F1 score=91.26%
[53]	MCNN	It was highly stable and achieved the highest accuracy.	Overfitting can degrade the classification efficiency.	Lung nodule dataset created by Japanese Society of Radiology Technology (JSRT) and Wisconsin Breast Cancer Database (WBCD)	**Lung nodule dataset:** Accuracy=99.86% **WBCD Dataset:** Accuracy=99.89%
[54]	EL-LFHBA	It achieved better performance.	It was highly complex in terms of both time and memory.	Chest X-ray and OCT datasets from Kaggle	**Chest X-ray Dataset:** Accuracy=87.1%; F1 score=86.19%; Sensitivity=87.1%; Precision=88.56% **OCT Dataset:** Accuracy=94.32%; F1 score=94.3%; Sensitivity=94.32%; Precision=94.93%
[55]	AIFM	It realized better accuracy and specificity.	It cannot handle overfitting problem.	Dataset from the Warwick-QU study on colorectal glands	Sensitivity=93.58%; Specificity=96.98%; Accuracy=96.18%;
[56]	MEEDNet	It achieved the highest accuracy.	It was entirely focused on narrowing the network's width, ignoring its depth.	Brain tumor MRI dataset with SARS-CoV-2 CT scan dataset	**SARS-CoV-2 CT Scan Dataset:** Accuracy=99.31%; Precision=99.31%; Sensitivity=99.32%; F1 score=99.31% **Brain Tumor Dataset:** Accuracy=98.89%; Precision=98.63%; Sensitivity=98.91%; F1 score=98.76%
[57]	TS-CNN	It was more flexible and	The accuracy was heavily	Custom blob dataset and open-	**Custom Blob Dataset:**

		easier to understand.	dependent on the global branch and vulnerable to noise, resulting in degraded performance as noise levels increased. This indicates an inability to handle noisy input images.	access skin lesion PAD-UFES-20 dataset	Accuracy=99%; Precision=99.1%; Sensitivity=99%; F1 score=98.99%; AUC=99.99% **PAD-UFES-20 Dataset:** Accuracy=72.6%; Precision=71.7%; Sensitivity=72.6%; F1 score=71.5%; AUC=90.12%
[58]	TPEvo-CNN	It was highly efficient for large-scale medical image datasets.	The accuracy was not high for small-scale datasets.	Colon cancer, radiology, COVID-19, and pneumonia databases	**COVID-19x_1:** Accuracy=98.2%; Precision=98.5%; Sensitivity=98%; F1 score=98% **COVID-19x_2:** Accuracy=97.3%; Precision=96%; Sensitivity=95.3%; F1 score=95.6% **COVID-19-CT:** Accuracy=95.4%; Precision=95.5%; Sensitivity=95.5%; F1 score=95% **COVID-19-Radiography:** Accuracy=97%; Precision=94.8%; Sensitivity=95.3%; F1 score=94.7% **Pneumonia Dataset:** Accuracy=79.4% **Skin Cancer Dataset:** Accuracy=83.4%
[59]	CTransCNN	It has a strong ability to generalize for medical image classification.	Its ability to handle dependencies was inefficient and accuracy was low because of inappropriate hyperparameter settings.	Chest X-ray11, NIH Chest X-ray14, and custom-made Traditional Chinese Medicine Tongue Dataset (TCMTD)	**Chest X-ray11:** AUC=83.37% **NIH Chest X-ray14:** AUC=78.47% **TCMTD:** AUC=86.56%
[60]	ICNN-Ensemble	It achieved maximum accuracy with minimal complexity in classifying medical images.	Its generalizability was limited.	Malaria cell images dataset	Accuracy=99.67%; Precision=0.966; Sensitivity=0.9; F1 score=0.966

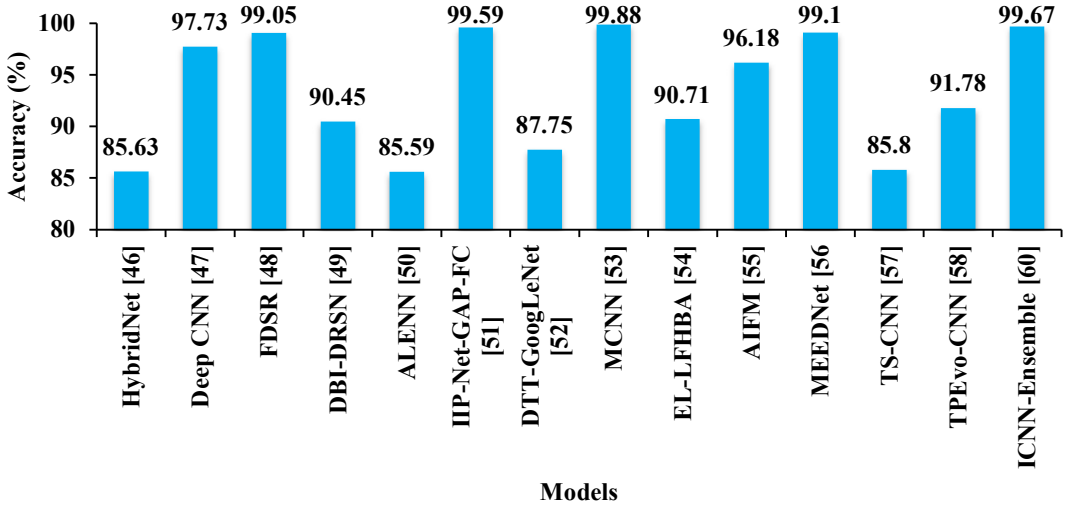

Figure 12. Comparison of Accuracy Values Achieved by Different DL Models on Various Medical Image Datasets

Figure 12 shows a comparison of the accuracy of various DL models on datasets of medical images. Every model's accuracy in image classification is displayed here. The MCNN [53] had the highest accuracy, but it suffered from overfitting. To address this, the ICNN-Ensemble [60] was developed, achieving the highest accuracy on the Malaria cell images dataset by reducing overfitting and prediction errors.

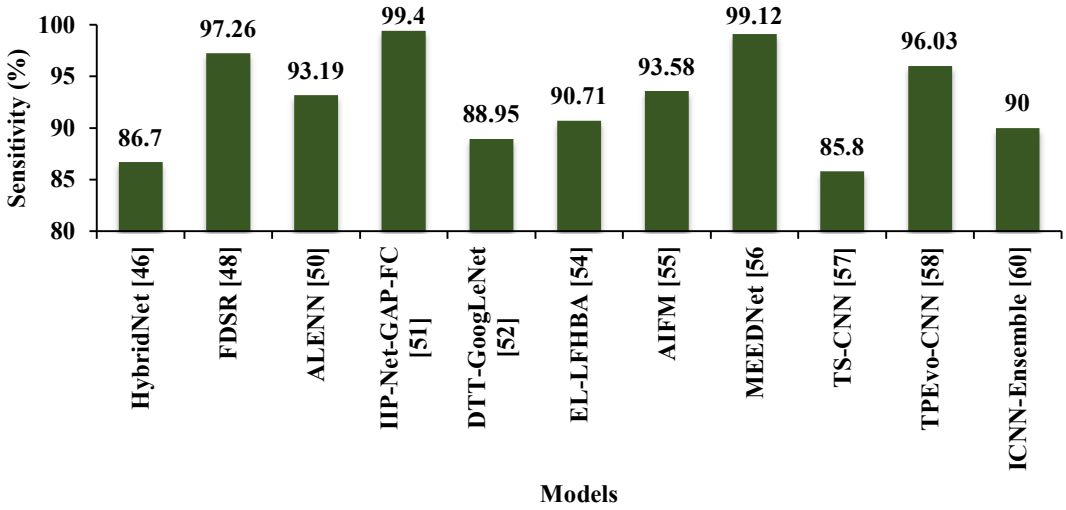

Figure 13. Comparison of Sensitivity Values Achieved by Different DL Models on Various Medical Image Datasets

A variety of medical picture datasets are used to test the susceptibility of various DL models (Figure 13). It measures a model's ability to correctly identify diseased cases in the dataset. The IIP-Net-GAP-FC [51] achieved the highest sensitivity on the Colonic Polyp image dataset dataset, while the ICNN-Ensemble [60] achieved 90% sensitivity on the Malaria cell images dataset.

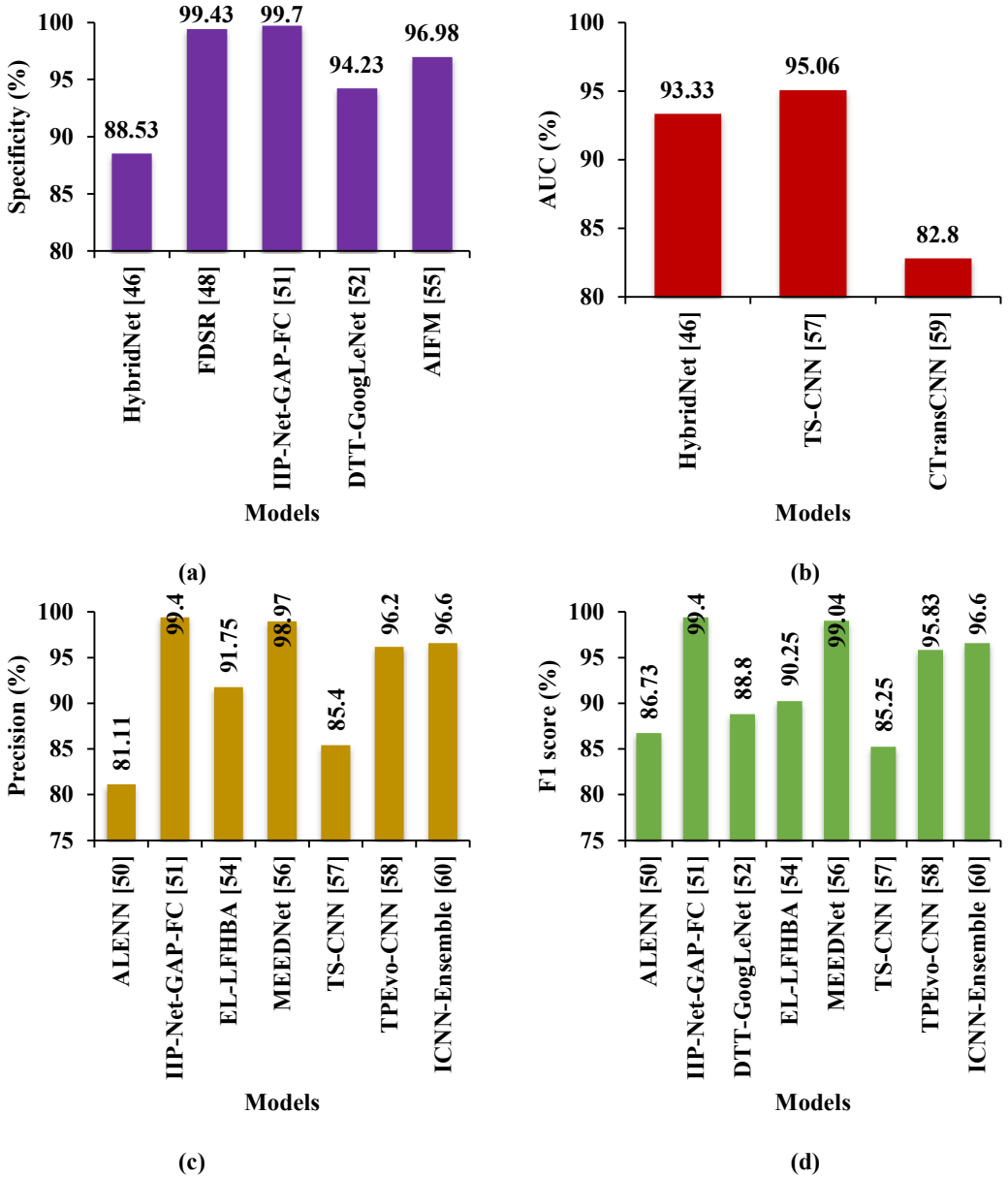

Figure 14. Comparison of (a) Specificity, (b) AUC, (c) Precision, and (d) F1 score Values Achieved by Different DL Models on Various Medical Image Datasets

The metrics for DL model performance on several medical imaging datasets are compared in Figure 14. It includes specificity, AUC, precision, and F1 score values, which measure the models' ability to identify negative cases, overall performance in distinguishing between positive and negative cases, avoidance of false positives, and a balanced evaluation of precision and sensitivity, respectively.

In Figure 14(a), models like FDSR, MEEDNet, and AIFM achieved high specificity values above 96%, indicating excellent performance in correctly identifying negative cases (e.g., healthy cases). The specificity value for ICNN-Ensemble is not explicitly provided. In Figure 14(b), the AUC values indicate that models like MCNN and HybridNet achieved AUC values above 0.9, indicating very good performance. However, the AUC value for ICNN-Ensemble is not explicitly provided.

Figure 14(c) shows that the IIP-Net-GAP-FC [51] achieved the maximum precision on the Colonic Polyp image dataset, while the ICNN-Ensemble model achieved a precision of 96.6% on the Malaria cell images dataset, indicating its ability to avoid false positives. In Figure 14(d), the IIP-Net-GAP-FC [51] achieved the maximum F1 score on the Colonic Polyp image dataset, while the ICNN-Ensemble model achieved an F1 score of 96.6%

on the Malaria cell images dataset, indicating a well-balanced performance in terms of both precision and sensitivity.

The findings highlight the need to keep in mind that various datasets and visual modalities may yield varying model performances. As far as medical picture classification goes, the ICNN-Ensemble model is head and shoulders above the competition, especially on the malaria cell images dataset. It achieved the highest accuracy, sensitivity, precision, and F1 score by utilizing the most effective ensemble combination to prevent the loss of APF. This model demonstrates exceptional capability in accurately classifying images in this dataset.

III. CONCLUSION

In this exhaustive work, we have examined the most recent DL methods for medical image classification. Automatic feature extraction from raw image data has demonstrated encouraging results using DL models, including CNNs. Among the various studies reviewed, the ICNN-Ensemble model demonstrated the highest performance in classifying medical images by leveraging ensemble learning to reduce prediction errors. However, while there have been performance improvements, there are trade-offs between accuracy, model complexity, generalizability, and computational cost. Since other datasets were not taken into account during the model's evaluation, issues with generalization and the complexity of the model's size still need to be resolved. Therefore, future research should focus on developing robust DL models capable of reliably classifying diverse medical images across different imaging modalities. It will be crucial to integrate domain knowledge, address data irregularities, and improve interpretability to facilitate the clinical adoption of AI-powered medical image classification systems.

REFERENCES

[1] Panayides, A. S., Amini, A., Filipovic, N. D., Sharma, A., Tsaftaris, S. A., Young, A., ... & Pattichis, C. S. (2020). AI in medical imaging informatics: current challenges and future directions. *IEEE Journal of Biomedical and Health Informatics*, *24*(7), 1837-1857.

[2] Chan, H. P., Hadjiiski, L. M., & Samala, R. K. (2020). Computer-aided diagnosis in the era of deep learning. *Medical Physics*, *47*(5), e218-e227.

[3] Azam, M. A., Khan, K. B., Salahuddin, S., Rehman, E., Khan, S. A., Khan, M. A., ... & Gandomi, A. H. (2022). A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. *Computers in Biology and Medicine*, *144*, 105253.

[4] Grgat, J., & Matijaš, T. (2021). Comparison of different radiographic image receptors. *Radiološki Vjesnik: Radiologija, Radioterapija, Nuklearna Medicina*, *45*(1), 2-10.

[5] Hussain, S., Mubeen, I., Ullah, N., Shah, S. S. U. D., Khan, B. A., Zahoor, M., ... & Sultan, M. A. (2022). Modern diagnostic imaging technique applications and risk factors in the medical field: A review. *BioMed Research International*, *2022*, 1-19.

[6] Sarker, S., Biswas, A., Al, N. M. A., Ali, M. S., Puppala, S., & Talukder, S. (2023). Case studies on X-ray imaging, MRI and nuclear imaging. *Data Driven Approaches on Medical Imaging*, 207-225.

[7] Li, Y., Ma, Y., Wu, Z., Xie, R., Zeng, F., Cai, H., ... & Wu, M. (2021). Advanced imaging techniques for differentiating pseudoprogression and tumor recurrence after immunotherapy for glioblastoma. *Frontiers in Immunology*, *12*, 790674.

[8] Wierzbicki, R., Pawłowicz, M., Job, J., Balawender, R., Kostarczyk, W., Stanuch, M., ... & Skalski, A. (2022). 3D mixed-reality visualization of medical imaging data as a supporting tool for innovative, minimally invasive surgery for gastrointestinal tumors and systemic treatment as a new path in personalized treatment of advanced cancer diseases. *Journal of Cancer Research and Clinical Oncology*, *148*(1), 237-243.

[9] Isaac, A., Dalili, D., Dalili, D., & Weber, M. A. (2020). State-of-the-art imaging for diagnosis of metastatic bone disease. *Der Radiologe*, *60*(Suppl 1), 1.

[10] Azagidi, A. S., Ibitoye, B. O., Makinde, O. N., Idowu, B. M., & Aderibigbe, A. S. (2020). Fetal gestational age determination using ultrasound placental thickness. *Journal of Medical Ultrasound*, *28*(1), 17.

[11] Saxena, U., & Soni, P. K. (2023). X-rays: A boon to radiology; a hazard to human health. *International Journal of Drug Research and Dental Science*, *5*(3), 19-28.

[12] Jung, H. (2021). Basic physical principles and clinical applications of computed tomography. *Progress in Medical Physics*, *32*(1), 1-17.

[13] Pelicano, A. C., Gonçalves, M. C., Godinho, D. M., Castela, T., Orvalho, M. L., Araújo, N. A., ... & Conceição, R. C. (2021). Development of 3D MRI-based anatomically realistic models of breast tissues and tumours for microwave imaging diagnosis. *Sensors*, *21*(24), 8265.

[14] Yu, Y., Jain, B., Anand, G., Heidarian, M., Lowe, A., & Kalra, A. (2023). Technologies for non-invasive physiological sensing: Status, challenges, and future horizons. *Biosensors and Bioelectronics: X*, *16*, 100420.

[15] Alavi, A., Saboury, B., Nardo, L., Zhang, V., Wang, M., Li, H., ... & Revheim, M. E. (2022). Potential and most relevant applications of total body PET/CT imaging. *Clinical Nuclear Medicine*, *47*(1), 43-55.
[16] Abdelrahman, L., Al Ghamdi, M., Collado-Mesa, F., & Abdel-Mottaleb, M. (2021). Convolutional neural networks for breast cancer detection in mammography: A survey. *Computers in Biology and Medicine*, *131*, 104248.
[17] Mahmood, T., Rehman, A., Saba, T., Nadeem, L., & Bahaj, S. A. O. (2023). Recent advancements and future prospects in active deep learning for medical image segmentation and classification. *IEEE Access*, *11*, 113623-113652.
[18] Geraldes, C. F. (2020). Introduction to infrared and Raman-based biomedical molecular imaging and comparison with other modalities. *Molecules*, *25*(23), 5547.
[19] Diaz, O., Kushibar, K., Osuala, R., Linardos, A., Garrucho, L., Igual, L., ... & Lekadir, K. (2021). Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools. *Physica Medica*, *83*, 25-37.
[20] Nayyef, R. H., & Al-Tamimi, M. S. (2022). A comparative study and overview on the magnetic resonance images skull stripping methods and their correspondence techniques. *International Journal of Nonlinear Analysis and Applications*, *13*(1), 3783-3802.
[21] Duarte-Salazar, C. A., Castro-Ospina, A. E., Becerra, M. A., & Delgado-Trejos, E. (2020). Speckle noise reduction in ultrasound images for improving the metrological evaluation of biomedical applications: An overview. *IEEE Access*, *8*, 15983-15999.
[22] Marfisi, D., Tessa, C., Marzi, C., Del Meglio, J., Linsalata, S., Borgheresi, R., ... & Giannelli, M. (2022). Image resampling and discretization effect on the estimate of myocardial radiomic features from T1 and T2 mapping in hypertrophic cardiomyopathy. *Scientific Reports*, *12*(1), 10186.
[23] Unberath, M., Gao, C., Hu, Y., Judish, M., Taylor, R. H., Armand, M., & Grupp, R. (2021). The impact of machine learning on 2d/3d registration for image-guided interventions: A systematic review and perspective. *Frontiers in Robotics and AI*, *8*, 716007.
[24] Kociołek, M., Strzelecki, M., & Obuchowicz, R. (2020). Does image normalization and intensity resolution impact texture classification?. *Computerized Medical Imaging and Graphics*, *81*, 101716.
[25] Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A review of deep-learning-based medical image segmentation methods. *Sustainability*, *13*(3), 1224.
[26] Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2021). A review of deep learning based methods for medical image multi-organ segmentation. *Physica Medica*, *85*, 107-122.
[27] Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved threshold based and trainable fully automated segmentation for breast cancer boundary and pectoral muscle in mammogram images. *IEEE Access*, *8*, 203097-203116.
[28] Mazouzi, S., & Guessoum, Z. (2021). A fast and fully distributed method for region-based image segmentation: Fast distributed region-based image segmentation. *Journal of Real-Time Image Processing*, *18*(3), 793-806.
[29] Abdel-Gawad, A. H., Said, L. A., & Radwan, A. G. (2020). Optimized edge detection technique for brain tumor detection in MR images. *IEEE Access*, *8*, 136243-136259.
[30] Cao, C., Zhou, C., Yu, J., Hu, K., & Xiao, F. (2020). A novel active contour model using oriented smoothness and infinite Laplacian for medical image segmentation. *Cognitive Internet of Things: Frameworks, Tools and Applications*, *810*, 311-321.
[31] Khan, Z., Yahya, N., Alsaih, K., Al-Hiyali, M. I., & Meriaudeau, F. (2021). Recent automatic segmentation algorithms of MRI prostate regions: A review. *IEEE Access*, *9*, 97878-97905.
[32] Kuwil, F. H. (2022). A new feature extraction approach of medical image based on data distribution skew. *Neuroscience Informatics*, *2*(3), 100097.
[33] Ghalati, M. K., Nunes, A., Ferreira, H., Serranho, P., & Bernardes, R. (2021). Texture analysis and its applications in biomedical imaging: A survey. *IEEE Reviews in Biomedical Engineering*, *15*, 222-246.
[34] Gomes Ataide, E. J., Ponugoti, N., Illanes, A., Schenke, S., Kreissl, M., & Friebe, M. (2020). Thyroid nodule classification for physician decision support using machine learning-evaluated geometric and morphological features. *Sensors*, *20*(21), 6110.
[35] Latha, S., Samiappan, D., & Kumar, R. (2020). Carotid artery ultrasound image analysis: A review of the literature. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, *234*(5), 417-443.
[36] Pizer, S. M., Marron, J. S., Damon, J. N., Vicory, J., Krishna, A., Liu, Z., & Taheri, M. (2022). Skeletons, object shape, statistics. *Frontiers in Computer Science*, *4*, 842637.
[37] Lei, M., Varghese, B., Hwang, D., Cen, S., Lei, X., Desai, B., ... & Duddalwar, V. (2021). Benchmarking various radiomic toolkit features while applying the image biomarker standardization initiative toward clinical translation of radiomic analysis. *Journal of Digital Imaging*, *34*, 1156-1170.

[38] Imani, M. (2021). Automatic diagnosis of coronavirus (COVID-19) using shape and texture characteristics extracted from X-Ray and CT-Scan images. *Biomedical Signal Processing and Control*, *68*, 102602.
[39] Chowdhury, T. H., Poudel, K. N., & Hu, Y. (2020). Time-frequency analysis, denoising, compression, segmentation, and classification of PCG signals. *IEEE Access*, *8*, 160882-160890.
[40] Abdel-Hamid, L. (2024). Multiresolution analysis for COVID-19 diagnosis from chest CT images: wavelet vs. contourlet transforms. *Multimedia Tools and Applications*, *83*(1), 2749-2771.
[41] Bansal, M., Kumar, M., & Kumar, M. (2021). 2D object recognition: a comparative analysis of SIFT, SURF and ORB feature descriptors. *Multimedia Tools and Applications*, *80*, 18839-18857.
[42] Mall, P. K., Singh, P. K., Srivastav, S., Narayan, V., Paprzycki, M., Jaworska, T., & Ganzha, M. (2023). A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities. *Healthcare Analytics*, 100216.
[43] Wei, R., & Mahmood, A. (2020). Recent advances in variational autoencoders with representation learning for biomedical informatics: A survey. *IEEE Access*, *9*, 4939-4956.
[44] Seo, H., Badiei Khuzani, M., Vasudevan, V., Huang, C., Ren, H., Xiao, R., ... & Xing, L. (2020). Machine learning techniques for biomedical image segmentation: an overview of technical aspects and introduction to state-of-art applications. *Medical Physics*, *47*(5), e148-e167.
[45] Rana, M., & Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. *Multimedia Tools and Applications*, *82*(17), 26731-26769.
[46] Huang, Z., Zhu, X., Ding, M., & Zhang, X. (2020). Medical image classification using a light-weighted hybrid neural network based on PCANet and DenseNet. *IEEE Access*, *8*, 24697-24712.
[47] Ashraf, R., Habib, M. A., Akram, M., Latif, M. A., Malik, M. S. A., Awais, M., ... & Abbas, Z. (2020). Deep convolution neural network for big data medical image classification. *IEEE Access*, *8*, 105659-105670.
[48] Ghasemi, M., Kelarestaghi, M., Eshghi, F., & Sharifi, A. (2020). FDSR: a new fuzzy discriminative sparse representation method for medical image classification. *Artificial Intelligence in Medicine*, *106*, 101876.
[49] Assad, M. B., & Kiczales, R. (2020). Deep biomedical image classification using diagonal bilinear interpolation and residual network. *International Journal of Intelligent Networks*, *1*, 148-156.
[50] Zhang, B., & Hu, X. (2021). A medical image classification model based on adversarial lesion enhancement. *Scientific Programming*, *2021*, 1-9.
[51] Wang, W., Hu, Y., Luo, Y., & Wang, X. (2021). Medical image classification based on information interaction perception mechanism. *Computational Intelligence and Neuroscience*, *2021*, 1-12.
[52] Yang, Y., Hu, Y., Zhang, X., & Wang, S. (2021). Two-stage selective ensemble of CNN via deep tree training for medical image classification. *IEEE Transactions on Cybernetics*, *52*(9), 9194-9207.
[53] An, F., Li, X., & Ma, X. (2021). Medical image classification algorithm based on visual attention mechanism-MCNN. *Oxidative Medicine and Cellular Longevity*, *2021*, 1-12.
[54] Abd Elaziz, M., Mabrouk, A., Dahou, A., & Chelloug, S. A. (2022). Medical image classification utilizing ensemble learning and levy flight-based honey badger algorithm on 6G-enabled internet of things. *Computational Intelligence and Neuroscience*, *2022*, 1-17.
[55] Mansour, R. F., Alfar, N. M., Abdel-Khalek, S., Abdelhaq, M., Saeed, R. A., & Alsaqour, R. (2022). Optimal deep learning based fusion model for biomedical image classification. *Expert Systems*, *39*(3), e12764.
[56] Zhu, H., Wang, W., Ulidowski, I., Zhou, Q., Wang, S., Chen, H., & Zhang, Y. (2023). MEEDNets: Medical image classification via ensemble bio-inspired evolutionary DenseNets. *Knowledge-Based Systems*, *280*, 111035.
[57] Ashwath, V. A., Sikha, O. K., & Benitez, R. (2023). TS-CNN: A Three-Tier Self-Interpretable CNN for Multi-Region Medical Image Classification. *IEEE Access*, *11*, 78402-78418.
[58] Ghosh, A., Jana, N. D., Das, S., & Mallipeddi, R. (2023). Two-phase evolutionary convolutional neural network architecture search for medical image classification. *IEEE Access*, *11*, 115280-115305.
[59] Wu, X., Feng, Y., Xu, H., Lin, Z., Chen, T., Li, S., ... & Zhang, S. (2023). CTransCNN: Combining transformer and CNN in multilabel medical image classification. *Knowledge-Based Systems*, *281*, 111030.
[60] Musaev, J., Anorboev, A., Seo, Y. S., Nguyen, N. T., & Hwang, D. (2023). ICNN-Ensemble: An improved convolutional neural network ensemble model for medical image classification. *IEEE Access*, *11*, 86285-86296.