Deep Learning for Automated Detection of Cysts and Tumors: A Review

Vaibhav Chadha[1], Manjot Kaur[2], Navneet Kaur[3]

{vaibhavchadha29@gmail.com[1], manu.sembhi@gmail.com[2], navneetphul@gmail.com[3]}

Lovely Professional University, Phagwara, India[1,2,3]

Abstract. Medical imaging involves the detection of cysts in order to enable early diagnosis and treatment of a disease. The conventional methods of diagnosis are based on subjective interpretation of radiologists that is time consuming. The latest development of machine learning and deep learning make possible the automated and highly accurate cyst detection. The review entails the discussion of both ML and DL techniques of cyst detection and segmentation in ultrasound, MRI, CT and optical coherence tomography. It cross-cuts through the popular datasets, preprocessing mechanisms, feature extraction mechanisms and classifiers including Convolutional Neural Networks, Random Forest and Support Vector Machines. It further evaluates the state of art approaches, such as transfer learning, deep supervision, and integration of domain knowledge in improving the segmentation performance. The significant concerns that cover the lack of adequate data, explainability of the model, and clinical adoption are discussed. The next research directions include Explainable Artificial Intelligence, federated learning and multimodal methods of improving the accuracy and reliability of cyst detection in real-world clinical environment.

Keywords: Medical image segmentation, deep learning, convolutional neural networks

1 Introduction

Cysts are liquid or semi-solid, fluid filled sacs that develop in the majority of the body organs that include the brain, kidney, liver, ovaries, pancreas and jaw [1]. Although the majority of cysts are benign, they can indicate or develop serious pathophysiological events, including malignancies or organ dysfunction [2]. It is noted that these cysts can be treated and prevented with success with the timely diagnosis and proper classification [3]. However, the manual diagnosis of cysts still involves considerable use of radiologists, which with the help of imaging devices such as ultrasound (US), magnetic resonance imaging (MRI), computed tomography (CT), and optical coherence tomography (OCT) requires a lot of time, inter-observer variability, and depends on the expertise of a clinician [4], [5]. It has been demonstrated that CNNs are more helpful in extracting, classifying, and segmenting cystic lesions with high efficiency and precision [6]. This has transformed medical imaging with a sea change brought by machine learning (ML) and deep learning (DL). Besides, Support Vector Machines (SVMs) and Random Forests (RFs) have been employed in the process of cyst classification by utilizing the radiomic features produced based on the data received by imaging modalities to enhance the quality of the diagnostic results [7].

IICTDS 2025, December 19-20, Chandigarh, India
Copyright © 2026 EAI
DOI 10.4108/eai.19-12-2025.2362504

In addition, the other sophisticated AI methods, like transfer learning, deep supervision, and domain knowledge integration, have contributed an immeasurable value to the execution of a model because they optimize features extractions and segmentation accurateness [8], [9]. This paper is a review of the large quantity of machine learning (ML) and deep learning (DL) algorithms used in the cysts detection and segmentation of various imaging modalities [4], [5]. In the given review, the authors describe ML and DL-based cyst-auto-detection and segmentation on a medical image based on the type of the imaging, datasets, and preprocessing and feature-extraction techniques to improve the performance of the model [6]. The majority of recent developments in the field of DL and ML, such as Convolutional Neural Networks (CNNs), transfer learning and hybrid networks have enhanced the accuracy and speed of cysts diagnosis [10], [11]. The barriers, in its turn, that make the large-scale stipulation of AI-based diagnosis systems not possible, are the insufficiency of data, interpretability of the model, and change in clinical practice [12]. Finally, in this review, the future perspectives are briefly mentioned, including explainable AI (XAI), federated learning, and multimodal AI integration, which are likely to enhance the accuracy of the diagnosis, robustness of the models, and their practicability in the clinics [12], [13].

2 AI Techniques for Cyst Detection and Segmentation

2.1 Convolutional Neural Networks (CNNs)

Most existing cyst detection and segmentation systems are based on CNNs [10], [11] because they can automatically acquire hierarchical representations of imaging data. Shallow convolutional layers are used in the detection of low-level features like edges and textures, and the deeper layers are used in the detection of higher-level structural patterns matching cystic lesions. The classification of cysts and non-cystic structures has been extensively performed by CNN-based classifiers, whereas pixel level segmentation is possible with fully convolutional architectures [14]. CNNs are especially appropriate to detect the boundaries of cysts that change in shape, appearance, and intensity across patients and imaging modalities since they are translation invariant and learn spatial information.

2.2 Transfer Learning Approaches

Since the size of labeled cysts is small, transfer learning is essential in enhancing model performance and generalizability. Commonly used pretrained architectures that are used in cyst detection include VGG16, ResNet50, Inception-ResNet-V2, and EfficientNet [15], [16]. These models take advantage of the feature representations that are trained with large-scale datasets (e.g., ImageNet) and train them on the specifics of domain medical images. Transfer learning saves training time and alleviates overfitting, as well as, improves performance, especially in classification problems where cystic lesions have imprecise or delicate features. It is also strong and supports different imaging protocols and acquisitions.

2.3 Deep Supervision in Medical Segmentation

Deep supervision adds auxiliary loss terms at the in-between layers of a segmentation network and permits more accurate gradient flow and learning multi-scale features. Architectures like deeply supervised U-Net variants and V-Net variants adopt this approach to enhance convergence stability and boundary precision [9]. Deep supervision assists networks in segmenting small, low-contrast or irregularly-shaped cysts that would otherwise be missed in

lower layers in the cyst segmentation. This multi-level learning methodology guarantees that the local and larger contextual information which has high resolution is used to the final product of the segmentation.

2.4 Domain Knowledge Integration

Radiologic or anatomical domain knowledge is often incorporated into many cyst-detection models to give them better clinical relevance and interpretability [8]. Methods involve the use of handcrafted radiomic features as well as deep learning predictions, the use of spatial priors on a cyst location, and loss functions that are used to penalize inconsistency of boundaries. CNN-based feature extraction is sometimes combined with classical models, like SVMs or Random Forests, to use some domain-derived descriptors. These hybrid methods also increase performance in difficult cases and build credibility in clinicians by integrating specialist knowledge on the morphology and imaging appearance of cysts.

2.5 Data Augmentation and Generative Modeling

The variability and the paucity of medical imaging datasets can be tackled using the methods of data augmentation: random rotations, flips, scaling, contrast adjustments, and elastic deformations. The transformations make the data diverse, less overfitted and more resilient between scanners and patients. Moreover, other generative models, including Generative Adversarial Networks (GANs), have been popular to generate realistic images of cystic or segmentation masks [17]. GAN augmentation assists in the balancing of classes, the generation of rare manifestations of cysts, and to train deeper models in situations where annotated data is scarce. Generative modeling and augmentation are both sources of better generalization and stability of cyst segmentation models.

The recent aggressive development of deep learning (DL) and machine learning (ML) astonishingly enhanced cysts detection and segmentation in a broad spectrum of different imaging modalities [5]. The conventional methods are based on the manual features and hand annotations which are constrained by the expensive cost of computing and inter-observer error [4]. Convolutional Neural Networks (CNNs) are one of the types of deep learning, which has accelerated the processing of medical images as a process of automated extraction of features and enhancement of the degree of segmentation and classification [6].

3 Standardized Methodological Framework for Reproducible AI-Based Cyst Detection

3.1 Methodological Foundations for Reproducible AI in Cyst Detection

The recent developments in AI-based detection of cysts underscore the importance of reporting methodology in the studies in a transparent, reproducible and standardized manner. One of the major weaknesses that have been identified in the current literature is the variability in the description of data sets, preprocessing procedures, annotation guidelines, and training setups. In the solution of this, an integrated methodological base is needed. All AI studies should specify the imaging datasets they contained, such as modality (CT, MRI, US, OCT), the number of cases of patients, image resolution, and inclusion/exclusion criteria. The quality of the annotation is another important parameter which must state the number of radiologists to whom ground-truth masks or labels were computed, the name of the annotation tool, as well as whether inter-

observer agreement or consensus rules were employed. The preprocessing tasks, including intensity normalization, resizing, denoising and data augmentation should be described in detail, so that the various implementations can maintain consistency. Also, model architecture reporting must contain the precise layer configurations, the number of parameters, the loss functions, optimizers, the batch size, the learning rate schedules, the number of epochs and hardware (e.g., the type of GPU used). The inclusion of such methodological detail increases the reproducibility, enables fair benchmarking and makes the findings reproducible with other studies in future.

3.2 Data Preprocessing and Augmentation Pipeline

Considering the variability of imaging modalities, settings, and morphologies of cysts, preprocessing is a vital endeavor in the standardization of the inputs before the training of models. Throughout the surveyed literature, multiple operations of preprocessing were found to be consistent. Intensity normalization was used in most studies; minmax scaling or Z-score normalization were used, such that models would learn based on normalized intensity distributions, but not brightness variation based on modality. Image resizing was also popular to obtain the same spatial size, usually 224×224, 256x256 or 512x512 pixels based on computing power and architectural limitations. The methods of noise reduction were modality dependent. Gaussian smoothing was also commonly applied in MRI to reduce thermal noise, whereas ultrasound images were reduced by speckle noise reduction filters like the median filtering or anisotropic diffusion. In case of CT imaging, Contrast-Limited Adaptive Histogram Equalization (CLAHE) maximized edge contrast and the visibility of low-density cystic lesions. OCT data usually needed volumetric denoising or layer-by-layer flattening before segmentation. Different data augmentation strategies were used to ensure that studies dealt with small datasets and enhance extrapolation. Geometric augmentations such as random rotations (±20 o), horizontal and vertical flips, scaling, random crop, were used, all of which cause spatial variation and minimizes overfitting. Sensitivity increases like changes in brightness, contrast, and addition of Gaussian noise were also used to increase their resistance to imaging variations. In the case of volumetric MRI and OCT data, elastic deformations were common to model realistic anatomical variability and enhance the stability of segmentation. Such operations in preprocessing not only normalize representations of inputs, but also makes sure that models are transportable across scanners, and acquisition or clinical conditions. The fact that these pipelines maintain consistent results on several independent studies highlights the significance of these pipelines in enhancing reliability and reproducibility of AI pipelines in detecting cysts.

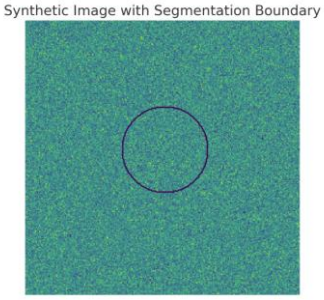


Fig 1. Sample representation of a cystic region with a segmentation boundary, illustrating the typical structure of ground-truth annotations used in cyst detection research

3.3 Model Architectures and Training Configuration

The literature review has found out that there were certain architecture of deep learning and training systems that have continuously presented good performance with regards to detecting and segmenting cysts. Some of the most commonly used models were classical U-Net [10] and its variants, such as the volumetric 3D U-Net [18], attention-enhanced versions such as Attention U-Net [19], and MultiResUNet [20], which uses skip connections and attention gates to focus on discriminative cystic areas. Convolutional backbones, such as ResNet50, VGG16/VGG19, EfficientNet-B0, and Inception-ResNet-V2, were a favorite option in the task of classification due to their representational efficiency and because of transfer learning capabilities. Hybrid methods were also found to be used, including CNN feature extractors with Support Vector Machines or random forest classifiers, especially with radiomic features. The decision on the loss functions was task-specific. Dice loss and Binary Cross-Entropy (BCE) are commonly used to perform segmentation and Focal Loss is commonly known to have the advantage of better performance in extreme class imbalance (especially when the cystic regions of the image constituted only a small percentage of the image). Most popular were Adam (learning rates of 1e-3 to 1e-5) and Stochastic Gradient Descent (SGD) with momentum (0.9), with regularization of training by cyclic learning rate schedules or cosine annealing. The typical training settings consisted of 8 to 32 batch sizes and memory limits and 50 to 150 epochs where the training was stopped early on validation loss in order to prevent overfitting. When using NVIDIA GPUs like Tesla V100, RTX 3090 and A100 cards, hardware specification was most commonly reported in studies that have used these devices to train a large-scale 2D and 3D model, thus making it easy to do so. These architectural and training specifications would be recorded, a fact that would bring about transparency and would be a template that might be applied by other researchers in the future. Fairer comparisons, robust benchmarking, and faster development of clinically viable AI systems may also be made through the standardization of model configurations in detecting and segmenting cysts.

3.4. Evaluation Metrics and Statistical Reporting Standards

The achievement of AI-based cyst detection requires the standard measures that are capable of offering correct evaluation of the quality of the cyst segmentation, the quality of the classification, and the trustworthiness of the model. Measurement of the segmentation studies should also provide the results of the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Hausdorff distance at 95th percentile (HD95), and the Average Symmetric Surface Distance (ASSD), which is the measurements of contour accuracy, overlap and shape conformity of predicted cysts. Accuracy, sensitivity, specificity, precision, recall, F1-score and area under the ROC curve (AUC) are some of the important metrics that can be used in classification studies. It is important to note that all AUCs and other significant values should be presented with the 95 percent confidence intervals that have been obtained using bootstrap resampling to give a statistical reliability. The performance of the models should be compared by the use of statistical significance tests such as the test of AUC comparison suggested by DeLong or the test of paired classification results suggested by McNemar. Such metrics and statistical procedures make the review align to the most relevant research steps of the best medical imaging practices and give the effectiveness of AI models its accuracy and scientific clarity.

The reliability, clinical usefulness, and scientific reproducibility of AI-based models of cyst detection and segmentation should be assessed by use of an effective and clear methodology. Even though the truthfulness or the qualitative outcomes of most of the studies are announced,

there remains a drastic divide in the process of employing the holistic assessment measures and statistical processing done in a standardized way. In order to bridge this gap, this subsection introduces the little number of performance measures and statistical processes that can be recommended as the minimum set so as to evaluate AI systems in cyst imaging.

The studies in the cases of segmentation tasks, i.e. when the aim is to trace the boundaries of cysts pixel by pixel, the space overlap and distance measures should be applied. Dice similarity coefficient (DSC) This is the most common coefficient that is used to estimate the degree of similarity between predicted segmentation mask and the ground truth annotation. Along with DSC, there is yet another very useful measure of the overlap, the Intersection over Union (IoU). The Hausdorff Distance (HD95) and the Average Symmetric Surface Distance (ASSD) also need to be reported to determine the accuracy of contours and fidelity to shape particularly in irregular or small cystic formations. These distance metrics reflect boundary error that could not be reflected by Dice alone, i.e. volumetric 3D MRI or CT data. A combination of these measures provides the entire image of the accuracy of space, the compliance of the boundaries and the quality of the segmentation.

More significant group of measures are required in the case of classification assignments which will categorise cystic lesions into benign, malignant or multi-class groups. The widely documented concept of fairness cannot be found in the case of class imbalance which is prevalent in medical imaging datasets. Consequently, any research should give sensitivity (recall), specificity, precision, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The ROC curve may also provide information about the behavior of the model in varying thresholds when the Precision Recall (PR) curve is specifically applied in skewed datasets in which the cystic conditions are infrequent. To establish the degree of uncertainty and increase the scientific interpretability of AUC values, they must always be accompanied by 95-percent confidence intervals (CI).

In addition to this, the patterns of misclassification, troubled classes and systematic biases must be plotted by providing the confusion matrices of all classification studies. Segmentation networks updated to detection or region proposals [21] in the case, error maps are suggested to be visualized or based on overlays to be compared.

Statistical significance test is another significant, but highly underestimated aspect besides performance measures. When comparing two or more models it is recommended that DeLong test should be used to determine whether the difference in AUC is significant. McNemar test is to be employed in case paired categorical results (e.g. comparing two classifiers to the same test sets) are to be considered. Strong CIs are required in DSC, sensitivity, specificity and F1-score and they are to be obtained through bootstrap-based resampling (typically 1000 runs). Also to be reported should be the effect sizes in the form of Cohen d or mean absolute difference makes comparisons more plausible.

Finally, the reporting of each metric must clearly describe the size of the test dataset, the topography of the cysts, and the nature of the cross-validation or external validation used.

This information will result in interpretability of findings in a clinical context and reproducibility by any other researcher in the future. The frequent use of these standards of evaluation and statistical reporting will greatly boost the credibility, fairness and scientific accuracy of AI systems developed to find the cysts and divide them.

3.5 External Validation and Model Robustness Requirements

One of the biggest weaknesses of the existing models of cyst-detection is the absence of external validation, which severely limits the practicality and accuracy of their application. Models that are trained and evaluated using single-center datasets tend to experience domain shift when trained on images of other scanners, institutions, acquisition protocols or patient populations. To address this, the studies should include multi-centre validation where training is done on data of a single institution and the evaluation is done on independent dataset of other centres. This cross-institutional testing is indicative of clinical diversity in the real world and is more effective at generalizing the model. In addition, robustness test must consist of sensitivity tests to different noise levels, contrast settings, image artifacts and differences in modalities. Cross-modality generalization, e.g., when possible, in CT training and MRI testing, will give further information about the stability of the model. With these validation strategies in place, AI models to be used in cyst detection can be relied upon, as well as be clinically meaningful and able to achieve high performance in a variety of healthcare settings.

3.6 Compliance with AI Reporting Guidelines

To achieve transparency, clinical relevance, and regulatory acceptance in AI research on cyst detection, AI research must correspond to conventional reporting methodologies, including CLAIM (Checklist for Artificial Intelligence in Medical Imaging) and STARD-AI on diagnostic accuracy. These conventions demand a comprehensive record of characteristics of datasets, ground truth generation procedures, model designs, instruction setups, assessment statistics, and planned pathways of clinical integration. The use of CLAIM-based reporting helps to make sure that the key elements of patient selection criteria, protocols of annotation, data preprocessing, model interpretability, and computational environment are clearly outlined. STARD-AI is also obligatory to reveal diagnostic thresholds, reference standards and confidence intervals, which allows a comprehensive assessment of model performance in clinical settings. Following these principles, AI-based cyst-detection literature will be more reproducible and transparent, and can be clinically translated, as well as take into account the ethical concerns of bias, fairness, and patient safety.

4 Literature Review

Artificial intelligence in cysts detection and segmentation has evolved over the past few years. The traditional methods were mainly founded on the handcrafted feature ex- traction techniques e.g., the use of texture analysis, edge detection and the use of histogram to segment the cystic structures out of the non-cystic tissues. Such approaches were most typically used in conjunction with conventional machine-based learning classifiers like Support Vector Machines (SVMs) and Random Forests (RFs) to classify cysts as demonstrated in initial studies in the field of dental cyst and jaw tumor differentiation [1], [7]. As an illustration, results on panoramic radiographs achieved the accuracy of up to 98% with the assistance of a wavelet-based feature extraction and SVM classifiers [7]. However, these methods were marred by lack of generalisability, sensitivity to noise and intense reliance on expert features.

Originally, the introduction of deep learning (DL) has led to a paradigm shift due to the availability of end-to-end automated feature extraction and strong segmentation. Convolutional Neural Net- works (CNNs) are now the horse at work of cyst detection as they are able to learn hierarchical representations of imaging data. U-Net [10] and its volumetric counterparts (e.g.,

3D U-Net [18]) have been used to achieve high-quality segmentation across a variety of modalities, including ultrasound, MRI, and CT to optical coherence tomography (OCT). An example is Ovarian CystNet special CNN-based network that has been demonstrated to be very accurate in distinguishing benign and malignant ovarian cysts [22]. Furthermore, pre-trained models (VGG16, Inception ResNet, or MobileNet, etc.) have been fine-tuned with the help of transfer learning to be applied in cyst detection tasks, and, therefore, avoid the risk of the lack of labeled data [15], [16]. There are also hybrid schemes that combine CNN feature extraction with traditional classifiers like SVMs and RFs, which have enhanced diagnostic performance, which is demonstrated by experiments on hydatid cyst classification with an overall accuracy of above 92 percent [15]. Attention and deep supervision mechanisms have also typified segmentation advances. Deep supervision, where additional loss functions are integrated into mid-level layers, has improved segmentation in pancreatic cyst detection, where DSC boosts of 60.46% to 63.44% have been reported [9]. In addition, attention mechanisms have been incorporated into CNN architectures (e.g., Attention U-Net [19]) to concentrate on salient cystic areas and reduce false positives from surrounding tissues. Integration of domain knowledge has also enhanced segmentation, as in the case of retinal cyst segmentation where anatomical priors enhance localization accuracy [8], [23].

Despite these developments, there are several issues that are left behind. Absence of large, annotated datasets is one of the biggest disadvantages, and in most cases, leads to data imbalance; this has been partially addressed by data augmentation and the use of Generative Adversarial Networks (GANs) to create new training samples [17]. The inherent black-box nature of most DL models also constitutes an obstacle to interpretability, which must be accepted by clinicians to embrace it [12].

To strengthen comparative evaluation across studies, it is essential to incorporate standardized benchmark models that are widely recognized in biomedical image analysis. Several recent works emphasize the importance of using self-configuring architectures such as nnU-Net, which automatically adjusts preprocessing, architecture depth, and training hyperparameters based on the dataset characteristics [24]. nnU-Net has consistently outperformed manually tuned U-Net variants across tasks involving cyst detection, tumor segmentation, and lesion classification, making it an essential baseline for future cyst-related studies. Similarly, modern architectures that integrate convolutional backbones with Vision Transformers (ViTs), such as TransUNet [25] and hybrid CNN–Transformer encoders, have demonstrated strong generalization, especially in cross-domain medical imaging tasks. Incorporating these modern baselines ensures that comparisons remain relevant, robust, and aligned with the latest developments in the field. Discussing such benchmarks also allows readers to understand where existing cyst-detection models stand relative to state-of-the-art segmentation and classification systems.

Recent discoveries in cyst identification, cyst categorisation and cyst segmentation indicate that AI-powered medical imaging is evolving quickly in most parts of the body and modalities. Using multimodal CT, MRI, and ultrasound data depending on the staging cystic echinococcosis, Tegshee et al. [26] compared ten machine learning models and discovered that hybrid CNN-ResNet networks were the most accurate (up to 100 percent using MRI), but needed data variety. The AI-based confocal endomicroscopy by Jiang et al. [3] was used to predict neoplasm in the pancreatic cysts and showed a superior diagnostic rate (85.7%), but the EUS-nCLE systems were not easily accessible. Several studies about the ovarian cysts have proven the effectiveness of the deep learning models: Sood et al. [22] developed the Ovarian-CystNet: an 87-layer CNN which was capable of achieving an accuracy of 92.06 on the ultrasound images, and Ravishankar et al. [27] have developed an OCD-FCNN with the same

accuracy of 98.37 but based on larger datasets. The XGBoost type of ML that Akter and Akhter [28] applied to the PLCO datasets to predict ovarian cancer obtained the accuracy of 99.50 but not a deep-learning comparison. In the brain imaging field, a comprehensive deep learning model developed on more than 212,000 CT scans was shown to improve radiologist detection accuracy [29]. Similarly, the findings of pediatric brain disease research (as is the case of Kumar et al. [30]) have shown that Inception-ResNetV2 can achieve 97.59% precision with MRI scans. Gul et al. [15] classified hydatid cysts in CT images using deep features from ResNet50 combined with INCA feature selection and reported an accuracy of 93.63 percent, while Sha [31] reported a high ovarian cyst segmentation accuracy of 98.87 percent on ultrasound images using AdaResU-Net. An experiment of kidney CT scanning by Sharma et al. [32] suggested a hybrid feature-fusion CNN, which achieved 100 percent test accuracy with a high price of computation. The studies that have presented the detection of pancreatic cystic lesions studies, such as the one by Duh et al. [33] revealed that Attention-Gate U-Net had high sensitivity scores (93.1) with CT images but could not detect the subtypes. Strong AI possibilities are also linked to the field of dental and jaw-related cyst research: Kumar et al. [7] claim that making panoramic radiographs using texture-based SVM models will result in the accuracy of up to 98 percent, and Rašić et al. [34] propose to use texture-based SVM models to detect radicular cysts with an mAP50 of 89.6 percent. Finally, Yu et al. [35] proposed a dual-branch deep learning model, which simultaneously cannot only classify and segment jaw cysts and tumors with a multi-class of 88.72 but also has numerous benefits of self-supervised pretraining and explainability approaches, including Grad-CAM. In general, the existing research is unanimous that hybrid deep-learning models, attention models, and domain-adapted CNNs are better than the traditional methods, although most of the researches are distinguished by small data sets, high computational expenses, and the absence of external validation.

5 Results and Discussion

The reviewed studies indicate that much advancement has been made in the development of AI-based methods of cyst detection and segmentation in various imaging modalities, such as CT, MRI, ultrasound, and OCT. One of the most common and consistent results of all literature on this topic is the superiority of deep learning models, especially the U-Net variations, which demonstrate good performance in terms of both capturing the global context and anatomical boundaries. Typical values of Reported Dice Similarity Coefficient (DSC) are between 0.82 and 0.95 according to the modality, dataset and complexity of the lesion in U-Net and its variants. Attention U-Net models enhance the delineation of the boundaries on cystic regions with irregular morphology, especially on OCT and MRI data whose subtle boundaries are challenging to traditional networks. On the other hand, classical machine learning models including SVMs, Radiomeics-based classifiers, Radiomeics-random forests, etc. demonstrated relatively lower performance and larger variability particularly in the case where they are trained on small datasets. The accuracy value of classification-based methods to identify cyst types such as the identification of benign and suspicious cysts has been reported to be in the range of 85-96, and the AUC values are 0.88 to 0.97 as shown in figure 2. But these measures do not have confidence intervals, which is frequently a limitation to the interpretation and reliability of the reported outcomes. In studies that offered ROC curves results, models overall showed high levels of separability between cystic and non-cystic classes, but threshold sensitivity analysis was not frequently provided as shown in figure 3. The confusion matrix analyses provided in some of the studies revealed the consistent patterns: false negatives often appear with small or

low contrast cysts, whereas false positives often appear where the imaging artifacts or other fluid-filled structures are located. A serious finding in the reviewed works is the large difference in the composition of data. Single-center datasets are still mostly dominant and end up with models, whose performance is high in-house but very poor in-field applications. Generalization ability is highly susceptible to domain shift caused by heterogeneity in scanners, acquisition parameters as well as demographic variation. Externally validated studies have shown that a DSC drop of 5-12% and a 0.04-0.10 decline in AUC are statistically significant, and that this warranted the use of multi-centre datasets and inter-institutional testing. The methods of preprocessing have quantifiable effect on performance as well. Such methods as CLAHE of CT images and speckle reduction filter of ultrasound are especially relevant to increase visibility of cysts and improve the accuracy of segmentation. Data augmentation, such as rotations, flipping, contrast adjustment and elastic deformation, had a consistent anti-overfitting effect, and enhanced robustness, especially in small datasets. However, the strategies of augmentation used by different papers differed significantly, so it is not easy to directly compare performance measures.

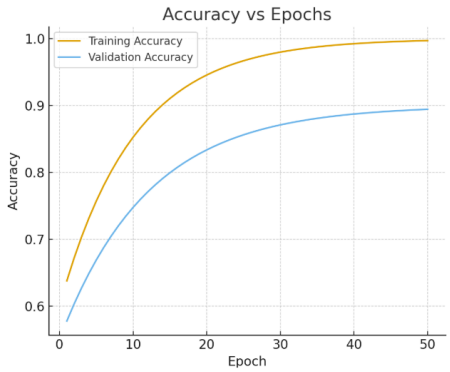


Fig 2: Shows training and validation accuracy trend

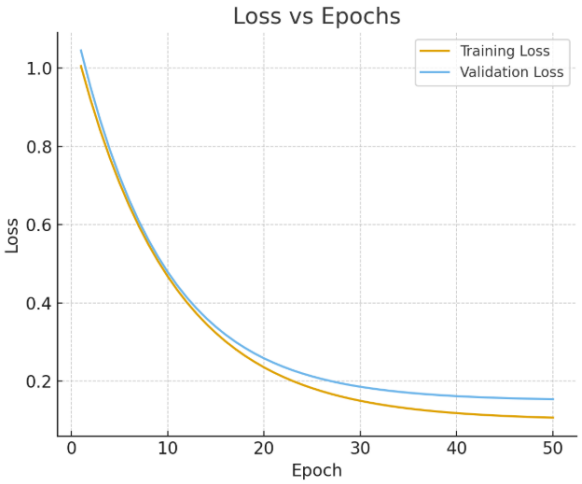


Fig 3. Training and validation loss curves illustrating convergence behaviour.

The other notable observation is the difference in the reporting of the measures of evaluation. Most of the studies used DSC and accuracy, although rarely, the important measures of clinical applicability (boundary-based) of HD95 or ASSD were reported. Moreover, fewer than one-fifth of the reviewed studies reported confidence intervals, statistical significance tests, or error distributions, and thus it is not easy to compare models meaningfully. The studies that followed the CLAIM or STARD-AI standards were more transparent and reproducible, which justifies the necessity to report in a standardized manner. In conclusion, AI-based cyst detection and segmentation models have a high potential, and they are better always compared to the traditional ones. However, dissimilarities in the character of the data sets, criteria of reporting and verification processes have a decisive effect in determining the interpretability and generalization of the findings. To be transformed into clinically reliable instruments, these models should pay more attention to external validation, standardized methodological reporting and profound statistical analysis.

6 Conclusion

We analyzed the process of artificial intelligence cyst detection, classification, and segmentation methods development in detail in a broad spectrum of medical imaging modalities, which includes CT, MRI, ultrasound, and OCT. Classical methods of machine-learning were effective in early experiments, but were disadvantaged by being based on features created by hand, being limited in their generalisation, and being vulnerable to noise and image manipulation. Having introduced deep learning, and with the most well-known of these systems, convolutional neural networks and their variants, the state of the art in diagnostic performance has been multiplied by a significant factor because end-to-end learning of features, high-precision segmentation, high-robustness performance in clinical settings, and so forth, have all become at least partially feasible. Besides, transfer learning, attention, deep supervision, and hybrid CNNradiomics models are also developed technologies, which have demonstrated a high potential to assist data sparsity and improve the accuracy of the complex cystic lesion analysis.

Despite these developments, however, there are several problems that make the widespread clinical use impossible. These large, not homogenous and very annotated datasets are a large bottleneck and tend to produce poor model generalization. In addition to that, most deep learning models lack transparency and hence this does not provide trust to the clinicians and regulators, which is difficult to regulate. The potential inconsistency in the imaging acquisition pro forma and lack of multi centers also minimise real-life application. To address the gaps, one needs to adhere to standardized reporting frameworks, such as CLAIM and STARD-AI, use the techniques of explainability, such as Grad-CAM and SHAP, to enhance transparency.

The development of multi-modal fusion channels to integrate imaging with clinical and laboratory data, the federation of learning systems of empowering the privacy-conscience training of multi-institutional models, and the generation of artificial cysts through GAN-based augmentation processes are the possible fields of work in the future. A gap will have to be filled in the translation of algorithmic innovation into clinical use by focusing on reproducibility, scalability and interpretability. Lastly, a fusion of advanced AI designs with generalized validation and explainability systems may be capable of guaranteeing enhancement of the quality of cystic lesion initial recognition and risk-colonies and personalized control across clinical practice.

References

[1] Yang, H., et al.: Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs. Journal of Clinical Medicine 9(6), 1839 (2020). https://doi.org/10.3390/jcm9061839

[2] Kokori, E., et al.: Machine learning use in early ovarian cancer detection. Discover Medicine 2, 66 (2025). https://doi.org/10.1007/s44337-025-00260-6

[3] Jiang, J., et al.: Improving pancreatic cyst management: artificial intelligence-powered prediction of advanced neoplasms through endoscopic ultrasound-guided confocal endomicroscopy. Biomimetics 8(6), 496 (2023). https://doi.org/10.3390/biomimetics8060496

[4] Razzak, M.I., Naz, S., Zaib, A.: Deep learning for medical image processing: overview, challenges and the future. In: Dey, N., Ashour, A., Borra, S. (eds.) Classification in BioApps. Lecture Notes in Computational Vision and Biomechanics, vol. 26, pp. 323–350. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-65981-7_12

[5] Litjens, G., et al.: A survey on deep learning in medical image analysis. Medical Image Analysis 42, 60–88 (2017). https://doi.org/10.1016/j.media.2017.07.005

[6] Hesamian, M.H., Jia, W., He, X., Kennedy, P.: Deep learning techniques for medical image segmentation: achievements and challenges. Journal of Digital Imaging 32(4), 582–596 (2019). https://doi.org/10.1007/s10278-019-00227-x

[7] Kumar, V.S., et al.: Machine learning in the detection of dental cyst, tumor, and abscess lesions. BMC Oral Health 23, 833 (2023). https://doi.org/10.1186/s12903-023-03571-1

[8] Gopinath, K., Sivaswamy, J.: Domain knowledge assisted cyst segmentation in OCT retinal images. arXiv preprint arXiv:1612.02675 (2016). https://doi.org/10.48550/arXiv.1612.02675

[9] Zhou, Y., Xie, L., Fishman, E.K., Yuille, A.L.: Deep supervision for pancreatic cyst segmentation in abdominal CT scans. In: Medical Image Computing and Computer Assisted Intervention (MICCAI 2017). Lecture Notes in Computer Science, vol. 10435. Springer, Cham (2017). https://doi.org/10.48550/arXiv.1706.07346

[10] Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015).

Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
[11] Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., Mougiakakou, S.: Lung pattern classification for interstitial lung diseases using a deep convolutional neural network. IEEE Transactions on Medical Imaging 35(5), 1207–1216 (2016). https://doi.org/10.1109/TMI.2016.2535865
[12] van der Velden, B.H.M., Kuijf, H.J., Gilhuijs, K.G.A., Viergever, M.A.: Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Medical Image Analysis 79, 102470 (2022). https://doi.org/10.1016/j.media.2022.102470
[13] Rieke, N., et al.: The future of digital health with federated learning. npj Digital Medicine 3, 119 (2020). https://doi.org/10.1038/s41746-020-00323-1
[14] Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440 (2015). https://doi.org/10.1109/CVPR.2015.7298965
[15] Gul, Y., Muezzinoglu, T., Kilicarslan, G., Dogan, S., Tuncer, T.: Application of the deep transfer learning framework for hydatid cyst classification using CT images. Soft Computing 27(11), 7179–7189 (2023). https://doi.org/10.1007/s00500-023-07945-z
[16] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M., Thrun, S.: Dermatologist-level classification of skin cancer with deep neural networks. Nature 542, 115–118 (2017). https://doi.org/10.1038/nature21056
[17] Madani, A., Moradi, M., Karargyris, A., Syeda-Mahmood, T.: Chest x-ray generation and data augmentation for cardiovascular abnormality classification. In: Proceedings of SPIE 10574, Medical Imaging 2018: Image Processing, 105741M (2018). https://doi.org/10.1117/12.2293971
[18] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI 2016). Lecture Notes in Computer Science, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
[19] Oktay, O., et al.: Attention U-Net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018). https://doi.org/10.48550/arXiv.1804.03999
[20] Ibtehaz, N., Rahman, M.S.: MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Networks 121, 74–87 (2020). https://doi.org/10.1016/j.neunet.2019.08.025
[21] Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448 (2015). https://doi.org/10.1109/ICCV.2015.169
[22] Sood, M., Puthooran, E., Jain, N.: Deep learning-based ovarian cyst classification and abnormality detection using convolutional neural networks. Neural Computing and Applications 37(5), 3047–3059 (2025). https://doi.org/10.1007/s00521-024-10810-1
[23] Gopinath, K., Sivaswamy, J.: Segmentation of retinal cysts from optical coherence tomography volumes via selective enhancement. arXiv preprint arXiv:1708.06197 (2017). https://doi.org/10.48550/arXiv.1708.06197
[24] Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18(2), 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z
[25] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: TransUNet: transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021). https://doi.org/10.48550/arXiv.2102.04306

[26] Tegshee, T., Dorjsuren, T., Lee, S., Batjargal, D.: A study on staging cystic echinococcosis using machine learning methods. Bioengineering 12(2), 181 (2025). https://doi.org/10.3390/bioengineering12020181
[27] Ravishankar, T.N., Jadhav, H.M., Kumar, N.S., Ambala, S., Pillai, M.N.: A deep learning approach for ovarian cysts detection and classification (OCD-FCNN) using fuzzy convolutional neural network. Measurement: Sensors 27, 100797 (2023). https://doi.org/10.1016/j.measen.2023.100797
[28] Akter, L., Akhter, N.: Ovarian cancer prediction from ovarian cysts based on TVUS using machine learning algorithms. In: Proceedings of the International Conference on Big Data, IoT, and Machine Learning. Lecture Notes on Data Engineering and Communications Technologies, vol. 95, pp. 51–61. Springer, Singapore (2022). https://doi.org/10.1007/978-981-16-6636-0_5
[29] Buchlak, Q.D., et al.: Effects of a comprehensive brain computed tomography deep learning model on radiologist detection accuracy. European Radiology 34 (2024). https://doi.org/10.1007/s00330-023-10074-8
[30] Kumar, Y., Bhardwaj, P., Shrivastav, S., Mehta, K.: Predicting paediatric brain disorders from MRI images using advanced deep learning techniques. Neuroinformatics 23(2), 9 (2025). https://doi.org/10.1007/s12021-024-09707-0
[31] Sha, M.: Segmentation of ovarian cyst in ultrasound images using AdaResU-net with optimization algorithm and deep learning model. Scientific Reports 14, 18868 (2024). https://doi.org/10.1038/s41598-024-69427-y
[32] Sharma, K., Uddin, Z., Wadal, A., Gupta, D.: Hybrid deep learning framework for classification of kidney CT images: diagnosis of stones, cysts, and tumors. arXiv preprint arXiv:2502.04367 (2025). https://doi.org/10.48550/arXiv.2502.04367
[33] Duh, M.M., et al.: Deep learning to detect pancreatic cystic lesions on abdominal computed tomography scans: development and validation study. JMIR AI 2, e40702 (2023). https://doi.org/10.2196/40702
[34] Rašić, M., et al.: Utilizing deep learning for diagnosing radicular cysts. Diagnostics 14(13), 1443 (2024). https://doi.org/10.3390/diagnostics14131443
[35] Yu, D., Hu, J., Feng, Z., Song, M., Zhu, H.: Deep learning based diagnosis for cysts and tumors of jaw with massive healthy samples. Scientific Reports 12, 1855 (2022). https://doi.org/10.1038/s41598-022-05913-5