IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 4 April 2025, accepted 25 April 2025, date of publication 2 May 2025, date of current version 13 May 2025.

*Digital Object Identifier 10.1109/ACCESS.2025.3566627*

SURVEY

# A Comprehensive Survey on Deep Learning in Abdominal Imaging: Datasets, Techniques, and Performance Metrics

**MARIEM BELLAL[1], (Member, IEEE), SANAA EL FKIHI[1], KORHAN CENGIZ[2], (Senior Member, IEEE), AND NIKOLA IVKOVIĆ[3], (Senior Member, IEEE)**

[1]Information Retrieval and Data Analytics Laboratory, ENSIAS, Mohammed V University in Rabat, Rabat 10000, Morocco
[2]Department of Electrical Engineering, Prince Mohammad Bin Fahd University, Al Khobar 31952, Saudi Arabia
[3]Faculty of Organization and Informatics, University of Zagreb, 42000 Varaždin, Croatia

Corresponding author: Mariem Bellal (mariemebellal@gmail.com)

**ABSTRACT** Integrating Deep Learning (DL) into abdominal imaging represents a significant leap forward in diagnosing and managing abdominal conditions, offering the potential to transform conventional medical practices. This comprehensive survey explores the application of DL techniques, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Generative Adversarial Networks (GAN), across various domains of abdominal imaging, including liver, spleen, kidney, and other structures such as subcutaneous adipose tissue (SAT), muscle, viscera, and bone. It discusses the critical role of performance metrics in evaluating model efficacy and clinical applicability. Furthermore, the paper highlights emerging trends in DL, such as integrating multimodal data and exploring unsupervised and semi-supervised learning techniques, which promise to address current limitations and pave the way for future advancements. Ethical considerations, including algorithmic bias, transparency in model development, and equitable patient care, are thoroughly examined to underscore the importance of ethical practices in deploying Artificial Intelligence (AI) technologies in healthcare.

**INDEX TERMS** Deep learning in medical imaging, traumatic abdominal injuries, medical image analysis, clinical decision support system, artificial intelligence.

## I. INTRODUCTION

Traumatic abdominal injuries represent a critical concern in emergency medicine, often resulting from high-impact events such as vehicular accidents, falls, or assaults. These injuries can range from minor lacerations to severe organ damage, affecting vital structures like the liver, spleen, kidneys, and intestines. Diagnosing these injuries is challenging due to the abdominal cavity's intricate anatomy and the diverse nature of potential injuries. Rapid and precise detection is essential for timely and effective medical intervention, directly influencing survival rates and long-term patient outcomes. Traditional diagnostic approaches, such as ultrasound, X-rays, Computed Tomography (CT) scans, and Magnetic Resonance Imaging (MRI), rely heavily on expert interpretation. While effective, these methods are often limited by factors such as specialist availability, subjective variability, and time constraints. These limitations highlight the urgent need for advanced technological solutions to enhance diagnostic accuracy and efficiency in abdominal imaging [1], [2], [3].

Deep Learning (DL), a sophisticated subset of Artificial Intelligence (AI), has emerged as a transformative technology in medical image analysis, particularly for traumatic abdominal injuries. Utilizing advanced neural network architectures, such as Convolutional Neural Networks (CNNs), DL has demonstrated exceptional capabilities in processing and interpreting complex imaging data. These algorithms excel at detecting subtle patterns and anomalies in medical images, offering accuracy and speed that can surpass the diagnostic

The associate editor coordinating the review of this manuscript and approving it for publication was Carmelo Militello.

capabilities of experienced radiologists. The application of DL in abdominal imaging promises to improve diagnostic accuracy, expedite the detection process, and provide objective, consistent analysis. These advancements are particularly valuable in high-pressure settings such as emergency departments, where rapid and accurate decision-making is crucial. Furthermore, DL models benefit from continuous learning, improving their diagnostic prowess as they are exposed to more data, potentially leading to breakthroughs in identifying novel diagnostic markers [4], [5], [6].

The efficacy of DL in medical imaging [7], [8], [9], mainly for traumatic abdominal injuries, hinges significantly on the quality and diversity of datasets used for training and validation. These datasets typically comprise a range of imaging modalities, including CT scans, MRIs, and X-rays, annotated with detailed information about the type and severity of injuries. The challenges in compiling these datasets are manifold: they must be sufficiently large and diverse to train robust models, accurately annotated to ensure reliability, and ethically sourced concerning patient privacy and consent. Additionally, the datasets must represent various injury types and severities to ensure the trained models are generalizable and effective across different clinical scenarios. The preprocessing of these images, such as normalization, augmentation, and segmentation, is pivotal in enhancing model performance [10], [11], [12].

To this end, DL technologies have marked a pivotal shift in medical imaging, offering unprecedented capabilities for analyzing, diagnosing, and managing a wide range of conditions. Particularly in abdominal imaging, the complexity and variability of structures within the abdominal cavity present significant challenges for traditional diagnostic methods. This survey aims to provide a comprehensive exploration of the integration of DL techniques in abdominal imaging, focusing on three key areas: datasets and data handling requirements for robust model development, advanced DL techniques, and the performance metrics used to evaluate model efficacy. By synthesizing findings from existing literature, we seek to highlight the advancements, challenges, and future directions in this rapidly evolving field, particularly in diagnosing and managing conditions affecting vital organs such as the liver, spleen, and kidneys. Specifically, this review has the following objectives:

**Analyze the application and implementation** of DL techniques across different areas of abdominal imaging.

**Synthesize findings from existing literature** to provide a cohesive understanding of the advancements, efficacy, and challenges faced in leveraging DL for abdominal imaging.

**Evaluate the impact of diverse datasets** for training and validating DL models in abdominal imaging. Emphasis is placed on the significance of dataset quality, annotation accuracy, and ethical considerations in dataset compilation.

**Identify existing challenges and ethical considerations**, reflecting on the hurdles in integrating DL technologies into clinical practice, including data privacy, algorithmic bias, and interdisciplinary collaboration.

**Explore emerging trends and suggest future research directions**, aiming to highlight novel approaches and technologies that could further enhance DL's capability in abdominal imaging.

This review aims to provide a detailed and forward-looking perspective on the intersection of DL and abdominal imaging. It highlights its transformative potential in improving diagnostic processes and patient outcomes across specific domains such as liver, spleen, and kidney imaging. By mapping out the current landscape and identifying avenues for future research, this survey seeks to contribute significantly to the field, guiding researchers, clinicians, and policymakers in harnessing AI's full potential in enhancing healthcare delivery.

The remainder of this paper is organized as follows: **Section II** presents the methodology, detailing the systematic approach used to select and analyze the relevant literature. **Section III** offers a comprehensive literature review, exploring the application of DL techniques in abdominal imaging and discussing key findings from current research. **Section IV** focuses on emerging trends and future directions, highlighting innovative DL applications poised to transform abdominal imaging and addressing the ethical considerations surrounding AI in medical imaging. It emphasizes the importance of tackling ethical challenges to ensure the responsible deployment of these technologies. **Section V** summarizes the principal insights from the review and discusses the implications for future research and clinical practice in abdominal imaging.

## II. METHODOLOGY

A meticulous and multi-faceted search strategy was employed to ensure a comprehensive and systematic review of the literature about DL applications in detecting and classifying traumatic abdominal injuries. The rationale for the selection of databases and the methodology adopted in conducting this review are detailed as follows (Figure 1):

**Database Selection Rationale:** The literature search spanned several major scientific databases, including Scopus, Web of Science, PubMed, IEEE Xplore, Google Scholar, ScienceDirect, and the Cochrane Library.

**Keywords and Search Strategy:** A combination of specific keywords and phrases, including "Deep Learning," "traumatic abdominal injuries," "medical imaging," "CNN," "abdominal trauma," "CT imaging in trauma," and "AI in radiology" was used. Boolean operators (AND, OR) facilitated a refined search strategy, capturing the most relevant studies without overwhelming the results with non-pertinent information. The formulation of keyword combinations was iteratively refined to balance comprehensiveness with specificity in search results.

**Inclusion and Exclusion Criteria:** Studies published in English from December 2012 to the present were included to focus on the latest advancements. This timeframe was chosen to reflect the rapid developments in DL technology and its application in medical imaging. Exclusions were made for non-peer-reviewed articles, studies outside the scope of traumatic abdominal injuries, and research where DL was not the primary focus. For example, studies focusing solely on traditional imaging techniques without AI integration were excluded.

**Cross-Reference Check Process:** The reference lists of retrieved articles were meticulously examined to uncover additional studies that might have been overlooked in the initial search. This process involved reviewing citations to two levels, ensuring a comprehensive capture of relevant literature. Priority was given to articles published in leading journals and conferences in medical imaging and AI.

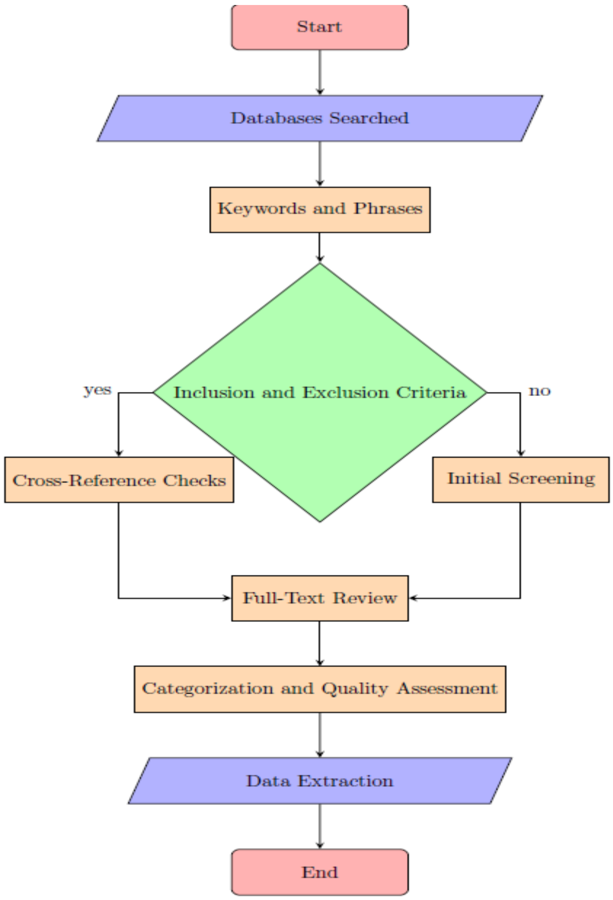


**FIGURE 1.** Methodology flowchart.

**Selection Process Description:** Titles and abstracts were screened initially, filtering out studies that did not directly address the application of DL in traumatic abdominal injury imaging. For instance, of the 152 articles initially screened, approximately 100 were selected for full-text review based on their relevance to the review's objectives. Articles passing the initial screening underwent a full-text review, with around 76 ultimately selected for inclusion based on detailed evaluation criteria.

**Quality Assessment Criteria:** The quality of each study was assessed based on criteria such as study design, sample size, statistical analysis, and clarity of reported results. High-quality studies typically demonstrate robust experimental design, clear definition of research objectives, comprehensive analysis, and well-articulated conclusions. Conversely, studies with small sample sizes, unclear methodologies, or vague outcomes were considered of lower quality.

**Data Extraction:** Relevant data from each selected study, including study objectives, DL techniques used, datasets, results, and conclusions, were systematically extracted using a standardized data extraction template. Mendeley and Microsoft Excel facilitated the organization and management of extracted data, ensuring accuracy and efficiency in the review process.

**Temporal Distribution:** To understand the evolution of research in this field, the temporal distribution of publications was analyzed. Figure 2 illustrates the number of publications per year from 2012 to the present, highlighting key trends and milestones in the application of DL to traumatic abdominal injuries. The analysis reveals a gradual increase in publications from 2012 to 2017, followed by rapid growth from 2018 to 2023, reflecting the maturation of the field and the integration of DL into clinical practice.

**Methodological Limitations:** The review's methodological limitations include potential biases in article selection and the exclusion of non-English studies, which might omit relevant research conducted in other languages. Despite efforts to conduct a comprehensive search, some relevant studies might have been missed due to the selection criteria or database coverage limitations.

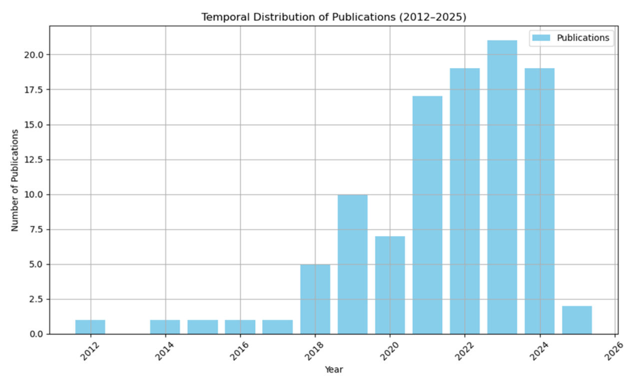


**FIGURE 2.** Temporal distribution of publications in deep learning for abdominal imaging (2012–2025).

Through this rigorous methodology, the review aims to provide a thorough and unbiased overview of the current state of DL in the diagnosis and management of traumatic abdominal injuries, highlighting both its achievements and areas needing further exploration.

## III. LITERATURE REVIEW

The advent of DL in the realm of abdominal imaging marks a pivotal shift towards precision medicine, offering groundbreaking approaches to diagnosing, managing, and treating abdominal pathologies. This section delves into the intricacies of datasets, the multifaceted landscape of DL applications, and the critical metrics for model evaluation, underscoring the innovative strides and challenges encountered in the field. Through an exploration of DL techniques, we uncover the transformative potential of neural networks in enhancing image analysis, from segmentation to anomaly detection, and the pivotal role of high-quality, diverse datasets in training models that accurately reflect clinical realities. Furthermore, we navigate the complexities of evaluating these models, emphasizing the need for a comprehensive suite of metrics that capture not just accuracy but also the precision, reliability, and clinical applicability of DL algorithms. This literature review aims to provide a solid understanding of the current state, significant achievements, and future directions of DL in abdominal imaging, setting the stage for continued innovation and its translation into enhanced patient care.

### A. DATASETS AND DATA HANDLING IN ABDOMINAL IMAGING

The effectiveness of DL in abdominal imaging heavily relies on the quality and diversity of the datasets used. These datasets are essential for training and validating the models that detect and classify traumatic abdominal injuries. Commonly used datasets in this field exhibit unique characteristics and inherent limitations that impact the performance of DL models [13], [14], [15].

In abdominal imaging, datasets generally consist of various modalities like CT scans, MRI, and X-rays, each offering different perspectives and details. Renowned datasets such as The Cancer Imaging Archive (TCIA), which provides a wide array of imaging data, including abdominal CT scans, are frequently utilized. Another significant source is the Medical Image Computing and Computer-Assisted Intervention (MICCAI) Society datasets, known for hosting challenges with accompanying datasets that often include abdominal images. These datasets are meticulously annotated, detailing the type and severity of injuries, which is crucial for training accurate models. However, challenges such as limited data size, variability in image quality, and representation of diverse injury types are common. This limited diversity can affect the generalizability of the models trained on these datasets [16], [17], [18].

Data preprocessing and augmentation are critical steps in preparing these datasets for use in DL models. Preprocessing involves techniques such as normalization, where the intensity values of the images are scaled to a standard range, and segmentation, where specific regions of interest, like organs or lesions, are isolated for more focused analysis. On the other hand, data augmentation artificially expands the dataset by applying various transformations to the images, such as rotation, scaling, and flipping. These techniques help in combating overfitting and improve the robustness of the models by providing a more comprehensive representation of potential variations in real-world data [19], [20], [21].

In abdominal imaging, the preprocessing and augmentation steps are tailored to address specific challenges. For instance, in CT scans, preprocessing might focus on enhancing contrast between different tissues, while augmentation practices might simulate common variations in injury presentations.

Recent advancements in dataset compilation and annotation have significantly contributed to the field. For example, Li et al. [22] 's comprehensive study systematically compiles a vast array of medical image datasets tailored for DL applications, covering a span from 2013 to 2020. The paper categorizes these datasets into four principal areas: head and neck, chest and abdomen, pathology, and blood, alongside other miscellaneous categories. This resource addresses the critical challenge of data scarcity in medical image analysis by providing an extensive reference for researchers to easily locate relevant datasets for clinical image analysis, method testing, and evaluation.

Similarly, Park et al. [23] present a comprehensive approach to annotate abdominal CT images from individuals without pancreatic disease, aiming to develop DL algorithms for the automatic recognition of a normal pancreas. Utilizing dual-phase contrast enhanced CT scans from potential kidney donors, the team manually annotated 22 structures per dataset, later confirmed by expert radiologists. A total of 1150 datasets from 575 subjects were annotated, leading to the development of a reliable annotation process for DL applications, demonstrating high fidelity in deep network predictions for multi-organ segmentation.

In the area of abdominal organ and lesion segmentation, several studies have introduced innovative methods and datasets.

Luo et al. [24] introduced the Whole Abdominal Organ Dataset (WORD), a large-scale, finely annotated dataset to advance DL-based medical image segmentation for abdominal organ analysis from CT images. This dataset, comprising 150 abdominal CT volumes with 16 organs annotated per volume, addresses the critical need for comprehensive training data in the field. The study evaluates several state-of-the-art segmentation methods on WORD, highlighting the potential to significantly reduce manual annotation efforts and improve segmentation accuracy, especially for challenging and complex organs.

Sadikine et al. [25] introduce a novel DL-based abdominal image segmentation approach incorporating semi-overcomplete shape priors through a convolutional auto-encoder (S-OCAE). Their method significantly improves the delineation of both abdominal organs and vessels in CT and MR images by leveraging undercomplete and overcomplete representations. The study demonstrates that this approach outperforms traditional U-Net and U-Net+CAE models regarding the Dice Similarity Coefficient, Absolute Volume Difference, Average Symmetric Surface Distance, and Hausdorff Distance.

The LiverHccSeg dataset, introduced by Gross et al. [26] is a publicly available multiphasic MRI dataset designed for liver and hepatocellular carcinoma (HCC) tumor segmentation. It features manual segmentations by two board-approved abdominal radiologists. The dataset aims to address the need for reliable data in liver and tumor segmentation, facilitating external validation and benchmarking of segmentation algorithms. With 17 cases for liver and 14 for HCC tumor segmentation, the dataset demonstrates high segmentation agreement for liver and variable agreement for tumors, offering a valuable resource for developing and testing machine learning algorithms in HCC patient care.

Addressing class imbalance and rare conditions remains a critical challenge in the field, with several innovative approaches emerging to tackle these issues.

Rezaei et al. [27] introduce the Ensemble-GAN framework, aimed at addressing the class imbalance issue in the semantic segmentation of abdominal CT and MR images. By utilizing a single generator with an ensemble of discriminators, the framework shows significant improvement in segmentation accuracy, particularly for organs and tumor regions within highly imbalanced datasets. The performance of Ensemble-GAN is evaluated on the CHAOS 2019 and LiTS 2017 datasets, demonstrating exceptional precision in the semantic segmentation of the spleen, liver, and both healthy left and right kidneys, with F1 scores reaching 0.93, 0.96, 0.90, and 0.94, in that order. Furthermore, in the simultaneous segmentation of liver lesions and the liver itself, the collective F1 scores were notably impressive, recorded at 0.83 and 0.94. Building on this, the same team, Rezaei et al. [28], further expanded their research to develop a novel Bayesian deep ensemble learning framework designed to refine the learning process for long-tail and out-of-distribution samples in medical imaging. By incorporating uncertainty estimation into its methodology, this framework adeptly navigates imbalanced datasets, enhancing classifier generalizability. Tested across a variety of medical imaging datasets, this approach outshines existing methods, particularly in recall and overall performance for semantic segmentation tasks. This advancement underscores the pivotal role of uncertainty estimation in improving the management of rare classes and boosting generalization to unseen conditions. Notably, in the semantic segmentation of high-resolution CT and MRI images, the framework's recall rates marked a significant improvement over previous methods, highlighting its efficacy in addressing the challenges presented by complex imaging datasets.

Specialized applications in abdominal imaging have also seen significant progress.

Baldazzi et al. [29] present an annotated dataset of real and synthetic signals for benchmarking post-processing techniques in non-invasive fetal electrocardiography (fECG). The dataset comprises 21 dual-channel, 15-second fECG signals from real patients and 40 synthetic signals, aiming to improve signal-to-noise ratio (SNR) enhancement methods. This unique dataset facilitates research in fECG signal processing, offering a valuable resource for developing and testing algorithms aimed at detecting fetal QRS complexes, crucial for fetal health monitoring.

Płotka et al. [30] investigate the efficacy of a DL approach in estimating fetal weight from ultrasound video scans of the fetal abdomen and gestational age, aiming to achieve accuracy comparable to that of traditional biometry-based estimations by clinical experts. The methodology involved training a DL model on a dataset of 800 retrospective ultrasound video scans from 700 pregnant women at gestational ages ranging from 15 6/7 to 41 0/7 weeks. The model was then tested on a prospective dataset of 100 scans from different pregnant women between 16 2/7 and 38 0/7 weeks of gestation. The results showed no systematic deviations between the DL model and human readers in the Bland-Altman analysis, with a mean absolute percentage error of $3.75\% \pm 2.00\%$ across all readers, and $2.59\% \pm 1.11\%$ when excluding junior readers. The intraclass correlation coefficients indicated excellent reliability, ranging between 0.9761 and 0.9865.

Finally, Tong et al. [31] present a robust and efficient method for abdominal CT segmentation using a two-stage approach: coarse localization and fine segmentation with shape constraints. This method, developed to address the challenges posed by varying organ shapes and intensity distributions in multi-center, multi-phase CT images, leverages a multi-scale attention network. It was evaluated using the FLARE challenge dataset, achieving high Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD) scores respectively 83.7% and 64.4%, indicating promising performance in robustness and efficiency.

Hansen et al. [32] evaluates the accuracy of three mortality risk scores for ruptured abdominal aortic aneurysms (rAAA) using an independent dataset from a community hospital. It includes a retrospective review of 38 patients, validating the Dutch Aneurysm Score, Harborview Medical Center score, and Vascular Surgery Group of New England score. The analysis found all three scores accurately predicted 30-day mortality, with the VSGNE score showing the highest accuracy. However, statistical significance in accuracy differences among the scores was not achieved. This suggests that each scoring system could be useful in predicting early mortality following rAAA repair, underscoring their potential in clinical decision-making and patient counseling.

To this end, the advancement of DL in abdominal imaging is intricately linked to the quality, diversity, and handling of datasets used to train and validate models. These models, crucial for detecting and classifying traumatic abdominal injuries, derive their effectiveness from datasets enriched with a variety of imaging modalities such as CT scans, MRI, and X-rays. Each modality brings forth unique perspectives, enriching the dataset and thereby the models trained on them. Renowned repositories like The Cancer Imaging Archive (TCIA) and datasets provided by the Medical Image Computing and Computer-Assisted Intervention (MICCAI) Society, with their comprehensive annotations, have been instrumental. However, the challenges of limited dataset sizes, variability in image quality, and the representation of diverse injury types pose significant hurdles, affecting model generalizability.

Data preprocessing and augmentation emerge as pivotal steps in making these datasets more conducive for DL applications. Techniques such as normalization and segmentation prepare the data by enhancing image quality and isolating regions of interest, respectively. Augmentation tactics like image rotation, scaling, and flipping are employed to simulate a broader spectrum of real-world variations, mitigating overfitting and bolstering model robustness. This tailored approach to dataset preparation ensures that DL models are well-equipped to replicate clinical scenarios with higher fidelity.

The meticulous compilation by Li et al. [22] the introduction of the Whole Abdominal Organ Dataset (WORD) by Luo et al. [24], and other notable contributions outlined in Table 1 provides a comparative analysis of how different studies handle data preprocessing, dataset challenges, and their impact on deep learning applications in abdominal imaging. It evaluates studies based on their methodology, key findings, and broader implications for improving AI in healthcare.

Furthermore, assembly of large and diverse datasets for abdominal imaging is fraught with challenges, pivotal among which are data privacy, anonymization, and ethical considerations. The intrinsic sensitivity of patient data necessitates stringent protocols to ensure privacy and confidentiality. While crucial for protecting patient identity, anonymization techniques must be carefully implemented to prevent the loss of clinically relevant information that could be vital for model training and validation. Ethical implications extend to ensuring equitable data representation, avoiding bias, and maintaining transparency in data usage, particularly in how datasets are collected, processed, and shared within the research community [34].

Moreover, the heterogeneity of abdominal imaging data, stemming from varied imaging modalities (CT, MRI, Ultrasound) and diverse patient demographics, adds another layer of complexity. Ensuring that datasets are large and representative of the wide array of pathologies encountered in abdominal imaging requires concerted efforts in data collection and curation. While a challenge, this heterogeneity is also an opportunity to develop robust models capable of generalizing across different conditions and patient populations.

Data preprocessing is a critical step in preparing imaging data for DL applications. In abdominal imaging, preprocessing techniques such as normalization, where pixel or voxel intensities are scaled to a common range, and segmentation, which involves isolating specific regions of interest (e.g., liver, spleen, kidneys), are fundamental.

These steps help enhance input data quality, making it more suitable for model training by reducing variability and focusing the model's attention on relevant features. Advanced preprocessing techniques, such as elastic deformations to simulate physiological movements or the simulation of varying contrast agent phases in CT or MRI scans, are particularly beneficial in abdominal imaging.

Furthermore, Data augmentation is an indispensable technique for addressing limited or imbalanced datasets in abdominal imaging.

By artificially enlarging the training dataset through transformations like rotation, flipping, scaling, or adding noise, augmentation can enhance the model's ability to generalize from the training data to unseen data. Moreover, innovative augmentation techniques such as GAN have been utilized to generate synthetic but realistic abdominal imaging data, which can help overcome the scarcity of certain pathologies in available datasets [15], [17].

Advanced preprocessing and augmentation techniques specific to abdominal imaging include elastic deformations, which simulate physiological movements, and the simulation of varying contrast agent phases in CT or MRI scans. These methods are particularly beneficial in abdominal imaging, where understanding and recognizing the variability in organ appearance under different conditions is crucial.

The constraints of limited or imbalanced datasets in abdominal imaging are significant, with certain conditions ns or pathologies being underrepresented. Techniques such as transfer learning, where a model pre-trained on a large dataset is fine-tuned on a smaller, specific dataset, have shown promise in leveraging external data sources to improve model performance. Similarly, few-shot learning approaches, which aim to train models effectively with limited examples, are gaining traction. As the field advances, the continuous refinement of data curation, preprocessing, and augmentation processes is essential to fully harnessing DL's potential in medical imaging. For instance, G-SET-DCL [35] introduces a guided sequential episodic training framework with dual contrastive learning to enhance colon segmentation under data scarcity. The model achieves high segmentation accuracy with minimal annotated data by leveraging anatomical continuity across consecutive CT slices and incorporating contrastive learning for feature discrimination. Innovative preprocessing and augmentation. Methods specific to abdominal imaging, such as elastic deformations and simulation of varying contrast agent phases, further enhance model training under the constraints of limited or imbalanced datasets [36], [37].

Furthermore, the availability of varied and extensive datasets is crucial for developing and validating DL models in abdominal imaging. Table 2 offers an overview of widely used medical imaging datasets relevant to abdominal imaging. Unlike Table 1 which focuses on how studies utilize and preprocess data, Table 2 categorizes datasets based on their characteristics, sources, modalities, and annotation types.

It is essential to acknowledge the inherent limitations of currently available datasets to enhance the robustness and generalizability of deep learning models in abdominal imaging. While datasets such as The Cancer Imaging Archive

**TABLE 1. Comparative analysis of data handling in abdominal imaging studies.**

| Paper | Study Focus | Dataset(s) Used | Key Findings | Implications |
|---|---|---|---|---|
| Li Et Al. [22] | Compilation of medical image datasets | Various (2013-2020) | Provides an extensive reference for locating relevant datasets | Facilitates AI advancements in healthcare by addressing data scarcity |
| Luo Et Al. [24] | Introduction of WORD for organ segmentation | Whole abdominal ORgan Dataset (WORD) | Sets a new benchmark for abdominal organ segmentation accuracy | Reduces manual annotation efforts, improves clinical application |
| Sadikine Et Al.[25] | Improving abdominal image segmentation | Six datasets (MSD, LiTS, KiTS19, CHAOS, 3D-IRCADb, DRIVE) | Outperforms traditional models in segmentation accuracy | Enhances diagnostic accuracy, beneficial for treatment planning |
| Bidaut [33] | Overview of data/image processing in abdominal imaging | Not specified | Highlights the evolution towards advanced imaging techniques | Enhances diagnostic accuracy and treatment planning |
| Gross Et Al. [26] | Liver and HCC tumor segmentation | LiverHccSeg dataset | Shows high segmentation agreement for liver and variable for tumors | Supports the development/testing of ML algorithms in HCC care |
| Baldazzi Et Al.[29] | Annotated dataset for non-invasive fECG | Real and synthetic fECG signals | Improves SNR enhancement methods for fetal health monitoring | Facilitates research in fetal QRS complex detection |
| Rezaei Et Al. [27] | Ensemble-GAN for addressing class imbalance in segmentation | CHAOS 2019, LiTS 2017 | Significant improvement in segmentation accuracy for imbalanced datasets | Demonstrates precision in organ and tumor segmentation |
| Rezaei Et Al. [28] | Bayesian deep ensemble learning framework | Various medical imaging datasets | Improves recall and overall performance in semantic segmentation | Enhances model generalizability, especially for rare classes |
| Płotka Et Al.[30] | DL approach for estimating fetal weight | Ultrasound video scans | Achieves accuracy comparable to traditional biometry by experts | Suggests the efficacy of DL in fetal weight estimation from ultrasound |
| Tong Et Al. [31] | Two-stage approach for abdominal CT segmentation | FLARE challenge dataset | Achieves high DSC and NSD scores, indicating robustness | Enhances automatic abdominal multi-organ segmentation |
| Park Et Al. [23] | Annotation of abdominal CT images for pancreas recognition | Dual-phase contrast enhanced CT scans | Develops a reliable annotation process for DL applications | Demonstrates high fidelity in deep network predictions for segmentation |
| Hansen Et Al. [32] | Evaluating mortality risk scores for rAAA | Independent dataset from a community hospital | Validates the accuracy of three mortality risk scores | Suggests the utility of scoring systems in clinical decision-making for rAAA |

(TCIA), CHAOS, LiTS, and WORD provide valuable imaging data, they often suffer from challenges such as limited sample diversity, annotation inconsistencies, and dataset biases. Many of these datasets predominantly feature imaging data from specific demographic groups, leading to algorithmic biases that may impact performance when applied to broader patient populations. Additionally, the variability in imaging protocols, scanner types, and resolution across datasets can introduce discrepancies, making it difficult for models trained on one dataset to generalize well to others. Furthermore, class imbalance, where certain pathologies or anatomical structures are underrepresented, presents a significant challenge in model training and validation. This imbalance can lead to biased predictions and decreased sensitivity for rare conditions, necessitating advanced techniques such as transfer learning, few-shot learning, and generative augmentation methods to compensate for these deficiencies. Addressing these limitations requires a concerted effort to develop more diverse, standardized, and representative datasets and implement robust validation strategies to ensure reliable AI-assisted diagnostics across varying clinical settings

### B. DL TECHNIQUES IN ABDOMINAL IMAGING

DL, a transformative branch of AI, has significantly impacted various sectors, most notably in medical imaging. This field involves using neural networks with multiple layers, enabling these models to learn and extract complex patterns from data in a hierarchical manner. Central to DL in image analysis are CNN which are particularly adept at processing pixel data and learning spatial hierarchies of features. Other techniques, like RNN and Autoencoders, play a role in sequential image analysis and unsupervised learning tasks such as dimensionality reduction. Additionally, transfer learning, which involves applying a pre-trained model to new problems, is increasingly popular in medical imaging due to often limited specialized datasets. Training these sophisticated models requires large datasets and substantial computational resources, with validation on separate datasets being crucial to avoid overfitting and ensure generalization to new, unseen data [13], [14], [54].

The integration of DL into abdominal imaging marks a significant leap forward in diagnosing and managing abdominal injuries and conditions. In abdominal imaging, CNNs are primarily employed to classify scans into categories and detect specific features or anomalies, such as organ injuries.

A critical application of DL is in image segmentation, where CNNs identify and delineate specific regions within an image, such as organs or tumors, which is vital for assessing the extent of an injury or disease. These models are also adept at anomaly detection, recognizing deviations from normal anatomy crucial for diagnosing traumatic injuries [55], [56], [57].

Moreover, DL extends its utility to enhancing the quality of medical images. Techniques for improving resolution and reducing noise contribute to clearer, more diagnostic-quality images. In some advanced instances, DL models are employed for predictive analytics, like forecasting disease progression or treatment responses, showcasing the versatility of these technologies. However, applying DL in abdominal imaging is not without challenges. Variability in datasets, the subtlety of certain injuries, and the imperative for high accuracy and reliability in clinical settings are ongoing concerns.

To address these, adaptations often involve customizing network architectures and fine-tuning models with specific medical datasets, integrating clinical insights into the development process. These efforts ensure that DL tools are not just technologically advanced but also clinically relevant and reliable [58], [59], [60].

Several studies have explored novel deep learning techniques and compared their effectiveness in abdominal imaging. These investigations highlight the evolution of DL methods, such as CNNs, GANs, and autoencoders, and their impact on tasks like classification, segmentation, and image enhancement. Comparative analyses demonstrate the strengths and limitations of different approaches, guiding the development of more robust and clinically applicable DL solutions.

Zhang and Qie [61] present a comprehensive review of the application of DL in medical imaging, discussing the strides DL has made in this field. They focus on DL techniques like CNNs, RNNs, and GANs, exploring their impact on disease diagnosis and treatment. The review highlights the future potentials and challenges of DL in medical imaging, emphasizing its transformative role. However, the article could benefit from deeper exploration into the ethical aspects of AI in medical imaging and more extensive real-world clinical applications to demonstrate the practical efficacy of these technologies.

Similarly, Litjens et al. [14] provide an in-depth survey of DL applications in medical image analysis, exploring the evolution and impact of DL methods like CNNs, RNNs, and GANs in this field. They cover various medical imaging tasks, including classification, detection, and segmentation across different anatomical areas. While the survey is comprehensive, it could benefit from a more detailed discussion on the practical challenges of implementing these technologies in clinical settings, such as integration with existing medical systems and data privacy concerns.

In contrast, Rehman and Khan [62] provide a detailed review of DL techniques for classifying and segmenting abdominal images, highlighting the significant progress and challenges in computer-aided diagnosis through imaging of organs like the liver, kidney, pancreas, and stomach. It covers advancements in imaging acquisition, preprocessing, and feature extraction and evaluates various DL models and datasets. The paper emphasizes the need for improved imaging devices, preprocessing algorithms, publicly available datasets, and the exploration of recent CNN architectures for future research in abdominal imaging.

Saleh et al. [63] explore a DL approach for automated morphometric analysis and weight estimation of Black Tiger Prawns from digital images. The novel approach involves a Kronecker product-based feature extraction module and a landmark localization module to predict key morphological points on the prawn body. The study evaluates this method on a large dataset of prawn images, demonstrating its superior performance in accuracy, robustness, and efficiency compared to existing methods. This research represents a significant advancement in aquacultural engineering by providing a more efficient and accurate method for prawn morphometrics and weight estimation.

DL has revolutionized image reconstruction techniques, significantly improved diagnostic quality, and reduced noise. For example, Jensen et al. [64] explore the efficacy of Deep Learning Image Reconstruction (DLIR) in enhancing the image quality of contrast-enhanced oncologic CT scans of the abdomen. The research, involving a quantitative and qualitative assessment, compares DLIR against 30% Adaptive Statistical Iterative Reconstruction V (ASIR-V) across various metrics, including attenuation, noise measurements, and lesion conspicuity. Results indicate that DLIR, especially at high strength, significantly improves contrast-to-noise ratio, overall image quality, and lesion visibility while reducing noise compared to 30% ASIR-V. The study concludes that DLIR represents a promising advancement in CT imaging, offering improved diagnostic confidence without compromising on image resolution, even though mild blurring increases with higher DLIR strengths.

This is in alignment with the theory posited by Jensen et al. [64], Li et al. [65] examines the feasibility of using a DLIR algorithm for low-dose abdominal CT scans, aiming to maintain diagnostic image quality while reducing radiation exposure. Involving 47 patients, the research compared DLIR at medium and high settings against traditional Adaptive Statistical Iterative Reconstruction-V (ASIR-V) at varying strengths. Findings indicate that DLIR, particularly at high settings, effectively reduces image noise and improves contrast noise ratio without compromising image quality, suggesting DLIR's potential to minimize radiation dose in abdominal CT imaging without adversely affecting diagnostic capabilities.

Also, van Stiphout et al. [66] present systematic reviews and meta-analyze the impact of DLIR on the densitometry

**TABLE 2. Overview of diverse medical imaging datas.**

| Dataset | Source | Characteristics | Modalities | Annotations | Limitations |
|---|---|---|---|---|---|
| Us Simulation & Segmentation [38], [39] | Kaggle | Real and simulated ultrasound datasets with abdominal organ annotations | US, CT | Manual abdominal organ annotations | Relies on manual annotations; affected by technological advances |
| Computed Tomography (CT) Of The Abdomen [40] | Kaggle | Insight into abdominal conditions via CT scans | CT | Disease types and image file links | Limited free dataset volume |
| Hello_World_Deep_Learning_Siim [41][42] | Kaggle | Chest and abdomen X-ray dataset | Radiology | Categorizes images into abdomen and chest | Small dataset size |
| Abdomen MRI [43] | Kaggle | Detailed annotations in high-quality MRI images | MRI | - | Limited pathological conditions |
| Visceral Adipose [44] | Physionet | Follow-up data on pregnant women's visceral adipose tissue (VAT) | Ultrasound | Comprehensive maternal and gestational data | Potential bias in GDM diagnosis; recall bias in BMI calculations |
| Kidney Tumor Segmentation Challenge (Kits 19) [45], [46], [47] | Kaggle | CT scans from nephrectomy patients for tumor segmentation | CT | Annotations delineate kidney tissue | May not represent all tumor variations; subjectivity in manual segmentation |
| Rsna 2023 Abdominal Trauma Detection [48], [49] | Kaggle | CT scans for classifying abdominal injuries | CT | Injury classifications and additional metadata | CT scan parameter variation; may omit non-targeted medical issues |
| St Annotations Of Adfecgdb Database [50][51] | IEEEDataPort | ST segment annotations from fetal ECG database | ECG | Annotations for recordings r01 to r12 | Limited to ST segments; may not cover all fetal ECG changes |
| Liver Vessel Segmentation [52], [53] | IEEEDataPort | CT scan dataset focusing on liver vessels, including masks | CT | 532 CT volumes with original images and vessel masks | Annotations only cover liver vessels |

and image quality of abdominal CT scans. Comparing DLR with filtered back-projection (FBP) and hybrid iterative reconstruction (IR), the study finds no significant difference in CT values, confirming the consistency of these reconstruction techniques in abdominal imaging. However, DLR significantly enhances signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR), implying better image quality without altering CT values.

The study by Kaga et al. [67] conducted a prospective study involving 59 patients to assess the effectiveness of DL image reconstruction (DLIR) in dynamic contrast-enhanced abdomen CT. The study compared the signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), and lesion conspicuity across low, medium, and high DLIR strengths and traditional adaptive statistical iterative reconstruction (ASiR-V). Results showed that high-strength DLIR significantly improved SNR and CNR but could reduce lesion conspicuity, particularly for smaller lesions (<5 mm), suggesting a trade-off between noise reduction and detail preservation in image reconstruction.

Njølstad et al. [68] conducted a study to evaluate the image quality of abdominal CT images reconstructed using a novel DLIR technique, comparing it to standard iterative reconstruction techniques. Their findings indicate that DLIR significantly improved image quality across various clinical image quality criteria, including visual grading criteria, image noise, and contrast-to-noise ratio.

Liu et al. [69] evaluated a novel DL-based reconstruction software, PixelShine, comparing its impact on objective and subjective image quality in low-dose abdominal CT against traditional methods (filtered back projection and iterative reconstruction). The study showed that PixelShine significantly reduced image noise by up to 48% compared to filtered back projection and up to 27% compared to iterative reconstruction while maintaining image information. This resulted in higher signal-to-noise and contrast-to-noise ratios, indicating the potential for radiation dose reduction in abdominal CT imaging without compromising image quality.

Barca et al. [70] investigate the performance of the Precise Image (PI) DL reconstruction (DLR) algorithm for abdominal CT imaging. The study used the Catphan-600 phantom at various doses, comparing PI with Filtered Back-Projection (FBP) and iDose4. Results indicated that PI's performance varied with reconstruction levels, showing significant noise reduction and alterations in noise texture, spatial resolution, and detectability index, depending on the selected level of reconstruction. The study concludes that while PI can effectively reduce image noise and maintain CT number accuracy, it also introduces significant changes in image texture and spatial resolution, which may impact clinical image interpretation.

Racine et al. [71] examine a novel DLIR algorithm for abdominal CT imaging. The study used a phantom at various doses and compared DLIR with Filtered Back-Projection (FBP) and adaptive statistical iterative reconstruction (ASiR-V). The findings revealed that DLIR significantly reduces noise while maintaining noise texture and slightly enhancing spatial resolution, outperforming ASiR-V in simulated clinical scenarios. This suggests that DLIR could be a valuable tool for enhancing image quality in abdominal CT scans.

In addition to image reconstruction, DL has significantly advanced segmentation tasks in abdominal imaging, enabling precise delineation of anatomical structures and pathologies.

Weston et al. [72] developed a fully automated, DL-based algorithm for segmenting abdominal CT images to analyze body composition, including subcutaneous and visceral adipose tissue, muscle, and bone. Utilizing a U-Net architecture, the model was trained on a dataset of 2,430 CT scans and validated on separate test and hepatocellular carcinoma patient datasets. The algorithm achieved high accuracy, comparable to expert manual segmentation, demonstrating potential for clinical application in body composition analysis with the advantage of automating time-consuming segmentation tasks. This approach also showed promise for generalization across different abdominal levels and 3D body composition analysis, indicating its potential utility in large-scale health and disease studies.

The article by Kaur et al. [73] provides a comprehensive systematic review of the evolution of multiorgan segmentation (MOS) techniques in abdominal CT images, transitioning from traditional methods to advanced DL models. It categorizes the segmentation approaches into atlas-based, statistical shape models, and DL models, highlighting the progression and effectiveness of each method. Additionally, the paper identifies organs that require more research focus and discusses publicly available datasets and their challenges.

Caradu et al. [74] evaluated a fully automated software, PRAEVAorta, for segmenting abdominal aortic aneurysm (AAA) in CT scans. Analyzing 100 CT angiographies, the study compares this automatic method with semiautomatic segmentation corrected manually by a senior and junior surgeon. The results demonstrate high accuracy of the automatic method, with significant time savings and potential for consistent clinical application in AAA assessment. The study underscores the effectiveness and efficiency of automated segmentation in enhancing AAA analysis in clinical practice.

Sadikine et al. [25] introduce a semi-overcomplete convolutional auto-encoder (S-OCAE) to improve DL-based segmentation of abdominal structures in medical imaging. It compares this method with standard convolutional autoencoders (CAE) and U-Net architectures. The S-OCAE incorporates undercomplete and overcomplete shape representations, offering enhanced detail and accuracy in segmenting various abdominal structures. The study demonstrates that this approach yields more realistic and precise segmentation results than traditional methods, making it a valuable tool for medical image analysis.

Halkoaho et al. [75] present a novel neural network-based method for automatically quantifying calcification in the abdominal aorta and its branches using contrast-enhanced CT angiography. It involves training two V-Net ensemble models for segmenting arteries and calcifications. The method showed high accuracy with a Dice score of 0.69 for calcification and 0.96 for aorta segmentations.

The integration of deep learning (DL) has significantly enhanced diagnostic accuracy and predictive analytics in abdominal imaging. By leveraging large datasets, DL models detect abnormalities, such as small bowel obstructions, with high sensitivity and specificity.

Kim et al. [76] develop a DL model with high accuracy to automatically identify small bowel obstructions (SBO) in plain abdominal radiographs. Utilizing an ensemble of five CNN trained via transfer learning on a dataset of 990 radiographs, the model achieved an impressive area under the AUC of 0.961, with sensitivities and specificities of 91% and 93%, respectively. This advancement suggests a significant potential for AI to support rapid diagnosis in emergency settings, potentially improving patient outcomes by facilitating quicker clinical decision-making.

Elhage et al. [77] developed and validated DL models (DLMs) using preoperative CT images to predict surgical complexity and outcomes in abdominal wall reconstruction (AWR). The study involved 369 patients and utilized 9,303 images to train three DLMs to predict surgical complexity, surgical site infection (SSI), and postoperative pulmonary failure. The models demonstrated high accuracy, particularly for predicting surgical complexity and SSIs, outperforming expert surgeon predictions. However, the model for predicting pulmonary failure showed less effectiveness. These findings suggest the potential of DLMs to enhance preoperative planning and patient outcomes by accurately predicting surgical complexities and complications.

Płotka et al. [30] investigated the use of DL to estimate fetal weight from ultrasound (US)video scans of the fetal abdomen. Utilizing a dataset of 900 ultrasound examinations, the DL model was trained on 800 retrospective scans and evaluated on 100 prospectively acquired scans. A comparison with manual measurements by six human readers showed that the DL model achieved a mean absolute percentage error of 3.75%, reflecting a similar performance to that of expert readers. The study concludes that DL is a promising approach for estimating fetal weight, potentially offering a more efficient and consistent method than traditional biometry.

Yonezawa et al. [78] focus on developing a DL model to create synthetic 2D digital subtraction angiograms from native abdominal angiograms. It aims to reduce artefacts commonly associated with patient movement during imaging. The model was trained and validated on a dataset of angiograms, showing that the generated images had fewer motion artefacts and maintained high image quality, comparable to traditional 2D-DSA images. This development suggests potential improvements in abdominal angiography, especially when patient movement is a concern.

Deep Learning has enhanced radiomics analysis by improving feature stability and discriminative power in abdominal imaging.

Michallek et al. [79] conducted a phantom study to assess the impact of DL reconstruction (AiCE) on radiomics feature stability and discriminative power in abdominal CT imaging, comparing it to filtered back projection (FBP), hybrid iterative reconstruction (AIDR 3D), and model-based iterative reconstruction (FIRST). The study found that AiCE significantly enhanced feature consistency, discriminative power, and repeatability, more than doubling the yield of radiomics features at clinically used doses, thereby suggesting AiCE's potential to improve the reliability of radiomics analysis in abdominal CT imaging.

The transformative impact of deep learning (DL) in abdominal imaging has emerged through extensive research and development, as evidenced by the studies summarized in Table 3. This table highlights key methodological details. Specifically, the term "N/A" (Not Applicable) is used when specific details were not explicitly mentioned in the original study or were irrelevant to the research scope. "Various" indicates that multiple datasets, methodologies, or imaging techniques were employed, with further details elaborated in the respective study descriptions. "Phantom study" refers to research using synthetic imaging data, typically generated from standardized anatomical models, to evaluate deep learning models under controlled conditions before clinical application. Additionally, methodology descriptors such as "Quantitative and qualitative assessment" refer to studies that evaluate performance using numerical metrics and subjective expert analysis. In contrast, "Comparative analysis" denotes studies that compare different deep learning techniques, datasets, or evaluation approaches to assess their relative effectiveness.

These collaborative efforts underscore the importance of adapting DL technologies to meet the rigorous accuracy and reliability standards required in medical diagnostics. By integrating these advancements into clinical workflows, researchers have not only expanded our understanding of DL's capabilities but also laid the groundwork for future innovations in abdominal imaging. The cumulative knowledge from these studies, as illustrated in Table 3, reinforces the need for continued refinement and validation of DL tools to ensure their safe and effective translation into real-world practice.

### C. PERFORMANCE METRICS AND MODEL EVALUATION

In the realm of DL applied to abdominal imaging, the nuanced evaluation of model performance transcends simple accuracy metrics to include a spectrum of metrics illuminating different facets of model efficacy. These metrics, pivotal in the context of traumatic abdominal injury detection, offer insights into the precision, reliability, and clinical applicability of DL algorithms [61], [80], [81].

**Accuracy (1)** is often perceived as the gateway metric for model evaluation, offering a quick snapshot of overall model performance by accounting for both correctly identified positives (injuries) and negatives (normal cases). Yet, its apparent simplicity belies its limitations in abdominal imaging, where the prevalence of non-injury cases can dwarf actual injuries, leading to potentially high accuracy scores even if the model is inadequately identifying true injury cases. This discrepancy underscores the importance of not relying solely on accuracy for model evaluation in such imbalanced datasets [13].

**Sensitivity (Recall) (2)** is indispensable for ensuring a model reliably detects injuries, minimizing the risk of missed diagnoses that could have dire consequences for patients. In abdominal imaging, where injuries can range from subtle indications to overt signs, sensitivity measures the model's ability to capture these variations. High sensitivity is crucial, but it must be balanced against the risk of increasing false positives, which could lead to unnecessary anxiety, further testing, and resource allocation [2], [82].

**Specificity (3)** counterbalances sensitivity by ensuring the model correctly identifies cases without injuries, thus avoiding the pitfalls of overdiagnosis. In abdominal imaging, where detecting conditions like hematomas, organ lacerations, or free fluid can be critical, specificity guards against the over-interpretation of normal variations as pathological findings. The challenge lies in maintaining high specificity without sacrificing sensitivity, necessitating sophisticated model tuning and validation against diverse datasets [83], [84], [85].

**Precision (4)** focuses on the accuracy of positive predictions. In abdominal imaging, precision affects clinical decision-making, influencing whether to proceed with further investigations or interventions based on model predictions. High precision indicates that when a model flags a potential injury, there is a high probability that the injury is indeed present, which is paramount for effective patient management and resource utilization. However, optimizing precision often requires addressing the imbalance between injury and

AQ:4

**TABLE 3. Comparative analysis of DL in abdominal imaging studies.**

| Paper | Focus | Dataset | Methodology | Key Contributions |
|---|---|---|---|---|
| Zhang Et Al. [61] | DL in medical imaging | Various | Review of CNNs, RNNs, GANs | Highlights DL's impact on diagnosis and treatment, suggests exploring ethical aspects and integration challenges |
| Litjens Et Al. [14] | DL applications in medical image analysis | Various | Survey of CNNs, RNNs, GANs | Comprehensive analysis across imaging tasks, calls for detailed discussion on clinical implementation challenges |
| Rehman And Khan [62] | DL for abdominal image classification and segmentation | NA | Review of DL models and datasets | Emphasizes the need for advanced imaging devices, preprocessing algorithms, and exploration of CNN architectures |
| Jensen Et Al [64] | DL Image Reconstruction (DLIR) in CT scans | Contrast-enhanced oncologic CT scans | Quantitative and qualitative assessment | Demonstrates DLIR's ability to improve image quality, contrast-to-noise ratio, and lesion visibility |
| Li Et Al.[65] | DLIR for low-dose abdominal CT scans | 47 patients | Comparative analysis | Shows DLIR's potential in maintaining image quality while reducing radiation exposure |
| Stiphout Et Al.[66] | Impact of DLR on abdominal CT densitometry and image quality | NA | Systematic review and meta-analysis | Confirms DLR's enhancement of SNR and CNR without altering CT values, supporting clinical implementation |
| Weston Et Al.[72] | Automated DL algorithm for body composition analysis in CT images | 2,430 CT scans | U-Net architecture | Achieves high accuracy in automated segmentation, demonstrating potential for large-scale health studies. The focus of the segmentation is fat and muscle tissue, the model does not specifically segment individual organs like the liver, kidneys, or intestines. |
| Kim Et Al. [77] | Identifying small bowel obstructions in radiographs | 990 abdominal radiographs | Ensemble of CNNs | Achieves high AUC, sensitivity, and specificity, indicating AI's potential in emergency diagnostics |
| Njølstad Et Al.[68] | Image quality assessment of abdominal CT with DLIR | NA | Comparative analysis | Finds DLIR significantly improves image quality, supporting its clinical use |
| Kaur Et Al. [73] | Evolution of multiorgan segmentation techniques | NA | Systematic review | Identifies DL as the optimal choice for accuracy in multiorgan segmentation |
| Elhage Et Al. [78] | Predicting surgical complexity in abdominal wall reconstruction | 369 patients | Development and validation of DLMs | DLMs predict surgical complexity and complications with high accuracy, enhancing preoperative planning |
| Kaga Et Al.[67] | Efficacy of DLIR in dynamic contrast-enhanced CT | 59 patients | Prospective study | Highlights trade-off between noise reduction and lesion conspicuity with DLIR, suggesting careful application |
| Michallek Et Al.[80] | Impact of AiCE on radiomics in abdominal CT | Phantom study | Comparative analysis | Shows AiCE enhances radiomics feature stability and discriminative power, suggesting improved analysis reliability |
| Liu Et Al [69] | PixelShine's impact on image quality in low-dose CT | NA | Comparative study | Indicates significant noise reduction and potential for radiation dose minimization without losing image quality |
| Płotka Et Al. [30] | DL for estimating fetal weight from ultrasound scans | 900 ultrasound examinations | DL model evaluation | Demonstrates DL's accuracy in fetal weight estimation, comparable to expert readers |
| Caradu Et Al. [74] | Automated segmentation of AAA in CT scans | 100 CT angiographies | Comparison of automated vs. semiautomatic segmentation | Shows automatic method's high accuracy and time efficiency, advocating for clinical use in AAA assessment |
| Barca Et Al. [70] | Performance of PI DLR algorithm for abdominal CT | Catphan-600 phantom | Comparative analysis | PI DLR shows significant noise reduction, though it alters image texture and resolution, impacting clinical interpretation |
| Sadikine Et Al [25] | Improved segmentation with semi-overcomplete CAE | NA | Method comparison | Yields more accurate segmentation, offering a valuable tool for medical image analysis |
| Yonezawa Et Al. [79] | Synthetic angiograms from DL model | Angiogram dataset | Model development and validation | Reduces motion artifacts in angiograms, maintaining high image quality |
| Saleh Et Al. [63] | Automated analysis for Black Tiger Prawns | Real and synthetic fetal electrocardiography (fECG) signals | Utilizes a novel feature extraction module for morphometrics | Offers a more accurate method for prawn weight estimation, benefiting aquacultural engineering |
| Halkoaho Et Al. [76] | Quantification of calcification in the abdominal aorta | Contrast-enhanced CT angiography | Two V-Net ensemble models | High accuracy in assessing atherosclerosis, enhancing cardiovascular risk management |
| Racine Et Al. [71] | Evaluating DLIR in abdominal CT imaging | Phantom study | Comparison with FBP and ASiR-V | Demonstrates DLIR's efficacy in noise reduction and image quality improvement |

non-injury cases and carefully calibrating the model's threshold for classifying positives [83], [86].

The **AUC-ROC** is a robust measure of the model's discriminative power, encapsulating its ability to differentiate between injury and non-injury cases across various decision thresholds. This metric is particularly informative in abdominal imaging, where distinguishing between subtle injuries and normal anatomical variations can be challenging. The AUC-ROC encapsulates the trade-offs between sensitivity and specificity, offering a comprehensive view of model performance. A model with a high AUC-ROC is adept at distinguishing between complex injury presentations and normal findings, a critical capability given the wide range of pathologies encountered in abdominal trauma [13], [80].

**F1 Score (5)** balances the precision-recall trade-off, providing a metric that captures the model's balance between identifying true injuries and minimizing false positives. This balance is vital in abdominal imaging, where the consequences of false negatives (missed injuries) are as significant as false positives (unnecessary interventions). The F1 Score is particularly useful in evaluating model performance in datasets where injury cases are rare compared to normal findings, ensuring that the majority class does not overly skew the model's performance [80].

For segmentation tasks crucial in abdominal imaging, such as delineating the extent of an injury or identifying specific anatomical structures, the **Dice Coefficient (6)** and **Jaccard Index (IoU) (7)** provide essential insights into the model's spatial accuracy. It is important to note that **F1 Score (5)** and **Dice Coefficient (6)** are mathematicaly equivalent, though reporting conventions differ by research group and both names remain prevalent in literature.

These metrics assess the overlap between the model's predictions and the actual ground truth, offering a direct measure of the model's effectiveness in segmenting images accurately. The precision in segmentation is critical for planning surgical interventions, assessing injury severity, and monitoring disease progression, making these metrics indispensable for evaluating models applied to abdominal imaging [83], [87], [88], [89], [90].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Sensitivity (Recall)} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{5}$$

$$\text{Dice Coefficient} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{6}$$

$$\text{Jaccard Index (IoU)} = \frac{TP}{TP + FP + FN} \tag{7}$$

where:

- True Positives (TP): The number of correctly identified positive cases.
- True Negatives (TN): The number of correctly identified negative cases.
- False Positives (FP): The number of negative cases incorrectly identified as positive.
- False Negatives (FN): The number of positive cases incorrectly identified as negative.

The choice of metrics and emphasis on each can vary depending on the clinical setting and the diagnosed condition. For example, in screening for abdominal cancers, a high sensitivity might be prioritized to ensure that as many cases as possible are identified early. Conversely, in conditions where overdiagnosis could lead to harmful interventions, specificity may take precedence [84].

Understanding the trade-offs between sensitivity and specificity is crucial in model selection and development. A sensitive model might generate many false positives, burdening the healthcare system and causing undue patient stress. On the other hand, a model with high specificity but low sensitivity could miss critical diagnoses, potentially endangering patients' lives. Moreover, Table 4 provides a comparative view of these metrics, illustrating their individual contributions and considerations in the broader context of model evaluation in abdominal imaging.

Recent advances in deep learning for abdominal imaging reflect a methodological diversification, with architectures optimized for clinical tasks ranging from segmentation and classification to reconstruction and predictive analytics. As synthesized in Table 5, performance metrics vary significantly across these domains, underscoring both technical progress and persistent standardization challenges. Segmentation and classification studies frequently report Dice scores and AUC-ROC values to quantify anatomical accuracy and diagnostic performance, alongside sensitivity and specificity. In contrast, reconstruction tasks prioritize technical outcomes such as noise reduction rates and signal-to-noise ratio (SNR) improvements, while predictive analytics rely on metrics like mean absolute percentage error (MAPE) and intraclass correlation coefficients (ICC) to assess prognostic accuracy. However, the prevalence of "N/A" entries in the table highlights critical gaps, where metrics are either unreported or deemed inapplicable, complicating cross-study comparisons. This variability, though adaptive to clinical priorities, emphasizes the need for harmonized evaluation protocols to ensure reproducibility and facilitate the translation of these tools into routine practice.

In conclusion, evaluating DL models in abdominal imaging demands a multi-faceted approach, leveraging a spectrum of metrics to capture the model's precision, reliability, and clinical applicability comprehensively. This thorough evaluation ensures that models are not only technically sound but also practically viable, aligning with the overarching goal of enhancing patient care through accurate and reliable imaging analysis. The transition from accuracy to more detailed

**TABLE 4.** Comparative overview of performance metrics in DL for abdominal imaging.

| Metric | Description | Importance in Medical Imaging | Considerations |
|---|---|---|---|
| Accuracy [13], [32][32], [62] | Proportion of true results (both true positives and true negatives) among the total number of cases examined. | Offers a general sense of model effectiveness. | Can be misleading in imbalanced datasets where one class dominates. |
| Sensitivity (Recall) [13], [85], [86] | Proportion of actual positives correctly identified (e.g., correctly identifying injuries). | Crucial for ensuring all relevant cases are identified; important where missing a positive case could be critical. | Higher sensitivity might lead to more false positives. |
| Specificity [84], [85], [86] | Proportion of true negatives correctly identified (e.g., correctly identifying non-injury cases). | Minimizes false positives; important for ensuring normal cases aren't misclassified. | Higher specificity might lead to more false negatives. |
| Precision [84], [87] | Ratio of true positives to all positives predicted by the model. | Indicates reliability of positive predictions; important where the cost of false positives is high. | Can be low even if the model has high accuracy, especially in imbalanced datasets. |
| AUC-ROC [13], [61], [87], [92] | Area Under the Curve of the Receiver Operating Characteristics (ROC); measures the model's ability to distinguish between classes. | Provides a single measure of model effectiveness in classification tasks. | Requires careful interpretation, especially in highly imbalanced datasets. |
| F1 Score [61], [87] | Harmonic mean of precision and recall. | Useful in scenarios with uneven class distributions; balances false positives and negatives. | More informative than accuracy in imbalanced datasets. |
| Dice Coefficient [84], [88], [91] | Measures the overlap between the predicted segmentation and the ground truth. | Essential in evaluating the spatial accuracy of segmentation tasks. | Particularly used in image segmentation, less relevant for classification tasks. |
| Jaccard Index (IoU) [88], [89], [90] | Intersection over Union (IoU); used for assessing the accuracy of a model in segmenting images. | Indicates how well the model delineates the boundaries of an object or lesion. | Highly relevant in segmentation, especially for delineating injury boundaries. |

metrics like sensitivity, specificity, precision, AUC-ROC, F1 Score, Dice Coefficient, and Jaccard Index represents a critical shift towards a more nuanced understanding of model performance, crucial for advancing diagnostic capabilities in abdominal imaging.

## IV. CHALLENGES, ETHICAL CONSIDERATIONS, AND FUTURE DIRECTIONS IN AI FOR ABDOMINAL IMAGING

As we peer into the future, several key trends are poised to redefine the landscape of medical imaging and patient care.

In abdominal imaging, where the early detection of conditions such as liver diseases, cancers, and inflammatory bowel diseases is crucial, the ability to draw on a rich tapestry of patient data could be game-changing.

The challenge lies in developing advanced algorithms to navigate and synthesize this complex data landscape.

Predictive analytics, fueled by AI, may soon offer insights into disease progression, response to therapies, and potential side effects, tailored to the unique profile of each patient [94]. The implications are profound, potentially leading to the identification of new diagnostic markers and improving data efficiency. Transfer learning, wherein models pre-trained on available datasets are adapted to new, yet similar tasks. Transfer learning, wherein models pre-trained on available datasets are adapted to new, yet similar data, exemplifies how we might navigate the challenges of data scarcity, accelerating the pace of innovation and model development in abdominal imaging [95], [96].

The path from research to clinical application is complex, fraught with regulatory hurdles, technological challenges, and ethical considerations. Ensuring patient safety, data privacy, and the trustworthiness of AI-driven decisions is paramount. Adapting regulatory frameworks to better accommodate the rapid pace of AI innovation will be crucial, as will efforts to integrate AI models seamlessly into existing healthcare systems and workflows. Such integration not only requires technological compatibility but also a deep understanding of clinical needs and processes. Additionally, fostering interdisciplinary collaboration stands as a cornerstone of successful AI deployment, ensuring that tools are not only technically advanced but also clinically relevant and ethically sound [97], [98].

Coupled with rigorous clinical validation, these efforts will help establish the reliability and utility of AI models, encouraging their broader acceptance and use in clinical practice [99], [100].

**TABLE 5.** Performance trends of deep learning techniques in abdominal imaging.

| Paper | Task | Methodology | Dataset | Performance Metrics |
|---|---|---|---|---|
| Weston et al [72] | Segmentation | U-Net | 2,430 CT scans | N/A |
| Rezaei et al.[27] | Segmentation | Ensemble-GAN | CHAOS 2019, LiTS 2017 | F1 scores: Spleen (0.93), Liver (0.96), Kidneys (0.90, 0.94), Liver lesions (0.83, 0.94) |
| Halkoaho et al.[76] | Segmentation | V-Net | Contrast-enhanced CT angiography | Dice scores: Calcification (0.69), Aorta (0.96) |
| Kim et al.[8] | Classification | Ensemble of CNNs | 990 abdominal radiographs | AUC: 0.961, Sensitivity: 91%, Specificity: 93% |
| Elhage et al.[78] | Classification | DL Models (DLMs) | 369 patients, 9,303 images | N/A |
| Jensen et al.[64] | Image Reconstruction | DLIR | Contrast-enhanced CT scans | N/A |
| Kaga et al.[67] | Image Reconstruction | DLIR | 59 patients | Improved SNR and CNR, reduced lesion conspicuity for small lesions (<5 mm) |
| Li et al.[65] | Image Reconstruction | DLIR | 47 patients | N/A |
| Liu et al. [93] | Image Reconstruction | PixelShine | N/A | Noise Reduction: 48%, Image Quality:Maintained |
| Płotka et al [30] | Predictive Analytics | DL model | 900 ultrasound scans | MAPE: 3.75%, ICC: 0.9761–0.9865 |
| Luo et al. [94] | Segmentation | WORD dataset | 150 abdominal CT volumes | N/A |
| Tong et al. [31] | Segmentation | Two-stage approach | FLARE challenge dataset | Dice Similarity Coefficient: 83.7%, Normalized Surface Dice: 64.4% |
| Sadikine et al.[25] | Segmentation | Semi-overcomplete CAE (S-OCAE) | Six datasets (MSD, LiTS, KiTS19, CHAOS, 3D-IRCADb, DRIVE) | N/A |

The integration of DL into abdominal imaging and broader healthcare practices is on the cusp of bringing about a paradigm shift towards more precise, personalized medical care. The emerging trends of multimodal data integration, innovative learning techniques, and strategic clinical deployment underscore the potential for AI to enhance diagnostic processes, treatment strategies, and patient outcomes. However, realizing this potential will necessitate ongoing collaboration across disciplines, careful navigation of regulatory landscapes, and a steadfast commitment to ethical principles. As we move forward, the promise of DL in transforming abdominal imaging and healthcare remains vast, matched only by the collective resolve to tackle the challenges and seize the opportunities that lie ahead.

In abdominal imaging, such biases could lead to misdiagnoses or overlooked conditions, disproportionately affecting certain groups of patients. For instance, if an AI system is trained predominantly on imaging data from one demographic group, it might perform less accurately on other demographic groups, thereby exacerbating existing health disparities [101]. To mitigate algorithmic bias, employing diverse and representative datasets in training AI models is crucial. Moreover, ongoing evaluation and recalibration of AI systems are necessary to identify and correct biases that may emerge over time. Collaboration among data scientists, clinicians, and ethicists is essential to develop effective methodologies that identify and address bias. A robust approach to mitigating bias in AI-driven abdominal imaging involves multiple strategies, including data diversification, algorithmic fairness techniques, and continuous model evaluation. Ensuring dataset representativeness is fundamental, requiring the inclusion of diverse patient demographics spanning different age groups, ethnicities, and socioeconomic backgrounds. Additionally, algorithmic adjustments such as reweighting techniques, adversarial debiasing, and fairness-aware loss functions can be applied to mitigate disparities in model predictions. XAI techniques also play a crucial role in bias detection, allowing researchers to analyze and adjust model behaviors when discrepancies arise. Beyond technical solutions, regulatory frameworks and institutional policies should mandate bias audits and promote the adoption of fair AI development practices. These steps collectively contribute to creating AI models that deliver equitable diagnostic outcomes for all patient populations.

Transparency in the development of AI models is another critical ethical consideration. It involves clear communication about how AI models are created, trained, and validated, including the sources of data used and the decision-making processes within the models. Transparency is vital not only for building trust among clinicians and patients but also for facilitating peer review and regulatory oversight. The integration of AI in medical imaging raises significant ethical and security concerns, particularly regarding data privacy, transparency, and secure communication.

As AI systems rely on large-scale patient datasets, ensuring confidentiality and integrity is crucial to prevent unauthorized access or data manipulation. Blockchain-based solutions, such as Blockchain-Based Secure, Interactive, and Fair Mobile Crowdsensing (BSIF), have been proposed to enhance data security, fairness, and transparency in AI-driven applications by leveraging decentralized and immutable data structures [102]. Additionally, the growing need for cross-technology communication security in medical AI systems has led to advancements in authentication mechanisms to prevent unauthorized access, as demonstrated in secure cross-technology communication frameworks [103]. Furthermore, robust cryptographic protocols are required to protect patient data and AI model integrity with the increasing use of cloud-based AI systems for medical imaging analysis. Recent research on collusion-resilient and maliciously secure cloud-assisted computation highlights how advanced cryptographic techniques can mitigate privacy risks and security vulnerabilities in cloud-assisted medical imaging applications [104]. Implementing such privacy-preserving techniques is critical for ensuring that AI in medical imaging adheres to ethical standards, minimizes bias, and maintains patient trust in AI-driven diagnostics. Openness in model development enables independent verification of an AI system's accuracy, reliability, and bias. It also allows healthcare professionals to understand the strengths and limitations of AI tools, guiding their appropriate and effective use in clinical practice. Efforts to enhance transparency include the adoption of open-source practices, detailed documentation of AI development processes, and the publication of validation studies in peer-reviewed journals [105], [106].

The promise of AI in medical imaging extends to enhancing patient care, improving diagnostic accuracy, and facilitating personalized treatment plans. However, to realize these benefits equitably, it is essential to ensure that improvements in patient care are accessible to all segments of the population. This challenge underscores the need to address healthcare access, technology infrastructure, and digital literacy disparities. Equitable AI deployment in abdominal imaging requires targeted efforts to include underrepresented populations in research and development processes, the adaptation of AI tools to low-resource settings, and initiatives to educate both healthcare providers and patients about AI technologies. Moreover, ethical AI use in healthcare demands a patient-centric approach, emphasizing consent, privacy, and the patient's autonomy in decision-making processes involving AI-assisted diagnoses or treatments [105], [107].

Furthermore, While much of the research on DL in abdominal imaging focuses on theoretical advancements, AI-driven solutions are already making their way into clinical practice, mainly through FDA-approved AI tools [108]. These tools have been validated for real-world use and are aiding radiologists in automating image interpretation, enhancing diagnostic accuracy, and improving workflow efficiency.

Several FDA-approved AI algorithms are designed to assist in abdominal CT and MRI analysis by automating segmentation, lesion detection, and quantitative analysis. Notable examples include:

- QCT Pro (Mindways Software) – FDA-approved for automated body composition analysis using CT scans, enabling precise measurement of visceral and subcutaneous fat.
- Aidoc – Approved for automated triage and detection of intra-abdominal hemorrhage in non-contrast CT scans, helping prioritize urgent cases.
- Zebra Medical Vision – Offers AI-based quantitative liver analysis, assisting in detecting hepatic steatosis and fibrosis.
- Arterys Cardio AI – Initially focused on cardiovascular imaging, it includes abdominal aorta measurements for aneurysm assessment in CT angiography.
- Riverain ClearRead CT—This product provides automated nodule detection in CT scans, which can assist in the detection of abdominal tumors.

These tools bridge the gap between research and clinical application, demonstrating how validated AI models can integrate into existing radiology workflows to support faster and more consistent diagnoses. Beyond improving diagnostic precision, they also help reduce radiologist workload, particularly in high-volume imaging centers where manual interpretation is time-intensive.

The regulatory approval of these AI models signifies a major step toward the widespread adoption of DL in abdominal imaging. However, continued clinical validation and real-world testing remain essential to ensure generalizability across diverse patient populations and imaging protocols.

As AI becomes increasingly integrated into medical imaging, addressing these ethical considerations is paramount to harnessing this technology's full potential for improving patient care. By tackling algorithmic bias, promoting transparency, and striving for equitable healthcare improvements, the medical community can ensure that AI serves as a force for good, enhancing the accuracy and accessibility of medical imaging across diverse patient populations.

## V. CONCLUSION

This comprehensive survey on DL in abdominal imaging has traversed the intricate landscape of AI technologies, datasets, performance metrics, and ethical considerations, unveiling the profound impact and potential challenges of integrating DL into medical diagnostics and treatment strategies. Through a detailed exploration of the current state of the art, this review highlights not only the technological advancements that have significantly enhanced the accuracy and efficiency of abdominal imaging but also underscores the critical role of data handling, model evaluation, and ethical governance in ensuring these innovations translate into equitable and effective patient care.

The advent of DL in abdominal imaging marks a pivotal shift towards precision medicine, where the convergence of diverse datasets, advanced algorithms, and interdisciplinary collaboration paves the way for diagnostic tools of unprecedented accuracy and reliability. The exploration of novel DL techniques, including integrating multimodal data and adopting unsupervised and semi-supervised learning models, stands at the forefront of this transformation, promising to address the enduring challenges of data scarcity and model generalizability.

However, the path to realizing AI's full potential in abdominal imaging is fraught with complexities. The ethical considerations discussed herein—ranging from algorithmic bias to the imperative for transparency and inclusivity—serve as a reminder of the responsibilities that accompany technological innovation. Addressing these ethical challenges is paramount, requiring a concerted effort from technologists, clinicians, ethicists, and policymakers to navigate the delicate balance between innovation and patient welfare.

As we look to the future, it is clear that DL will continue to reshape the landscape of abdominal imaging, offering new avenues for early detection, diagnosis, and personalized treatment planning. Yet, successfully integrating these technologies into clinical practice will depend on their technical prowess and alignment with ethical standards, regulatory requirements, and the overarching goal of improving patient outcomes.

In conclusion, this survey underscores the transformative potential of DL in abdominal imaging while highlighting the multifaceted challenges that must be addressed to harness this potential fully. By fostering an environment of collaboration, transparency, and ethical vigilance, we can ensure that the advancements in AI not only advance the frontiers of medical science but also do so in a manner that is equitable, responsible, and ultimately beneficial to all patients. The journey ahead is complex but immensely promising, heralding a new era of precision medicine where DL technologies play a pivotal role in enhancing healthcare delivery and patient care.

## GLOSSARY OF KEY TERMS

**Deep Learning (DL)**

A subset of artificial intelligence (AI) that uses multiple layers of artificial neural networks to learn patterns from data. DL helps analyze and interpret images for diagnosis and treatment planning in medical imaging.

**Convolutional Neural Networks (CNNs)**

A type of deep learning model designed for image analysis. CNNs automatically detect patterns such as edges, shapes, and textures in medical images, making them highly effective for organ segmentation and anomaly detection tasks.

**Recurrent Neural Networks (RNNs)**

A type of deep learning model specialized for sequential data processing. While commonly used for speech and text analysis, RNNs can also be applied in medical imaging to track changes over time in scans.

**Generative Adversarial Networks (GANs)**

A deep learning technique consisting of two competing networks—a generator and a discriminator. GANs can create realistic synthetic medical images, which help train AI models when accurate data is scarce.

**Dice Similarity Coefficient (DSC)**

A metric used to evaluate how accurately a model segments an image by measuring the overlap between predicted and actual regions. A higher DSC value indicates better segmentation performance. This coefficient is also called F1 score.

**Hausdorff Distance**

A mathematical measure used to assess the accuracy of image segmentation. It calculates the greatest distance between points on the predicted and actual boundary, with a lower value indicating better performance.

**Transfer Learning**

A method where a pre-trained model, developed for one task, is adapted to a new but related task. In medical imaging, this helps train AI models more efficiently using smaller datasets.

**Few-Shot Learning**

A machine learning approach that enables models to learn effectively with very limited training data. It is useful in medical imaging when only a few annotated scans are available.

**Explainable AI (XAI)**

A set of techniques that make deep learning models more interpretable, allowing clinicians to understand how AI makes decisions and ensuring transparency in medical diagnostics.

**AUC-ROC (Area Under the Curve - Receiver Operating Characteristic)**

A performance metric for classification models that measures their ability to distinguish between different conditions (e.g., disease vs. no disease). A higher AUC-ROC score indicates better model accuracy.

**Jaccard Index (Intersection over Union, IoU)**

A metric used to evaluate segmentation models by measuring how much the predicted and actual segmented regions overlap. Higher IoU values indicate more accurate segmentations.

**Normalization and Standardization**

Preprocessing techniques that adjust pixel intensity values in images to a common scale, improving model performance and consistency across different datasets.

**Image Augmentation**

A technique used to artificially expand training datasets by applying transformations such as rotation, flipping, contrast adjustment, and noise addition to images. This helps models generalize new data better.

**Phantom Study**

A research method using synthetic imaging data or simulated anatomical models to test AI algorithms before applying them to real clinical images.

**Semi-supervised and Unsupervised Learning**

**Semi-Supervised Learning:** A training approach that combines a small amount of labeled data with a larger amount of unlabeled data to improve model performance.

**Unsupervised Learning:** A method where models learn patterns in data without labeled examples, useful for identifying hidden structures in medical images.

**Algorithmic Bias in AI Models**

A situation where AI models produce unfair or inaccurate predictions due to imbalanced training data. For example, if a model is trained mostly on images from one demographic group, it may not work well for others.

**Multimodal Data Integration**

Combining different types of medical data—such as imaging scans, electronic health records (EHRs), and genetic information—improves diagnostic accuracy and personalized treatment planning.

## REFERENCES

[1] E. Talbot, S. Evans, N. Hellenthal, D. Monie, P. Campbell, and S. Cooper, "Abdominal and pelvic vascular injury: A national trauma data bank study," *Amer. Surgeon*, vol. 85, no. 3, pp. 292–293, Mar. 2019.

[2] E. Reitano, S. P. B. Cioffi, C. Airoldi, O. Chiara, G. La Greca, and S. Cimbanassi, "Current trends in the diagnosis and management of traumatic diaphragmatic injuries: A systematic review and a diagnostic accuracy meta-analysis of blunt trauma," *Injury*, vol. 53, no. 11, pp. 3586–3595, Nov. 2022, doi: 10.1016/j.injury.2022.07.002.

[3] R. Van Vugt, "Thoracoabdominal computed tomography in blunt high-energy trauma patients," M.S. thesis, Radboud Univ. Nijmegen, Nijmegen, The Netherlands, Jun. 2014.

[4] L. Smyth et al., "WSES guidelines on blunt and penetrating bowel injury: Diagnosis, investigations, and treatment," *World J. Emergency Surg.*, vol. 17, no. 1, pp. 1–15, Mar. 2022, doi: 10.1186/s13017-022-00418-y.

[5] L. Shaddock and T. Smith, "Potential for use of portable ultrasound devices in rural and remote settings in Australia and other developed countries: A systematic review," *J. MultidisciplinaryHealthcare*, vol. 15, pp. 605–625, Mar. 2022.

[6] V. Nisenblat, P. M. Bossuyt, C. Farquhar, N. Johnson, and M. L. Hull, "Imaging modalities for the non-invasive diagnosis of endometriosis," *Cochrane Database Systematic Rev.*, vol. 2016, no. 2, Feb. 2016, Art. no. CD009591, doi: 10.1002/14651858.CD009591.pub2.

[7] A. Urushibara, T. Saida, K. Mori, T. Ishiguro, K. Inoue, T. Masumoto, T. Satoh, and T. Nakajima, "The efficacy of deep learning models in the diagnosis of endometrial cancer using MRI: A comparison with radiologists," *BMC Med. Imag.*, vol. 22, no. 1, pp. 1–14, Dec. 2022, doi: 10.1186/s12880-022-00808-3.

[8] J. Kim, J. Hong, and H. Park, "Prospects of deep learning for medical imaging," *Precis. Future Med.*, vol. 2, no. 2, pp. 37–52, Jun. 2018, doi: 10.23838/pfm.2018.00030.

[9] C. Singh, "Medical imaging using deep learning models," *Eur. J. Eng. Technol. Res.*, vol. 6, no. 5, pp. 156–167, Aug. 2021, doi: 10.24018/ejers.2021.6.5.2491.

[10] M. A. Weigelt, H. A. Lev-Tov, M. Tomic-Canic, W. D. Lee, R. Williams, D. Strasfeld, R. S. Kirsner, and I. M. Herman, "Advanced wound diagnostics: Toward transforming wound care into precision medicine," *Adv. Wound Care*, vol. 11, no. 6, pp. 330–359, Jun. 2022.

[11] D. Stengel, G. Rademacher, A. Ekkernkamp, C. Güthoff, and S. Mutze, "Emergency ultrasound-based algorithms for diagnosing blunt abdominal trauma," *Cochrane Database Systematic Rev.*, vol. 2015, no. 9, Sep. 2015, Art. no. CD004446, doi: 10.1002/14651858.CD004446.pub4.

[12] K. Hanna, S. Asmar, M. Ditillo, M. Chehab, M. Khurrum, L. Bible, M. Douglas, and B. Joseph, "Readmission with major abdominal complications after penetrating abdominal trauma," *J. Surgical Res.*, vol. 257, pp. 69–78, Jan. 2021.

[13] R. Aggarwal, V. Sounderajah, G. Martin, D. S. W. Ting, A. Karthikesalingam, D. King, H. Ashrafian, and A. Darzi, "Diagnostic accuracy of deep learning in medical imaging: A systematic review and meta-analysis," *NPJ Digit. Med.*, vol. 4, no. 1, Apr. 2021, doi: 10.1038/s41746-021-00438-z.

[14] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017, doi: 10.1016/j.media.2017.07.005.

[15] A. Kebaili, J. Lapuyade-Lahorgue, and S. Ruan, "Deep learning approaches for data augmentation in medical imaging: A review," *J. Imag.*, vol. 9, no. 4, p. 81, Apr. 2023.

[16] F. Garcea, A. Serra, F. Lamberti, and L. Morra, "Data augmentation for medical imaging: A systematic literature review," *Comput. Biol. Med.*, vol. 152, Jan. 2023, Art. no. 106391, doi: 10.1016/j.compbiomed.2022.106391.

[17] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *J. Med. Imag. Radiat. Oncol.*, vol. 65, no. 5, pp. 545–563, Aug. 2021.

[18] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.

[19] A. Serra, "Data augmentation for medical image analysis: A systematic literature review," Master's thesis, Dipartimento di Automatica e Informatica, Politecnico di Torino, Turin, Italy, 2021.

[20] E. Goceri, "Medical image data augmentation: Techniques, comparisons and interpretations," *Artif. Intell. Rev.*, vol. 56, no. 11, pp. 12561–12605, Nov. 2023.

[21] O. O. Abayomi-Alli, R. Damaševič ius, A. Qazi, M. Adedoyin-Olowe, and S. Misra, "Data augmentation and deep learning methods in sound classification: A systematic review," *Electronics*, vol. 11, no. 22, p. 3795, Nov. 2022.

[22] J. Li, G. Zhu, C. Hua, M. Feng, B. Bennamoun, P. Li, X. Lu, J. Song, P. Shen, X. Xu, L. Mei, L. Zhang, S. A. A. Shah, and M. Bennamoun, "A systematic collection of medical image datasets for deep learning," *ACM Comput. Surv.*, vol. 56, no. 5, pp. 1–51, May 2024, doi: 10.1145/3615862.

[23] S. Park, L. C. Chu, E. K. Fishman, A. L. Yuille, B. Vogelstein, K. W. Kinzler, K. M. Horton, R. H. Hruban, E. S. Zinreich, D. F. Fouladi, S. Shayesteh, J. Graves, and S. Kawamoto, "Annotated normal CT data of the abdomen for deep learning: Challenges and strategies for implementation," *Diagnostic Interventional Imag.*, vol. 101, no. 1, pp. 35–44, Jan. 2020, doi: 10.1016/j.diii.2019.05.008.

[24] S. Luo, L. Pan, Y. Jian, Y. Lu, and S. Luo, "CTBANet: Convolution transformers and bidirectional attention for medical image segmentation," *Alexandria Eng. J.*, vol. 88, pp. 133–143, Feb. 2024, doi: 10.1016/j.aej.2024.01.018.

[25] A. Sadikine, B. Badic, J.-P. Tasu, V. Noblet, P. Ballet, D. Visvikis, and P.-H. Conze, "Improving abdominal image segmentation with overcomplete shape priors," *Computerized Med. Imag. Graph.*, vol. 113, Apr. 2024, Art. no. 102356, doi: 10.1016/j.compmedimag.2024.102356.

[26] M. Gross, S. Arora, S. Huber, A. S. Kücükkaya, and J. A. Onofrey, "LiverHccSeg: A publicly available multiphasic MRI dataset with liver and HCC tumor segmentations and inter-rater agreement analysis," *Data Brief*, vol. 51, Dec. 2023, Art. no. 109662, doi: 10.1016/j.dib.2023.109662.

[27] M. Rezaei, J. J. Näppi, C. Lippert, C. Meinel, and H. Yoshida, "Generative multi-adversarial network for striking the right balance in abdominal image segmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 15, no. 11, pp. 1847–1858, Nov. 2020, doi: 10.1007/s11548-020-02254-4.

[28] M. Rezaei, J. Nappi, B. Bischl, and H. Yoshida, "Bayesian uncertainty estimation for detection of long-tail and unseen conditions in abdominal images," *Proc. SPIE*, vol. 12033, pp. 284–290, Apr. 2022, doi: 10.1117/12.2612955.

[29] G. Baldazzi, E. Sulas, M. Urru, R. Tumbarello, L. Raffo, and D. Pani, "Annotated real and synthetic datasets for non-invasive foetal electrocardiography post-processing benchmarking," *Data Brief*, vol. 33, Dec. 2020, Art. no. 106399, doi: 10.1016/j.dib.2020.106399.

[30] S. S. Płotka, M. K. Grzeszczyk, P. I. Szenejko, K. Żebrowska, N. A. Szymecka-Samaha, T. Łęgowik, M. A. Lipa, K. Kosińska-Kaczyńska, R. Brawura-Biskupski-Samaha, I. Išgum, C. I. Sánchez, and A. Sitek, "Deep learning for estimation of fetal weight throughout the pregnancy from fetal abdominal ultrasound," *Amer. J. Obstetrics Gynecol. MFM*, vol. 5, no. 12, Dec. 2023, Art. no. 101182, doi: 10.1016/j.ajogmf.2023.101182.

[31] N. Tong, Y. Xu, J. Zhang, S. Gou, and M. Li, "Robust and efficient abdominal CT segmentation using shape constrained multi-scale attention network," *Phys. Medica*, vol. 110, Jun. 2023, Art. no. 102595, doi: 10.1016/j.ejmp.2023.102595.

[32] S. K. Hansen, P. J. Danaher, B. W. Starnes, H. W. Hollis, and B. T. Garland, "Accuracy evaluations of three ruptured abdominal aortic aneurysm mortality risk scores using an independent dataset," *J. Vascular Surg.*, vol. 70, no. 1, pp. 67–73, Jul. 2019, doi: 10.1016/j.jvs.2018.10.095.

[33] L. Bidaut, "Data and image processing for abdominal imaging," *Abdominal Imag.*, vol. 25, no. 4, pp. 341–360, Mar. 2000, doi: 10.1007/s002610000003.

[34] K. Magudia, C. P. Bridge, K. P. Andriole, and M. H. Rosenthal, "The trials and tribulations of assembling large medical imaging datasets for machine learning applications," *J. Digit. Imag.*, vol. 34, no. 6, pp. 1424–1429, Dec. 2021, doi: 10.1007/s10278-021-00505-7.

[35] S. F. Harb, A. Ali, M. Yousuf, S. Elshazly, and A. Farag, "G-SET-DCL: A guided sequential episodic training with dual contrastive learning approach for colon segmentation," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 20, no. 2, pp. 279–287, Jan. 2025, doi: 10.1007/s11548-024-03319-4.

[36] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: A literature review," *BMC Med. Imag.*, vol. 22, no. 1, pp. 1–13, Dec. 2022, doi: 10.1186/s12880-022-00793-7.

[37] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of CNN and transfer learning in medical imaging: Advantages, challenges, future scope," *Sustainability*, vol. 15, no. 7, p. 5930, Mar. 2023.

[38] *U.S. Simulation & Segmentation*. Accessed: Mar. 14, 2024. [Online]. Available: https://www.kaggle.com/datasets/ignaciorlando/ussimandsegm

[39] X. Liang, H. Zhang, L. Lin, and E. Xing, "Generative semantic manipulation with mask-contrasting GAN," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, in Lecture Notes in Computer Science, 2018, pp. 574–590, doi: 10.1007/978-3-030-01261-8_34.

[40] *Computed Tomography (CT) of the Abdomen*. Accessed: Mar. 14, 2024. [Online]. Available: https://www.kaggle.com/datasets/trainingdatapro/computed-tomography-ct-of-the-abdomen

[41] *Hello_World_Deep_Learning_SIIM*. Accessed: Mar. 14, 2024. [Online]. Available: https://www.kaggle.com/datasets/wfwiggins203/hello-world-deep-learning-siim

[42] P. Lakhani, D. L. Gray, C. R. Pett, P. Nagy, and G. Shih, "Hello world deep learning in medical imaging," *J. Digit. Imag.*, vol. 31, no. 3, pp. 283–289, Jun. 2018, doi: 10.1007/s10278-018-0079-6.

[43] *Abdomen MRI | Object Detection*. Accessed: Mar. 14, 2024. [Online]. Available: https://www.kaggle.com/datasets/imtkaggleteam/abdomen-mri-object-detection

[44] A. da Silva Rocha, L. von Diemen, D. Kretzer, S. Matos, J. R. Bernardi, and J. A. Magalhães, "Visceral adipose tissue measurements during pregnancy," PhysioNet, Cambridge, MA, USA, Tech. Rep. Version 1.0.0, 2020.

[45] *Kidney Tumor Segmentation Challenge(KITS 19)*. Accessed: Mar. 14, 2024. [Online]. Available: https://www.kaggle.com/datasets/sabahesaraki/kidney-tumor-segmentation-challengekits-19

[46] N. Heller, N. Sathianathen, A. Kalapara, E. Walczak, K. Moore, H. Kaluzniak, J. Rosenberg, P. Blake, Z. Rengel, M. Oestreich, J. Dean, M. Tradewell, A. Shah, R. Tejpaul, Z. Edgerton, M. Peterson, S. Raza, S. Regmi, N. Papanikolopoulos, and C. Weight, "The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes," 2019, *arXiv:1904.00445*.

[47] N. Heller et al., "The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101821, doi: 10.1016/j.media.2020.101821.

[48] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, M. Bach, and M. Segeroth, "TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images," *Radiol., Artif. Intell.*, vol. 5, no. 5, Sep. 2023, doi: 10.1148/ryai.230024.

[49] *RSNA 2023 Abdominal Trauma Detection | Kaggle*. Accessed: Mar. 14, 2024. [Online]. Available: https://www.kaggle.com/competitions/rsna-2023-abdominal-trauma-detection/data

[50] R. Martinek, R. Kahankova, R. Jaros, K. Barnova, A. Matonia, M. Jezewski, R. Czabanski, K. Horoba, and J. Jezewski, "Non-invasive fetal electrocardiogram extraction based on novel hybrid method for intrapartum ST segment analysis," *IEEE Access*, vol. 9, pp. 28608–28631, 2021, doi: 10.1109/ACCESS.2021.3058733.

[51] *ST Annotations of ADFECGDB Database | IEEE DataPort*. Accessed: Mar. 14, 2024. [Online]. Available: https://ieee-dataport.org/documents/st-annotations-adfecgdb-database

[52] Z. Gao, Q. Zong, Y. Wang, Y. Yan, Y. Wang, N. Zhu, J. Zhang, Y. Wang, and L. Zhao, "Laplacian salience-gated feature pyramid network for accurate liver vessel segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 10, pp. 3059–3068, Oct. 2023, doi: 10.1109/TMI.2023.3273528.

[53] *Liver Vessel Segmentation | IEEE DataPort*. Accessed: Mar. 14, 2024. [Online]. Available: https://ieee-dataport.org/documents/liver-vessel-segmentation

[54] S. Umirzakova, S. Ahmad, L. U. Khan, and T. Whangbo, "Medical image super-resolution for smart healthcare applications: A comprehensive survey," *Inf. Fusion*, vol. 103, Mar. 2024, Art. no. 102075, doi: 10.1016/j.inffus.2023.102075.

[55] I. C. Passos, P. Ballester, J. V. Pinto, B. Mwangi, and F. Kapczinski, "Big data and machine learning meet the health sciences," in *Personalized Psychiatry: Big Data Analytics in Mental Health*. Cham, Switzerland: Springer, 2019, pp. 1–13.

[56] A. Meyer, D. Zverinski, B. Pfahringer, J. Kempfert, T. Kuehne, S. H. Sündermann, C. Stamm, T. Hofmann, V. Falk, and C. Eickhoff, "Machine learning for real-time prediction of complications in critical care: A retrospective study," *Lancet Respiratory Med.*, vol. 6, no. 12, pp. 905–914, Dec. 2018.

[57] N. Noorbakhsh-Sabet, R. Zand, Y. Zhang, and V. Abedi, "Artificial intelligence transforms the future of health care," *Amer. J. Med.*, vol. 132, no. 7, pp. 795–801, Jul. 2019.

[58] T. Jaiswal and S. Jaiswal, "Deep learning in medicine," *Int. J. Trend Sci. Res. Develop.*, vol. 3, no. 4, pp. 212–217, Jun. 2019.

[59] F. Commandeur, P. J. Slomka, M. Goeller, X. Chen, S. Cadet, A. Razipour, P. McElhinney, H. Gransar, S. Cantu, R. J. H. Miller, A. Rozanski, S. Achenbach, B. K. Tamarappoo, D. S. Berman, and D. Dey, "Machine learning to predict the long-term risk of myocardial infarction and cardiac death based on clinical risk, coronary calcium, and epicardial adipose tissue: A prospective study," *Cardiovascular Res.*, vol. 116, no. 14, pp. 2216–2225, Dec. 2020.

[60] A. Rajkomar et al., "Scalable and accurate deep learning with electronic health records," *NPJ Digit. Med.*, vol. 1, no. 1, p. 18, May 2018.

[61] H. Zhang and Y. Qie, "Applying deep learning to medical imaging: A review," *Appl. Sci.*, vol. 13, no. 18, p. 10521, Sep. 2023, doi: 10.3390/app131810521.

[62] A. Rehman and F. G. Khan, "A deep learning based review on abdominal images," *Multimedia Tools Appl.*, vol. 80, no. 20, pp. 30321–30352, Aug. 2021, doi: 10.1007/s11042-020-09592-0.

[63] A. Saleh, M. M. Hasan, H. W. Raadsma, M. S. Khatkar, D. R. Jerry, and M. Rahimi Azghadi, "Prawn morphometrics and weight estimation from images using deep learning for landmark localization," *Aquacultural Eng.*, vol. 106, Aug. 2024, Art. no. 102391, doi: 10.1016/j.aquaeng.2024.102391.

[64] A. J. R. Simpson, C. Y. Hsu, M. I. Ojaghi, A. B. Labrique, R. Haque, P. M. Akhter, Y. Chen, and P. J. Hotez, "Abdominal CT by use of new deep," *Amer. J. Roentgenol.*, vol. 215, pp. 1–8, Jul. 2020.

[65] L.-L. Li, H. Wang, J. Song, J. Shang, X.-Y. Zhao, and B. Liu, "A feasibility study of realizing low-dose abdominal CT using deep learning image reconstruction algorithm," *J. X-Ray Sci. Technol.*, vol. 29, no. 2, pp. 361–372, Mar. 2021, doi: 10.3233/xst-200826.

[66] J. A. van Stiphout, J. Driessen, L. R. Koetzier, L. B. Ruules, M. J. Willemink, J. W. T. Heemskerk, and A. J. van der Molen, "The effect of deep learning reconstruction on abdominal CT densitometry and image quality: A systematic review and meta-analysis," *Eur. Radiol.*, vol. 32, no. 5, pp. 2921–2929, May 2022, doi: 10.1007/s00330-021-08438-z.

[67] T. Kaga, Y. Noda, K. Fujimoto, T. Suto, N. Kawai, T. Miyoshi, F. Hyodo, and M. Matsuo, "Deep-learning-based image reconstruction in dynamic contrast-enhanced abdominal CT: Image quality and lesion detection among reconstruction strength levels," *Clin. Radiol.*, vol. 76, no. 9, pp. 710.e15–710.e24, Sep. 2021, doi: 10.1016/j.crad.2021.03.010.

[68] T. Njølstad, A. Schulz, J. C. Godt, H. M. Brøgger, C. K. Johansen, H. K. Andersen, and A. C. T. Martinsen, "Improved image quality in abdominal computed tomography reconstructed with a novel deep learning image reconstruction technique–initial clinical experience," *Acta Radiologica Open*, vol. 10, no. 4, Apr. 2021, Art. no. 205846012110083, doi: 10.1177/20584601211008391.

[69] Z. Liu, H. Zhao, X. Fang, and D. Huo, "Abdominal computed tomography localizer image generation: A deep learning approach," *Comput. Methods Programs Biomed.*, vol. 214, Feb. 2022, Art. no. 106575.

[70] P. Barca, S. Domenichelli, R. Golfieri, L. Pierotti, L. Spagnoli, S. Tomasi, and L. Strigari, "Image quality evaluation of the precise image CT deep learning reconstruction algorithm compared to filtered back-projection and iDose4: A phantom study at different dose levels," *Phys. Medica*, vol. 106, Feb. 2023, Art. no. 102517, doi: 10.1016/j.ejmp.2022.102517.

[71] D. Racine, F. Becce, A. Viry, P. Monnin, B. Thomsen, F. R. Verdun, and D. C. Rotzinger, "Task-based characterization of a deep learning image reconstruction and comparison with filtered back-projection and a partial model-based iterative reconstruction in abdominal CT: A phantom study," *Phys. Medica*, vol. 76, pp. 28–37, Aug. 2020, doi: 10.1016/j.ejmp.2020.06.004.

[72] A. D. Weston, P. Korfiatis, T. L. Kline, K. A. Philbrick, P. Kostandy, T. Sakinis, M. Sugimoto, N. Takahashi, and B. J. Erickson, "Automated abdominal segmentation of CT scans for body composition analysis using deep learning," *Radiology*, vol. 290, no. 3, pp. 669–679, Mar. 2019, doi: 10.1148/radiol.2018181432.

[73] H. Kaur, N. Kaur, and N. Neeru, "Evolution of multiorgan segmentation techniques from traditional to deep learning in abdominal CT images—A systematic review," *Displays*, vol. 73, Jul. 2022, Art. no. 102223, doi: 10.1016/j.displa.2022.102223.

[74] C. Caradu, B. Spampinato, A. M. Vrancianu, X. Bérard, and E. Ducasse, "Fully automatic volume segmentation of infrarenal abdominal aortic aneurysm computed tomography images with deep learning approaches versus physician controlled manual segmentation," *J. Vascular Surg.*, vol. 74, no. 1, pp. 246–256.e6, Jul. 2021, doi: 10.1016/j.jvs.2020.11.036.

[75] J. Halkoaho, O. Niiranen, E. Salli, T. Kaseva, S. Savolainen, M. Kangasniemi, and H. Hakovirta, "Quantifying the calcification of abdominal aorta and major side branches with deep learning," *Clin. Radiol.*, vol. 79, no. 5, pp. e665–e674, May 2024, doi: 10.1016/j.crad.2024.01.023.

[76] D. Kim, H. Wit, M. Thurston, M. Long, G. Maskell, M. Strugnell, D. Shetty, I. Smith, and N. Hollings, "An artificial intelligence deep learning model for identification of small bowel obstruction on plain abdominal radiographs," *Brit. J. Radiol.*, vol. 94, no. 1122, Jun. 2021, doi: 10.1016/j.jocn.2024.01.032.

[77] S. A. Elhage, E. B. Deerenberg, S. A. Ayuso, K. J. Murphy, J. M. Shao, K. W. Kercher, N. J. Smart, J. P. Fischer, V. A. Augenstein, P. D. Colavita, and B. T. Heniford, "Development and validation of image-based deep learning models to predict surgical complexity and complications in abdominal wall reconstruction," *JAMA Surg.*, vol. 156, no. 10, p. 933, Oct. 2021, doi: 10.1001/jamasurg.2021.3012.

[78] H. Yonezawa, D. Ueda, A. Yamamoto, K. Kageyama, S. L. Walston, T. Nota, K. Murai, S. Ogawa, E. Sohgawa, A. Jogo, D. Kabata, and Y. Miki, "Maskless 2-Dimensional digital subtraction angiography generation model for abdominal vasculature using deep learning," *J. Vascular Interventional Radiol.*, vol. 33, no. 7, pp. 845–851.e8, Jul. 2022, doi: 10.1016/j.jvir.2022.03.010.

[79] F. Michallek, U. Genske, S. M. Niehues, B. Hamm, and P. Jahnke, "Deep learning reconstruction improves radiomics feature stability and discriminative power in abdominal CT imaging: A phantom study," *Eur. Radiol.*, vol. 32, no. 7, pp. 4587–4595, Jul. 2022, doi: 10.1007/s00330-022-08592-y.

[80] D. Müller, I. Soto-Rey, and F. Kramer, "Towards a guideline for evaluation metrics in medical image segmentation," *BMC Res. Notes*, vol. 15, no. 1, pp. 1–8, Dec. 2022, doi: 10.1186/s13104-022-06096-y.

[81] M. Moassefi, P. Rouzrokh, G. M. Conte, S. Vahdati, T. Fu, A. Tahmasebi, M. Younis, K. Farahani, A. Gentili, T. Kline, F. C. Kitamura, Y. Huo, S. Kuanar, K. Younis, B. J. Erickson, and S. Faghani, "Reproducibility of deep learning algorithms developed for medical imaging analysis: A systematic review," *J. Digit. Imag.*, vol. 36, no. 5, pp. 2306–2312, Jul. 2023, doi: 10.1007/s10278-023-00870-5.

[82] C.-T. Cheng, H.-S. Lin, C.-P. Hsu, H.-W. Chen, J.-F. Huang, C.-Y. Fu, C.-H. Hsieh, C.-N. Yeh, I.-F. Chung, and C.-H. Liao, "The three-dimensional weakly supervised deep learning algorithm for traumatic splenic injury detection and sequential localization: An experimental study," *Int. J. Surg.*, vol. 109, no. 5, pp. 1115–1124, May 2023, doi: 10.1097/js9.0000000000000380.

[83] D.-M. Koh, N. Papanikolaou, U. Bick, R. Illing, C. E. Kahn, J. Kalpathi-Cramer, C. Matos, L. Martí-Bonmatí, A. Miles, S. K. Mun, S. Napel, A. Rockall, E. Sala, N. Strickland, and F. Prior, "Artificial intelligence and machine learning in cancer imaging," *Commun. Med.*, vol. 2, no. 1, pp. 1–14, Oct. 2022, doi: 10.1038/s43856-022-00199-0.

[84] *Sensitivity and Specificity | Radiology Reference Article | Radiopaedia*. Accessed: Mar. 16, 2024. [Online]. Available: https://radiopaedia.org/articles/sensitivity-and-specificity

[85] T.-W. Loong, "Understanding sensitivity and specificity with the right side of the brain," *BMJ*, vol. 327, no. 7417, pp. 716–719, Sep. 2003, doi: 10.1136/bmj.327.7417.716.

[86] S. Weinstein, N. A. Obuchowski, and M. L. Lieber, "Clinical evaluation of diagnostic tests," *Amer. J. Roentgenol.*, vol. 184, no. 1, pp. 14–19, Jan. 2005, doi: 10.2214/ajr.184.1.01840014.

[87] *The Current Status and Future Potential of AI in Medical Imaging*. Accessed: Mar. 16, 2024. [Online]. Available: https://insights.omnia-health.com/technology/current-status-and-future-potential-ai-medical-imaging

[88] *Jaccard Coefficient—An Overview | ScienceDirect Topics*. Accessed: Mar. 16, 2024. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/jaccard-coefficient

[89] E. Trucco, A. McNeil, S. McGrory, L. Ballerini, M. R. K. Mookiah, S. Hogg, A. Doney, and T. MacGillivray, "Validation," in *Computational Retinal Image Analysis: Tools, Applications and Perspectives*. Amsterdam, The Netherlands: Elsevier, Jan. 2019, pp. 157–170, doi: 10.1016/B978-0-08-102816-2.00009-5.

[90] *Dice Similarity Coefficient | Radiology Reference Article | Radiopaedia*. Accessed: Mar. 16, 2024. [Online]. Available: https://radiopaedia.org/articles/dice-similarity-coefficient

[91] S. Tigges, "Diagnostic test metrics and receiver operating characteristic curves: The graphic novel," *J. Amer. College Radiol.*, vol. 19, no. 9, pp. 1034–1041, Sep. 2022, doi: 10.1016/j.jacr.2022.06.011.

[92] J. Liu, R. Jiang, X. Liu, F. Zhou, Y. Chen, and C. Shen, "Large-scale clustering on 100 M-scale datasets using a single T4 GPU via recall KNN and subgraph segmentation," *Neural Process. Lett.*, vol. 56, no. 1, Feb. 2024, doi: 10.1016/j.jacr.2022.06.015.

[93] X. Luo, W. Liao, J. Xiao, J. Chen, T. Song, X. Zhang, K. Li, D. N. Metaxas, G. Wang, and S. Zhang, "WORD: A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from CT image," *Med. Image Anal.*, vol. 82, Nov. 2022, Art. no. 102642, doi: 10.1016/j.media.2022.102365.

[94] S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, and M. P. Lungren, "Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines," *NPJ Digit. Med.*, vol. 3, no. 1, Oct. 2020, doi: 10.1038/s41746-020-00341-z.

[95] K. Raza, "A tour of unsupervised deep learning for medical image analysis," *Current Med. Imag. Rev.*, vol. 17, no. 9, pp. 1059–1077, 2021, doi: 10.2174/1573405617666210127154257.

[96] S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, "Self-supervised learning for medical image classification: A systematic review and implementation guidelines," *NPJ Digit. Med.*, vol. 6, no. 1, Apr. 2023, doi: 10.1038/s41746-023-00811-0.

[97] M. Mittermaier, M. Raza, and J. C. Kvedar, "Collaborative strategies for deploying AI-based physician decision support systems: Challenges and deployment approaches," *NPJ Digit. Med.*, vol. 6, no. 1, Aug. 2023, doi: 10.1038/s41746-023-00889-6.

[98] L. Petersson, I. Larsson, J. M. Nygren, P. Nilsen, M. Neher, J. E. Reed, D. Tyskbo, and P. Svedberg, "Challenges to implementing artificial intelligence in healthcare: A qualitative interview study with healthcare leaders in Sweden," *BMC Health Services Res.*, vol. 22, no. 1, pp. 1–16, Dec. 2022, doi: 10.1186/s12913-022-08215-8.

[99] K. Borys, Y. A. Schmitt, M. Nauta, C. Seifert, N. Krämer, C. M. Friedrich, and F. Nensa, "Explainable AI in Medical Imaging: An overview for clinical practitioners–saliency-based XAI approaches," *Eur. J. Radiol.*, vol. 162, May 2023, Art. no. 110787.

[100] K. Borys, Y. A. Schmitt, M. Nauta, C. Seifert, N. Krämer, C. M. Friedrich, and F. Nensa, "Explainable AI in medical imaging: An overview for clinical practitioners–beyond saliency-based XAI approaches," *Eur. J. Radiol.*, vol. 162, May 2023, Art. no. 110786.

[101] J. R. Geis, A. Brady, C. C. Wu, J. Spencer, E. Ranschaert, J. L. Jaremko, S. G. Langer, A. B. Kitts, J. Birch, W. F. Shields, R. van den Hoven van Genderen, E. Kotter, J. W. Gichoya, T. S. Cook, M. B. Morgan, A. Tang, N. M. Safdar, and M. Kohli, "Ethics of artificial intelligence in radiology: Summary of the joint European and North American multisociety statement," *Insights Imag.*, vol. 10, no. 1, Dec. 2019, doi: 10.1016/j.juro.2023.01.020.

[102] W. Wang, Y. Yang, Z. Yin, K. Dev, X. Zhou, X. Li, N. M. F. Qureshi, and C. Su, "BSIF: Blockchain-based secure, interactive, and fair mobile crowdsensing," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 12, pp. 3452–3469, Dec. 2022.

[103] W. Wang, D. Niyato, Z. Xiong, and Z. Yin, "Detection and authentication for cross-technology communication," *IEEE Trans. Veh. Technol.*, vol. 74, no. 2, pp. 3157–3171, Feb. 2025.

[104] Z. Liu, W. Wang, Y. Ye, N. Min, Z. Cao, L. Zhou, and Z. Liu, "Collusion-resilient and maliciously secure Cloud- assisted two-party computation scheme in mobile cloud computing," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 7019–7032, 2024.

[105] M. Mittermaier, M. M. Raza, and J. C. Kvedar, "Bias in AI-based models for medical applications: Challenges and mitigation strategies," *NPJ Digit. Med.*, vol. 6, no. 1, Jun. 2023, doi: 10.1109/ACCESS.2023.10601195.
[106] K. N. Vokinger, S. Feuerriegel, and A. S. Kesselheim, "Mitigating bias in machine learning for medicine," *Commun. Med.*, vol. 1, no. 1, pp. 3–5, Aug. 2021, doi: 10.1038/s43856-021-00028-w.
[107] F. Liao, S. Adelaine, M. Afshar, and B. W. Patterson, "Governance of clinical AI applications to facilitate safe and equitable deployment in a large health system: Key elements and early successes," *Frontiers Digit. Health*, vol. 4, pp. 1–8, Aug. 2022, doi: 10.3389/fdgth.2022.931439.
[108] *FDA Publishes List of AI-enabled Medical Devices | UW Radiology*. Accessed: Feb. 5, 2025. [Online]. Available: https://rad.washington.edu/news/fda-publishes-list-of-ai-enabled-medical-devices/

**MARIEM BELLAL** (Member, IEEE) received the bachelor's degree in mathematics and computer science from the University of Nouakchott Al Aasriya, in 2018, and the master's degree in data science and big data from ENSIAS, in 2020. She is currently a Researcher in artificial intelligence with the National School of Computer Science and Systems Analysis (ENSIAS), Mohammed V University in Rabat. Her research focuses on the application of deep learning in medical imaging, with a particular interest in detecting and classifying traumatic abdominal injuries. She has participated in several international conferences and events, including the Affordable AI Summer School with the University of Bonn and GITEX Africa 2024 in Marrakesh, gaining valuable insights into the latest advances in AI, the IoT, and cybersecurity.

**SANAA EL FKIHI** received the bachelor's degree (License en Sciences) in computer science, in 2002, and the master's degree in computer science and telecommunications, in 2004, and the joint Ph.D. degree in computer science from Mohammed V University in Rabat, Morocco, and the University of Science and Technology of Lille, in 2008. She is currently a Full Professor of computer engineering with the Higher National School of Computer Science and Systems Analysis (ENSIAS), Mohammed V University in Rabat. Her current research interests include computer vision, pattern recognition, machine learning, and deep learning.

**KORHAN CENGIZ** (Senior Member, IEEE) was born in Edirne, Turkey, in 1986. He received the B.Sc. degree in electronics and communication engineering from Kocaeli University and Business Administration, the B.Sc. degree from Anadolu University, Turkey, in 2008 and 2009, respectively, the M.Sc. degree in electronics and communication engineering from Namik Kemal University, Turkey, in 2011, and the Ph.D. degree in electronics engineering from Kadir Has University, Turkey, in 2016. Since October 2023, he is at the world top 2% scientist list (2023, 2024). Since January 2025, he is an Associate Research Professor at the Department of Electrical Engineering, Prince Mohammad Bin Fahd University, KSA. Also, since August 2024, he has been an Associate Professor at Biruni University, Turkiye. Moreover since June 2024, he has been an Adjunct Professor at the University of Jordan, Jordan. From September 2022 to August 2024, he worked as an Associate Professor with the Department of Computer Engineering, Istinye University, Istanbul, Turkey. From April 2023 to April 2024, he was a Senior Researcher with the Department of Information Technologies, University of Hradec Kralove, Kralove, Czech Republic. He is the author of more than 60 SCI / SCI-E articles including IEEE Internet of Things Journal, IEEE Access, *Expert Systems with Applications*, *Knowledge Based Systems*, and *ACM Transactions on Sensor Networks*, five international patents, more than 20 book chapters, and one book in Turkish. He is editor of more than 30 books. His research interests include wireless sensor networks, wireless communications, statistical signal processing, indoor positioning systems, the Internet of Things, power electronics, and 5G. He is the Associate Editor of IEEE Transactions on Intelligent Transportation Systems, IEEE Internet of Things Journal. He serves several reviewer positions for IEEE Internet of Things Journal, IEEE Sensors Journal and IEEE Access. He serves several book editor positions in IEEE, Springer, Elsevier, Wiley, and CRC. He presented 40+ keynote talks in reputed IEEE and Springer Conferences about WSNs, the IoT, and 5G. He is a Professional Member of ACM. His awards and honors include the Tubitak Priority Areas Ph.D. Scholarship, the Kadir Has University Ph.D. Student Scholarship, best presentation award in ICAT 2016 Conference and best paper award in ICAT 2018 Conference.

**NIKOLA IVKOVIĆ** (Senior Member, IEEE) was born in Zagreb, Croatia, in 1979. He received the M.S. degree in computing and the Ph.D. degree in computer science from the Faculty of Electrical Engineering and Computing, University of Zagreb. His doctoral thesis was in the area of swarm and evolutionary computation. He was the Head of the Department of Computing and Technology, Faculty of Organization and Informatics, University of Zagreb where he is currently an Associate Professor. His research interests include computational intelligence and optimization, computer networks, security, and formal methods. He teaches computer networks, operating systems, and computer architecture related courses. He was a member of committees for creating new university study programs. He is a member of two research laboratories–Artificial Intelligence Laboratory and the Laboratory for Generative Programming and Machine Learning, both at the Faculty of Organization and Informatics, University of Zagreb. He serves as a regular reviewer for high-quality scientific journals and takes part in a number of international conference committees. He gave several invited talks at international scientific conferences and guest lectures at different universities in Europe and Asia. He joined the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronics Engineers (IEEE) in 2008.