Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253218

ABDOMINAL IMAGING

ORIGINAL ARTICLE

Detection of common bile duct dilatation on magnetic resonance cholangiopancreatography by deep learning

Hilal Er Ulubaba[1]
Rukiye Çiftçi[2]
İpek Atik[3]
Osman Furkan Karakuş[4]

[1]İnönü University Faculty of Medicine, Department of Radiology, Malatya, Türkiye

[2]Gaziantep İslam Science and Technology University Faculty of Medicine, Department of Anatomy, Gaziantep, Türkiye

[3]Gaziantep İslam Science and Technology University Faculty of Engineering and Natural Sciences, Department of Electrical Electronics Engineering, Gaziantep, Türkiye

[4]Yıldız Technical University, Department of Computer Engineering, İstanbul, Türkiye

PURPOSE
This study aims to detect common bile duct (CBD) dilatation using deep learning methods from artificial intelligence algorithms.

METHODS
To create a convolutional neural network (CNN) model, 77 magnetic resonance cholangiopancreatography (MRCP) images without CBD dilatation and 70 MRCP images with CBD dilatation were used. The system was developed using coronal maximum intensity projection reformatted 3D-MRCP images. The ResNet50, DenseNet121, and visual geometry group models were selected for training, and detailed training was performed on each model.

RESULTS
In the study, the DenseNet121 model showed the best performance, with a 97% accuracy rate. The ResNet50 model ranked second, with a 96% accuracy rate.

CONCLUSION
CBD dilatation was detected with high performance using the DenseNet CNN model. Once validated in multicenter studies with larger datasets, this method may help in diagnosis and treatment decision-making.

CLINICAL SIGNIFICANCE
Deep learning algorithms can aid clinicians and radiologists in the diagnostic process once technical, ethical, and financial limitations are addressed. Fast and accurate diagnosis is crucial for accelerating treatment, reducing complications, and shortening hospital stays.

KEYWORDS
Artificial intelligence, bile duct dilatation, choledocholithiasis, convolutional neural network, magnetic resonance cholangiopancreatography

The common bile duct (CBD) is approximately 7–10 cm long and 3–6 mm in diameter, and its diameter increases slightly with age. The CBD functions as a duct that stores bile produced in the liver and empties it into the small intestine. Bile, a digestive fluid, plays an important role in breaking down fats during the digestive process. CBD dilation refers to a condition in which the bile ducts dilate beyond the normal range. This condition can be caused by a variety of factors, including bile duct obstruction, the presence of gallstones, infections affecting the bile ducts, or other diseases that block the flow of bile, causing the bile ducts to dilate.[1] CBD dilatation is a common clinical symptom that can result from various conditions, including pancreatobiliary tumors, choledocholithiasis, and periampullary diverticula.[2,3] Non-pathological causes of bile duct dilatation, such as advanced age, previous surgical interventions, and chronic narcotic use, are also widely recognized.[4] Although patients with CBD dilatation may present with colicky pain, fever, jaundice, and other clinical

Corresponding author: Hilal Er Ulubaba

E-mail: erhilal44@yahoo.com

Received 08 January 2025; revision requested 24 January 2025; last revision received 25 February 2025; accepted 05 April 2025.

Epub: 05.05.2025

Publication date: 06.11.2025

DOI: 10.4274/dir.2025.253218

You may cite this article as: Er Ulubaba H, Çiftçi R, Atik İ, Karakuş OF. Detection of common bile duct dilatation on magnetic resonance cholangiopancreatography by deep learning. *Diagn Interv Radiol.* 2025;31(6):532-538.

symptoms, a considerable proportion of patients remain asymptomatic. Nevertheless, the degree of dilation in the bile duct often serves as an essential indicator for assessing disease severity and guiding treatment selection.[5] Bile duct dilatation is a condition that can negatively affect the digestive system and can cause greater health problems if left untreated. Early diagnosis and appropriate treatment are important for preventing this problem and maintaining health.[6]

Diagnostic modalities for CBD dilatation include transabdominal ultrasonography (US), computed tomography (CT), magnetic resonance cholangiopancreatography (MRCP), and endoscopic retrograde cholangiopancreatography (ERCP). Among these, MRCP plays a well-established role in investigating biliary disorders and serves as a non-invasive alternative to ERCP.[7]

Treatment often becomes necessary in cases of CBD dilatation caused by stones. Early detection of CBD dilatation is crucial for enabling the timely initiation of treatment modalities. Artificial intelligence applications (AIAs) are increasingly utilized across various domains and are gradually being integrated into healthcare.[7] They facilitate disease diagnosis for physicians and shorten the time to treatment initiation. This can provide opportunities for early treatment and improve treatment success rates.[8]

In the healthcare industry, machine learning techniques are becoming increasingly common. As the term suggests, machine learning allows algorithms to learn and extract meaningful representations from data in a semi-automatic manner. Early diagnosis and treatment of biliary dilatation enable the detection of abnormalities in the biliary tract in a shorter time and shorten the treatment period.

Deep learning is one of the artificial learning approaches based on convolutional neural network (CNN) models, also known as multi-layer neural networks. These models are advanced feedforward neural networks widely used in image analysis, natural language processing, and other complex image classification problems. They are uniquely capable of identifying and interpreting patterns from images and text.[9,10]

Main points

- Common bile duct (CBD) dilation was detected with high performance from magnetic resonance cholangiopancreatography images.
- CBD dilation was also detected with high accuracy using the DenseNet deep learning algorithm.
- Artificial intelligence can assist clinicians and radiologists in the early and accurate diagnosis of bile duct dilation.

Deep learning is widely used in automatic image segmentation and also in medical image processing. Due to its superior performance, deep learning has emerged as the most popular technology.[11,12] Considering the importance of deep learning, this paper aims to analyze bile duct dilatation using ResNet50, DenseNet121, and visual geometry group (VGG) models.

This study aims to determine the diagnostic performance of deep learning algorithms in detecting CBD dilatation caused by bile duct stones using MRCP images.

Methods

This was a retrospective study, and ethical approval was obtained from Gaziantep Islam Science and Technology University in 2024 with decision number 2024/465. In this study, MRCP images of individuals aged 18–65 years were retrospectively analyzed. The images were obtained using a 1.5 Tesla magnetic resonance imaging system (Siemens Vision-Symphony Upgrade, Erlangen, Germany) and reviewed by a radiologist with approximately 13 years of clinical experience. The presence of stone-associated CBD dilatation was evaluated using coronal maximum intensity projection (MIP) reformatted 3D-MRCP images. A total of 147 MRCP images were included in this study, comprising 77 images from the normal group and 70 from patients with bile duct dilatation caused by choledocholithiasis. The dataset consisted of 147 MRCP images collected from [source]. The images were categorized based on the presence of stone-associated bile duct dilatation. Preprocessing steps included contrast enhancement, normalization, and resizing to 224 × 224 pixels. The exclusion criteria included individuals with cholecystectomy, images with artifacts in which the CBD was not visible, patients with tumors in the bile ducts, and those without a confirmed diagnosis of choledocholithiasis through clinical evaluation or ERCP.

MRCP artifacts, technical and reconstruction-related artifacts, gas-related artifacts, and other fixed fluids (such as those in the duodenum or ascites) may result from the overlap of the CBD. Poor spatial resolution may also limit interpretation due to artifacts. Examination of thin sections and multiple planes reduces these issues. However, coronal MIP reformatted 3D-MRCP images were used. Therefore, if the CBD was not visible in coronal MIP reformatted 3D-MRCP images, those images were excluded.

Patients with detected stones and dilatation on MRCP along with a confirmed diagnosis of stones via ERCP were assigned to the CBD dilatation group. Patients with normal bile ducts on MRCP who clinically improved without requiring ERCP or were found to have no stones on ERCP were assigned to the normal group.

To enhance generalization, five-fold cross-validation was employed. In this approach, the dataset was divided into five subsets, with four subsets used for training and one for validation in each iteration. This ensured that the model learned from different portions of the data, enhancing robustness. To mitigate overfitting, dropout layers (rate: 0.3) were applied, and L2 weight regularization was used. Additionally, early stopping was employed to prevent excessive training on the same patterns.

Study design

The images, stored in Digital Imaging and Communications in Medicine format, were transferred to a personal workstation using the Horos Medical Image Viewer (Version 3.0; Horos Project, Annapolis, MD, USA). The images were analyzed using a deep learning approach, a subset of AI. Initially, all images underwent preprocessing steps such as cropping, padding, smoothing, and resizing to fit the input layer of the deep learning architecture. The processed images were then fed into the input layer of the architecture. In the subsequent step, features were extracted from the fully connected layers of the architecture. These features were saved in a file and used as training and test data for classifier methods.[13]

Pre-trained CNN models were employed as feature extractors. The features were obtained from the fully connected layers of each model, with a total of 1,000 features per model. Training and testing processes were conducted using all features, as well as subsets selected through feature selection, to analyze the impact of feature selection on system performance. Various numbers of features were tested to assess their effect on system success.

CNNs operate on specific-sized patches of images, processing them piece by piece through comparison. Filters are spatially

small along the width and height but extend through the entire depth of the input image. They are designed to detect specific feature types in the input image. In the convolutional layer, the filter/kernel is moved to every possible position over the input matrix. Element-wise multiplication is performed between the filter and the patch of the input image, followed by summation. This process is repeated for every possible position of the filter on the input image matrix, enabling feature detection at any location in the image.[14]

Deep learning algorithms

This study employed a deep learning-based process to classify images into normal and choledocholithiasis-associated CBD dilatation groups. In the initial stage, raw datasets were prepared, consisting of normal and CBD dilatation images. During the data preprocessing phase, the images were resized, and data augmentation techniques were applied. The ResNet50, DenseNet121, and VGG models were selected for the training process, with detailed training conducted on each model. The models' performance was evaluated using metrics such as accuracy, F1 score, and the confusion matrix. Finally, the results were compared, and the model with the best performance was identified. The process flowchart is presented in Figure 1.

Dataset

This study utilized a dataset comprising a total of 147 images categorized into two classes: normal and stone-associated CBD dilatation. Of these, 77 images belonged to the normal group, and 70 images were classified as CBD dilatation. The images were subjected to various preprocessing techniques to prepare them for deep learning models. During preprocessing, all images were resized to 224 × 224 pixels, and pixel values were normalized between 0 and 1. The dataset was split into two subsets: 80% for training and 20% for testing. Additionally, to enhance model generalization and mitigate overfitting, five-fold cross-validation was applied, ensuring that different subsets of the data were used for training and validation in each iteration.

Data augmentation

To increase the diversity of the dataset and prevent model overfitting during training, data augmentation techniques were applied. Data augmentation techniques included rotation (–20° to +20°), horizontal and vertical flipping, zoom (0–10%), brightness adjustment, and contrast normalization. These transformations improved model robustness. The following data augmentation methods were used:

1. Horizontal flipping: Random flipping of images along the horizontal axis.

2. Rotation: Random rotation of images between 0° and 30°.

3. Brightness adjustment: Random modification of image brightness levels.

4. Random cropping: Cropping specific portions of the images.

These data augmentation techniques were dynamically applied during each epoch to enhance variability during training. By applying augmentation, the total number of images increased to 735.

Image examples with data augmentation

Figure 2 presents examples of images processed with data augmentation techniques.

1. Original image: This represents the raw image with no data augmentation applied.

2. Horizontal flipping: The image is flipped along the horizontal axis, swapping the left and right sides.

3. Rotation: The image is rotated randomly between 0° and 30°.

4. Brightness adjustment: The brightness level of the image is either increased or decreased.

5. Random cropping: A random portion of the image is cropped and then resized to the original dimensions.

Models

In this study, we selected three widely used CNN architectures—ResNet50, DenseNet121, and VGG16—to evaluate their effectiveness in detecting stone-associated CBD dilatation. Each model was chosen based on its specific strengths in feature extraction, classification accuracy, and computational efficiency.

• **ResNet50:** This model was selected for its residual learning framework, which effectively addresses the vanishing gradient problem in deep networks. The use of residual connections allows for deeper architectures while maintaining efficient feature propagation, making it well-suited for complex medical image classification.

• **DenseNet121:** This model was incorporated due to its dense connectivity mechanism, which enables improved gradient flow and feature reuse, leading to enhanced model efficiency. Compared with ResNet50, DenseNet121 requires fewer parameters while maintaining high classification accuracy, making it advantageous for medical imaging tasks where computational efficiency is crucial.

• **VGG16:** VGG16 was included as a baseline model, as it has been extensively used in medical image analysis studies. Its simple yet effective architecture allows for a controlled comparison against more advanced CNNs while providing insights into the relative advantages of deep feature extraction techniques.

By selecting these three architectures, this study provided a comprehensive evaluation of different deep learning models, ensuring a balanced comparison of accuracy, computational cost, and real-world applicability for CBD dilatation detection.

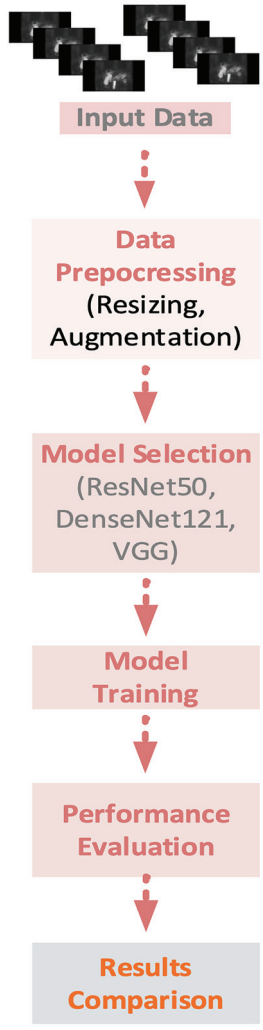

Figure 1. Flux diagram. VGG, visual geometry group.

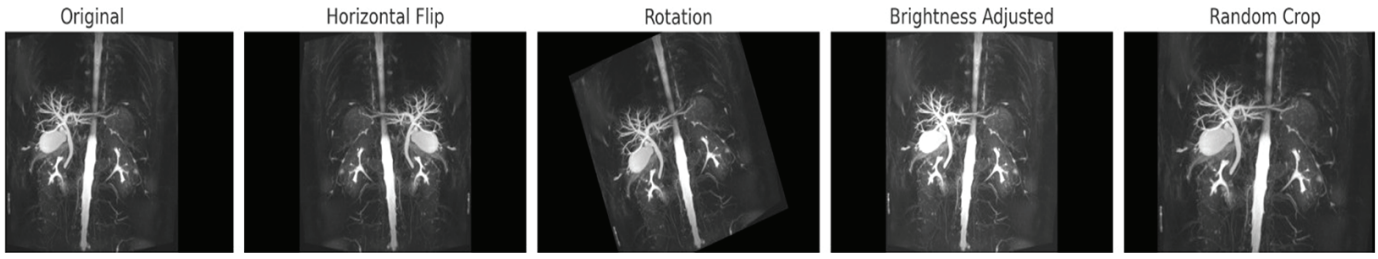

Figure 2. Augmented MRCP images.

Training process

The models were trained for a total of 100 epochs, with training and validation losses monitored during each epoch. The Adam optimization algorithm was used to update model weights. The cross-entropy loss function, suitable for multi-class classification problems, was employed for loss calculation. The learning rate was set at 0.001, and the mini-batch size was chosen as 32. To ensure optimal model performance, key hyperparameters were carefully selected and fine-tuned during the training phase. The models were trained using the Adam optimizer with a learning rate of 0.0001, where beta1: 0.9 and beta2: 0.999. The batch size was set to 16, balancing computational efficiency with stable convergence. The models were trained for 50 epochs, applying categorical cross-entropy as the loss function due to its effectiveness in multi-class classification problems.

Additionally, learning rate decay was incorporated using a step decay strategy, gradually reducing the learning rate to prevent the model from getting stuck in local minima and to enhance convergence stability. These hyperparameter choices were made based on preliminary experiments to optimize model accuracy while preventing overfitting. The detailed tuning of these parameters contributed to the improved generalizability of the deep learning models for detecting CBD dilatation.

Performance evaluation

The models' performance was evaluated using metrics commonly employed in classification problems, such as precision, recall, and F1 score. These metrics analyze the prediction success of the model from different perspectives:

• **Precision:** This is the ratio of true positive (TP) predictions to the total positive predictions. It measures the impact of false positive (FP) predictions.

• **Recall:** This is the ratio of correctly classified positive samples to the total actual positive samples. It measures the impact of false negative (FN) predictions.

• **F1 Score:** This metric aims to balance precision and recall, representing their harmonic mean. It is particularly useful when there is an imbalance in data distribution between classes.

The formulas for these metrics are provided in Equations 1–5, and the evaluation results are presented in Table 1.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

In Formulas 1–5, the following abbreviations are used:

• **TP**: The number of correctly predicted positive cases.

• **FP**: The number of cases incorrectly predicted as positive.

• **FN**: The number of cases incorrectly predicted as negative.

• **TN (true negatives)**: The number of correctly predicted negative cases.

These formulas are derived from standard classification metric definitions and are widely used in deep learning-based medical image classification.[14]

Results

To address learning challenges in deep networks, this model was evaluated using the 50-layer ResNet50 architecture, which employs residual connections. The ResNet model achieved high performance with a minimum loss value of 0.0926, precision of 0.9642, recall of 0.9705, and F1 score of 0.9663.

The VGG16 architecture, characterized by a fixed structure of 3 × 3 filters, was also evaluated. Its minimum loss value was measured at 0.6752. However, it demonstrated lower performance than the other models, with a precision of 0.2166, recall of 0.5102, and F1 score of 0.3023.

The DenseNet121 model, which connects each layer to all preceding layers to enhance learning efficiency, achieved a minimum loss value of 0.1384. Its performance metrics were a precision of 0.9722, recall of 0.9615, and F1 score of 0.9657.

The loss values throughout the training process are presented in Figure 3, and the changes in precision, recall, and F1 score are illustrated in Figure 4.

The analysis revealed that the deep learning-based ResNet, VGG, and DenseNet models demonstrated high performance in classifying normal and dilated images, with the ResNet and DenseNet models showing low loss values and high precision, recall, and F1 scores. Specifically, DenseNet emerged as an effective model in terms of learning efficiency. In contrast, the VGG model performed relatively poorly compared with the others. The data augmentation techniques and training processes used in this study supported the models' general feature learning and enhanced classification success.

Discussion

In this study, which aimed to determine the diagnostic performance of CBD dilation in patients with choledocholithiasis from MRCP images using the CNN method, the highest accuracy was achieved with the DenseNet algorithm at 97%. MRCP can effectively show changes in the bile ducts; however, MRCP images can have many artifacts and low resolution, which may complicate diagnosis. In contrast, deep learning methods

Table 1. Performance evaluation metrics

	Model	Accuracy	Precision	Recall	Specificity	F1-score
ResNet	ResNet50	0.9682	0.9642	0.9705	0.9661	0.9663
VGG	VGG16	0.3923	0.2166	0.5102	0.3513	0.3023
DenseNet	DenseNet121	0.9701	0.9722	0.9615	0.9772	0.9657

VGG, visual geometry group.

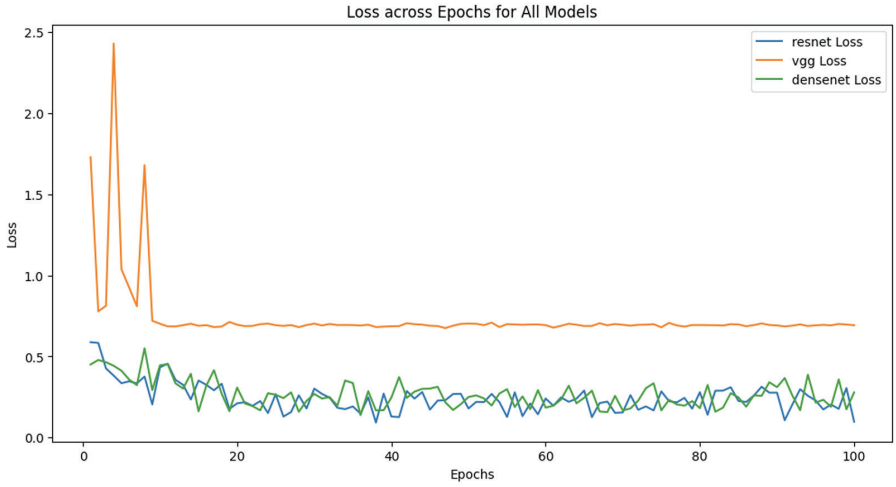

Figure 3. Loss changes during the training process of models.

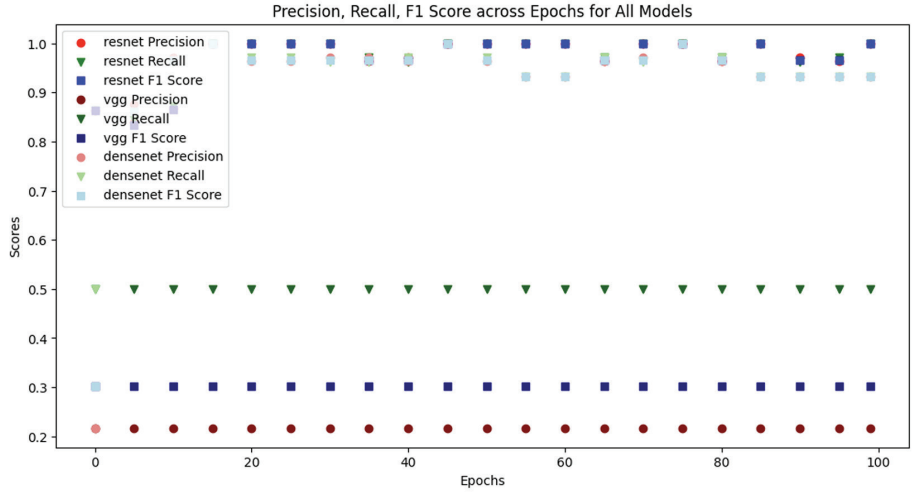

Figure 4. Performance evaluation metrics of the models.

can assist clinicians and radiologists quickly and independently in the diagnostic process. Our study also included images with artifacts and low resolution in which the CBD could be seen. These images can help the clinician determine the indication for ERCP treatment. Early detection of stone-related CBD dilation is crucial in accelerating treatment, preventing complications related to stones, and reducing hospitalization and intensive care unit stay times. This also reduces complications that may arise from prolonged hospital stays.

To better contextualize the effectiveness of deep learning models in detecting stone-associated CBD dilatation, we conducted a comparative analysis with existing non-deep learning diagnostic techniques. Traditional methods, such as radiologist-based manual assessment, have long been the gold standard in clinical practice due to their high accuracy and interpretability. However, these assessments are highly dependent on the radiologist's expertise and experience, making them prone to inter-observer variability and subjective interpretation.

Another conventional approach involves feature engineering-based machine learning models, such as support vector machines and random forest classifiers, which extract handcrafted features from MRCP images. Although these methods can improve classification accuracy, their reliance on manually selected features limits their ability to generalize across different datasets and imaging conditions. Additionally, conventional image processing techniques, such as thresholding and edge detection, have been explored for bile duct segmentation and anomaly detection but often struggle with complex variations in anatomy and image artifacts.

In contrast, our proposed deep learning approach leverages CNNs to automatically extract high-level features, eliminating the need for manual feature selection and minimizing human bias. Unlike traditional methods, CNN-based models can learn hierarchical representations of CBD dilatation patterns directly from raw images, leading to superior classification accuracy and robustness. Furthermore, deep learning models have the potential for real-time application, enabling automated and consistent diagnoses without requiring extensive human intervention. Although CNN models may lack the inherent interpretability of traditional techniques, methods such as gradient-weighted class activation mapping (Grad-CAM) visualization can enhance explainability by highlighting the regions of interest in MRCP images.

Overall, our comparative analysis highlights the advantages of deep learning models in terms of automation, scalability, and reproducibility while recognizing the strengths of traditional diagnostic approaches in interpretability and clinical trustworthiness. Future research could explore hybrid approaches that combine deep learning with traditional radiology methods to further enhance diagnostic performance. As AI rapidly advances and integrates into the healthcare field, this high-performing method can help reduce the workload of radiologists and improve diagnostic accuracy in collaboration with clinicians. Despite the high diagnostic value of the current model, a small margin of error remains. This is because the deep learning model evaluates based on a single, predetermined image, unlike radiologists.[9,15]

There are several studies in the literature that use AIAs to analyze bile duct pathologies from MRCP images. Ringe et al.[16] detected primary sclerosing cholangitis with 95% sensitivity using machine learning algorithms on coronal MIP 3D-MRCP images. Hou et al.[17] detected choledocholithiasis with 95% accuracy from thick-section 2D MRCP images. Sun et al.[18] also detected choledocholithiasis with 93.48% accuracy using 3D-MRCP images. In our study, a CNN model was developed

for the diagnosis of stone-related CBD dilation using 3D-MRCP images, achieving high precision, recall, and F1 score. Among the three models used, DenseNet demonstrated the highest performance, with precision, recall, and F1 score values of 0.9722, 0.9615, and 0.9657, respectively.

There are some limitations in our study. The first is that it only includes patients with stone-related CBD dilation and a normal group. Including patients with CBD dilation due to other pathologies, such as tumors and strictures, would have made the detection of stone-related dilation more valuable. Another limitation is that it is a single-center study and includes a small number of patients. Additionally, we did not differentiate CBD dilatation according to severity, so we did not specifically determine the accuracy rate for patients with a borderline CBD diameter. Therefore, we do not know if our method will perform well on MRCP images obtained from different centers with various devices and parameters. These limitations should be addressed, and larger studies should be conducted with this method.

Although we think that deep learning will be useful in diagnosis, there are many technical, ethical, and financial challenges to overcome. For example, incorrect predictions can be made due to image artifacts and noise. Deep learning practitioners need to manually correct image noise to create an effective dataset. This process is both time-consuming and costly. Another challenge is the requirement for large datasets to build an effective deep learning system. This is both costly and may raise ethical concerns. Additionally, it may not be possible to create large datasets for infrequent diseases.[19]

In conclusion, a high-performance method for diagnosing CBD dilatation in patients with choledocholithiasis using the DenseNet CNN model from MRCP images has been described. However, the number of patients in our study was small, and only patients with choledocholithiasis were included. We hope to validate this method with a larger multicenter dataset in future studies. Additionally, future studies could include images of other pathologies, such as tumors, stenosis, and inflammation, that cause biliary tract dilatation to examine whether differential diagnosis can be made in addition to dilatation detection using this method. After addressing these deficiencies, the DenseNet CNN model may become a valuable tool for detecting CBD dilatation and determining its pathogenesis in the future.

To evaluate the impact of artifacts on model performance, we analyzed images with varying levels of noise and low resolution. The results indicate that although the model performed well on high-quality images, accuracy was slightly reduced in images with major artifacts. This finding highlights the need for artifact-specific preprocessing methods.

To further evaluate the reliability of our model, we calculated confidence intervals for the reported performance metrics and conducted statistical analyses using the bootstrap method. Additionally, five-fold cross-validation was applied to assess the model's generalizability across different data partitions. These analyses confirmed the robustness and consistency of our model, strengthening its potential for clinical applications.

Performance comparisons among ResNet50, DenseNet121, and VGG16 revealed that DenseNet121 exhibited superior sensitivity and classification accuracy. DenseNet121 consistently outperformed the other models, particularly in sensitivity, indicating a higher ability to correctly identify stone-associated CBD dilatation.

This study highlights the potential of deep learning in detecting stone-associated CBD dilatation but has limitations. The single-center dataset limits generalizability, requiring multicenter validation. Future research should include other etiologies, such as tumors and strictures, for broader clinical use. Image artifacts and MRCP variations remain challenges, necessitating improved preprocessing. Additionally, explainability methods such as Grad-CAM and multi-modal imaging integration (CT, US) could enhance diagnostic accuracy and clinical applicability.

Ethical considerations

The integration of deep learning models into clinical decision-making raises important ethical considerations, particularly regarding bias, human–AI collaboration, and regulatory challenges. One of the primary concerns in AI-driven diagnostics is algorithmic bias, which may arise if the training dataset is not sufficiently diverse or representative of different patient populations. A biased dataset can lead to disparities in diagnostic accuracy, particularly among underrepresented demographic groups. Therefore, future studies should prioritize multicenter and demographically diverse datasets to ensure fair and unbiased AI decision-making.

As for human–AI collaboration, although deep learning models demonstrate high accuracy in detecting CBD dilatation, they should be regarded as assistive tools rather than replacements for radiologists. The final clinical decision should always involve expert verification, ensuring that AI-generated diagnoses are contextualized within the broader clinical picture. Methods such as Grad-CAM visualization and other explainability techniques can enhance trust in AI models by providing interpretable insights into the decision-making process.

Additionally, the deployment of AI in medical imaging is subject to regulatory and transparency challenges. Ensuring compliance with healthcare standards and AI governance frameworks is essential for widespread clinical adoption. Future research should focus on developing transparent, explainable AI models that align with ethical and legal standards, thereby increasing both clinician and patient trust in automated diagnostic systems. By addressing these ethical concerns, we aim to contribute to the responsible development and implementation of AI-driven radiology applications, ensuring that deep learning models are used in a manner that prioritizes fairness, safety, and clinical efficacy.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Toy Ş, Şenol D, Çiftçi R, Sevgi S, Seçgin Y, Yıldırım İO. Morphometric examination of the hepatobiliary duct system in healthy individuals and patients with cholelithiasis: a radio-anatomic magnetic resonance cholangiopancreatography study. *Cukurova Med J*. 2023;48(2):706-714. [Crossref]
2. Ham JH, Yu JS, Choi JM, Cho ES, Kim JH, Chung JJ. Peri-ampullary duodenal diverticulum: effect on extrahepatic bile duct dilatation after cholecystectomy. *Clin Radiol*. 2019;74(9):735. [Crossref]
3. Holm AN, Gerke H. What should be done with a dilated bile duct? *Curr Gastroenterol Rep*. 2010;12(2):150-156. [Crossref]
4. Benjaminov F, Leichtman G, Naftali T, Half EE, Konikoff FM. Effects of age and cholecystectomy on common bile duct diameter as measured by endoscopic ultrasonography. *Surg Endosc*. 2013;27(1):303-307. [Crossref]
5. Radmard AR, Khorasanizadeh F, Poustchi H, et al. Prevalence and clinical outcomes of

common bile duct dilation in patients who use opium. *Am J Med Sci*. 2018;356(1):39-46. [Crossref]

6. Abi Nader C, Vetil R, Wood LK, et al. Automatic detection of pancreatic lesions and main pancreatic duct dilatation on portal venous CT scans using deep learning. *Invest Radiol*. 2023;58(11):791-798. [Crossref]
7. Sugimoto Y, Kurita Y, Kuwahara T, et al. Diagnosing malignant distal bile duct obstruction using artificial intelligence based on clinical biomarkers. *Sci Rep*. 2023;13(1):3262. [Crossref]
8. Diaz-Martinez J, Pérez-Correa N. Postcholecystectomy duodenal ınjuries, their management, and review of the literature. *Euroasian J Hepatogastroenterol*. 2024;14(1):44-50. [Crossref]
9. Çiftçi R, Dönmez E, Kurtoğlu A, Eken Ö, Samee NA, Alkanhel RI. Human gender estimation from CT images of skull using deep feature selection and feature fusion. *Sci Rep*. 2024;14(1):16879. [Crossref]
10. Luna LH. Test of the minimum supero-inferior femoral neck diameter as a sex predictor in a contemporary documented osteological collection from Portugal. *Forensic Sci Res*. 2024:owae045. [Crossref]
11. Shoaib MA, Lai KW, Chuah JH, et al. Comparative studies of deep learning segmentation models for left ventricle segmentation. *Front Public Health*. 2022;10:981019. [Crossref]
12. Wu X, Zhang YT, Lai KW, Yang MZ, Yang GL, Wang HH. A novel centralized federated deep fuzzy neural network with multi-objectives neural architecture search for epistatic detection. *IEEE Transactions on Fuzzy Systems*. 2025;33(1):94-107. [Crossref]
13. Atik I. Classification of electronic components based on convolutional neural network Architecture. *Energies*. 2022;15(7):2347. [Crossref]
14. Atik I. COVID-19 case forecast with deep learning BiLSTM approach: the Turkey case. Int. Int J Mech Eng. 2022;7(1):6307-6314. [Crossref]
15. Kurtoğlu A, Eken Ö, Çiftçi R, et al. The role of morphometric characteristics in predicting 20-meter sprint performance through machine learning. *Sci Rep*. 2024;14(1):16593. [Crossref]
16. Ringe KI, Vo Chieu VD, Wacker F, et al. Fully automated detection of primary sclerosing cholangitis (PSC)-compatible bile duct changes based on 3D magnetic resonance cholangiopancreatography using machine learning. *Eur Radiol*. 2021;31(4):2482-2489. [Crossref]
17. Hou JU, Park SW, Park SM, Park DH, Park CH, Min S. Efficacy of an artificial neural network algorithm based on thick-slab magnetic resonance cholangiopancreatography images for the automated diagnosis of common bile duct stones. *J Gastroenterol Hepatol*. 2021;36(12):3532-3540. [Crossref]
18. Sun K, Li M, Shi Y, et al. Convolutional neural network for identifying common bile duct stones based on magnetic resonance cholangiopancreatography. *Clin Radiol*. 2024;79(7):553-558. [Crossref]
19. Dhar T, Dey N, Borra S, Sherratt RS. Challenges of deep learning in medical image analysis-improving explainability and trust. *IEEE Transactions on Technology and Society*. 2023;4(1):68-75. [Crossref]