SoftwareX 34 (2026) 102668

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

Original Software Publication

MedFakeDet: multi-organ medical deepfake detection and inpainting localization system via effective deep learning models and density maps

Mehmet Karakose [a], T. Göktuğ Altundoğan [b,*], Mert Çeçen [a]

[a] Firat University, Elazig, Türkiye
[b] Celal Bayar University, Manisa, Türkiye

ARTICLE INFO

Keywords:
Medical deepfake detection
Tumor inpaint detection
Synthesis detection
UNET

ABSTRACT

The ability of deepfake technologies to produce highly realistic images has become a significant security concern for healthcare systems and insurance auditing processes, either through the generation of fake medical images or the realistic manipulation of original ones. This study employs effective deep learning methods to detect medical images generated by deepfake technologies and to localize inpainting-based manipulations. Two different datasets were constructed for this purpose: the MedFake dataset, which contains brain, lung, kidney, and chest images generated using multiple synthesis approaches; and the MedFakeInpaint dataset, which includes localized inpainting manipulations applied to brain images. For deepfake synthesis detection, an 8-class ResNet-based model was trained to jointly classify the organ type of a medical image and its real or fake status. For manipulation localization, a U-Net-based model was trained to predict density maps corresponding to manipulated regions within the image. Bounding boxes were subsequently derived from the predicted density maps to explicitly represent the localized manipulated regions. The performance of the proposed approach was evaluated using F1-score, and IoU-based localization metrics under strict training, validation, and test data splits with varying threshold values. The developed models were integrated into a backend system using FastAPI, and a lightweight web application was implemented to connect the provided endpoints with end users. MedFakeDet is released as a reproducible software package, including pretrained models and inference scripts. When the developed models were evaluated under different ablation configurations and threshold values, it was observed that in some configurations, the F1 score reached levels of up to 99%.

Metadata

Nr	Code metadata description	*Metadata*
C1	Current code version	v1.0.0
C2	Permanent link to code/ repository used for this code version	https://github.com/tgoktug94/MedDetFake
C3	Permanent link to reproducible capsule	https://github.com/tgoktug94/MedDetFake
C4	Legal code license	MIT License
C5	Code versioning system used	None
C6	Software code languages, tools and services used	Python 3.8+, PyTorch, FastAPI, Uvicorn, Torchvision, Pillow, NumPy, OpenCV
C7	Compilation requirements, operating environments and dependencies	*OS:Linux/Windows/MacOS; Python≥3.8 GPU Optional Dependencies listed in requirements.txt*
C8	If available, link to developer documentation/manual	-
C9	Support email for questions	turan.altundogan@cbu.edu.tr

1. Motivation and significance

Medical imaging consists of various methods that support the diagnosis and treatment of diseases by producing high-resolution images of different parts of the body [1]. Advanced techniques, ranging from MRI to PET, allow for the study of diseases at the cellular or molecular level [2]. These images play a critical role both in clinical decision-making and in obtaining rare medical data [3]. The increase in data volume also enhances the performance of image analysis algorithms. Deep learning-based image generation, in particular, reduces data scarcity and imbalance, improving model generalization [4]. In recent years, the use of deep learning applications in the medical field has rapidly expanded [5–7]. However, medical image generation and fusion remain

* Corresponding author.
E-mail addresses: mkarakose@firat.edu.tr (M. Karakose), turan.altundogan@cbu.edu.tr (T.G. Altundoğan).

https://doi.org/10.1016/j.softx.2026.102668
Received 13 January 2026; Received in revised form 4 April 2026; Accepted 17 April 2026
Available online 22 April 2026

quite challenging and costly [8]. For these reasons, various methods for medical deep-fake generation have been developed in the literature. Some approaches are based on Generative Adversarial Networks (GANs). Han et al. generated realistic brain MRI images using GANs and reported that expert doctors had difficulty distinguishing these images [9]. Chuquicusma and his team performed DCGAN training on lung CT scans [10]. In another study, a two-stage GAN architecture was used to generate MRI images with and without tumors, achieving superior performance with 97.48 % accuracy on BRATS 2016 [11]. A GAN-based iris DCGAN framework was also developed for generating fake iris images and deceiving biometric systems [12].

In addition, diffusion-based methods are also being used. In one study, lung CT images were generated using DDPM, and anatomical accuracy was found to be high [13]. Pneumonia-specific X-ray images were synthesized using a U-Net-based DDPM, and successful results were obtained for both anatomical and pathological findings [14]. Another study used DDPM to generate X-ray, MRI, and ultrasound images; the generated images were subjected to a Turing test with doctors, and experts reported only 42–55 % recall in distinguishing real from fake [15]. This clearly demonstrates the importance of medical deepfake detection. Since fake medical images can lead to serious problems such as insurance fraud [16], the detection of these images is a critical area of research. Some studies have applied contrast enhancement and denoising to improve image quality [17]. Detection accuracy was increased by data augmentation and resizing [16]. The iDCGAN study used preprocessing such as resizing, normalization, and denoising [12]. CNN-based methods are generally used in the feature extraction phase [17]. Another study reported AUC values of 89.9 % and 63.4 % on the BraTS2020 and LUNA16 datasets, respectively [18]. Furthermore, 95 % accuracy was achieved using SVM, Random Forest, and Decision Trees on CT scans with tumor addition/removal using CT-GAN [19].

Medical images are critical data that directly influence clinical decisions in diagnosis and treatment processes, making their reliability of healthcare and insurance systems of paramount importance. In recent years, medical images created using GAN and DDPM-based deepfake generation methods have become extremely realistic visually, and in some studies, Turing test results have shown that even expert physicians cannot distinguish them from real images. This situation demonstrates that fake or manipulated medical images can pose serious risks such as insurance fraud, misdiagnosis, and damage to clinical credibility. Therefore, it is crucial not only to detect such images but also to reliably determine the location of the manipulation. The software developed in this study directly addresses this need by offering both synthetic medical image detection and the determination of local manipulations based on density maps within a single system. The fact that U-Net-based density map approaches produce successful results even in high-density and complex problems (e.g., crowd counting) is the primary reason for choosing this technique for detecting local manipulations. Furthermore, the fact that the developed system is presented with an API-based and easily integrable architecture makes it possible to use the approach practically in real-world medical monitoring applications.

2. Software description

2.1. Software architecture

The architectural structure of the software developed in this study is presented in Fig. 1. As seen in Fig. 1, the proposed system consists of two main components:

- Medical Deepfake Synthesis Detection (MDSD)
- Medical Deepfake Manipulation Detection and Localization (MDMDL)

Both components are designed as separate API services and integrated into the software's backend architecture. The main reason for choosing an API-based architecture is to ensure that the developed models can be easily integrated into different medical imaging systems and security-focused applications. Furthermore, the developed APIs are presented to the end-user via a simple and user-friendly web interface.

The MDSD component aims to detect medical images produced using deepfake technologies. This component uses a single unified 8-class ResNet-based model that supports multi-organ structures. Separate models have not been trained for each organ type. Instead, the model

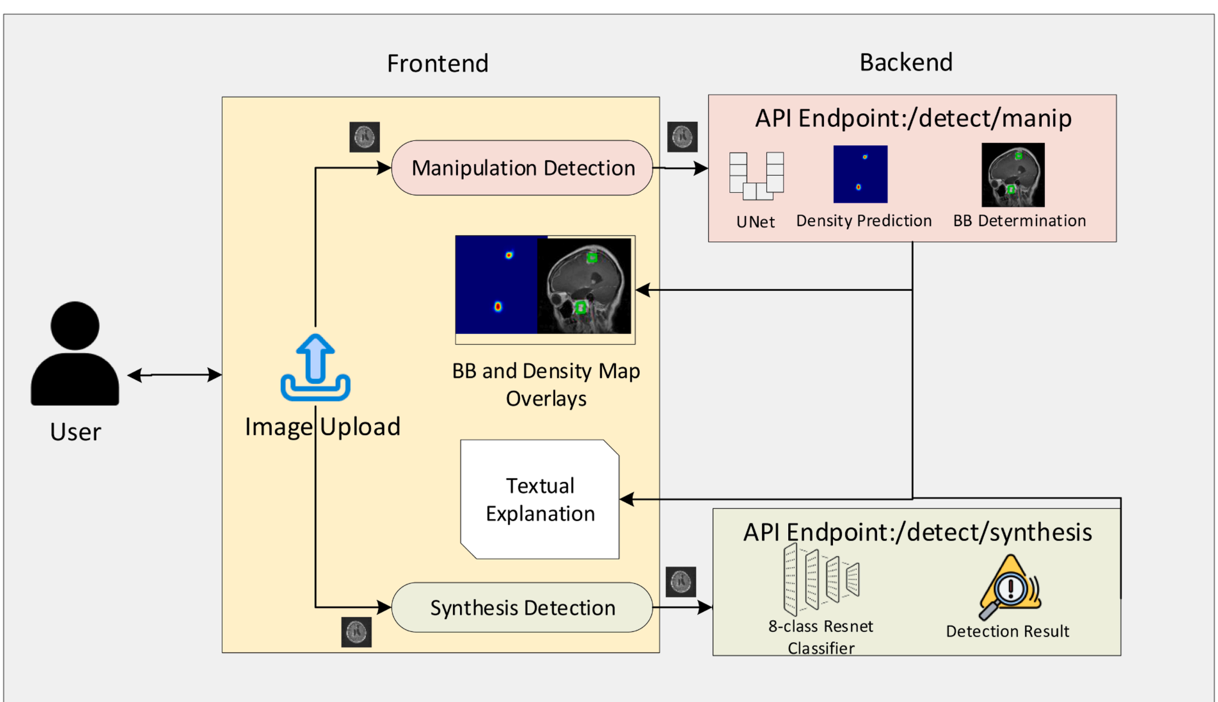


Fig. 1. Use-case diagram of the developed software in this study.

jointly learns to classify both the organ category and the real/fake status of a given medical image within a single architecture. This design enables the model to capture shared representations across different organ types while maintaining discrimination capability. The API only takes the image file as a parameter from the user and returns the classification result and related confidence scores as output. The frontend includes components that enable image loading and the textual display of results.

The MDMDL component aims to detect and locate local manipulations performed on medical images. For this purpose, a U-Net-based density map estimation model has been developed. The model generates a density map representing manipulated regions from the input image. Thresholding and bounded component analysis are applied to the resulting density map, and the manipulated regions are displayed on the image in bounding box format. The API output provides the user with (i) the estimated density map, (ii) bounding box coordinates, and (iii) explanatory text regarding manipulation detection. The web interface includes components designed to display these outputs visually and intuitively.

The backend layer of the web application was developed using the Python programming language and the FastAPI library. The frontend layer utilizes HTML, CSS, and JavaScript technologies. Model training and inference processes were performed using the Python programming language and the PyTorch deep learning library. All components were designed with a modular and extensible software structure to support reproducibility.

2.2. Software functionalities

The software developed in this study offers an integrated functional structure targeting two different deepfake threats performed on medical images: (i) detection of synthetically generated medical images and (ii) both detection and spatial localization of local manipulations performed on original medical images. Since these two problems have different characteristics in terms of visual similarity level, artifact density, and spatial distribution, two complementary but differentiated deep learning approaches have been adopted within the software.

The first fundamental function of the developed system is the detection of synthetically generated medical images using deepfake technologies. For this purpose, a multi-class RESNET-based classifier was trained using the MedFake dataset. The MedFake dataset includes real brain, lung, kidney, and chest images, as well as deepfake samples derived from these images. Synthetic images were obtained using both DDPM-based diffusion models and different GAN-based generation methods, rather than adhering to a single generation paradigm. This approach prevents the classifier from learning only the statistical traces specific to a particular generative model, instead enabling it to learn visual and statistical distortions common to different deepfake production methods.

The RESNET architecture used in synthesis detection is structured based on the transfer learning principle. The RESNET backbone, previously trained on natural images, has been re-adapted for medical images, and the upper layers have been retrained to perform organ classification and real/fake discrimination. Thanks to the transfer learning approach, low-level edge, texture, and frequency-based features learned in the early layers of the network are preserved; deeper layers are optimized to distinguish subtle artifacts left by DDPM and GAN-based synthesis processes. The model output is an eight-class probability vector that simultaneously expresses both the organ type and whether the image is real or fake for each medical image. A multi-class cross-entropy loss function was used during model training, and this function (1) is defined as follows:

$$L_{CE} = -\sum_{c=1}^{C} y_c \log(\hat{y}_c) \tag{1}$$

Here, C represents the number of classes, y_c represents the actual class label, and $\hat{y}_c$ represents the class probability predicted by the model. The second and more challenging function of the software is the detection of local manipulations performed on medical images and the determination of the spatial distribution of these manipulations on the image. The local manipulation detection problem is more complex than synthetic image detection, mainly because manipulations affect only a small region of the image and largely preserve global statistics. Therefore, in this study, the manipulation detection problem is treated not as a classification problem, but as a density map estimation problem.

A density map is a continuous-valued spatial function that represents the probability or manipulation intensity of each pixel. This approach, commonly used in problems such as crowd counting and object density estimation, has been adapted for the first time in this study to the problem of localizing medical deepfake manipulations. Ground-truth density maps were generated for each manipulation region in the MedFakeInpaint dataset. These maps were calculated using two-dimensional Gaussian kernels placed at the center points of the manipulation bounding boxes. The true density map (2) for an image is defined as follows:

$$D(x,y) = \sum_{k=1}^{N} \mathscr{G}(x,y|\mu_k,\sigma) \tag{2}$$

Here, N represents the number of manipulations on the image, μ_k is the center coordinate of the k-th manipulation region, and G is the two-dimensional Gaussian function. This representation allows modeling not only the presence of manipulations but also their spatial distribution and intensity.

A U-Net-based architecture was used for density map estimation. Thanks to its symmetrical structure consisting of encoder and decoder blocks and the skip-connection mechanism between these blocks, U-Net can learn both high-level semantic information and low-level spatial details simultaneously. In the encoder section, abstract representations are obtained from the image through sequential convolution and pooling operations, while in the decoder section, these representations are brought to the original resolution through upsampling operations. Skip-connection structures enable the transfer of fine detail information from early layers to decoder layers, contributing to more precise learning of small and low-contrast manipulation regions.

The model output is the predicted density map $\hat{D}(x,y)$, and the model is trained to match this map as closely as possible to the actual density map D(x,y). The baseline regression loss used for this purpose is defined by the following mean squared error function:

$$L_{den} = \frac{1}{HW} \sum_{x=1}^{H} \sum_{y=1}^{W} \left(\hat{D}(x,y) - D(x,y)\right)^2 \tag{3}$$

Here, H and V represent image dimensions. During the training process, the model learns to generate high density values in regions with manipulation and to keep the density close to zero in regions without manipulation. A binary mask is obtained by applying a specific threshold value τ to the estimated density maps, and then connected component analysis is performed. As a result of this process, each connected component is considered as a separate manipulation region, and the minimum encompassing rectangles for these regions are calculated and presented in bounding box format. This approach allows for the direct counting of the number of manipulations on an image and the spatial evaluation of each manipulation. Fig. 2(a) shows the diagram of the RESNET-based multi-organ deepfake classifier model, and Fig. 2(b) shows the diagram of the UNET-based inpainting localization model.,

The software developed in this study is designed with an API-based architecture in line with the goals of modularity and extensibility. The backend layer is developed using Python and FastAPI, and deepfake synthesis detection, manipulation detection, and localization functions are offered as independent services. The frontend layer, developed with HTML, CSS, and JavaScript technologies, offers a lightweight and

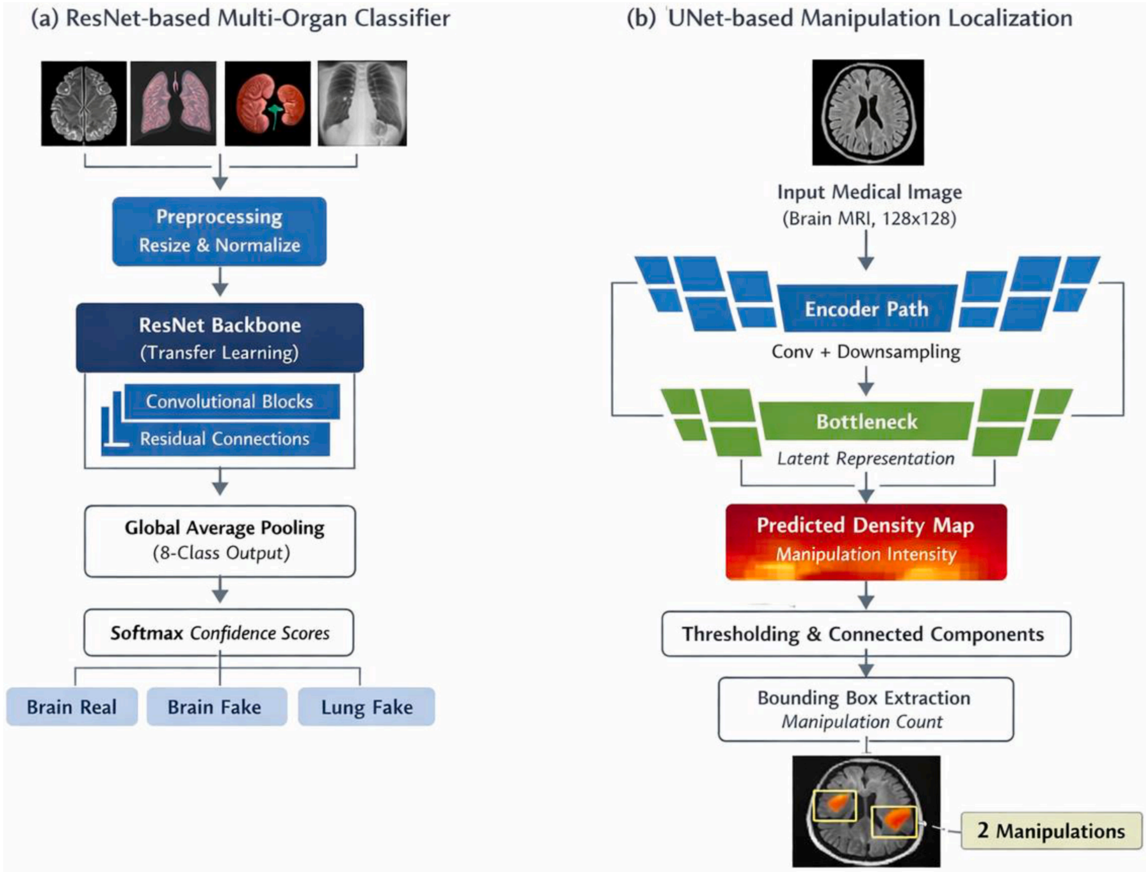


Fig. 2. Block diagram of neural models used in this study (a: ResNet based Multi-Organ classifier, b:UNet-based Manipulation Localization).

intuitive interface where users can visually and textually review image upload, detection results and model behaviours (Density maps and GradCam results).

3. Illustrative examples

In this study, DDPM-based synthesis and local manipulation models were trained using medical images of four different organs: brain, lung, kidney, and breast. The data used in this study were generated under controlled synthetic conditions. Synthesized images were produced from random noise using DDPM and GAN-based methods, while manipulations were applied to clinically meaningful tumor bounding box regions in tumor images and to randomly selected regions in tumor-free images to increase variability. To avoid data leakage, identical samples were not shared across training, validation, and test sets. Different manipulation configurations and the inclusion of both manipulated and non-manipulated samples reduce the likelihood of trivial memorization. However, the dataset may not fully represent the full diversity of real-world clinical scenarios, and the results should be interpreted within a controlled experimental context. Fig. 3 presents examples of real images and synthetic images produced using DDPM and GAN-based methods. Examination of the images shows that the DDPM-based method better preserves anatomical structures and produces results closer to the real images compared to GAN-based approaches.

The developed models are integrated into web interfaces using an API-based architecture. Screenshots of these interfaces are presented in Fig. 4.

4. Impact

Deepfake technology's ability to generate and manipulate such realistic images can be used as an effective tool both in cyberattacks on hospital networks and in document forgery applications for insurance companies. This has increased the importance of deepfake detection approaches in medical image processing. In this study, two models were developed that enable the detection of both synthesized images from scratch and manipulated images. These models were integrated with API technology, making them easily integratable into the systems of organizations facing threats from this technology. Additionally, simple and user-friendly web interfaces enabling user interaction with the API were developed as part of the study. The points discussed so far address the social impacts of the study. A comprehensive evaluation of the developed models is necessary to quantify these impacts. In this context, Table 1 provides quantitative information about the datasets used in model training.

Table 2 presents an ablation study comparing the classifier model developed for the detection of deep fake images synthesized from scratch with different CNN encoder architectures. Table 2 shows that the best test result was achieved with the RESNET architecture.

The ROC curve graph and confusion matrix of the trained UNET-based localization model and detection examples of DDPM-based local manipulations are given in Fig. 5. Examination of the provided graphs shows that the trained model performed the classification process with very high performance. Table 3 presents the performance of the multi-task UNET model developed for manipulation detection, localization, and density map generation. Within the scope of ablation analysis, a comparison was made with the single-task version of the model trained solely for segmentation purposes. Furthermore, to evaluate the performance of the proposed approach against different problem formulations, object detection models based on YOLOv8 and YOLOv11, trained on the same dataset, were also included in the comparison. Table 4 presents a comparison between the proposed approach and studies reported in the

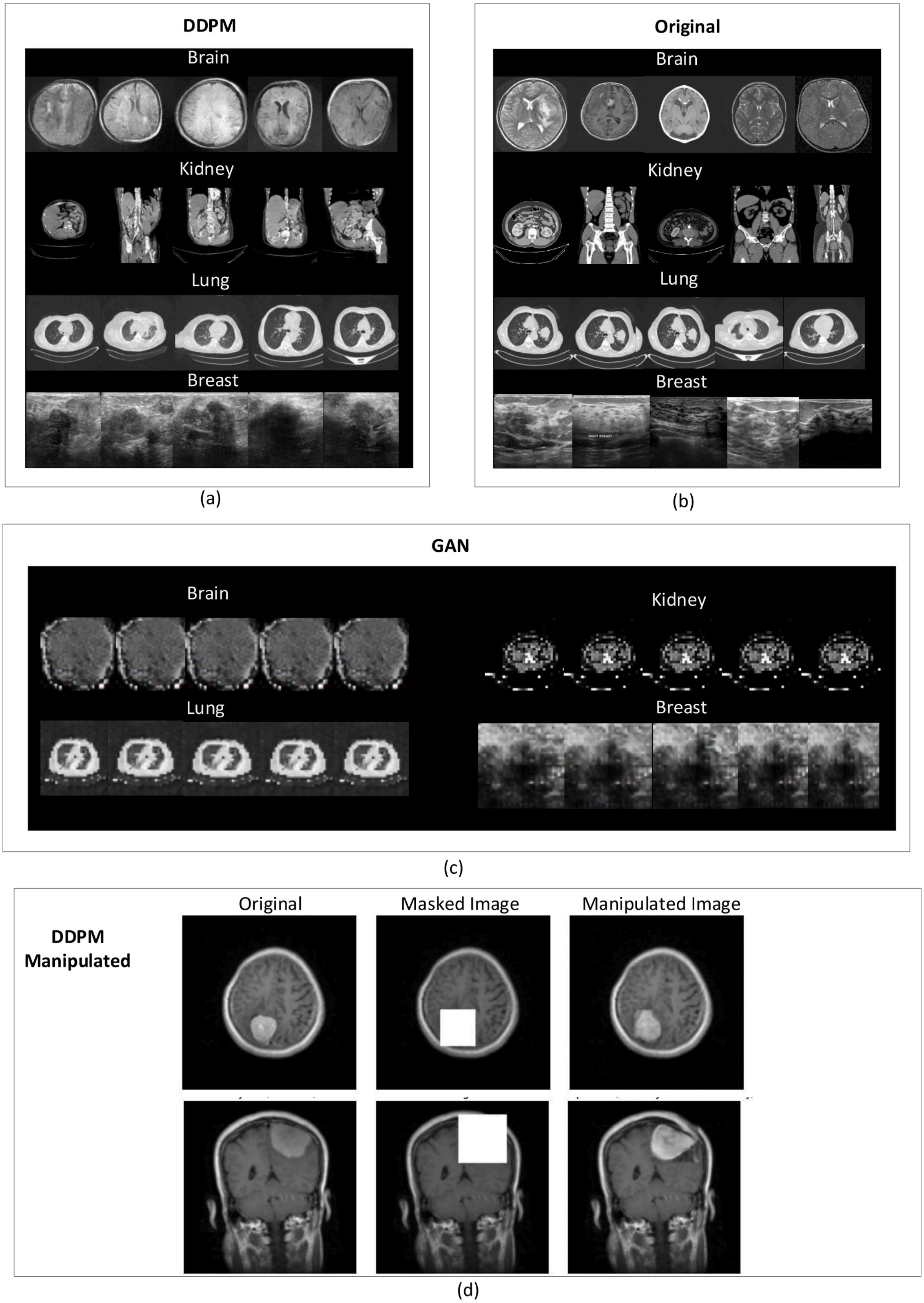
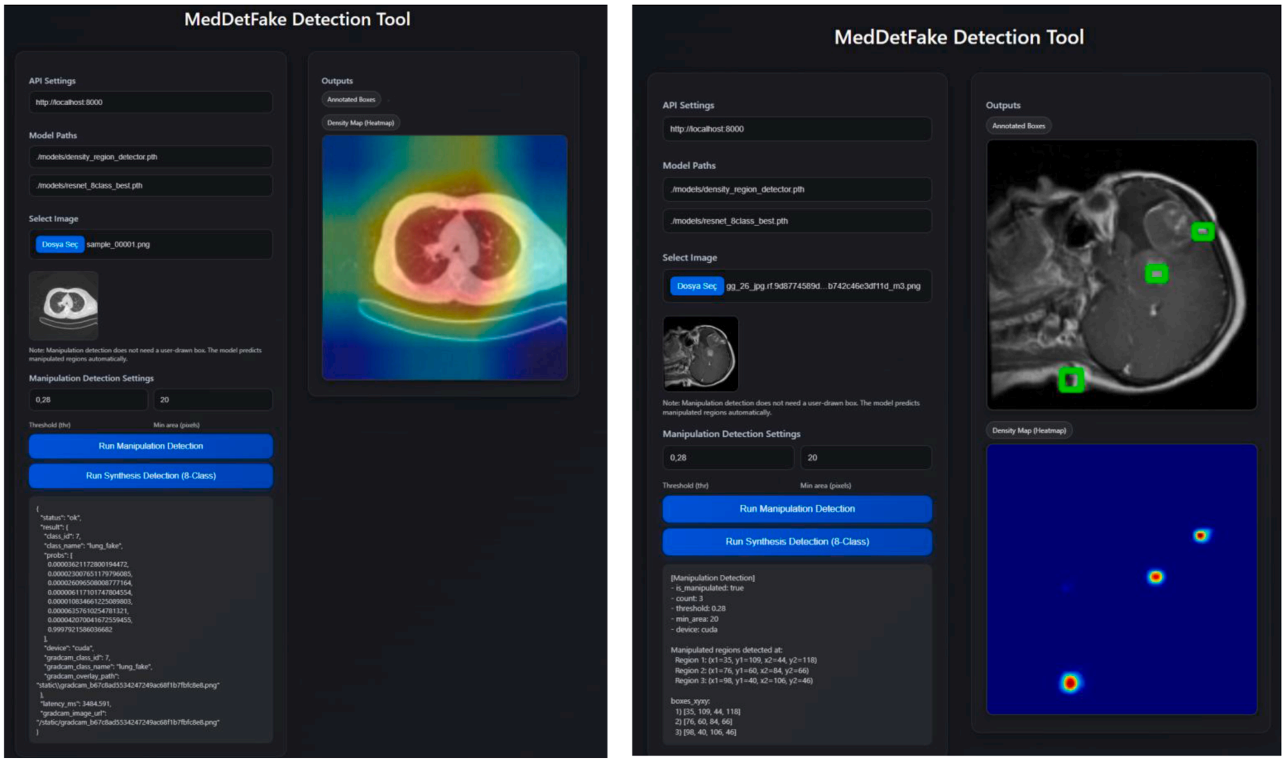


Fig. 3. Illustrative samples from dataset used in this study. (a: Deep fake images generated by DDPM, b:Orignal images, c: Deep fake images generated by GAN, d: DDPM manipulated images.).



(a) (b)

Fig. 4. Web interfaces developed within the scope of the study.

Table 1
Dataset sources, purposes, and train/val/test splits used in this study.

Dataset	Fake / Real	Source	Purpose	Train	Validation	Test
Brain	Real	Publicly available [20]	DDPM Synthesis	950	101	-
Lung	Real	Publicly available [21]	DDPM Synthesis	6000	377	-
Kidney	Real	Publicly available [22]	DDPM Synthesis	2700	239	-
Breast	Real	Publicly available [23]	DDPM Synthesis	925	100	-
Tumor Detection	Real	Publicly available [24]	DDPM Manipulating	6500	350	350
Synthetic Medical Images	Fake	DDPM / GAN on [20–24]	Deep fake detection for synthetic images	10500	1000	500
Real Medical Images	Real	Publicly available [20–24]	Deep fake detection for synthetic images	10500	1000	500
Manipulated Medical Images	Fake	DDPM on [20,24]	Deep fake detection for manipulated images	8000	1000	1000

Table 2
Ablation study on CNN Encoders for deep-fake synthesized image detection.

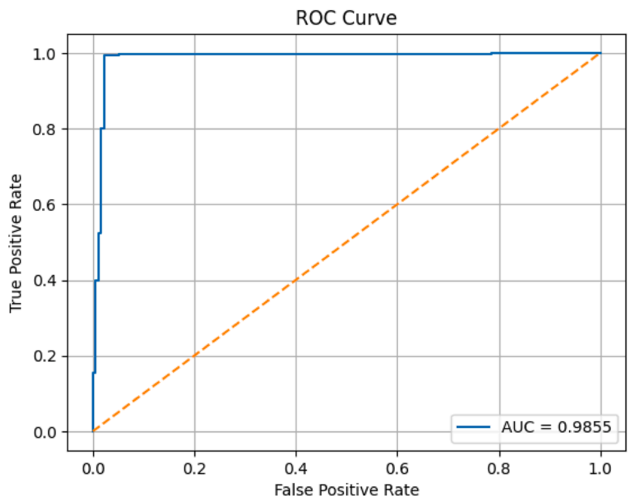
,	Parameter Count	FLOPs (GFLOPs)	Train-F1	Validation-F1	Test-F1
Lightweight-Custom-CNN	246056	0.07	0.8877	0.8731	0.8742
VGG-19	139603016	6.49	0.9553	0.9600	0.9599
InceptionV3	24365808	0.74	0.9778	0.9703	0.9724
LeNet	1628556	0.16	0.9324	0.9155	0.9245
RESNET	11180616	0.595	**0.9990**	**0.9986**	**0.9878**

literature. It should be noted that these comparisons are not conducted under identical datasets or experimental conditions, and therefore are intended to provide a contextual reference rather than a direct benchmark. Within this study, the most fair and controlled evaluations are the ablation experiments performed under the same dataset and training protocol.

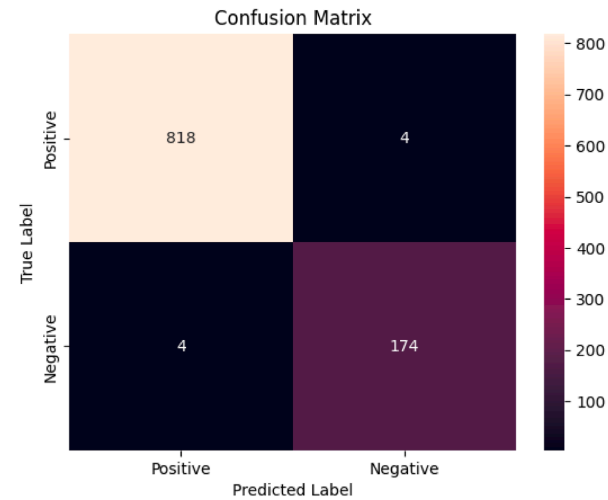
One of the key factors determining the system's feasibility is the latency and performance values of the developed modules. In this context, the API application was run on a laptop with a 3.2 GHz 14th generation Intel i5 processor, 6 GB RTX 4050 GPU, and 16 GB RAM. The results presented in Table 5 show that the modules offer acceptable performance on this hardware.

5. Conclusions

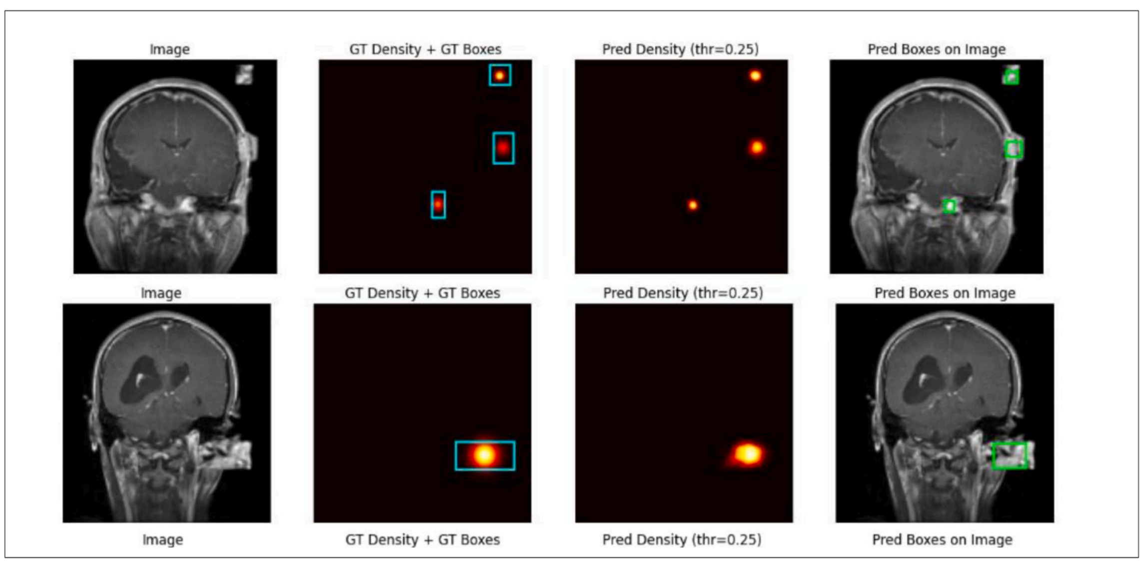
This study develops MedFakeDet, an API-based and reproducible software system that addresses the problems of deepfake synthesis detection and local manipulation detection/localization in medical images. The proposed system aims to reliably detect highly realistic medical images created with advanced deepfake production techniques.The multi-organ support ResNet-based classifier developed for synthesis detection showed the highest performance in comparative experiments with different CNN architectures, achieving a 98.78 % F1 score in the test set. These results demonstrate that the transfer learning-based ResNet architecture has a very strong representational ability in distinguishing advanced DDPM and GAN-based forged images.The proposed density map-based UNET model for manipulation detection and localization was evaluated under different threshold values. The experiments showed that the threshold value directly affects performance; the best result was obtained at a threshold value of 0.28, achieving a



(a)



(b)



(c)

Fig. 5. (a) ROC curve of the UNet-Density model, (b) Confusion matrix of the UNet-Density model, and (c) qualitative examples of DDPM-based local manipulations, along with their corresponding ground truth density maps, ground truth bounding boxes, and UNet predictions.

Table 3
Quantitative comparison of density-map-based multi-task learning, single-task segmentation, and detection-based approaches for manipulation presence detection.

Model	Validation-selected Decision Threshold	TP	TN	FP	FN	Recall	Precision	F1-score
YOLO-v8	0.37	567	196	4	233	0.7088	0.9930	0.8283
YOLO-v11	0.35	594	194	2	210	0.7388	**0.9966**	0.8486
UNET (Single-Task)	0.22	444	176	3	377	0.5408	0.9933	0.7008
UNET-Density (Multi-task)	0.28	**818**	**174**	**4**	**4**	**0.9951**	0.9951	**0.9951**

99.51 % F1 score. Thanks to density maps, manipulated areas were not only detected but also displayed spatially in bounding box format.All developed models were serviced with a modular backend architecture based on FastAPI and presented to the end user via a user-friendly web interface. This approach enables easy integration of the system into hospital information systems, insurance audit platforms, and security-focused medical applications. In conclusion, MedFakeDet offers a powerful and applicable solution in the field of medical deepfake detection, both for academic and practical use, with its high accuracy rates, spatial detection capability, and software-oriented design.

CRediT authorship contribution statement

Mehmet Karakose: Supervision, Resources, Project administration, Investigation. **T. Göktuğ Altundoğan:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology. **Mert Çeçen:** Investigation, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial

Table 4
Comparison of proposed methods with existing literature.

Study	Purpose	Recall	Precision	F1 / Accuracy
[25]	Synthesis Detection	-	-	91.57 %
[26]	Synthesis Detection	-	-	96.0 %
[27]	Synthesis Detection	-	-	97.79 %
[28]	Synthesis Detection	98.87 %	96.69 %	97.77 %
[29]	Manipulation Detection	93.9 %	94.4 %	94 %
[30]	Manipulation Detection	72.84 %	**100 %**	84.29 %
This Study	Synthesis Detection	**98.62 %**	**97.98 %**	**98.78 %**
This Study	Manipulation Detection	**99.51 %**	**99.51 %**	**99.51 %**

Table 5
Measured and estimated inference performance of system modules.

Module	Model (Encoder)	Latency (ms)	Performance Note
Synthesis Detection	RESNET	257.21	High accuracy (preferred)
Synthesis Detection	Lightweight-Custom-CNN	221.34	Faster but lower accuracy
Manipulation Detection	UNET-Density (Multi-Task)	284.92	High accuracy (preferred)
Manipulation Detection	UNET (Single-Task)	281.55	Faster but lower accuracy

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported in part by The Scientific and Technological Research Council of Turkey (TUBITAK) under Grant 122E676, and in part by the FIRAT University Scientific Research Projects Unit (FUBAP) under Grant MF.24.55

References

[1] Kumar R, Malik R. A review on generative adversarial networks used for image reconstruction in medical imaging. Proc IEEE Int Conf Reliab Infocom Technol Optim (ICRITO) 2021:1–5. https://doi.org/10.1109/ICRITO51393.2021.9596157.
[2] Eigenschink P, Reutterer T, Vamosi S, Vamosi R, Sun C, Kalcher K. Deep generative Models for synthetic data: a survey. IEEE Access 2023;11:47304–20. https://doi.org/10.1109/ACCESS.2023.3275134.
[3] Pezoulas VC, Zaridis DI, Mylona E, Androutsos C, Apostolidis K, Tachos NS, et al. Synthetic data generation methods in healthcare: a review on open-source tools and methods. Comput Struct Biotechnol J 2024;23:2892–910. https://doi.org/10.1016/J.CSBJ.2024.07.005.
[4] Islam T, Hafiz MS, Jim JR, Kabir MM, Mridha MF. A systematic review of deep learning data augmentation in medical imaging: recent advances and future research directions. Healthc Anal 2024;5:100340. https://doi.org/10.1016/J.HEALTH.2024.100340.
[5] Xia J, Wang Z, Yang D, Li R, Liang G, Chen H, et al. Performance optimization of support vector machine with oppositional grasshopper optimization for acute appendicitis diagnosis. Comput Biol Med 2022;143:105206. https://doi.org/10.1016/J.COMPBIOMED.2021.105206.
[6] Xia J, Yang D, Zhou H, Chen Y, Zhang H, Liu T, et al. Evolving kernel extreme learning machine for medical diagnosis via a disperse foraging sine cosine algorithm. Comput Biol Med 2022;141:105137. https://doi.org/10.1016/J.COMPBIOMED.2021.105137.
[7] Hu J, Liu Y, Heidari AA, Bano Y, Ibrohimov A, Liang G, et al. An effective model for predicting serum albumin level in hemodialysis patients. Comput Biol Med 2022; 140:105054. https://doi.org/10.1016/J.COMPBIOMED.2021.105054.
[8] Goceri E. Medical image data augmentation: techniques, comparisons and interpretations. Artif Intell Rev 2023;56:12561–605. https://doi.org/10.1007/S10462-023-10453-Z. 2023 5611.
[9] Han C, Hayashi H, Rundo L, Araki R, Shimoda W, Muramatsu S, Furukawa Y, Mauri G, Nakayama H. GAN-based synthetic brain MR image generation. In: Proc IEEE Int Symp Biomed Imaging (ISBI); 2018. p. 734–8. https://doi.org/10.1109/ISBI.2018.8363678.
[10] Chuquicusma MJM, Hussein S, Burt J, Bagci U. How to fool radiologists with generative adversarial networks? A visual turing test for lung cancer diagnosis. In: Proc IEEE Int Symp Biomed Imaging (ISBI); 2018. p. 240–4. https://doi.org/10.1109/ISBI.2018.8363564.
[11] Han C, Rundo L, Araki R, Nagano Y, Furukawa Y, Mauri G, et al. Combining noise-to-image and image-to-image GANs: brain MR image augmentation for tumor detection. IEEE Access 2019;7:156966–77. https://doi.org/10.1109/ACCESS.2019.2947606.
[12] Kohli N, Yadav D, Vatsa M, Singh R, Noore A. Synthetic iris presentation attack using iDCGAN. In: Proc IEEE Int Joint Conf Biometrics (IJCB); 2017. p. 674–80. https://doi.org/10.1109/BTAS.2017.8272756.
[13] Krishna A., Wang G., Mueller K. Multi-Conditioned Denoising Diffusion Probabilistic Model (mDDPM) for medical image synthesis. arXiv, 2024. doi:10.48550/arXiv.2409.04670.
[14] Mahaulpatha P., Abeywardane T., George T. DDPM based X-ray image synthesizer. arXiv, 2024. doi:10.48550/arXiv.2401.01539.
[15] Ju Z., Zhou W. VM-DDPM: vision mamba diffusion for medical image synthesis. arXiv 2024. doi:10.48550/arXiv.2405.05667.
[16] Karaköse M, Yetiş H, Çeçen M. A new approach for effective medical deepfake detection in medical images. IEEE Access 2024;12:52205–14. https://doi.org/10.1109/ACCESS.2024.3386644.
[17] Min-Jen T, Cheng-Tao C. Convolutional neural network for detecting deepfake palmprint images. IEEE Access 2024;12:103405–18. https://doi.org/10.1109/ACCESS.2024.3433497.
[18] Georgescu MI. Masked autoencoders for unsupervised anomaly detection in medical images. Procedia Comput Sci 2023;225:969–78. https://doi.org/10.1016/j.procs.2023.10.084.
[19] Solaiyappan S, Wen Y. Machine learning based medical image deepfake detection: a comparative study. Mach Learn with Appl 2022;8:100298. https://doi.org/10.1016/J.MLWA.2022.100298.
[20] Brain tumor MRI dataset. Kaggle. Available online: https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset (accessed April 4, 2026).
[21] Lung cancer MRI images. Kaggle. Available online: https://www.kaggle.com/datasets/xiaopengzhang12/lung-cancer-mri-images (accessed April 4, 2026).
[22] Kidney MRI dataset. Roboflow Universe. Available online: https://universe.roboflow.com/bhuv-7sdq9/kidney-ql5tq (accessed April 4, 2026).
[23] Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in Brief 2020;28:104863. https://doi.org/10.1016/j.dib.2019.104863.
[24] Brain tumor MRI dataset. Roboflow Universe. Available online: https://universe.roboflow.com/abhay-fm64z/brain-tumor-1fz61 (accessed April 4, 2026).
[25] Grabovski FM, Yasur L, Amit G, Mirsky Y. Back-in-time diffusion: unsupervised detection of medical deepfakes. ACM Trans Intell Syst Technol 2025;16(6):123. https://doi.org/10.1145/3744656.
[26] Sharafudeen M, Vinod Chandra SS. Medical deepfake detection using 3-dimensional neural learning. Lect Notes Comput Sci (Incl Subser Lect Notes Artif Intell Lect Notes Bioinform) 2023;13739:169–80. https://doi.org/10.1007/978-3-031-20650-4_14. LNAI.
[27] Karaköse M, İlhan İ, Yetiş H, Ataş S. A new approach for deepfake detection with the Choquet Fuzzy Integral. Appl Sci 2024;14:7216. https://doi.org/10.3390/APP14167216.
[28] Alsabbagh A.R., Al-Kadi O. Comparative analysis of deep convolutional neural networks for detecting medical image deepfakes. arXiv 2024. doi:10.48550/arXiv.2406.08758.
[29] Aruna S, Narayan S. Detection of GAN-manipulated medical images through deep learning techniques. In: Proc Int Conf Adv Mod Age Technol Health Eng Sci (AMATHE); 2024. p. 1–6. https://doi.org/10.1109/AMATHE61652.2024.10582065.
[30] Zhang J, Huang X, Liu Y, Han Y, Xiang Z. GAN-based medical image small region forgery detection via a two-stage cascade framework. PLoS One 2024;19:e0290303. https://doi.org/10.1371/JOURNAL.PONE.0290303.