Approaches for GAN-manipulated medical image detection

🔹 1. CNN-based End-to-End Forgery Classifier (simpler + faster)

Instead of separate **U-Net + SVM**, you can directly train a **deep CNN classifier** to detect tampering:

- Use **EfficientNetV2 / ResNet50 / DenseNet121** pretrained on ImageNet.
- Fine-tune on CT-GAN + LIDC-IDRI datasets.
- Train for **binary classification (authentic vs tampered)** or **multi-class (TB, TM, FB, FM)**.
- Add **contrastive learning loss** (e.g., ArcFace) to enhance separability.

✅ Advantages:

- No handcrafted LBP.
- Faster training and deployment.
- Transfer learning improves generalization.

🔹 2. Frequency-Domain Detection (robust against GAN artifacts)

GAN forgeries often leave **frequency artifacts** (unnatural textures, checkerboard noise).

- Convert CT images to **FFT / DCT spectrum**.
- Train CNN/Transformer on **frequency-domain + spatial-domain** jointly.
- Paper: “GAN Fingerprints” methods show **>98% detection accuracy** in natural images.

✅ Advantages:

- Harder for attackers to remove frequency cues.
- Complementary to spatial domain detection.

🔹 3. Vision Transformer (ViT / Swin) for Forgery Detection

Transformers are state-of-the-art in medical imaging tasks:

- Use **Swin Transformer / Vision Transformer** for patch-level analysis of CT scans.
- Combine **multi-head self-attention** to capture global inconsistencies.
- Pretrain with **Masked Autoencoders (MAE)** on large medical datasets (TCIA, NIH).

✅ Advantages:

- Better capture of **long-range dependencies** than CNNs.
- Can directly localize tampered regions.

🔹 4. Hybrid Spatial + Frequency + Self-Supervised Pretraining

A very strong modern solution:

- Train a **dual-stream model**:
  - **Spatial stream** (CNN/ViT on CT images).
  - **Frequency stream** (CNN/ResNet on FFT/DCT spectrum).
- Fuse embeddings → classification head.
- Use **self-supervised pretraining** (SimCLR, BYOL, DINOv2) to avoid overfitting small datasets.

✅ Advantages:

- Strong robustness against unseen GAN models.
- Works even when manipulations are subtle.

🔹 5. ROI-aware Multi-task Network

Instead of classifying full scans:

- Use **nodule detector (YOLOv8 or Faster R-CNN)** to localize suspicious nodules.
- Pass ROI to **forgery detector network**.
- Multi-task learning: simultaneously predict **nodule type (benign/malignant)** and **authenticity (real/fake)**.

✅ Advantages:

- More explainable for radiologists.
- Reduces false positives since only ROIs are analyzed.

🔹 6. GAN-fingerprint & Contrastive Learning

- GANs leave **unique noise fingerprints** (model-specific).
- Extract **noise residuals** (via high-pass filters / SRM filters).
- Train contrastive learning model to **separate real vs GAN fingerprints**.
- Extend to classify which GAN type (CycleGAN, CT-GAN, StyleGAN).

✅ Advantages:

- Detects even **high-quality deepfakes**.
- Scales to new attack types.