

---

# CT-3DFLOW : LEVERAGING 3D NORMALIZING FLOWS FOR UNSUPERVISED DETECTION OF PATHOLOGICAL PULMONARY CT SCANS

---

**Aissam Djahnine**

Claude Bernard University Lyon 1  
Philips Health Technology Innovation  
Paris, France  
aissam.djahnine@philips.com

**Alexandre Popoff**

Philips Health Technology Innovation  
Paris, France  
alexandre.popoff@philips.com

**Emilien Jupin-Delevaux**

Department of Radiology  
Hospices Civils de Lyon  
Lyon, France  
emilien.jupin-delevaux@chu-lyon.fr

**Vincent Cottin**

Claude Bernard University Lyon 1  
National Reference Center for Rare Pulmonary Diseases, Louis Pradel Hospital  
Lyon, France  
vincent.cottin@chu-lyon.fr

**Olivier Nempont**

Philips Health Technology Innovation  
Paris, France  
olivier.nempont@philips.com

**Loic Bousset**

Claude Bernard University Lyon 1  
Department of Radiology  
Hospices Civils de Lyon  
Lyon, France  
loic.bousset@chu-lyon.fr

## ABSTRACT

Unsupervised pathology detection can be implemented by training a model on healthy data only and measuring the deviation from the training set upon inference, for example with CNN-based feature extraction and one-class classifiers, or reconstruction-score-based methods such as AEs, GANs and Diffusion models. Normalizing Flows (NF) have the ability to directly learn the probability distribution of training examples through an invertible architecture. We leverage this property in a novel 3D NF-based model named CT-3DFlow, specifically tailored for patient-level pulmonary pathology detection in chest CT data. Our model is trained unsupervised on healthy 3D pulmonary CT patches, and detects deviations from its log-likelihood distribution as anomalies. We aggregate patches-level likelihood values from a patient's CT scan to provide a patient-level 'normal'/'abnormal' prediction. Out-of-distribution detection performance is evaluated using expert annotations on a separate chest CT test dataset, outperforming other state-of-the-art methods.

**Keywords** Unsupervised Anomaly Detection · Normalizing Flow · GLOW · Pulmonary CT

## 1 Introduction

Normalizing flow (NF) models, such as NICE[1], RealNVP[2], and GLOW[3] learn the distribution of the observed data by transforming it into a tractable distribution using invertible and differentiable mappings. This allows for exact log-likelihood computation during inference, and anomaly detection methods based on normalizing flows, such as DifferNet [4], CFLOW-AD [5], and FastFlow [6] have achieved high performance on industrial datasets. Although it is possible to apply normalizing flows (NF) on both images and CNN-extracted features, at the risk of ignoring the 3D context offered by volumetric data. Applying NF models on 3D data is possible but raises technical challenges. In theory, a 3D NF model would need to be trained on whole 3D volumes to learn the distribution of normal scans, and the log-likelihood of a sample scan would then be used to classify it as pathological or not. The corresponding memory requirements would however be far superior to 2D approaches, making training more challenging. To our knowledge, advancements in 3D anomaly detection models are scarce. The PET-3DFLOW model [7] operates at the feature level via an encoder-NF-decoder scheme, and computes an anomaly score based on a weighted combination of negative log-likelihood and reconstruction error, producing anomaly maps via reconstruction — a potentially sub-optimal method [8, 9, 10, 9, 11]. We propose here a novel approach for anomaly detection and localization in chest CT scans using a 3D patch-based normalizing flow model, whose architecture is a 3D extension of GLOW [3].

## 2 Methods

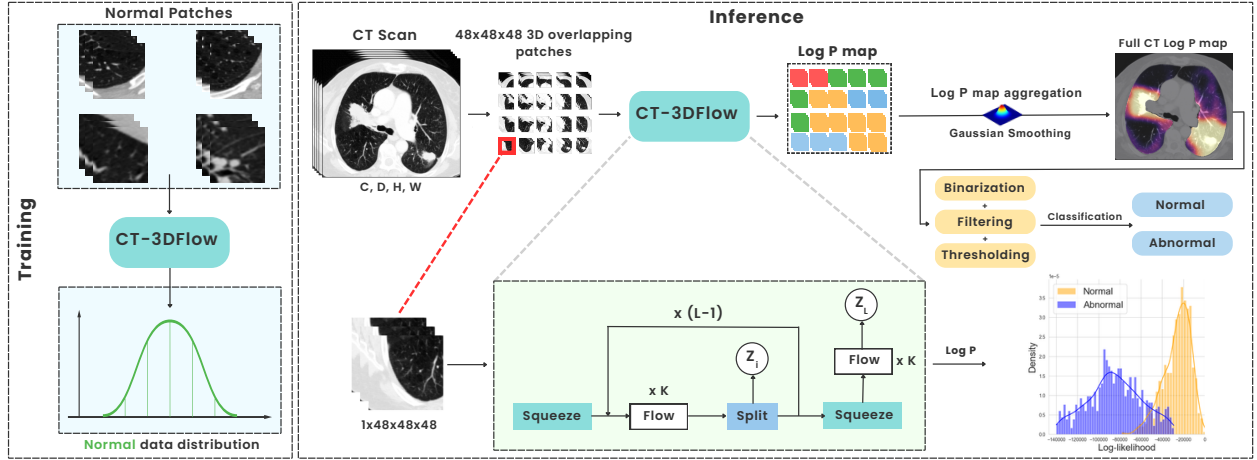


Figure 1: Overview of the CT-3DFlow anomaly detection framework. The model, whose architecture is shown in the middle, is trained on normal 48x48x48 CT sub-volumes. During inference, overlapping patches are processed and aggregated to generate a full CT Log P map, which is binarized, filtered, and thresholded for classification.

We use a dataset of 822 patients for which 3D chest CT scans were obtained from hospital data at a single institution (Hospices Civils de Lyon). It consists of a total of 570 normal and 252 abnormal CT scans (470 scans for training, 111 scans for the validation set, and 291 scans for the test set). These scans are pre-processed by resampling to 2 mm resolution, automatically segmenting the lung in 3D with DL [12] to guide patch-selection, clipping voxel values to the clinically relevant range of  $[-1020HU, +200HU]$ , and normalizing them to  $[-0.5, +0.5]$ .

After training a NF model on 500,000 normal 3D CT 48x48x48-sized patches, we predict abnormalities, consisting of four distinct chest-related abnormalities carefully selected by an expert radiologist based on the prevalence of the most common pulmonary pathologies, at patient-level in a sample CT scan by applying the model on 48x48x48 patches spanning the whole volume (with 10-voxel overlap). Each patch gives a log-likelihood value (indicated as Log P) which serves as a proxy for an anomaly score. These Log P values are aggregated with gaussian smoothing to give a full CT Log P map, which is post-processed to give a binary patient prediction. This involves binarization, filtering, and thresholding. The optimal threshold value  $T$  (between 0.5 and 20  $cm^3$ ) being obtained from the ROC curve on the validation set (50 normal and 61 abnormal). Finally, we use our test set (50 normal, 191 abnormal) to assess the method’s performance.

In our experiments, we used a GLOW-based NF architecture which we adapted to 3D with  $L = 4$  blocks, each one having  $K = 64$  flows. Each flow consists in an activation normalization layer, an invertible 1x1x1 convolution and

Table 1: Comparison of patient-level classification performance with several anomaly detection methods.

Method	AUROC	F1	Accuracy
AE [13]	0.608	0.538	0.556
VAE [10]	0.647	0.607	0.598
f-AnoGan [11]	0.821	0.783	0.765
GANomaly [14]	0.772	0.74	0.694
DifferNet [4]	0.901	0.861	0.843
FastFlow [6]	0.942	0.912	0.903
CFLOW-AD [5]	0.732	0.739	0.706
AnoDDPM [15]	0.924	0.891	0.873
<b>Ours</b>	<b>0.952</b>	<b>0.94</b>	<b>0.924</b>

an affine coupling layer. The network is trained for 50,000 iterations on 2 NVIDIA A100 SXM4 GPUs, using the maximum likelihood estimation objective with a batch size of 10, using Adam optimizer ( $lr = 10^{-4}$ , weight decay  $10^{-5}$ ). We compare, on a test set, our CT-3DFlow model against several reconstruction-based models (autoencoder AE, and variational autoencoder VAE), GAN-based models, NF-based models (DifferNet, FastFlow, CFLOW-AD), and one diffusion-based model (AnoDDPM). We adopted Area Under the Receiver Operating Characteristic curve (AUROC), F1-score (F1), and accuracy (ACC) as metrics to evaluate the patient-level classification performance of our model.

### 3 Results and Perspectives

As shown on Table 1, the CT-3DFlow model outperforms the state-of-the-art methods by a noticeable margin in terms of AUC, F1 and ACC. However, further validation of the aggregation methods is required due to the inherent differences between 2D and 3D approaches. Our 3D patch-based NF model demonstrates the superiority of 3D flow-based model over state-of-the-art 2D methods, while mitigating the problems associated to full volume normalizing flows. Further work to generalize this promising approach on other modalities and different organs is envisioned in the future.

### References

- [1] Laurent Dinh, David Krueger, and Yoshua Bengio. NICE: Non-linear Independent Components Estimation, April 2015. arXiv:1410.8516 [cs].
- [2] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP, February 2017. arXiv:1605.08803 [cs, stat].
- [3] Durk P Kingma and Prafulla Dhariwal. Glow: Generative Flow with Invertible 1x1 Convolutions. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.
- [4] Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same same but DifferNet: Semi-supervised defect detection with normalizing flows. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1906–1915, 2021. ISSN: 2642-9381.
- [5] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. CFLOW-AD: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1819–1828, 2022. ISSN: 2642-9381.
- [6] Jiawei Yu1, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. FastFlow: Unsupervised anomaly detection and localization via 2d normalizing flows. 2021.
- [7] Zhe Xiong, Qiaoqiao Ding, Yuzhong Zhao, and Xiaoqun Zhang. PET-3dflow: A normalizing flow based method for 3d PET anomaly detection. In Wenjian Qin, Nazar Zaki, Fa Zhang, Jia Wu, Fan Yang, and Chao Li, editors, *Computational Mathematics Modeling in Cancer Analysis*, Lecture Notes in Computer Science, pages 91–100. Springer Nature Switzerland, 2023.
- [8] Mahdyar Ravanbakhsh, Moin Nabi, Enver Sangineto, Lucio Marcenaro, Carlo Regazzoni, and Nicu Sebe. Abnormal event detection in videos using generative adversarial nets. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1577–1581, September 2017. ISSN: 2381-8549.

- [9] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery, March 2017. arXiv:1703.05921 [cs].
- [10] Diederik P. Kingma and Max Welling. Auto-Encoding Variational Bayes, December 2022. arXiv:1312.6114 [cs, stat].
- [11] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, May 2019.
- [12] Salim A. Si-Mohamed, Mouhamad Nasser, Marion Colevray, Olivier Nempont, Pierre-Jean Lartaud, Anna Vlachomitrou, Thomas Broussaud, Kais Ahmad, Julie Traclet, Vincent Cottin, and Loic Boussel. Automatic quantitative computed tomography measurement of longitudinal lung volume loss in interstitial lung diseases. 32(6):4292–4303, 2022.
- [13] Jaroslav Kopčan, Ondrej Škvarek, and Martin Klimo. Anomaly detection using autoencoders and deep convolution generative adversarial networks. 55:1296–1303, 2021.
- [14] Samet Akcay, Amir Atapour-Abarghouei, and Toby P. Breckon. GANomaly: Semi-supervised anomaly detection via adversarial training. In C. V. Jawahar, Hongdong Li, Greg Mori, and Konrad Schindler, editors, *Computer Vision – ACCV 2018*, Lecture Notes in Computer Science, pages 622–637. Springer International Publishing, 2019.
- [15] Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris G. Willcocks. AnoDDPM: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 649–655, 2022. ISSN: 2160-7516.