Appendix A: Description of the variables used in our experiment

Variable Name	Possible Values	Variable Description*
Last Status	Deceased, Discharged	For the selected visit the status of the patient
Age Splits	[18,59], (59, 74], (74, 90]	Age intervals (in years) at time of admission
Gender Concept Name	FEMALE, MALE	Documented gender in the EHR (Electronic Health Record)
Visit Concept Name	Inpatient Visit Outpatient Visit Emergency Room Visit	For the selected visit the type of the visit
Is ICU	True, False	Patient admitted to the ICU based on documented room charges
Was Ventilated	Yes, No	The patient had invasive ventilation
Acute Kidney Injury	Yes, No	Had an increase in serum creatinine of 0.3 mg/dL within 48 hours
Length of Stay	Numeric Value (days)	Number of calendar days in the facility
Oral Temperature	Numeric Value (°C)	
Oxygen Saturation	Numeric Value (%)	Oxygen saturation in Arterial blood by Pulse oximetry
Respiratory Rate	Numeric Value (/min)	
Heart Rate	Numeric Value (/min)	
Systolic Blood Pressure	Numeric Value (mmHg)	

* The descriptions are extracted from: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=89096912

Appendix B. Training of encoder/decoder models on RSNA dataset

1. Principal component analysis (PCA)

PCA is a statistical technique that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. This method is widely used for dimensionality reduction of data. We utilized the "**sklearn.decomposition.PCA"** module from the scikit-learn library (https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html). The parameter **n_components** was set to 128.

2. Autoencoder (AE)

An AE is a neural network used for the unsupervised learning of efficient coding. The aim of an AE is to learn the representation (encoding) of a set of data, typically for dimensionality reduction [S1]. We have specified 128 dimensions for the latent code as follows:

Encoder

Layer	Activation	Output Shape
(Input Image)		256 × 256 × 1
Convolution 1× 1	LReLU	256 × 256 × 8
Convolution 3× 3	LReLU	256 × 256 × 16
Downsampling		128 × 128 × 16
Convolution 3× 3	LReLU	128 × 128 × 32
Downsampling		64 × 64 × 32
Convolution 3× 3	LReLU	64 × 64 × 64
Downsampling		32 × 32 × 64
Convolution 3× 3	LReLU	32 × 32 × 128
Downsampling		16 × 16 × 128
Convolution 3× 3	LReLU	16 × 16 × 256
Downsampling		8 × 8 × 256
Convolution 3× 3	LReLU	8 × 8 × 512
Downsampling		4 × 4 × 512
Linear		128

Decoder

Layer	Activation	Output Shape
(Latent vector)		128
Linear	LReLU	4 × 4 × 512
Upsampling		8 × 8 × 512
Convolution 3×3	LReLU	8 × 8 × 256
Upsampling		16 × 16 × 256
Convolution 3× 3	LReLU	16 × 16 × 128
Upsampling		32 × 32 × 128
Convolution 3× 3	LReLU	32 × 32 × 64
Upsampling		64 × 64 × 64
Convolution 3× 3	LReLU	64 × 64 × 32
Upsampling		128 × 128 × 32
Convolution 3× 3	LReLU	128 × 128 × 16
Upsampling		256 × 256 × 16
Convolution 3× 3	LReLU	256 × 256 × 8
Convolution 1× 1	Tanh	256 256 × 1

3. Auto-encoding generative adversarial networks (αGAN)

An αGAN merges AE with GAN, aiming to improve upon the AE's capability by producing sharper and more realistic images through adversarial training [S2, S3]. We have specified 128 dimensions for the latent code as follows. The generator and encoder of the Auto-encoding GAN are identical to those of the decoder and encoder described in the AE, respectively.

Generator

Layer	Activation	Output Shape
(Same with decoder of the auto-encoder we used)		

Encoder

Layer	Activation	Output Shape
(Same with encoder of the auto-encoder we used)		

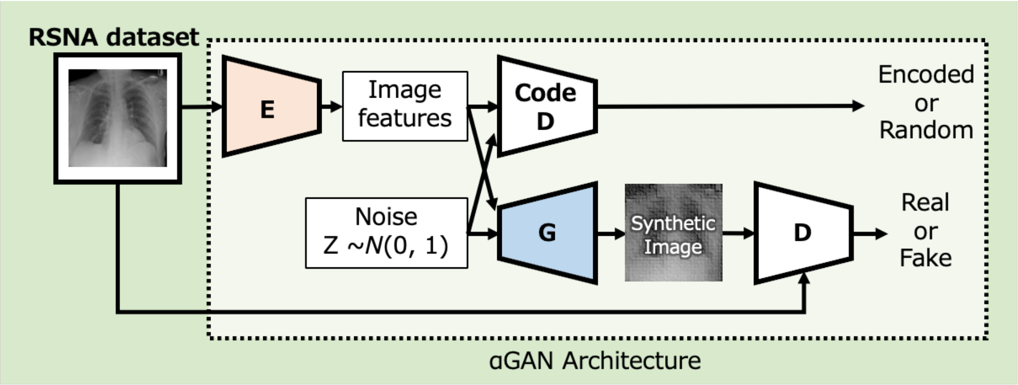
Discriminator

Layer	Activation	Output Shape
(Input Image)		256 × 256 × 1
Convolution 1× 1	LReLU	256 × 256 × 8
Convolution 3× 3	LReLU	256 × 256 × 16
Downsampling		128 × 128 × 16
Convolution 3× 3	LReLU	128 × 128 × 32
Downsampling		64 × 64 × 32
Convolution 3× 3	LReLU	64 × 64 × 64
Downsampling		32 × 32 × 64
Convolution 3× 3	LReLU	32 × 32 × 128
Downsampling		16 × 16 × 128
Convolution 3× 3	LReLU	16 × 16 × 256
Downsampling		8 × 8 × 256
Convolution 3× 3	LReLU	8 × 8 × 512
Downsampling		4 × 4 × 512
Linear		1

Code Discriminator

Layer	Activation	Output Shape
(Latent vector)		128
Linear	LReLU	1500
Linear		1

4. Outline of the αGAN model



This Figure outlines the architecture of the αGAN system, which includes four key components. The encoder (E) processes real images from the RSNA dataset, encoding them into latent representations. The generator (G), which uses either latent codes or Gaussian noise, synthesizes images that mimic real images. The discriminator (D) then assesses these images, distinguishing between the genuine images from the dataset and the fabricated images created by G. Finally, the code discriminator (Code D) distinguishes between the actual Gaussian distribution and the latent codes produced by the encoder.

Appendix C. Encoders for metric learning

We used the Torchxrayvision library (available at https://github.com/mlmed/torchxrayvision) to encode the images [S4]. Within this library, there is a model pretrained on a large dataset using densenet121 [S5] as a base, designed to classify 18 outcomes (Atelectasis, Cardiomegaly, Consolidation, Edema, Effusion, Emphysema, Enlarged Cardio-mediastinum, Fibrosis, Fracture, Hernia, Infiltration, Lung Lesion, Lung Opacity, Mass, Nodule, Pleural Thickening, Pneumonia, Pneumothorax). We reset the weights of the final fully connected layer of the pre-trained model before utilization. The images were resized to 224×224 pixels before input. To encode the tabular data, after converting the categorical variables into dummy variables, we employed a simple two-layer, fully connected network as follows:

Encoder for tabular data

Layer	Activation	Output Shape
(Number of table columns)		16
Linear	LReLU	32
Linear		18

For contrastive learning training, we used a loss function based on cosine similarity.

Appendix D: Structure of the prediction model for *Last Status*

We combined the images and tabular data using the following structure. As in Appendix C, we reset the weights of the final fully connected layer of the densenet121 model from TorchXrayvision.

Image Encoder

Layer	**Activation**	**Output Shape**
(Input image)		224 × 224 × 1
Pretrained densenet121 from torchxrayvision		
(Output dims)		18

Table Encoder

Layer	**Activation**	**Output Shape**
(Number of table columns)		15
Linear	LReLU	32
Linear		18

Classifier

Layer	**Activation**	**Output Shape**
Concatenate		36
Linear		1

Appendix E: Classification and regression performance

The first column served as the dependent variable, and models were trained using all the table data and images as independent variables, excluding the dependent variable itself and "Last Status." The model was evaluated on the test set of pDS. "Last Status" was excluded because it represents outcome information, indicative of future states. Comparisons were made using pDS : sDS = 1 : 0 as the reference and P values were adjusted using the Holm method. The type of test varied depending on the metric: McNemar's test for accuracy (ACC), DeLong's test for area under the curve (AUC), and Wilcoxon signed-rank test for mean absolute error (MAE). P-values less than 0.05 were considered statistically significant.

		Ratio between pDS and sSD (pDS : sDS)					
Dependent Variables	**Metrics**	**0 : 1**	**0.25 : 0.75**	**0.5 : 0.5**	**0.75 : 0.25**	**1 : 0**	**pDS + sDS**
Age Splits	ACC	**0.52**	0.59	**0.59**	0.62	0.66	0.66
Gender Concept Name*	AUC	**0.75**	**0.76**	0.86	0.91	0.90	**0.96**
Visit Concept Name**	AUC	**0.77**	0.95	0.94	0.95	0.95	0.94
Is ICU	AUC	0.86	0.90	0.89	0.90	0.90	088
Was Ventilated	AUC	0.92	0.91	0.95	0.93	0.94	0.96
Acute Kidney Injury	AUC	0.76	0.73	0.77	0.78	0.78	0.78
Length of Stay (days)	MAE	**6.29**	5.17	**5.60**	5.19	5.12	5.10
Oral Temperature (%)	MAE	1.38	1.52	1.32	1.41	1.39	**1.30**
Oxygen Saturation (%)	MAE	**4.53**	4.16	4.11	4.11	3.90	4.02
Respiratory Rate (/min)	MAE	3.90	3.88	**4.44**	3.92	3.83	3.93
Heart Rate (/min)	MAE	**17.01**	15.89	15.16	16.15	15.37	15.39
Systolic Blood Pressure (mmHg)	MAE	**18.80**	17.94	17.98	17.62	17.86	17.23

ACC: Accuracy, AUC: area under the receiver operating characteristic curve, MAE: mean absolute error

Cells highlighted in color indicate statistically significant performance changes compared to that with pDS : sDS = 1: 0, with orange denoting improvement and blue indicating a decrease in performance.

* Gender Concept Name: Binary classification (MALE/FEMALE).

** Visit Concept Name: Treated as binary because there are no cases with the "Outpatient Visit" category in the pDS test set.

S1. Bank D, Koenigstein N, Giryes R: Autoencoders. Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook. Springer International Publishing, Cham, pp 353–374, 2023

S2. Rosca M, Lakshminarayanan B, Warde-Farley D, Mohamed S: Variational approaches for auto-encoding generative adversarial networks. arXiv [stat.ML], 2017

S3. Nakao T, Hanaoka S, Nomura Y, Murata M, Takenaga T, Miki S, Watadani T, Yoshikawa T, Hayashi N, Abe O: Unsupervised deep anomaly detection in chest radiographs. J Digit Imaging 34:418–427, 2021

S4. Cohen JP, Viviano JD, Bertin P, et al (06–08 Jul 2022) TorchXRayVision: A library of chest X-ray datasets and models. In: Konukoglu E, Menze B, Venkataraman A, Baumgartner C, Dou Q, Albarqouni S (eds) Proceedings of The 5th International Conference on Medical Imaging with Deep Learning. PMLR, pp 231–249

S5. Chen Y, Kang X, Shi YQ, Wang ZJ: A multi-purpose image forensic method using densely connected convolutional neural networks. J Real Time Image Process 16:725–740, 2019