ConvNeXt Based Lung Cancer Survival Prediction using Dual Energy CT

Advisor: Ruey-Feng Chang, Ph.D.
張瑞峰教授
Student: Chao-Hsun Yang
楊朝勛

Contents

Contents

Lung Cancer

Leading Cause of Mortality

- Patients with the same clinical and pathological status may experience different survival times.
- Considering survival as a new and distinct marker is essential in making informed decisions regarding further treatment for patients.

Estimated Deaths

Males			Females		
Lung & bronchus	67,160	21%	Lung & bronchus	59,910	21%
Prostate	34,700	11%	Breast	43,170	15%
Colon & rectum	28,470	9%	Colon & rectum	24,080	8%
Pancreas	26,620	8%	Pancreas	23,930	8%
Liver & intrahepatic bile duct	19,000	6%	Ovary	13,270	5%
Leukemia	13,900	4%	Uterine corpus	13,030	5%
Esophagus	12,920	4%	Liver & intrahepatic bile duct	10,380	4%
Urinary bladder	12,160	4%	Leukemia	9,810	3%
Non-Hodgkin lymphoma	11,780	4%	Non-Hodgkin lymphoma	8,400	3%
Brain & other nervous system	11,020	3%	Brain & other nervous system	7,970	3%
All Sites	**322,080**	**100%**	**All Sites**	**287,740**	**100%**

Survival Prediction

Machine Learning

- Support vector machine on the Surveillance, Epidemiology, and End Results (SEER) data.
- Random forest classifier on radiomic features extracted from computed tomography (CT) images.

Deep Neural Network (DNN)

- Four hidden layers, each with 40 neurons, with rectifed linear unit (ReLU) on clinical data.

3D Convolution Neural Network

Study Objective:

- 3D CNNs can capture spatial relationships and contextual information across multiple slices.

Our approach:

- Leverage 3D CNNs to extract features from Dual Energy CT.
- Integrate the extracted features with clinical data for survival prediction.

Contents

Dataset

Dual Energy CT

- From Department of Medical Imaging at National Taiwan University Hospital;
- July 2018 to October 2022;
- 236 patients in total (40-140 keV, with an interval of 10 keV);
- Image size: slices × 512 × 512 ($pixel^3$) in DICOM format.

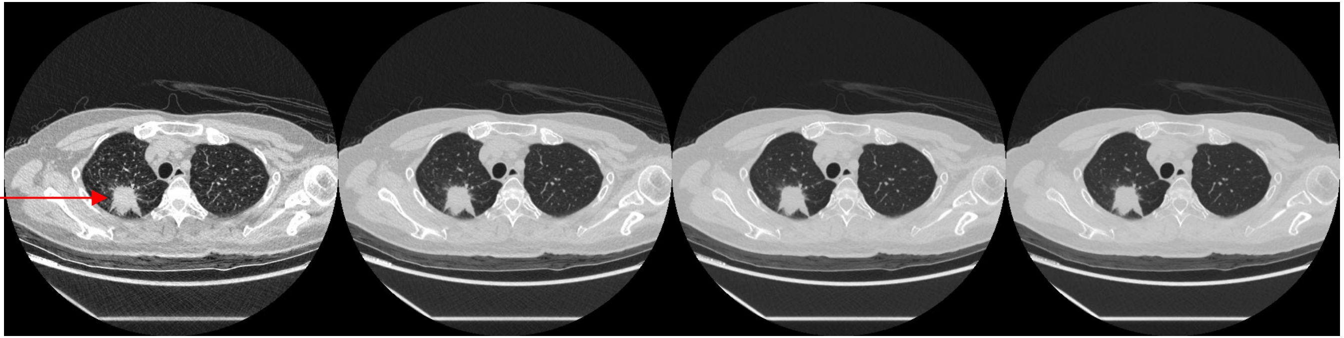
Clinical data

- Clinical and pathological information.

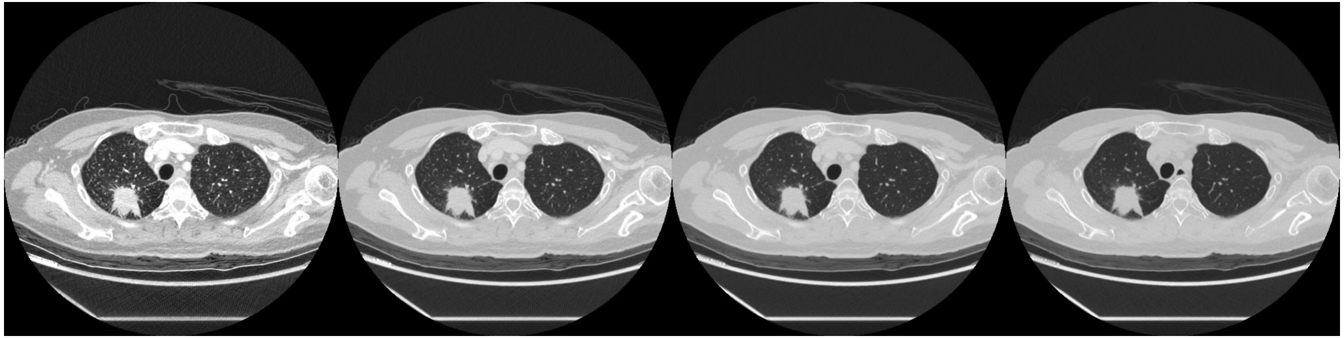
n = 236	
Deceased	Survivor
34	202

Dual Energy CT

Tumor



C-40 (keV) | C-70 | C-100 | C-140 | w/o contrast



C+40 | C+70 | C+100 | C+140 | Contrast

Clinical Data

Clinical information

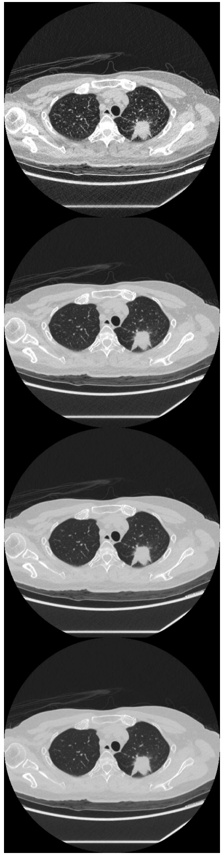
- gender, port-a/resection, differentiation, lymphovascular invasion (LVI), clinical staging, smoking, pack per day (PPD), hypertension (HTN), diabetes mellitus (DM), family history of lung cancer, complications, forced vital capacity (FVC), forced expiratory volume in one second (FEV1), FEV1/FVC

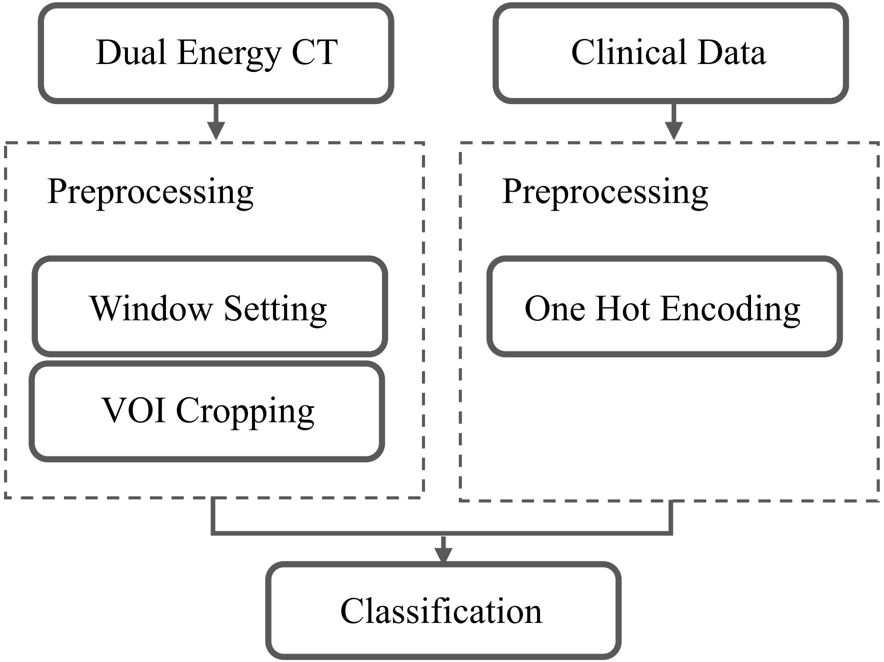
Pathological information

- tumor size, tumor location, anaplastic lymphoma kinase (ALK), ROS-1, epidermal growth factor receptor (EGFR), EGFR mutation, pathological staging, metastasis, recurrence

Contents

Survival Prediction Workflow





Clinical information

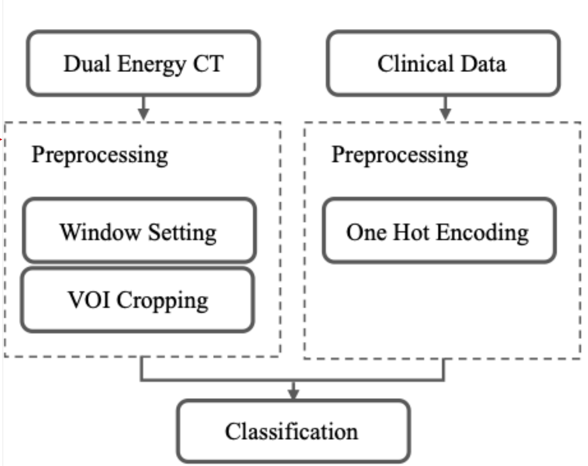
- gender, port-a/resection, differentiation, LVI, clinical staging, smoking, PPD, HTN, DM, family history of lung cancer, complications, FVC, FEV1, FEV1/FVC

Pathological information

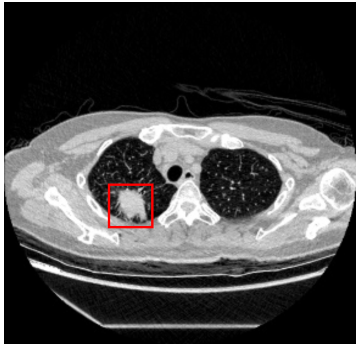
- tumor size, tumor location, tumor size, ALK, ROS-1, EGFR, EGFR mutation, pathological staging, metastasis, recurrence

Dual Energy CT

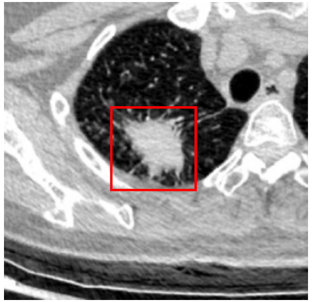
- Set the window center and window width to focus on the relevant organ structures like lung.
- Perform a Volume of Interest (VOI) cropping, where we center the images around the VOI and cut them into 128×224×224 dimensions from the original size.



Window center = -300
Window width = 1400



VOI cropping



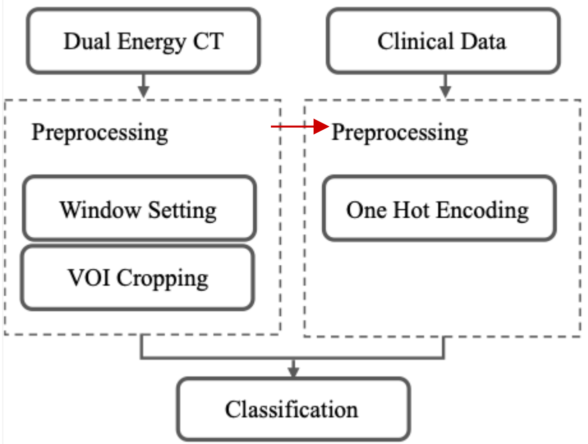
128×224×224

- Air with a Hounsfield Unit (HU) value of -1000
- Water with a HU value of 0
- Bones with HU values ranging from 400 to 1000

Clinical Data

- One hot encoding used to convert categorical variables, such as text or non-ordinal data, into numerical representations.
- Obtain a total of 150 clinical features.

Data Id	Size (cm)	Metastasis
1	1.8	Bone
2	2.4	-
3	11.2	Brain, bone
4	2.5	-

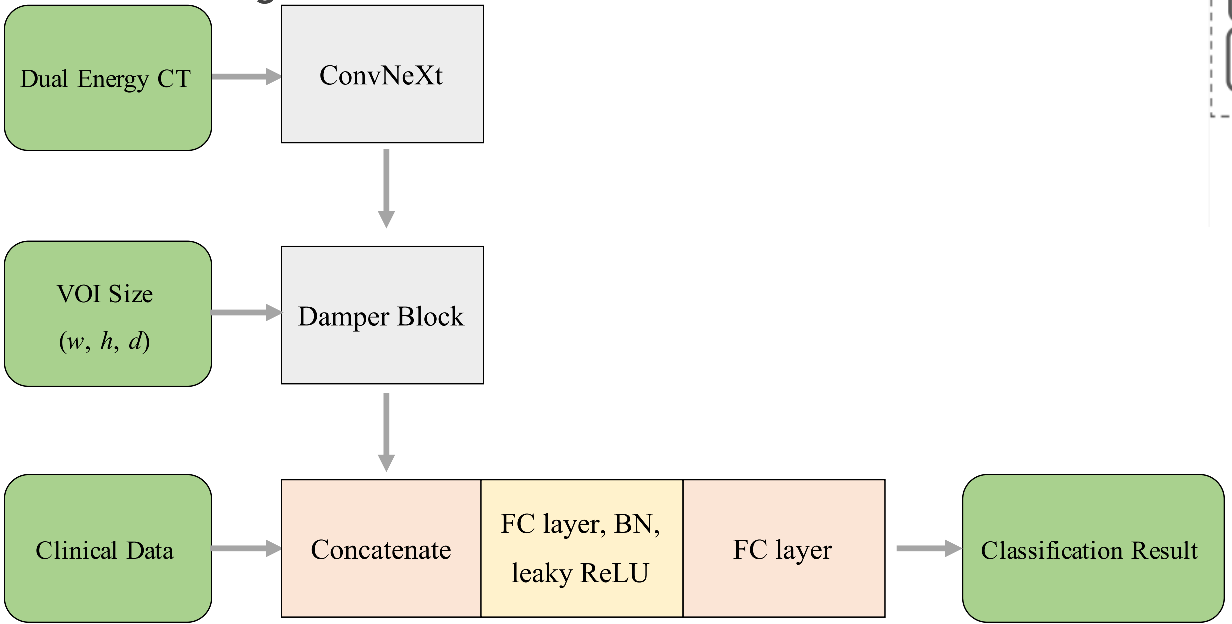


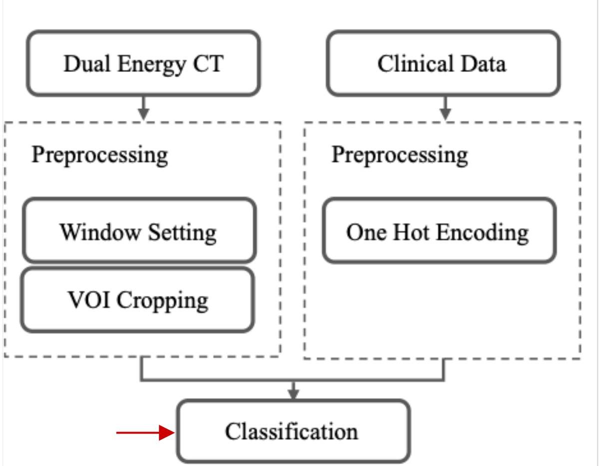
Data Id	Size<1	Size<2	Size<3	Size>=3
1	0	1	0	0
2	0	0	1	0
3	0	0	0	1
4	0	0	1	0

Data Id	Lymph Node	Brain	Bone	Liver	Kidney	Others	No Metastasis
1	0	0	1	0	0	0	0
2	0	0	0	0	0	0	1
3	0	1	1	0	0	0	0
4	0	0	0	0	0	0	1

Classification

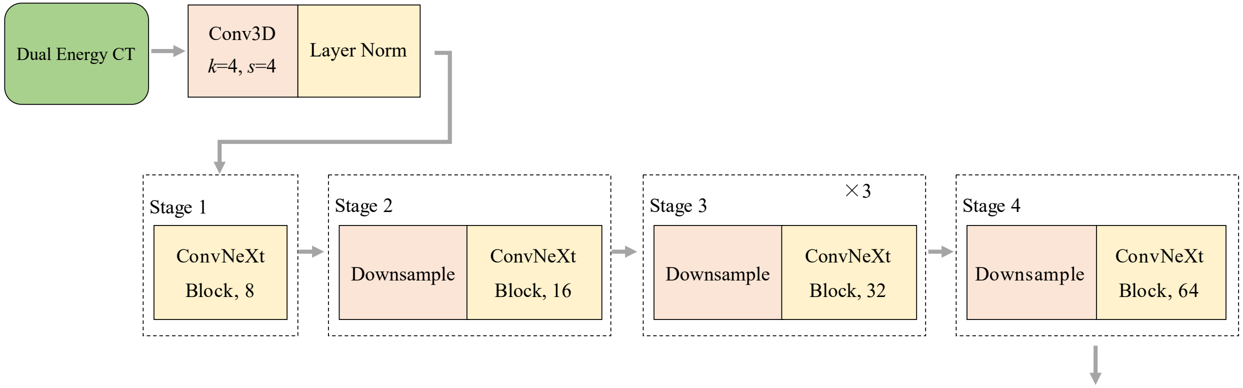
- Use ConvNeXt as the backbone.
- Bring in damper block to integrate the VOI size as an effect parameter.
- Combine image features with clinical data.

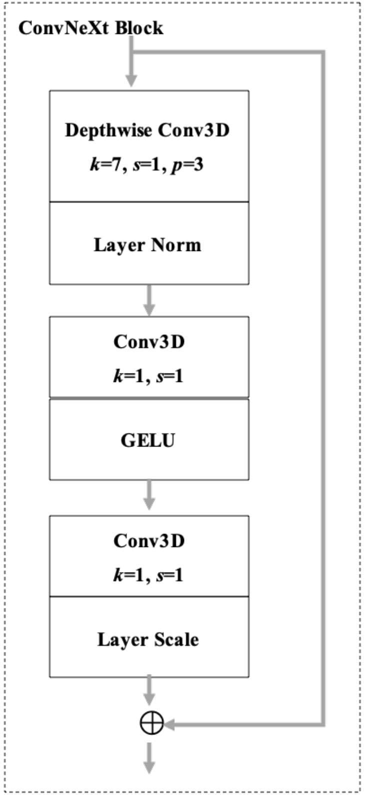


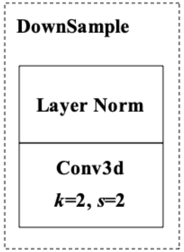


ConvNeXt

- Derive from ResNet and ResNeXt and mimic certain aspects of Vision Transformers.
- Reduce number of ConvNeXt blocks.
- Reduce channel numbers in each ConvNeXt block.

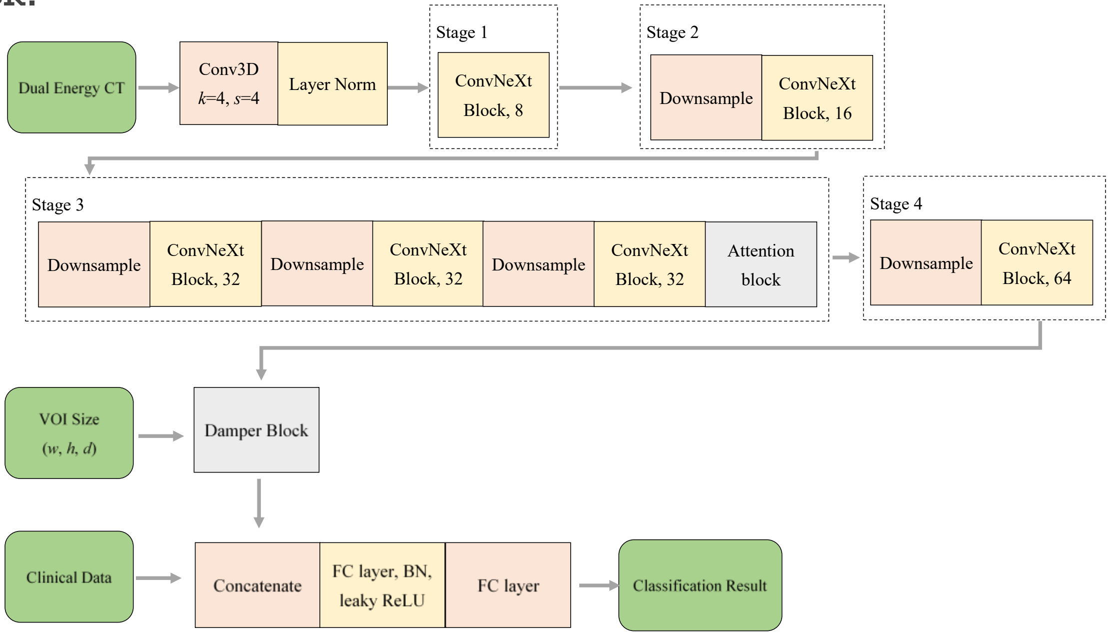


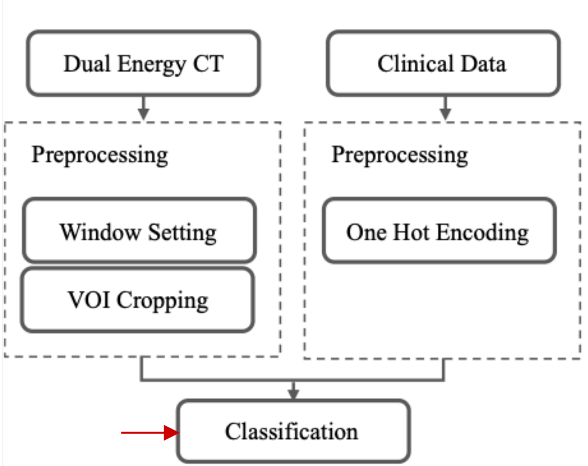




Classification

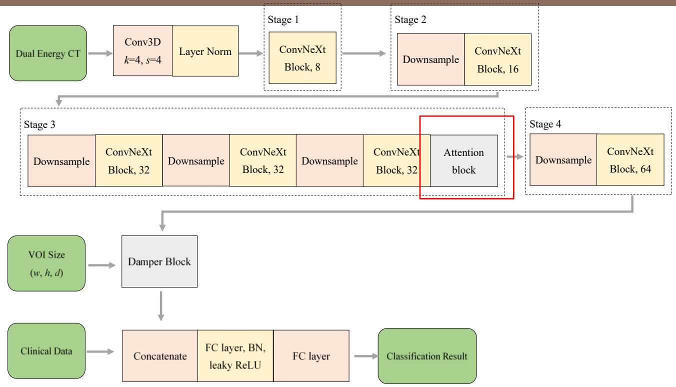
- Employed channel attention block to distinguish features with varying levels of importance within the channels.
- Combine SE block and GCT block to one channel attention block.

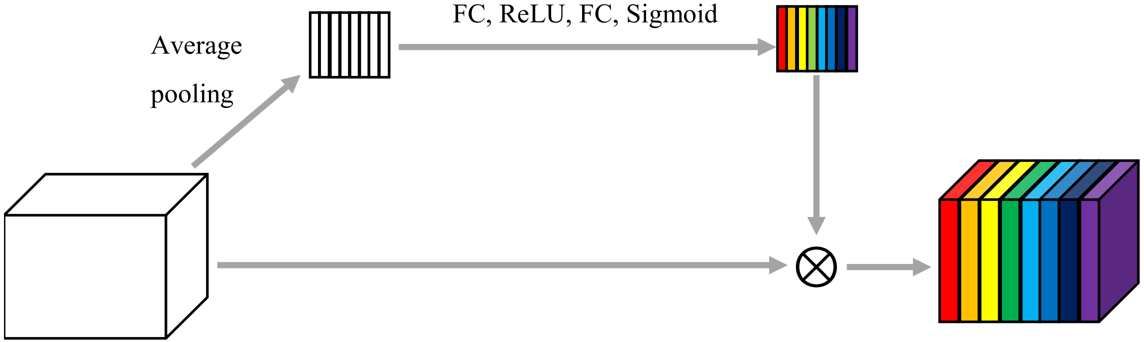




Squeeze and Excitation (SE)

- A channel attention block.
- Allow the network to focus on informative channels and suppress less useful ones.





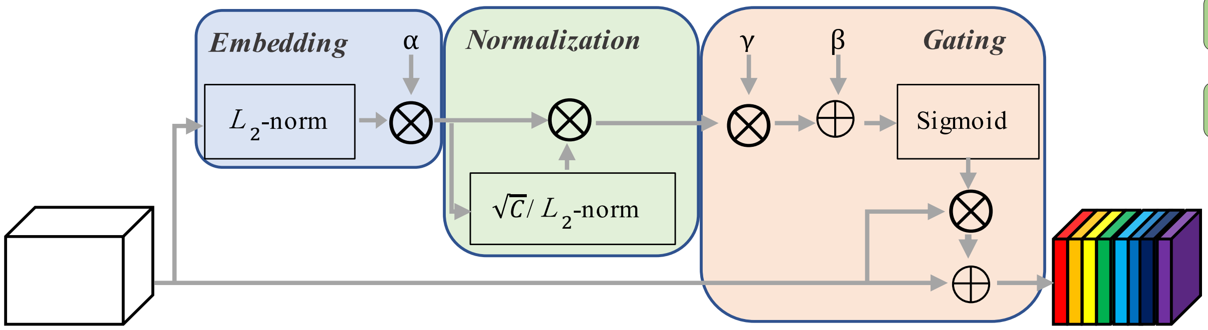
Squeeze: $y_c = \mathrm{F}_{sq}(\mathrm{x}) = \frac{1}{D \times H \times W} \sum_{i=1}^{D} \sum_{j=1}^{H} \sum_{k=1}^{W} x_c(i,j,k)$

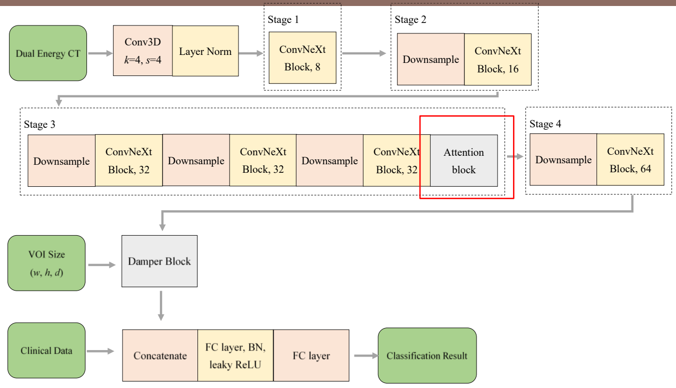
Excitation: $\mathrm{z} = \mathrm{F}_{ex}(\mathrm{y}) = \sigma(\mathrm{W}_2 \delta(\mathrm{W}_1 \mathrm{y}))$

Weight recalibration: $\tilde{\mathrm{x}} = \mathrm{F}_{scale}(\mathrm{x}, \mathrm{z}) = z_c x_c$

Gated Channel Transformation (GCT)

- A channel attention block.
- Enhance the channel-wise feature representation, similar with SE block.





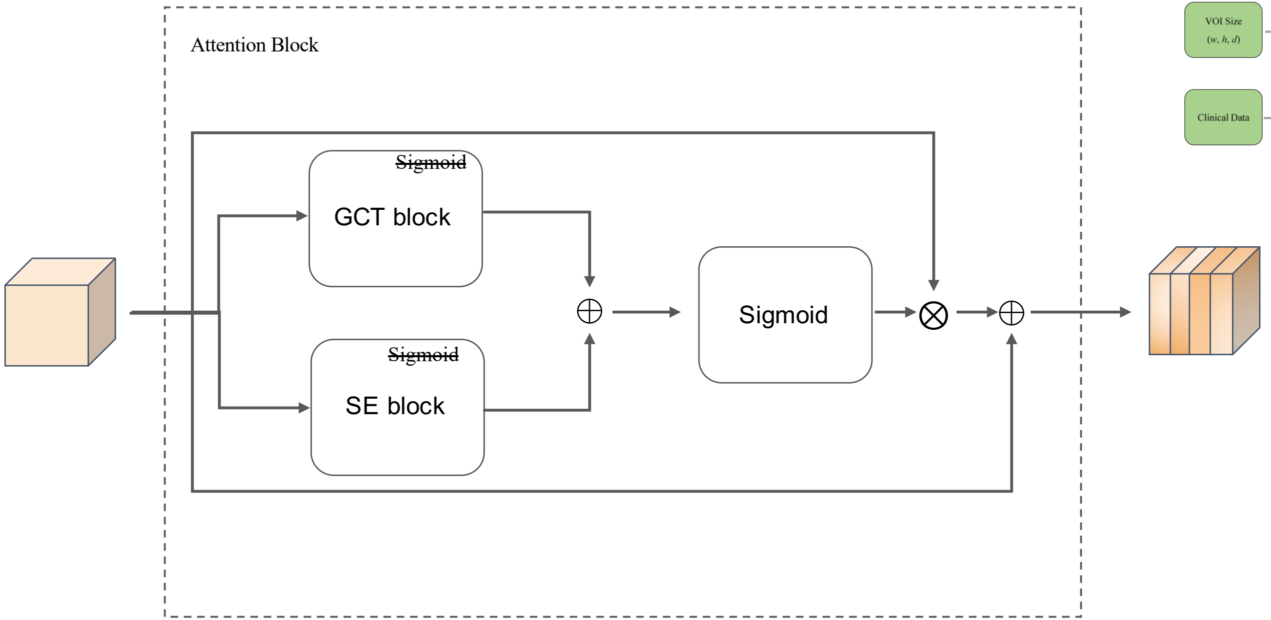
Embedding: $y_c = \mathrm{F}_{embedding}(\mathbf{x}) = \alpha_c \|x_c\|_2 = \alpha_c\left\{\left[\sum_{i=1}^{D}\sum_{j=1}^{H}\sum_{k=1}^{W} x_c(i,j,k)^2\right] + \epsilon\right\}^{\frac{1}{2}}$

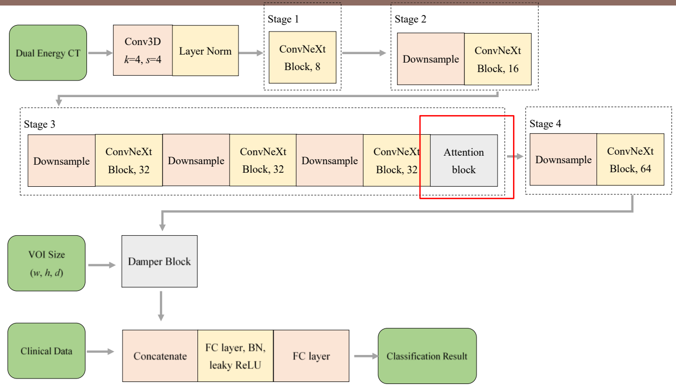
Normalization: $\mathrm{z} = \mathrm{F}_{normalization}(\mathbf{y}) = \frac{\sqrt{C}y_c}{\|\mathbf{y}\|_2} = \frac{\sqrt{C}y_c}{[(\sum_{c=1}^{C} y_c^2)+\epsilon]^{\frac{1}{2}}}$

Gating: $\tilde{\mathrm{x}} = \mathrm{F}_{gating}(\mathbf{x}, \mathbf{z}) = x_c[1 + \sigma(\gamma_c z_c + \beta_c)]$

Proposed Attention Block

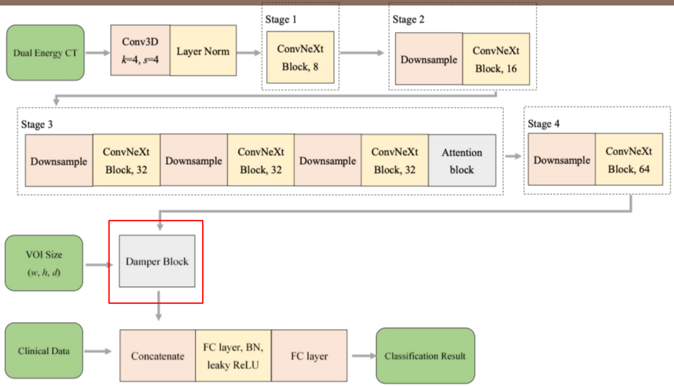
- Combine SE block and GCT block to one channel attention block.
- Allow the network to focus on informative channels and suppress less useful ones.

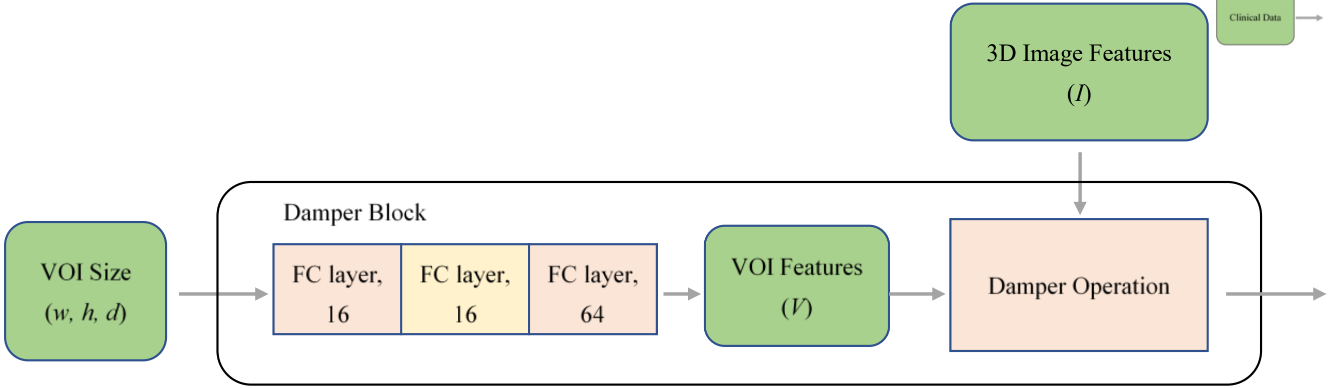




Damper block

- In the proposed CNN model, the size of input images is fixed. Thus, the output features do not capture information about the size of the VOI.
- Aim to normalize the output features by considering the size of the VOI.





Damper Operation: $\tilde{x} = I - \frac{Sum(I.*V)}{n(I.*V)}$

Focal Loss

- Help the model pay greater attention to challenging samples.
- γ allows the reduction of the loss for easily predictable samples.
- α assigns different weights to each class, further reducing the loss.
- The accumulated loss for easily predictable samples will not be too large.

$$p_t = f(p, y) = \begin{cases} p, & y = 1 \\ 1 - p, & y = 0 \end{cases}$$

$$\mathrm{CL}(p_t) = -\log(p_t)$$

$$\text{Focal Loss} = \mathrm{FL}(p_t) = -\alpha(1 - p_t)^{\gamma}\log(p_t)$$

- p: predicted probability
- y: label
- γ: modulating factor
- α: modulating factor

Contents

Dual Energy CT with Four Different keV

- Select 40, 70, 100, and 140 keV as the experimental settings and compared the results with and without contrast agents.
- "-" : without contrast, while "+" : contrast

keV	ACC (%)	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	AUC
C-40	85.60±6.42	**82.86±11.95**	86.19±7.85	54.38±18.51	96.70±2.32	0.8871±0.0523
C-70	85.18±6.17	76.19±8.91	86.64±7.17	52.72±18.70	95.63±1.48	0.8867±0.0571
C-100	85.18±6.17	**82.86±11.95**	85.69±7.63	53.29±18.25	96.69±2.32	0.8887±0.0495
C-140	84.75±6.25	73.33±7.22	86.64±7.17	51.81±18.97	95.09±1.27	0.8853±0.0571
C+40	85.59±6.09	73.33±12.42	87.62±8.11	56.94±25.07	95.24±1.89	0.8900±0.0467
C+70	85.59±6.09	73.81±11.42	**87.64±8.24**	57.34±24.93	95.23±1.92	0.8742±0.0654
C+100	85.60±6.42	80.00±16.29	86.69±8.78	**57.71±25.29**	96.28±2.82	0.8858±0.0486
C+140	**86.03±6.45**	**82.86±11.95**	86.69±7.84	55.29±18.35	**96.72±2.33**	**0.8908±0.0470**

Different Attention Blocks

- We compared the performances using different attention blocks.
- Include Bottleneck Attention Module (BAM) and Convolutional block Attention Module (CBAM).

	ACC (%)	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	AUC
SE	85.59±6.09	73.33±7.22	87.62±7.29	53.81±18.06	95.16±1.19	0.8889±0.0499
GCT	84.32±6.47	82.38±6.21	84.67±8.51	51.98±18.63	96.66±0.88	0.8788±0.0583
BAM	85.62±7.48	70.48±17.63	**88.19±9.39**	**59.72±28.09**	94.79±2.68	0.9002±0.0381
CBAM	84.75±5.48	76.19±13.47	86.14±8.01	51.99±17.96	95.78±2.37	0.8712±0.0550
SE + GCT	**86.03±6.45**	**82.86±11.95**	86.69±7.84	55.29±18.35	**96.72±2.33**	0.8908±0.0470
BAM + GCT	84.77±7.50	76.19±8.91	86.19±8.61	54.72±26.88	95.60±1.52	**0.9037±0.0410**
CBAM + GCT	84.76±5.01	70.48±2.13	87.17±5.82	51.21±18.38	94.61±0.29	0.8740±0.0574

Positions of Attention Block

- We placed the Attention block at different stages of the ConvNeXt blocks.
- We primarily placed the attention block in the third and fourth stages because these stages capture more detailed features.

ConvNeXt Block						ACC (%)	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	AUC
Stage 1	Stage 2	Stage 3			Stage 4						
						84.32±6.47	79.52±12.55	85.17±9.47	**55.31±25.77**	96.24±1.81	**0.8926±0.0405**
				✓		**86.03±6.45**	**82.86±11.95**	86.69±7.84	55.29±18.35	**96.72±2.33**	0.8908±0.0470
		✓	✓	✓		85.59±6.09	76.19±8.91	87.12±7.17	53.61±18.04	95.66±14.59	0.8845±0.0438
					✓	85.59±5.71	76.67±12.42	87.14±7.33	54.23±16.85	95.71±2.14	0.8836±0.0458
				✓	✓	84.75±5.48	77.14±16.29	86.19±7.85	52.83±16.93	95.74±2.86	0.8858±0.0507
✓	✓	✓	✓	✓	✓	84.75±6.25	64.76±7.22	**88.14±8.12**	55.14±26.14	93.71±1.14	0.8830±0.0496

Ablation Studies

- Comparison between the model with and without the damper block.

Damper Block	ACC (%)	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	AUC
Without	85.15±5.26	76.67±7.22	86.60±5.83	50.75±9.85	95.60±1.59	**0.9163±0.0404**
With	**86.03±6.45**	**82.86±11.95**	**86.69±7.84**	**55.29±18.35**	**96.72±2.33**	0.8908±0.0470

- Comparison between Cross Entropy Loss and Focal Loss.

	ACC (%)	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	AUC
Cross Entropy Loss	85.19±6.14	80.00±16.29	86.21±8.33	54.05±17.63	96.26±2.87	**0.8954±0.0358**
Focal Loss	**86.03±6.45**	**82.86±11.95**	**86.69±7.84**	**55.29±18.35**	**96.72±2.33**	0.8908±0.0470

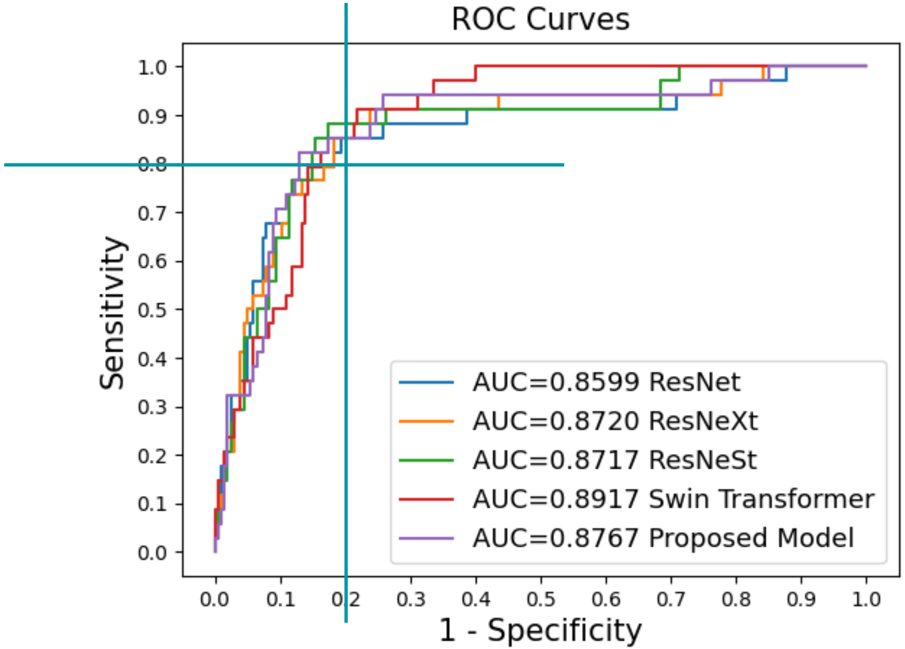
Comparison with Other Models

- We compared different CNN models.
- The proposed model achieves the highest performance.

CNNs	ACC (%)	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	AUC
ResNet	84.32±8.06	79.52±7.45	85.14±9.19	51.53±16.94	96.04±1.70	0.8747±0.0428
ResNeXt	84.77±4.47	76.67±12.42	86.21±6.90	52.23±16.58	95.73±1.98	0.8787±0.0508
ResNeSt	84.30±6.18	79.52±16.11	85.10±6.47	49.05±11.99	96.12±3.20	0.8785±0.0319
Swin Transformer	83.48±5.03	**82.86±11.95**	83.71±6.99	48.00±10.41	96.66±2.21	**0.8940±0.0495**
Proposed Model	**86.03±6.45**	**82.86±11.95**	**86.69±7.84**	**55.29±18.35**	**96.72±2.33**	0.8908±0.0470

ROC curve

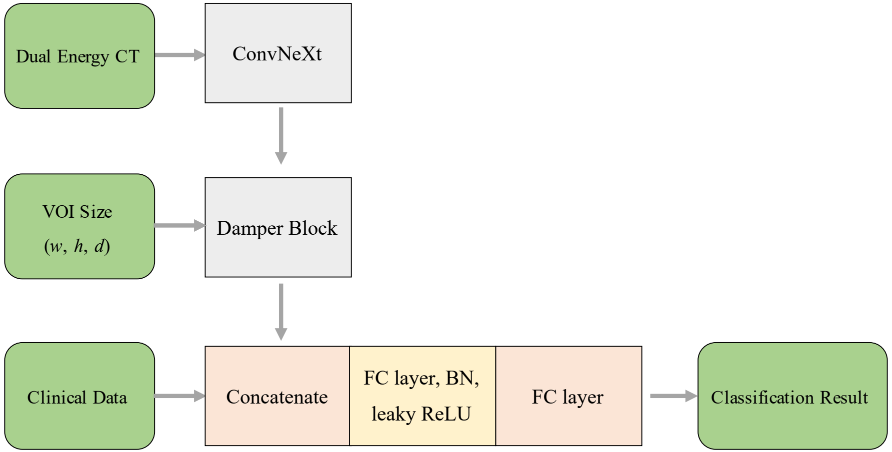
- Swin Transformer achieves the highest AUC.
- We manly focus on sensitivity > 0.8 and (1 – specificity) < 0.2.



Contents

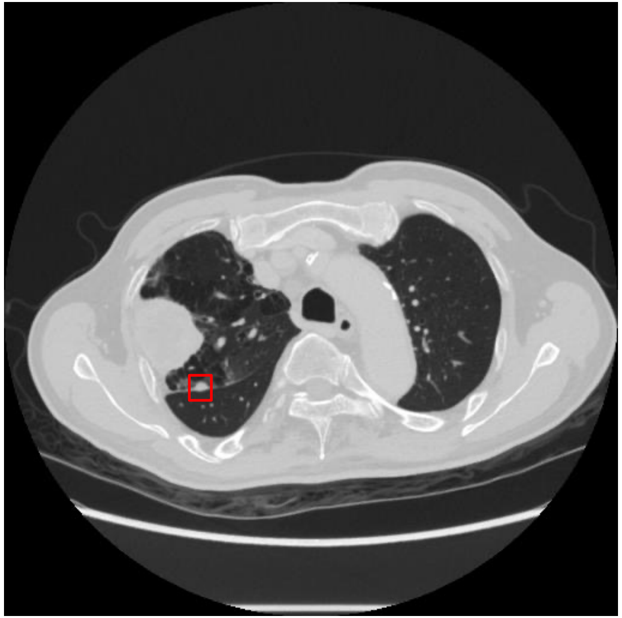
Discussion

- Utilize ConvNeXt and incorporated channel attention block in filtering image features.
- Bring in damper block to combines detailed image features with tumor size features.
- Merge the features from the images with clinical and pathology information.

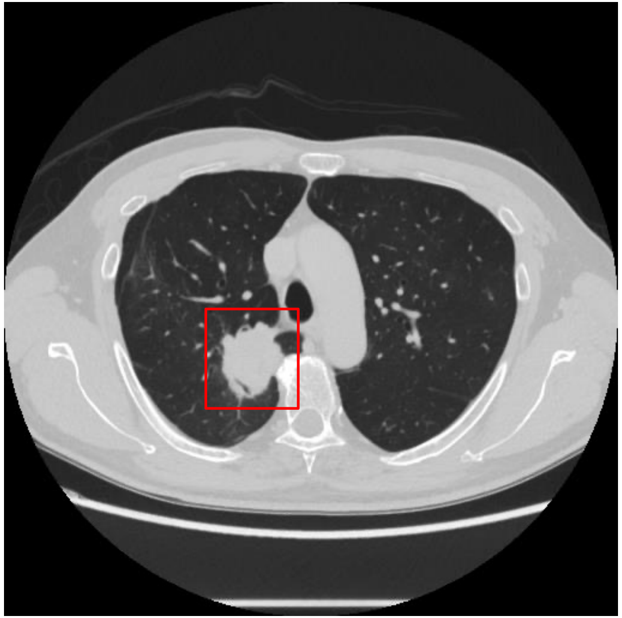


Cross Entropy Loss vs Focal Loss

- Two samples that model predicts correctly with focal loss, but predicts wrongly with cross entropy loss.



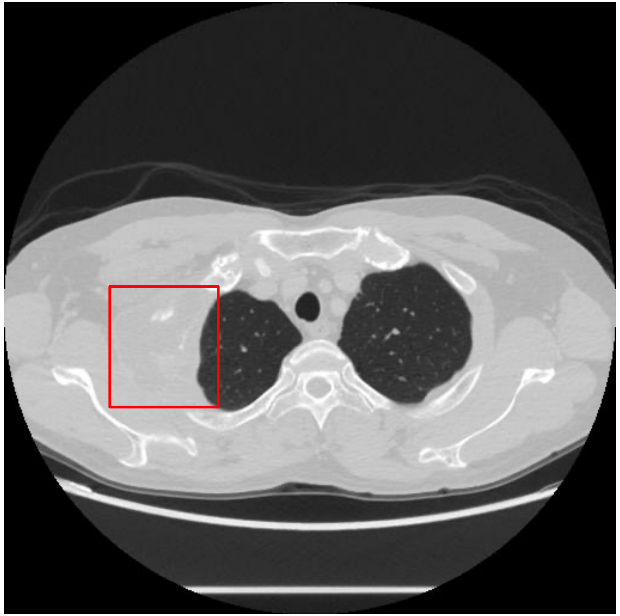
Deceased, brain metastasis

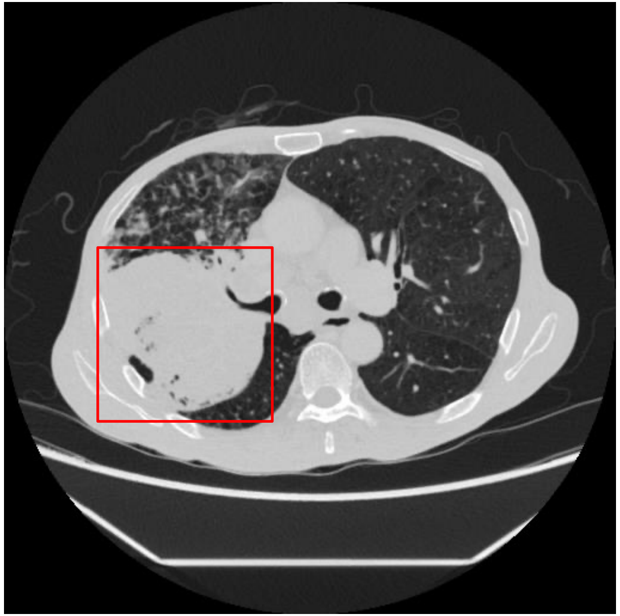


Survivor

Model With or Without Damper Block - Deceased

- Two deceased samples are predicted correctly from model with damper block, but wrongly from model without damper block.
- Incorporating damper block, model predicts survival outcome correctly for the data that tumor size > 5 cm.

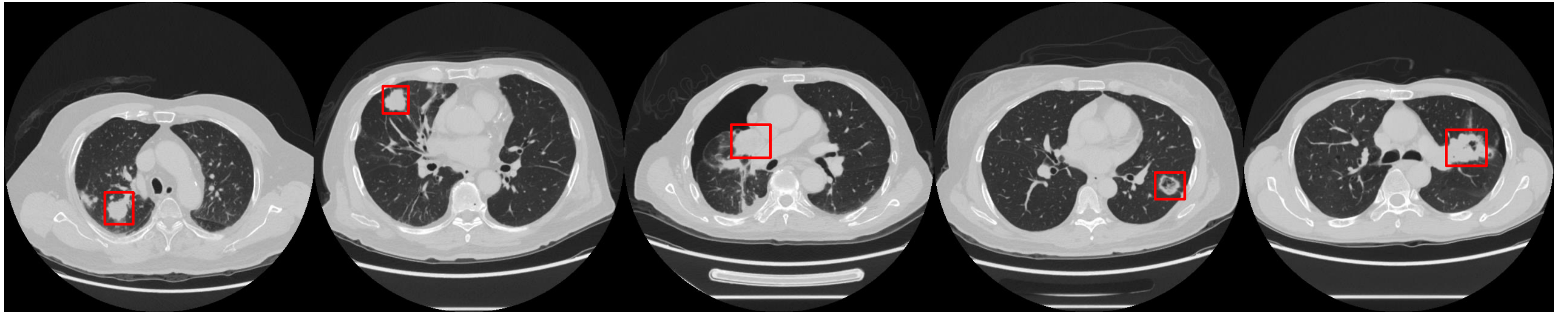




Model With or Without Damper Block - Survivor

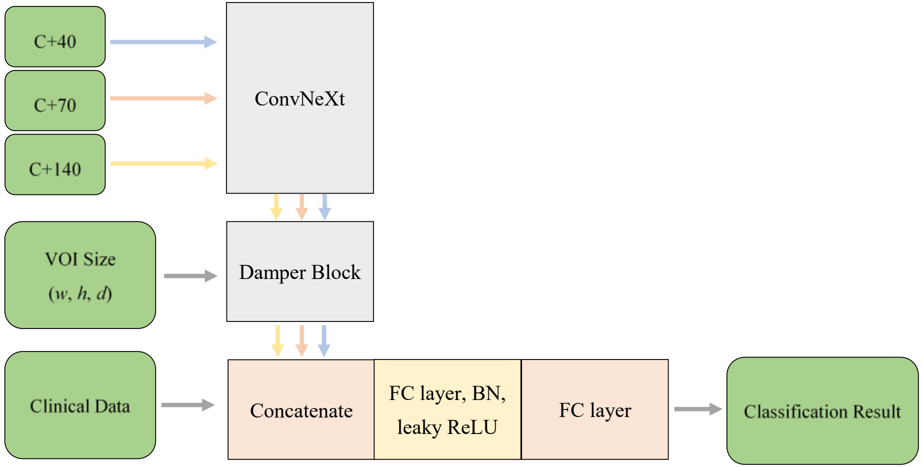
- Model predicts wrongly after adding damper block.

- Model predicts correctly after adding damper block.



Train the model simultaneously using 40, 70, and 140 keV

- Train the model simultaneously using 40, 70, and 140 keV with contrast agent enhancement.



keV	ACC (%)	SEN (%)	SPEC (%)	PPV (%)	NPV (%)
C+40, C+70, C+140	84.75±6.14	76.19±16.29	86.14±8.33	54.75±17.63	95.83±2.87
C+140	**86.03±6.45**	**82.86±11.95**	**86.69±7.84**	**55.29±18.35**	**96.72±2.33**

Comparison with Other Papers

- We compared the performances of different papers.
- The Dataset is not the same.

CNNs	ACC (%)	SEN (%)	SPEC (%)	PPV (%)	NPV (%)	AUC
Agrawal, A., et al.	85.50 (48952/57254)	-	-	-	-	0.8870
Lai, Y-H., et al.	75.44 (129/171)	-	-	-	-	0.8163
He, B., et al.	84.00 (63/75)	83.78 (31/37)	84.21 (32/38)	83.78 (31/37)	84.21 (32/38)	-
Chang Sin-You	80.78 (353/437)	73.57 (103/140)	84.18 (250/297)	68.67 (103/150)	87.11 (250/287)	0.8425
Proposed Model	86.02 (203/236)	82.35 (28/34)	86.63 (175/202)	50.91 (28/55)	96.69 (175/181)	0.8767

Contents

Conclusion

The study:

- Design a CNN to predict the 3-year survival outcomes of lung cancer patients.
- We achieved favorable results by making certain architectural adjustments and integrating image and clinical, pathological data.
- Address the issue of label imbalance in the dataset.

Future work:

- Continuous data collection and ongoing improvement to the proposed model are necessary.

Thank you.