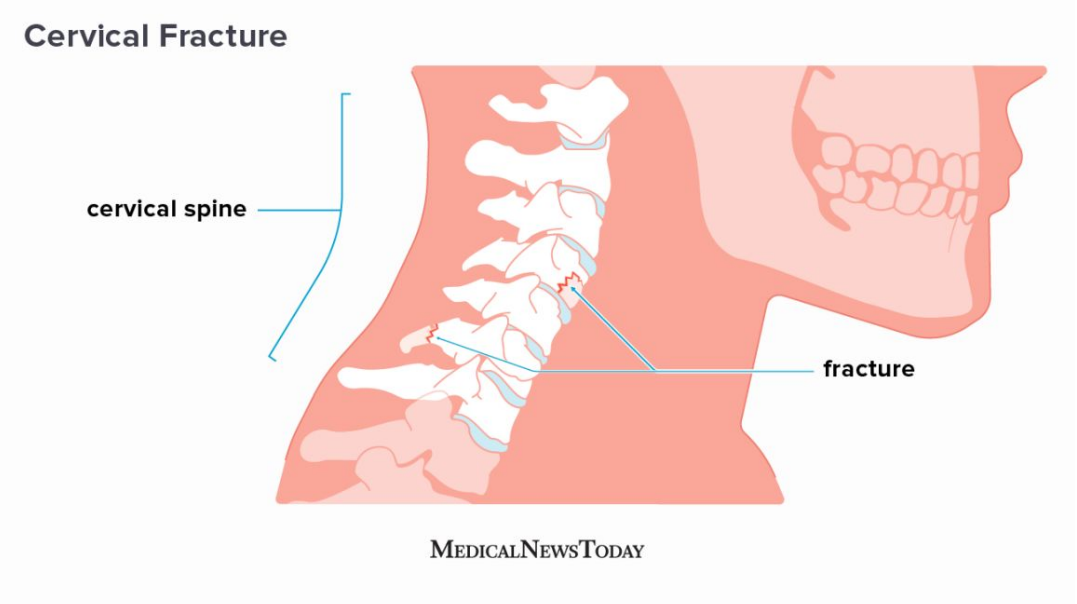
Cervical Spine Fracture Detection

Team presentation 1

W210 Capstone Research Proposal
Fall 2022 - Section 3
Weijia Li, Jane Hung, Minjie Xu, Fengyao Luo

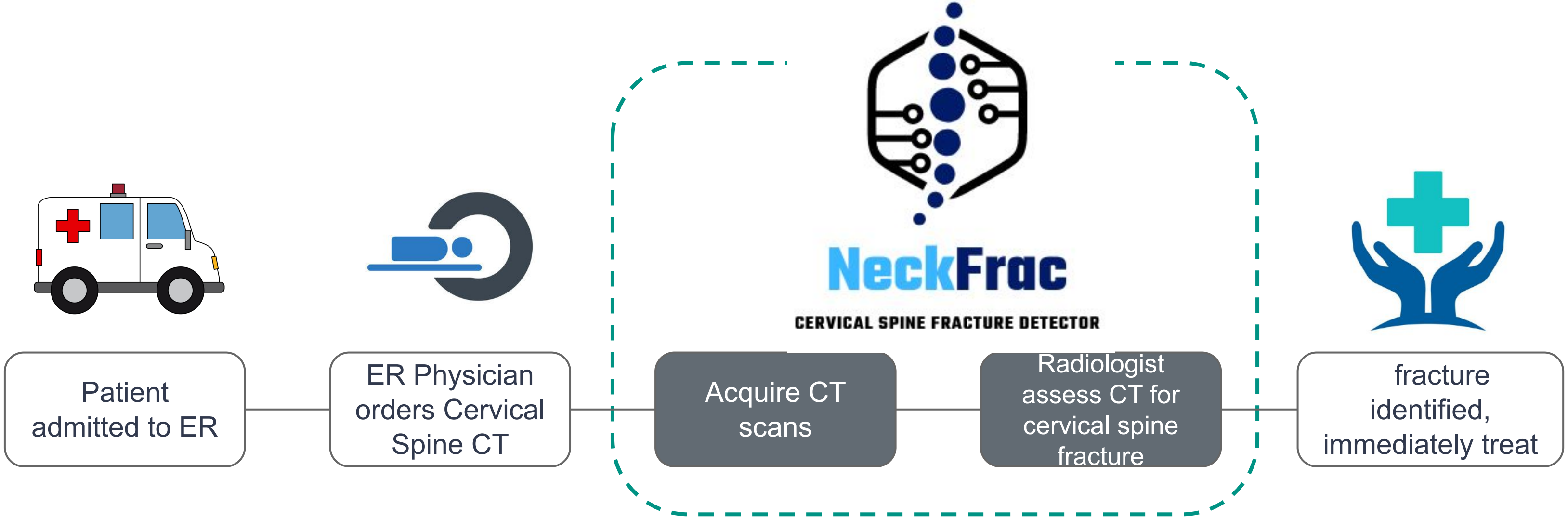
Background



- Cervical Spine Fractures (broken neck)
- **1.5 million** vertebral compression fractures occur
- 3 millions patients per year
- only **25–33%** of incident radiographically identified vertebral fractures are clinically diagnosed
- Quickly detecting and determining the location of vertebral fractures is essential for prevent paralysis after trauma.
- RSNA Cervical Spine Fracture AI Challenge

Mission

Quicker, better, more accurate diagnosis to save lives.



- Traditionally, this process takes 30 mins on average with the first priority.
- With **NeckFrac**, this process will take less than 10 mins.

Competitor Analysis

Direct Competitor

Emerging			Legacy
	aidoc ALWAYS ON AI	Medtronic Global Healthcare Technology Leader	
	one medical amazon	 Radiologists	

Indirect Competitor

Literature Review

	A	B	C	D	E	F	G	H
1	**Author**	**Title**	**Journal**	**Dataset & Sample Size**	**Main Method**	**Metric(s)**	**Best Results and Model Used**	**Reference**
2	Miyamura, Satoshi, et al.	Bone density measurements from CT scans may predict the healing capacity of scaphoid waist fractures	The Bone & Joint Journal	34 scaphoid waist fracture CTs; 17 age-matched scaphoid waist fracture CTs that healed (union group),), and 17 age-matched control CTs without injury (control group).	Correlation (2D and 3D)	Spearman's correlation coefficient (R) ROC curve	2D measurements were highly correlated to 3D bone density measurements (Spearman's correlation coefficient (R) = 0.85 to 0.95); Optimal cutoffs of 90.8% and 116.3% for distal and proximal fragments (ROC curve); Sensitivity of 1.00 if either cutoff is met and 0.82 when both cutoffs are met.	Miyamura, S., Lans, J., He, J. J., Murase, T., Jupiter, J. B., & Chen, N. C. (2020). Bone density measurements from CT scans may predict the healing capacity of scaphoid waist fractures. The Bone & Joint Journal, 102(9), 1200-1209.
3	Schreiber, Joseph J., et al.	Use of computed tomography for assessing bone mineral density	Neurosurgical focus	CT scans from 80 patients with trauma, 20 patients with adjacent-segment fractures along with 20 matched nonfracture controls	Correlation (Hounsfield unit (HU) and bone mineral density (BMD))	Interclass correlation Pearson correlation coefficient	Interclass correlation coefficients of 0.964 and 0.975 (reliability test); The Pearson correlation coefficients were 0.44 and 0.48 for BMD and T-score	Schreiber, J. J., Anderson, P. A., & Hsu, W. K. (2014). Use of computed tomography for assessing bone mineral density. Neurosurgical focus, 37(1), E4.
4	Salehinejad, Hojjat, et al.	Deep Sequential Learning For Cervical Spine Fracture Detection On Computed Tomography Imaging	2021 IEEE 18th International Symposium on Biomedical Imaging	An annotated dataset of 3,666 CT scans (729 positive and 2,937 negative cases)	deep convolutional neural network (DCNN) with bidirectional long-short term memory (BLSTM) layer	TPR, TNR, PPV, NPV, F1, Acc, MCC, AUC	Classification accuracy of 70.92% and 79.18% on the balanced (104 positive and 104 negative cases) and imbalanced (104 positive and 419 negative cases) test datasets	Salehinejad, H., Ho, E., Lin, H. M., Crivellaro, P., Samorodova, O., Arciniegas, M. T., ... & Colak, E. (2021, April). Deep sequential learning for cervical spine fracture detection on computed tomography imaging. In 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI) (pp. 1911-1914). IEEE.
5	Small, J. E., et al.	CT Cervical Spine Fracture Detection Using a Convolutional Neural Network	American Journal of Neuroradiology	665 examinations	C-spine, an FDA-approved convolutional neural network developed by Aidoc to detect cervical spine fractures on CT	k coefficients, sensitivity/specificity, PPV, and NPV	Convolutional neural network accuracy in cervical spine fracture detection was 92% (95% CI, 90%–94%), with 76% (95% CI, 68%–83%) sensitivity and 97% (95% CI, 95%–98%) specificity; compared with the radiologist accuracy of 95% (95% CI, 94%–97%), with 93% (95% CI, 88%–97%) sensitivity and 96% (95% CI, 94%–98%) specificity.	Small, J. E., Osler, P., Paul, A. B., & Kunst, M. (2021). Ct cervical spine fracture detection using a convolutional neural network. American Journal of Neuroradiology, 42(7), 1341-1347.
6	Murata, Kazuma, et al.	Artificial intelligence for the detection of vertebral fractures on plain spinal radiography	Scientific Reports by Nature	A total of 300 patients in the registry (150 patients with vertebral fracture (VF) and 150 without VF)	DCNN	accuracy, sensitivity, specificity	The DCNN achieved accuracy, sensitivity, and specificity rates of 86.0% [95% confidence interval (CI) 82.0–90.0%], 84.7% (95% CI 78.8–90.5%), and 87.3% (95% CI 81.9–92.7%), respectively.	Murata, K., Endo, K., Aihara, T., Suzuki, H., Sawaji, Y., Matsuoka, Y., ... & Yamamoto, K. (2020). Artificial intelligence for the detection of vertebral fractures on plain spinal radiography. Scientific Reports, 10(1), 1-8.
7	Li, Yi-Chu, et al.	Can a Deep-learning Model for the Automated Detection of Vertebral Fractures Approach the Performance Level of Human Subspecialists?	Clinical Orthopaedics and Related Research	Patients older than 60 years were treated for vertebral fractures treated in Taipei Veterans General Hospital	CNN, ResNet34, ResNet50, DenseNet121, DenseNet169, and DenseNet201	accuracy, sensitivity, specificity, interobserver reliability (kappa value), ROC curve, AUC	accuracy (93% [773 of 830] of vertebrae), sensitivity (91% [129 of 141]), and specificity (93% [644 of 689]) for detecting vertebral fractures of the lumbar spine. The interobserver reliability (kappa value) from model and human observers were 0.72 (95% CI 0.65 to 0.80; p < 0.001) and 0.77 (95% CI 0.72 to 0.83; p < 0.001), respectively. The AUCs for Grades 1, 2, and 3 vertebral fractures were 0.919, 0.989, and 0.990, respectively.	Li, Y. C., Chen, H. H., Lu, H. H. S., Wu, H. T. H., Chang, M. C., & Chou, P. H. (2021). Can a deep-learning model for the automated detection of vertebral fractures approach the performance level of human subspecialists?. Clinical Orthopaedics and Related Research®, 479(7), 1598-1612.
8	Sebro, R., et al.	Utilizing machine learning for opportunistic screening for low BMD using CT scans of the cervical spine	Journal of Neuroradiology	A retrospective study of 253 patients (177 training/validation and 76 test) with unenhanced CT scans of the cervical spine and Dual-energy x-ray Absorbtiometry (DXA) studies	Random Forest (RF), XGBoost, Naïve Bayes (NB), and Support Vector Machines (SVM)	correlations, ROC curve, AUC	A CT attenuation threshold of 305.2 HU at C3 had the highest accuracy (0.763, AUC=0.814) to detect femoral neck BMD T-scores ≤-1 and a CT attenuation threshold of 323.6 HU at C3 had the highest accuracy (0.774, AUC=0.843) to detect osteopenia/osteoporosis. The SVM classifier (AUC=0.756) had higher AUC than the RF (AUC=0.692, P=0.224), XGBoost (AUC=0.736; P=0.814), NB (AUC=0.622, P=0.133) and CT threshold of 305.2 HU at C3 (AUC=0.704, P=0.531) classifiers to identify patients with femoral neck BMD T-scores <-1. The SVM classifier (accuracy=0.816) was more accurate than using the CT threshold of 305.2 HU at C3 (accuracy=0.671)	Sebro, R., & De la Garza-Ramos, C. (2022). Utilizing machine learning for opportunistic screening for low BMD using CT scans of the cervical spine. Journal of Neuroradiology.
9	Yeh, Lee-Ren, et al.	A deep learning-based method for the diagnosis of vertebral fractures on spine MRI: retrospective training and validation of ResNet	European Spine Journal	A total of 190 patients, 50 with malignant and 140 with benign fractures	ResNet50	TP, TN, FN, FP, Sensitivity, Specificity, Accuracy	The accuracy achieved by using the ResNet50 deep learning model for the identified abnormal vertebral segment was 92%. Compared to the first-year resident's reading, the model improved the sensitivity from 78 to 94% (p < 0.001) and the specificity from 61 to 91% (p < 0.001).	Yeh, L. R., Zhang, Y., Chen, J. H., Liu, Y. L., Wang, A. C., Yang, J. Y., ... & Su, M. Y. (2022). A deep learning-based method for the diagnosis of vertebral fractures on spine MRI: retrospective training and validation of ResNet. European Spine Journal, 1-9.
10	Burns, J. E., et al.	Vertebral Body Compression Fractures and Bone Density: Automated Detection and Classification on CT Images	Radiology	A CT study set of 150 patients (mean age, 73 years; age range, 55–96 years; 92 women, 58 men) with (n = 75) and without (n = 75) compression fractures was assembled	Prototype fully automated spinal segmentation and fracture detection software	Sensitivity, FP, ROC curve	Sensitivity for detection or localization of compression fractures was 95.7% (201 of 210; 95% confidence interval [CI]: 87.0%, 98.9%), with a false-positive rate of 0.29 per patient. Additionally, sensitivity was 98.7% and specificity was 77.3% at case-based receiver operating characteristic curve analysis.	Burns, J. E., Yao, J., & Summers, R. M. (2017). Vertebral body compression fractures and bone density: automated detection and classification on CT images. Radiology, 284(3), 788.
11	Ahammad, Sk Hasane, et al.	A Hybrid CNN-Based Segmentation and Boosting Classifier for Real Time Sensor Spinal Cord Injury Data	IEEE Sensors Journal	Experimental results are simulated on SCI image database taken from Orange image diagnostic center using portable spinal cord MRI scanner and	CNN+RF, CNN+NN, CNN+Linear SVM, CNN+Ensemble,	TP, accuracy, error rate	Experimental results proved that the present model has better performance than the existing spinal cord injury detection models in terms of true positive rate; TP = 0.9859, Accuracy = 0.9894, and Error rate = 0.019 are concerned.	Ahammad, S. H., Rajesh, V., Rahman, M. Z. U., & Lay-Ekuakille, A. (2020). A hybrid CNN-based segmentation and boosting classifier for real time sensor spinal cord injury data. IEEE Sensors Journal

+ Papers | Other paper/webpage/technical report | Competitors | Explo

Datasets

Train.csv

2019 patients in train set (balanced at patient_overall level)

	StudyInstanceUID	patient_overall	C1	C2	C3	C4	C5	C6	C7
0	1.2.826.0.1.3680043.6200	1	1	1	0	0	0	0	0
1	1.2.826.0.1.3680043.27262	1	0	1	0	0	0	0	0
2	1.2.826.0.1.3680043.21561	1	0	1	0	0	0	0	0
3	1.2.826.0.1.3680043.12351	0	0	0	0	0	0	0	0
4	1.2.826.0.1.3680043.1363	1	0	0	0	0	1	0	0

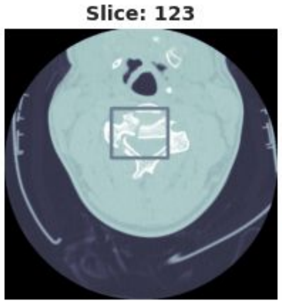
Train Image / Metadata

No missing data, Image size varies, need to resize to 512x512

- Patient ID
- Slice Number
- Image Size
- Slice Thickness
- Image Position Patient
- Image Orientation Patient

Bounding Box

235 patients (12% of train set) have the bounding box



- Most of patient have 15-25 bounding boxes
- Patients rarely have 100 bounding boxes, distribution right skewed

Segmentation

87 patients (4% of train set) have segmentation labelled

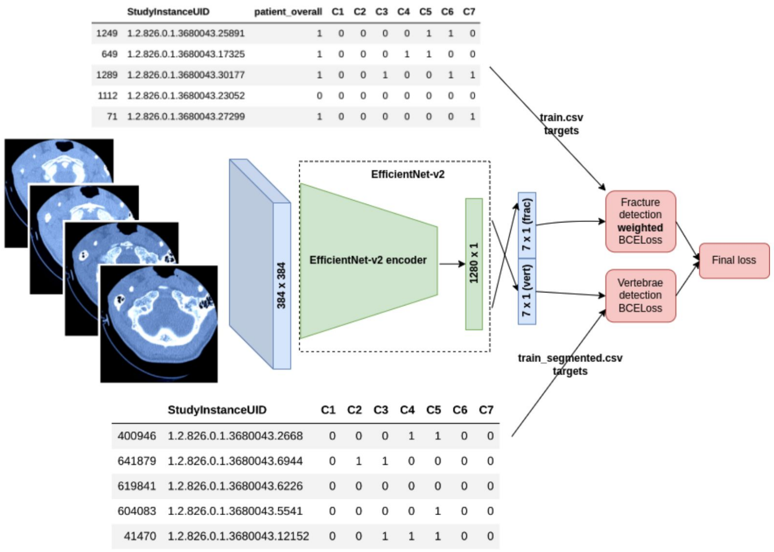
- Image has overlap of adjacent vertebrae
- Need to predict C1-C7 for rest of train set

	C1	C2	C3	C4	C5	C6	C7
150	1	1	0	0	0	0	0

Existing work on Kaggle

C1-C7 Multi-Label Classification

- Random Forest: 0.88 accuracy
- EfficientNet-V2: 0.95 accuracy

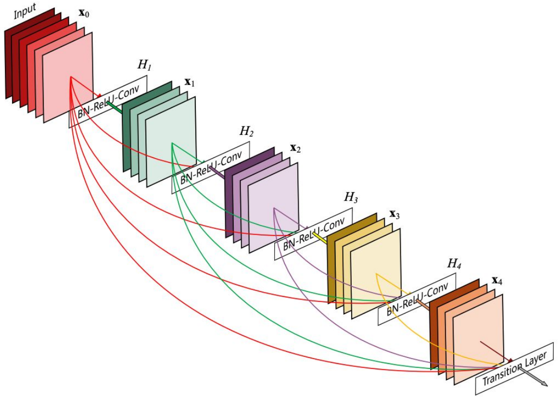


	StudyInstanceUID	patient_overall	C1	C2	C3	C4	C5	C6	C7
1249	1.2.826.0.1.3680043.25891	1	0	0	0	0	1	1	0
649	1.2.826.0.1.3680043.17325	1	0	0	0	1	1	0	0
1289	1.2.826.0.1.3680043.30177	1	0	0	1	0	0	1	1
1112	1.2.826.0.1.3680043.23052	0	0	0	0	0	0	0	0
71	1.2.826.0.1.3680043.27299	1	0	0	0	0	0	0	1

	StudyInstanceUID	C1	C2	C3	C4	C5	C6	C7
400946	1.2.826.0.1.3680043.2668	0	0	0	1	1	0	0
641879	1.2.826.0.1.3680043.6944	0	1	1	0	0	0	0
619841	1.2.826.0.1.3680043.6226	0	0	0	0	0	0	0
604083	1.2.826.0.1.3680043.5541	0	0	0	0	1	0	0
41470	1.2.826.0.1.3680043.12152	0	0	1	1	1	0	0

Fracture Classification

- DenseNet 121: 0.75 loss
- EfficientNet-V2: 0.49 loss



Evaluation Metrics

Product Use Case

- Evaluation time
- Accuracy
- Recall (Sensitivity)
- Precision (Positive Predictive Value)
- F1 score

AI versus ground truth

Ground Truth

AI	+	-	
+	109	17	*Positive Predictive Value: 87%*
-	34	505	*Negative predictive Value: 94%*

Sensitivity: 76%. Specificity: 97%. Accuracy: 92%.

Evaluation time: 3-8 min range

Kaggle

- Accuracy
- Weighted multi-label logarithmic loss function

Radiologist versus ground truth

Ground Truth

Rad	+	-	
+	133	20	*Positive Predictive Value: 87%*
-	10	502	*Negative Predictive Value: 98%*

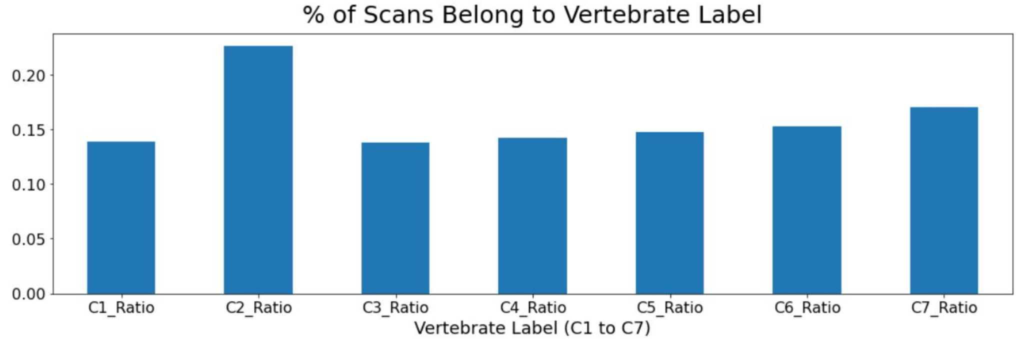
Sensitivity: 93%. Specificity: 96%. Accuracy: 96%.

Evaluation time: 33-43 min range

Initial Modeling
C1-C7 Classification

	StudyInstanceUID	Slice	C1	C2	C3	C4	C5	C6	C7
1362	1.2.826.0.1.3680043.12281	170	0	0	1	1	0	0	0
1363	1.2.826.0.1.3680043.12281	171	0	0	1	1	0	0	0
1364	1.2.826.0.1.3680043.12281	172	0	0	1	1	0	0	0
1365	1.2.826.0.1.3680043.12281	173	0	0	0	1	0	0	0
1366	1.2.826.0.1.3680043.12281	174	0	0	0	1	0	0	0
1367	1.2.826.0.1.3680043.12281	175	0	0	0	1	0	0	0
1368	1.2.826.0.1.3680043.12281	176	0	0	0	1	0	0	0
1369	1.2.826.0.1.3680043.12281	177	0	0	0	1	0	0	0
1370	1.2.826.0.1.3680043.12281	178	0	0	0	1	0	0	0
1371	1.2.826.0.1.3680043.12281	179	0	0	0	1	0	0	0

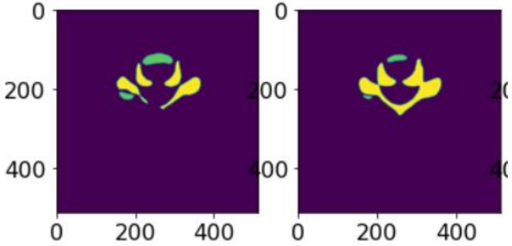
Imbalance dataset



Baseline Accuracy

```
Accuracy of C1: 86.26 %
Accuracy of C2: 77.78 %
Accuracy of C3: 86.45 %
Accuracy of C4: 85.97 %
Accuracy of C5: 85.63 %
Accuracy of C6: 85.18 %
Accuracy of C7: 83.52 %

Overall accuracy: 84.4 %
```



Initial Modeling - C1-C7 Classification Model Selection

	StudyInstanceUID	Slice	ImageHeight	ImageWidth	SliceThickness	ImagePositionPatient_x	ImagePositionPatient_y	ImagePositionPatient_z	C1	C2	C3	C4	C5	C6	C7
500	1.2.826.0.1.3680043.10921	72	512	512	0.625	-67.4	-85.0	-43.375	1	1	0	0	0	0	0
501	1.2.826.0.1.3680043.10921	73	512	512	0.625	-67.4	-85.0	-44.000	1	1	0	0	0	0	0
502	1.2.826.0.1.3680043.10921	74	512	512	0.625	-67.4	-85.0	-44.625	1	1	0	0	0	0	0
503	1.2.826.0.1.3680043.10921	75	512	512	0.625	-67.4	-85.0	-45.250	1	1	0	0	0	0	0
504	1.2.826.0.1.3680043.10921	76	512	512	0.625	-67.4	-85.0	-45.875	1	1	0	0	0	0	0
505	1.2.826.0.1.3680043.10921	77	512	512	0.625	-67.4	-85.0	-46.500	1	1	0	0	0	0	0

- **Train and test split based on patient ID**
- **Oversampling on training dataset**
- **Next step to add image to the model with EffNetV2**

Model	**Accuracy (%)**	**Parameters**
Baseline	**86.26**	Assign label to majority vote
Decision Tree	81.35	{'criterion': 'gini', 'max_depth': 32, 'min_samples_leaf': 4, 'min_samples_split': 32}
Random Forest	85.10	{'max_depth': 64, 'n_estimators': 128}
XGBoost	83.59	{'learning_rate': 0.05, 'max_depth': 9, 'n_estimators': 140}
KNN	81.70	{'leaf_size': 1, 'metric': 'chebyshev', 'n_neighbors': 1, 'p': 1, 'weights': 'uniform'}

Initial Modeling – Baseline Fracture Prediction

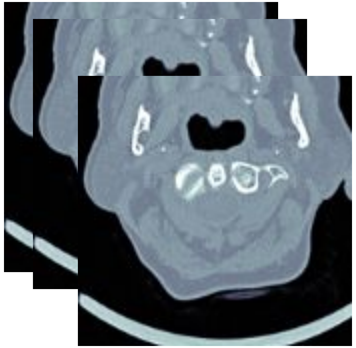
Train.csv

meta_train.csv

Segmentation.csv

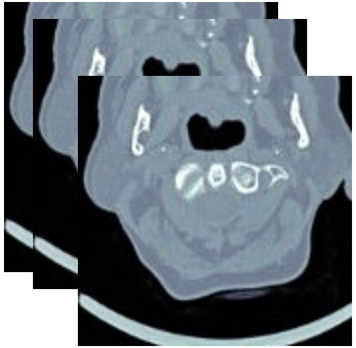
join

	StudyInstanceUID	ImageHeight	ImageWidth	SliceThickness	Slice	ImagePositionPatient_x	ImagePositionPatient_y	ImagePositionPatient_z	C1	C2	...	C6	C7	patient_overall	C1_fracture	C2_fracture
0	1.2.826.0.1.3680043.10633	512	512	1.0	377	-68	98	126.099976	0	0	...	0	0	0	0	0
1	1.2.826.0.1.3680043.10633	512	512	1.0	257	-68	98	186.099976	0	0	...	1	0	0	0	0
2	1.2.826.0.1.3680043.10633	512	512	1.0	240	-68	98	194.599976	0	0	...	0	0	0	0	0
3	1.2.826.0.1.3680043.10633	512	512	1.0	317	-68	98	156.099976	0	0	...	0	1	0	0	0
4	1.2.826.0.1.3680043.10633	512	512	1.0	12	-68	98	308.599976	0	0	...	0	0	0	0	0

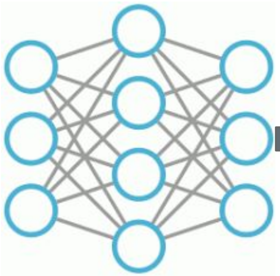


(height, width, channel)

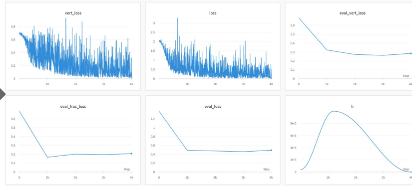
Transpose (img, (2,0,1))



(channel, height, width)

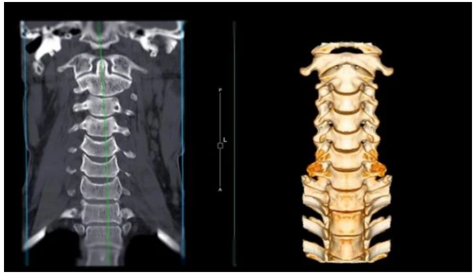


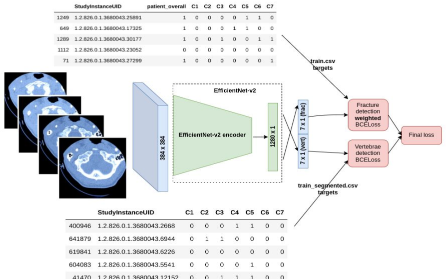
Efficient Net V2



Wandb

Challenges





- **Lack of domain knowledge**

- **Data**
  - Data quality
    - Low resolution images will impact the model’s performance and reduce the accuracy of the model.
  - Data size
    - large dataset might take up too much resources
  - Data assumption
    - The patient's’ demographic information is uniformly distributed (e.g. ethnicity, age, height, weight, gender)
    - the patient's’ physical condition is similar

- **Methodology**
  - Learning curve on deep learning models used in image processing
  - Working with 3D datasets (sagittal view vs. axial view)

Next Steps

- Understand Medical image
- Sagittal views
- Submit baseline score to Kaggle
- C1-C7 Classification with EffNetV2
- Feature Engineering
- Migrate to AWS
- Data Pipeline