Abdominal Trauma Classification from CT Scans

Wei Xin Lim[a], Wai Lam Hoo[a]

[a]*Faculty of Computer Science and Information Technology, Universiti Malaya, Kuala Lumpur, 50603, Wilayah Persekutuan Kuala Lumpur, Malaysia*

Abstract

Abdominal trauma, particularly from blunt force, presents critical challenges in emergencies due to potential multi-organ injury. Timely assessment is critical for patient outcomes, yet the current reliance on radiologists for interpreting CT scans is both time-consuming and burdensome, posing challenges in clinical practice. This study aims to develop an automated system that is capable of localizing abdominal organs and classify the severity of the trauma based on computed tomography (CT) scans. Dataset of RSNA Abdominal Trauma Detection Challenge 2023 was used to train and experiment with the 2.5D approach in conjunction with Focal Loss and Label Smoothing of different parameters. The outcomes were evaluated and concluded that using only Focal Loss with Gamma (γ) value set to 1 outperforms the baseline model, achieving 0.88 and 0.211 in AUROC score and RSNA Trauma Metric respectively. A Streamlit and FastAPI-powered dashboard was then developed to optimize healthcare providers' workload, offering personalized diagnostic support with the integrated, best performing model at its backend.

Keywords: abdominal trauma, abdominal trauma classification, computed tomography, 2.5D approach, computer vision

1. Introduction

Abdominal trauma presents a significant clinical challenge, especially in emergency situations, as it encompasses a broad range of injuries from minor contusions to life-threatening hemorrhage and organ damage (Barrett and Smith, 2012). Blunt force trauma, which is among the most common types of traumatic injury from motor vehicle accidents can injure multiple organs like the liver, kidneys and spleen, resulting in lesions or bleeding that is potentially fatal (Errol Colak et al., 2023). Patient outcomes are closely correlated with the speed and accuracy of assessing the extent and severity of abdominal trauma. For hemodynamically stable patients with abdominal trauma, findings from Computed Tomography (CT) scans are crucial for healthcare experts to decide on medical procedures to control potentially fatal hemorrhage and optimize resuscitation strategies (Cheng et al., 2023; Errol Colak et al., 2023; Latif et al., 2023).

However, these potentially lethal injuries are sometimes still overlooked in current practice, as interpreting CT scans for abdominal trauma can be complex and time-consuming, especially with multiple injuries or subtle bleeding (Cheng et al., 2023). Timely diagnosis is of paramount importance as patients with abdominal trauma require urgent surgery and often cannot be diagnosed clinically by physical exam, symptoms or laboratory test (Errol Colak et al., 2023). The current medical imaging standard also places a substantial burden on radiologists for result interpretation and analysis, which not only prolongs the decision-making process but also heightens the risk of incorrect interpretations, as evidenced by a missed lung cancer case in New Zealand in 2022 (Murphy, 2022).

Application of Deep Learning (DL) in medical imaging has been on going for several years in practice and led to a number of substantial improvements for the healthcare providers and surge of studies by the researchers on more use cases. Numerous researches have revealed the capability of DL for clinical physicians to diagnose accurately using medical images like pathological slides, plain films and even surgical planning. There is also study that provides evidence of outstanding performance of DL algorithms diagnosing patients with potential splenic injury by reconstructing the 3D model of the patients' spleen from CT scan images (Cheng et al., 2023). Farzaneh *et al.* also demonstrated the use of DL in automating the detection and assessment of liver trauma by combining different DL methods. Considering these recent breakthroughs, the integration of cutting-edge technology and medical expertise has the potential to transform the diagnosis and treatment of abdominal trauma significantly.

To solve the problems stated above, this project aims to achieve the following objectives:

1. To develop an automated system that is capable of localizing abdominal organs and classify the severity of the trauma
2. To evaluate existing approaches for classifying severity of abdominal trauma

Email addresses: u2102798@siswa.um.edu.my (Wei Xin Lim), wlhoo@um.edu.my (Wai Lam Hoo)

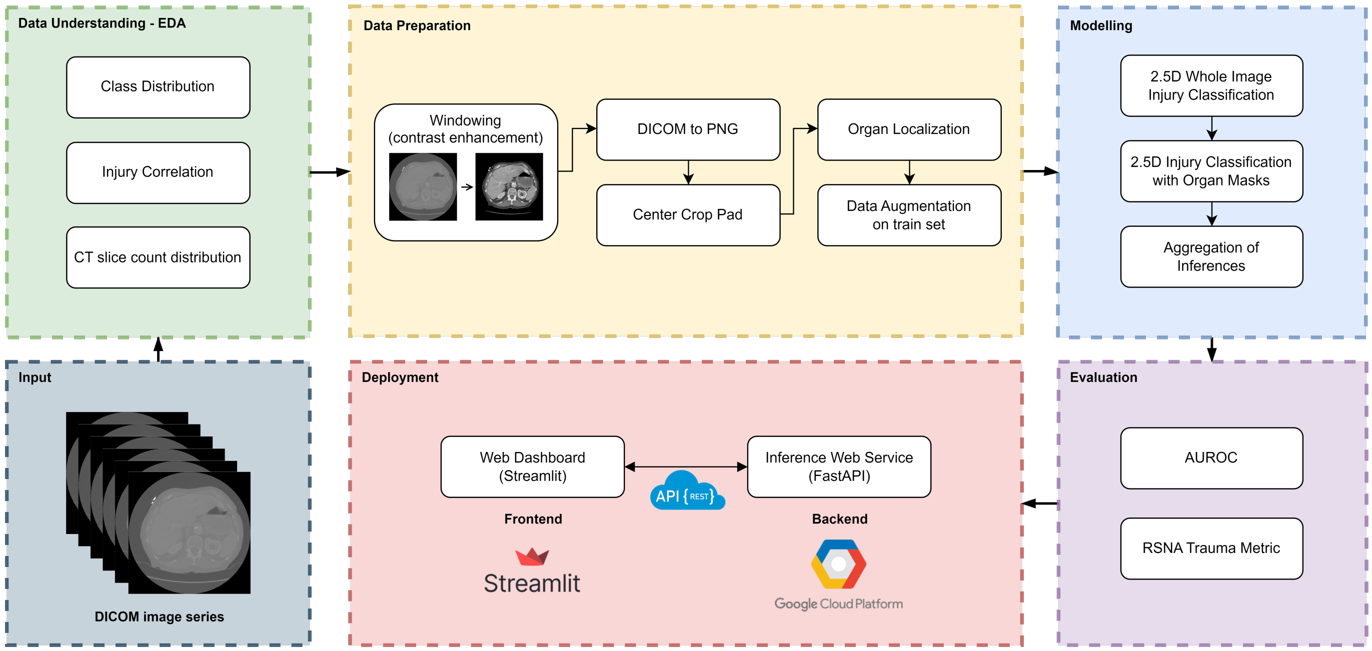


Figure 1: Project Framework

3. To implement a dashboard that optimizes the workload of healthcare provider and offer personalized decision and diagnostic support

2. Literature Review

2.1. Deep Learning for Medical Imaging Diagnosis

In recent years, numerous research studies have highlighted the feasibility and great potential of deep learning-assisted diagnostic tools for clinical physicians. These studies have shown the power of amalgamating both deep learning methods and medical images, like CT scans, in solving complex diagnostic problems in clinical settings and potentially surpassing human experts in the field. For instance, a pilot study demonstrated the effectiveness of deep learning models based on the LeNet architecture in predicting surgical complexity and postoperative outcomes from preoperative CT scans for abdominal wall reconstruction. These models not only showed promising results but also exceeded the predictive accuracy of expert surgeons in determining surgical complexity, highlighting the significant potential of deep learning in medical imaging analysis and surgical planning (Elhage et al., 2021). Different deep learning approaches were proposed by researchers to predict COVID-19 prevalence using lung CT scans (Amyar et al., 2020; Ghaderzadeh et al., 2021). For instance, Ghaderzadeh et al. (2021) introduced a model based on a NASNet CNN feature extractor to detect COVID-19 in patients at an early stage of the disease from multiple lung CT scans. Amyar et al. (2020) implemented a CNN+Unet multi-task deep learning model to detect COVID-19 and segment the lesions. Multi-task learning models are capable of combining several pieces of information from different tasks to improve the performance of the model and its ability to generalize better, leveraging useful information contained in multiple related tasks (Zhang and Yang, 2022).

2.2. Abdominal Trauma Detection

In the context of abdominal trauma-related injuries, model architectures using approaches that encompass both 2D and 3D methods have been introduced. Dreizin et al. (2020) proposed a multiscale attentional network (MSAN) for the quantitative assessment of traumatic hemoperitoneum, which demonstrated superior performance compared to traditional fully convolutional networks like the 3D U-net (Pawlowski et al., 2017) and Recurrent Saliency Transformation Network (Xie et al., 2020) in segmenting the hemoperitoneum region. Meanwhile, Cheng et al. (2023) developed an experimental two-step 3D weakly supervised deep learning approach specialized for detecting splenic injury and achieved state-of-the-art results. However, the dataset used for training the splenic injury detection model is limited, raising concerns about its robustness in real-world clinical scenarios, especially when dealing with subdividing the grade of injury. Farzaneh et al. (2018) implemented a hybrid approach of both machine learning and deep learning to detect traumatic injuries to the kidney. In her subsequent work, a two-step approach based on U-net and a deep convolutional network was developed to detect traumatic injuries to the liver. The model is generalizable to patients with pre-existing liver conditions, including fatty livers and congestive hepatopathy, showing robustness to various clinical scenarios (Farzaneh et al., 2022). Despite numerous research efforts in the detection of abdominal trauma injuries, there has yet to be a study related to performing inference on traumatic injury on all major organs in the abdominal cavity, thus presenting an opportunity for research to develop a holistic approach that is capable to simultaneously evaluate multiple abdominal organs and predict the injury level.

3. Methodology

The life cycle of this project is heavily influenced by the CRISP-DM methodology (Wirth and Hipp, 2000). This methodology widely recognized and accepted by different industry to provide a structured framework for data science projects as well as assuring higher prevalence of successful project outcomes (Europe, 2017). The project framework is summarized in Figure 1.

3.1. Business Understanding

Business understanding is the very first phase of CRISP-DM. In this phase, a clear understanding of the business problem is developed by defining the problem statements and objectives. These 2 components have already been highlighted in previous sections. The purpose of this phase is to assure that the project is aligned with the business objectives and at the same time have a clear image of success criteria (O'Hara et al., 2023; Stefanovskyi, 2023).

3.2. Data Understanding

3.2.1. Data Acquisition

This project is inspired by RSNA Abdominal Trauma Detection Challenge 2023, an open global data science competition took place on Kaggle in 2023. Thus, the competition dataset provided by Radiological Society of North America (RSNA) is utilized to develop the deep learning algorithm. The dataset is a collaborative work of 22 research institutes from 14 different countries. It is 460GB in size and comprises of 4711 CT studies. Each study is made up of 100 to 1000 consecutive slice of CT images, encompassing Axial, Coronal and Sagittal views. The difference in the views can be observed in Figure 2.

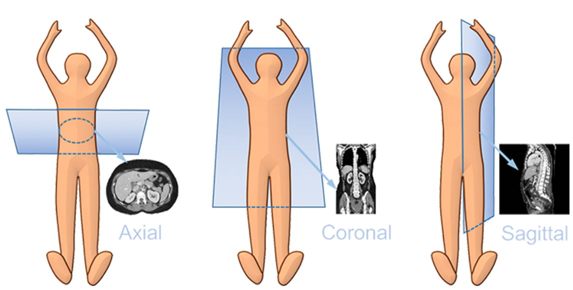


Figure 2: Difference between Axial, Coronal and Sagittal Views

Each CT study is also labelled with relevant clinical information such as patient ID, CT scan series ID and the volume of the aorta in hounsfield units. 3-targeted labels (healthy, low, high) are assigned to kidney, liver and spleen while binary target labels (healthy, injury) are assigned to bowels and extravasation injury., contributing to a total of 11 target labels. One or more organs can be injured in a study, hence making this dataset a multi-label classification dataset. The target labels are stored in another CSV file and are identified by the patient ID and series ID. Other medical issues of the patient such as broken bones and cancer are not covered in this dataset. The metadata of the dataset is summarized in Table 1.

RSNA Challenge Dataset	
Type	CT
# focus	abdomen
# size	460.67 GB
# study	4711
# slices	100 to 1000
# binary targets	2 (bowel, extravasation)
# 3-targeted	3 (kidney, liver, spleen)
# total targets	11

Table 1: The metadata of the RSNA Challenge dataset

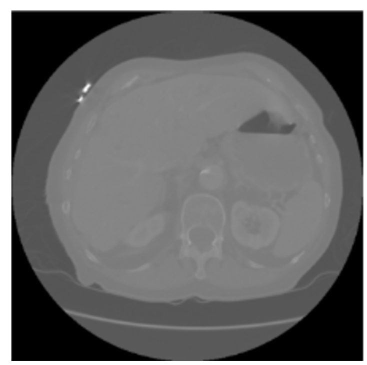


Figure 3: Sample CT image from the dataset

3.2.2. Exploratory Data Analysis (EDA)

To better understand the underlying patterns and relationships exist within the dataset, exploratory data analysis was conducted on the available image dataset. This analysis included an investigation into the distribution of target classes. From Table 2, it is discovered that the dataset is highly imbalanced. For instance, in case of extravasation, positive (injured) class forms only 2.21% of the dataset. Such imbalance is common in medical imaging context as the positive cases are difficult to obtain due to the fact that the data collection is often compounded by patients consent and privacy issues.

Organs	Injury Class	Percentage (%)
bowel	healthy	93.63
	injured	6.37
extravasation	healthy	97.79
	injured	2.21
kidney	healthy	93.95
	low	3.74
	high	2.31
liver	healthy	89.92
	low	8.15
	high	1.93
spleen	healthy	88.43
	low	6.71
	high	4.86

Table 2: Class distributions of target labels

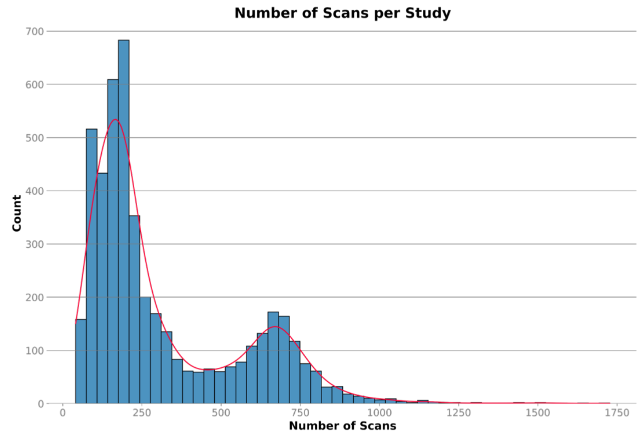


Figure 5: Distribution of number of scans in each study.

From Figure 4, correlations between various injuries revealed predominantly weak associations. Notably, the strongest correlations identified were between kidney and liver injuries, as well as between spleen injuries and extravasation, each with a correlation coefficient of 0.18. This low degree of correlation underscores the independence of these injuries from one another.

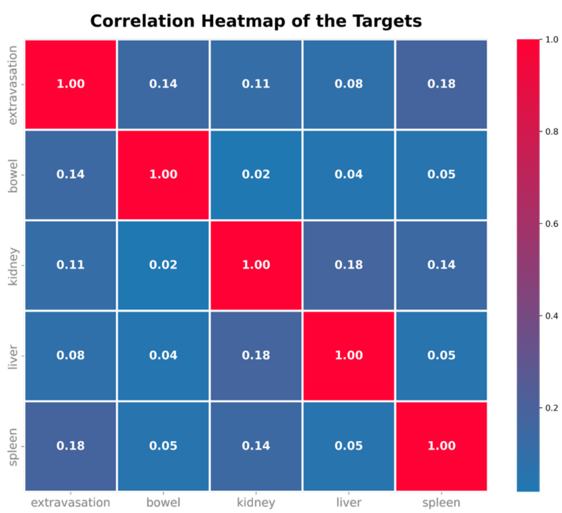


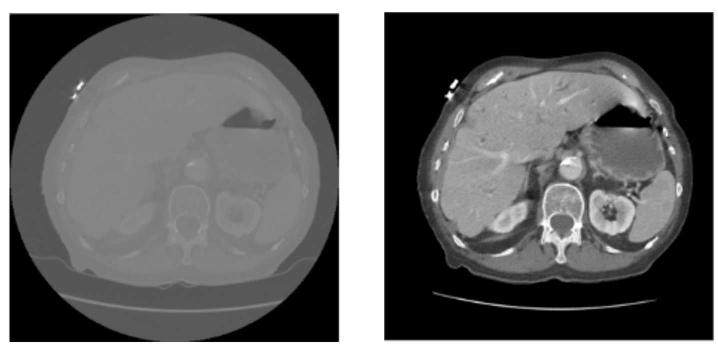
Figure 4: Correlation of the organ injuries.

Figure 5 shows a significant concentration of studies have around 168 and 668 number of scans. This notable pattern persists even when the data is grouped into large bin intervals of 50, suggesting the voxel length shall be set to a value around 668.

3.3. Data Preparation

Data preparation is one of the most imperative and time-consuming steps. It involves data selection, modification and splitting of dataset into train, validation and test set (IBM, 2023). The quality of this phase will have significant influences on the modelling outcome (Kerner et al., 2022).

Windowing technique will first be performed on the image data. Windowing is also known as contrast enhancement, it is a method commonly applied in analyzing CT scans, which is utilized to improve the contrast for the specific type of tissue or abnormality under examination (Xue et al., 2012). In particular, soft tissue windowing is applied to reveal subtle indications on the abdominal organs and tissue. Figure 6 shows the effect of windowing.



(a) (b)

Figure 6: Windowing effect. (a) Before windowing (b) After windowing

After windowing, the DICOM images are stored in PNG format in width and height of 512 pixels for ease of access. Center Crop Pad operation is then applied to the image data to minimize noises from surrounding cavity and better focus on the region of interests, resulting in images with width and height of 384 pixels. To further concentrate on large abdominal organs as well as reducing computational load (Cheng et al., 2023), kidney, liver and spleen are localized with a pre-trained segmentation model based on ResNet-18 (He et al., 2015). Masks for the organs were created and stored for training purposes. Strong image augmentation technique including Horizontal Flip, Shift-Scale-Rotate, color and blur augmentation, elastic transform and cut-mix are applied on train set as an effort to improve model's generalizability (Cheng et al., 2023; Farzaneh et al., 2018, 2022).

3.4. Modelling

3.4.1. Model Architecture

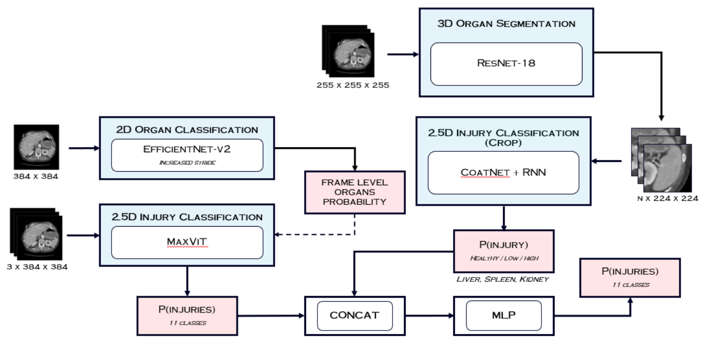


Figure 7: Multi-stage model proposed by Theo Viel

The multi-stage model architecture proposed by Theo Viel was utilized (Theo, 2024). The inference pipeline is visualized in Figure 7. Theo Viel's approach can be broken up into 3 stages as shown in Figure 8, 9 and 10:

Stage 1 - 2.5D Whole Image Injury Classification

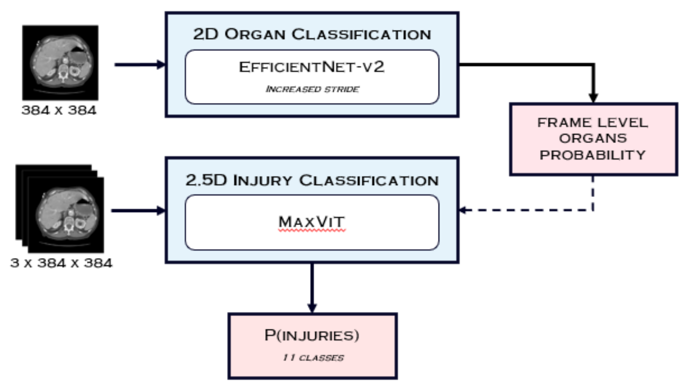


Figure 8: 2.5D Whole Image Injury Classification

The first stage starts by utilizing the EfficientNet-v2 model (Tan and Le, 2021) for frame-level organ detection in a series of CT images. This process of identifying organs at the frame level serves as a foundational step. Subsequently, these frame-level organ predictions are amalgamated with the corresponding CT images grouped in 3 to serve as inputs for the MaxViT model (Tu et al., 2022), which is tasked with the 2.5D classification of injuries. The primary objective of integrating frame-level organ predictions alongside the CT image series in the MaxViT model input is to furnish a more detailed organ-specific context. This integration is anticipated to significantly enhance the overall performance of the model. The output of this stage are the probabilities for the 11 target labels.

Stage 2 - 2.5D Injury Classification with Organ Mask

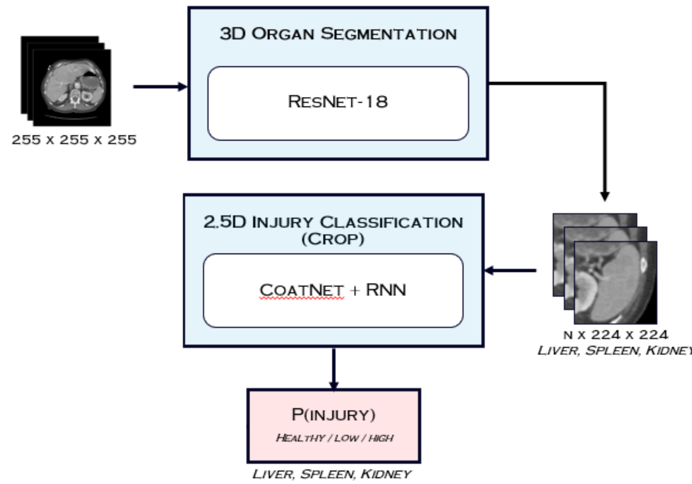


Figure 9: 2.5D Injury Classification with Organ Mask

The second stage then perform 3D organ segmentation of liver, spleen and kidney using aforementioned pretrained ResNet-18 segmentation model. After the 3 organs are segmented, they are fed into CoatNet (Dai et al., 2021) + RNN model for injury classification. This stage is to further improve the model performance on predicting the injury class of liver, spleen and kidney.

Stage 3 - Aggregation of Inferences

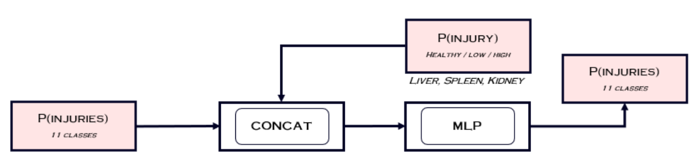


Figure 10: Aggregation of inferences

The last stage act as an aggregation layer, using RNN and MLP together to aggregate the inference results from the previous stages, and optimize the performance metrics directly.

3.4.2. Loss Function

This project experimented with 2 different loss functions; standard cross entropy loss following Theo Viel and Focal Loss. As mentioned in the previous sections, the dataset suffered from extreme class imbalance. Focal Loss is a technique designed to address the class imbalance issue by adding a modulating factor $(1 - p_t)^\gamma$ to the standard cross-entropy loss (Lin et al., 2020). This factor dynamically scales the loss: it decreases as the confidence in the correct class increases, which effectively reduces the loss contribution from easy examples. Consequently, it helps to focus the model training more on the hard, misclassified examples. The parameter γ in the Focal Loss function allows for fine-tuning of how much focus is placed on hard examples compared to easy ones. This approach has been shown to improve the performance of dense object detectors and classification models significantly, particularly in scenarios with many easy negative (background) examples (Petmezas et al., 2022; Wu et al., 2019).

3.4.3. Regularization

As an effort to avoid overconfidence and improve the generalizability of the model, regularization technique - Label Smoothing was experimented in this project. Label Smoothing adjusts the target labels to be less confident, thereby preventing the model from becoming too certain about its predictions (Szegedy et al., 2015). The idea is to replace the 0s and 1s in the label vector with values slightly more than 0 and slightly less than 1, respectively. The formula for Label Smoothing is given by:

$$y_{\text{smoothed}} = (1 - \epsilon) \cdot y + \epsilon / K \tag{1}$$

Here, y is the original label, ϵ is the smoothing parameter (a small constant), and K is the number of classes. The choice for the value of ϵ is crucial, as a higher value indicates more pronounced label softening. This approach challenges the model with a more complex learning task, fostering better generalization and lessening the likelihood of overfitting.

3.4.4. Experiment Setup

In this study, a series of experiments is conducted to evaluate the performance of the MaxVit model in 2.5D whole image injury classification (first stage), focusing on different combinations of the Focal Loss parameter, γ, and the Label Smoothing parameter, ϵ, where $\gamma \in \{0, 1, 2, 3\}$ for Focal Loss and $\epsilon \in \{0, 0.05, 0.1\}$ for Label Smoothing. All experiments were performed on an NVIDIA A6000 GPU, extending over 5 epochs, using pre-trained weights provided by Theo Viel. Original implementation was adhered for all other parameters and configurations to ensure consistency in the experimentation framework.

3.5. Evaluation

The model's predicted labels will be evaluated using the AUROC score and the RSNA Trauma Metric. The AUROC score is particularly robust in scenarios with class imbalance (Huang and Ling, 2005), as it is derived from the True Positive Rate (TPR, or Sensitivity) and the False Positive Rate (FPR, or 1 - Specificity). These rates are independently calculated for each class, ensuring that the score is not biased by the distribution of classes, thereby avoiding skewing by the majority class. RSNA Trauma Metric (abbreviated as RSNA T.M.), which is the metric used in the competition. It measures the overall capability of the model to classify the organ injuries by calculating the mean log loss from each target label (Dane, 2023).

3.6. Deployment

The best performing model will then be selected and deployed as a web service. A web dashboard will also be deployed to enable end-to-end user interaction with the model via graphical interface. Further information will be discussed in Section 4.2.

4. Results & Discussion

4.1. Experiment Result

Gamma, γ	Epsilon, ϵ	AUROC	RSNA T.M.
-	-	0.870	0.232
0	0	0.861	***0.213**
	0.05	0.818	0.289
	0.1	0.833	0.301
1	0	***0.880**	***0.211**
	0.05	0.807	0.265
	0.1	0.782	0.285
2	0	0.859	0.234
	0.05	0.810	0.301
	0.1	0.776	0.321
3	0	0.863	0.258
	0.05	0.801	0.353
	0.1	0.795	0.359

Table 3: Experiment result. ***bolded** score indicates outperforming baseline model

From the Table 3, the optimal results were achieved when the Focal Loss parameter, Gamma (γ), was set to 1, and without the employment of Label Smoothing (Epsilon, $\epsilon = 0$). This configuration surpassed the baseline model by improving the AUROC and RSNA Trauma Metric score by 1% and 2% respectively. These enhancements can be attributed to the influence of Focal Loss, particularly its role in better guiding the model's learning focus towards hard-to-classify examples, which, in this case, the positive instances, thereby improving the discriminative capability of the model (increased AUROC) and its predictive accuracy (reduced RSNA Trauma Metric).

Notably, in the experiment with 5 additional epochs, without both Focal Loss and Label Smoothing, there was an improvement in the RSNA Trauma Metric to 0.213 but a decrease in AUROC to 0.861. This outcome implies that the additional training may have improved performance of the majority class but insufficient to improve the minority class, leading to a lower AUROC. It also suggests potential overfitting, causing the model to start losing its generalization ability.

Experiments involving Label Smoothing showed significant degradation in AUROC and RSNA Trauma Metric by 7 to 10% and 3 to 13% respectively. The degradation is more obvious when in conjunction with Focal Loss. This behavior can be explained by the excessive dampening the model's learning capacity by the combinations, leading to underfitting.

4.2. Data Product

A web dashboard, under the name of Project **R**apid **A**bdominal **T**rauma **S**creening (Project **RATS**) is developed and deployed. It is hosted on https://project-rats.streamlit.app.

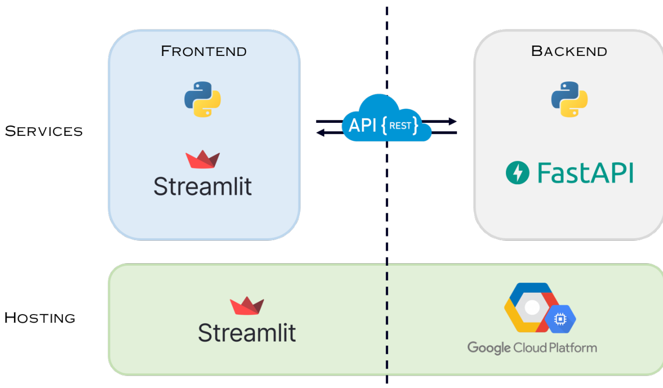


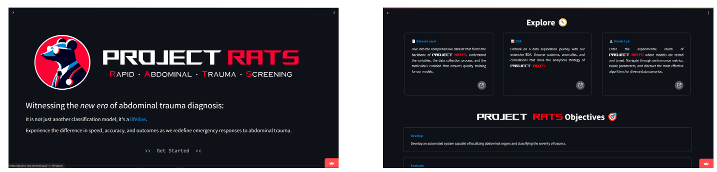
Figure 11: Project RATS deployment architecture diagram

The data product is broken down into 2 modules: backend and frontend. The backend is developed with FastAPI (Ramírez, 2023), a modern and high-performance web framework for building APIs with Python. The backend is used to host the model for inference tasks. It is hosted on L4 GPU instance on Google Cloud Platform ("GPUs on Compute Engine — Compute Engine Documentation", 2024). The frontend, which is the web dashboard, is developed with Streamlit ("Streamlit Documentation", 2024), an open-source Python library for creating and sharing beautiful, custom web apps for data science and machine learning with ease. The web dashboard aims to provide easy-to-use graphical interface for the user to upload CT scan series and perform injury inference with web browser. The Streamlit web dashboard is deployed on Streamlit Community Cloud. The frontend and backend are connected via REST API (RedHat, 2020), which facilitates seamless interaction between the user interface and the model's inference engine, ensuring efficient communication and data exchange for real-time analysis and results presentation on the dashboard.

4.2.1. Web Dashboard

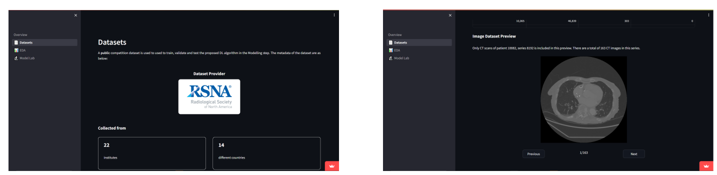
The developed web dashboard has 4 pages:

- **Landing Page**: The home page of the website. This page displays the basic introduction to Project RATS, including the objectives.
- **Dataset Page**: This page shows the information of the dataset used to train the model. Previews of the dataset is provided.
- **EDA Page**: This page covers the exploratory data analysis conducted on the dataset. Charts and explanation of findings are provided.
- **Model Lab Page**: This page hosts the interface to the inference service. Users are allowed to select preferred model and upload the CT scans series for injury inference.



(a) (b)

Figure 12: Project RATS Landing Page



(a) (b)

Figure 13: Project RATS Dataset Page

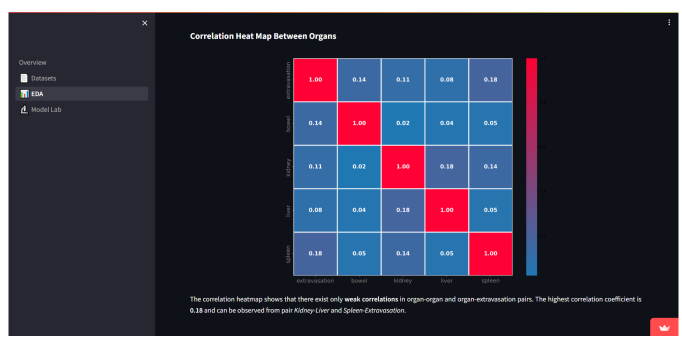


Figure 14: Project RATS EDA Page

5. Conclusion

In summary, an automatic organs localization and injury classification system is developed and validated to be feasible, using a

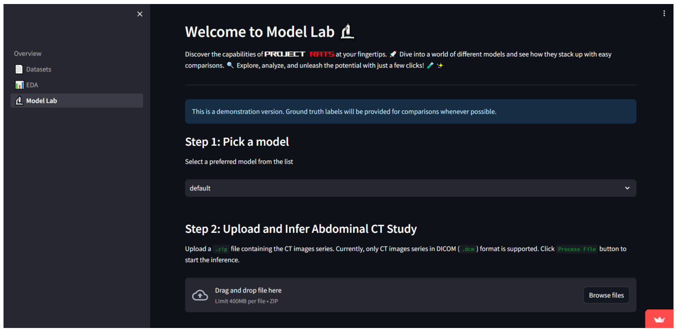


Figure 15: Project RATS Model Lab Page

novel multi-stage model architecture proposed by Theo Viel. The existing approaches for classifying abdominal trauma severity is evaluated. Various experiments targeting MaxViT model from the 2.5D whole image injury classification stage have been carried out and concluded that model using only Focal Loss with Gamma (γ) set to 1 has improved its performance, achieving 0.88 and 0.211 in AUROC and RSNA Trauma Metric respectively, outperforming the baseline model that used that classic standard cross-entropy loss. A Streamlit and FastAPI powered dashboard has been developed to optimize healthcare providers' workload, offering personalized diagnostic support with the integrated, fine-tuned model at its backend.

5.1. Social Impact

The impacts of this project to the society are significant. Notably, a faster and more accurate prediction offered by the automatic organ localization and injury classification system can lead to a quicker treatment decisions, thus improving the patients' outcomes. Besides, an optimized work flow in clinical settings can be achieved via this project as the workload of the radiologists are reduced with this system, allowing these valuable professional re-sources to be allocated to a more complex and critical cases. To-gether with the system and dashboard, this project democratizes the access to critical diagnostic support, especially in regions with a shortage of skilled radiologists.

5.2. Limitations

Inevitably, this project faced numerous limitations. Firstly, limited access to advanced computational resources posed challenge in developing the complex deep learning models, especially for processing high volume of medical imaging data and training the model. Not only extra time were required, but it also restricted the comprehensive development and in-depth testing phases, which are crucial for ensuring robustness and reliability in medical applications. Secondly, the model's low explainability remains a significant concern. As it only output the probabilities of the 11 injury classes without additional information, it is difficult to interpret and understand the model's decision-making process. This can hinder its clinical adoption and trust among healthcare professionals, which may potentially lead to failure of project. Third and lastly, a limited set of evaluation metrics was used, namely the AUROC and RSNA Trauma Metric. Solely relying on 2 metrics only does not fully capture the model's overall performance and effectiveness in various clinical scenarios.

5.3. Future Direction

In upcoming times, the model architecture in this project can be simplified to reduce the complexity of the architecture and thus reducing the computation resources needed to develop. However, the trade-off between model complexity and performance must be carefully balanced, as oversimplification might compromise model's generalizability. Besides, visualization of the self-attention layer and the CNN filters can be done as an effort to promote model's transparency and credibility. More evaluation metrics and statistical testing shall be included as part of the work flow to increase the model's robustness in different settings.

6. References

Amyar, A., Modzelewski, R., Li, H., & Ruan, S. (2020). Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: Classification and segmentation. *Computers in Biology and Medicine*, *126*, 104037. https://doi.org/10.1016/j.compbiomed.2020.104037

Barrett, C., & Smith, D. (2012). Recognition and management of abdominal injuries at athletic events. *International Journal of Sports Physical Therapy*, *7*, 448–451. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3414076/

Cheng, C.-T., Lin, H.-S., Hsu, C.-P., Chen, H.-W., Huang, J.-F., Fu, C.-Y., Hsieh, C.-H., Yeh, C.-N., Chung, I.-F., & Liao, C.-H. (2023). The three-dimensional weakly supervised deep learning algorithm for traumatic splenic injury detection and sequential localization: An experimental study. *International Journal of Surgery*, *109*, 1115–1124. https://doi.org/10.1097/js9.0000000000000380

Dai, Z., Liu, H., Le, Q. V., & Tan, M. (2021). Coatnet: Marrying convolution and attention for all data sizes. *CoRR*, *abs/2106.04803*. https://arxiv.org/abs/2106.04803

Dane, S. (2023, July). Rsna abdominal trauma detection metric. *kaggle.com*. Retrieved November 6, 2023, from https://www.kaggle.com/code/metric/rsna-trauma-metric/notebook

Dreizin, D., Zhou, Y., Fu, S., Wang, Y., Li, G., Champ, K., Siegel, E. L., Wang, Z., Chen, T., & Yuille, A. (2020). A multi-scale deep learning method for quantitative visualization of traumatic hemoperitoneum at ct: Assessment of feasibility and comparison with subjective categorical estimation. *Radiology*, *2*, e190220–e190220. https://doi.org/10.1148/ryai.2020190220

Elhage, S. A., Deerenberg, E. B., Heniford, B. T., Murphy, K. J., Shao, J. M., Kercher, K. W., Smart, N. A., Fischer, J. P., Augenstein, V. A., & Colavita, P. D. (2021). Development and validation of image-based deep learning models to predict surgical complexity and complications in abdominal wall reconstruction. *JAMA Surgery*, *156*, 933–933. https://doi.org/10.1001/jamasurg.2021.3012

Errol Colak, H.-M. L., Robyn Ball, S. J., Kirti Magudia, B. M., Savvas Nicolaou, L. P., Jeff Rudie, G. S., & Maryam Vazirabad, J. M. (2023, July). Rsna 2023 abdominal trauma detection. Retrieved October 22, 2023, from https://kaggle.com/competitions/rsna-2023-abdominal-trauma-detection

Europe, S. V. (2017). Building and applying predictive models in ibm spss modeler training webinar. *Smart Vision - Europe*. https://www.sv-europe.com/crisp-dm-methodology/

Farzaneh, N., Reza, M., Patel, H., Wood, A., Gryak, J., Fessell, D., & Najarian, K. (2018). Automated kidney segmentation for traumatic injured patients through ensemble learning and active contour modeling. *Europe PMC (PubMed Central)*. https://doi.org/10.1109/embc.2018.8512967

Farzaneh, N., Stein, E. B., Soroushmehr, R., Gryak, J., & Najarian, K. (2022). A deep learning framework for automated detection and quantitative assessment of liver trauma. *BMC Medical Imaging*, *22*. https://doi.org/10.1186/s12880-022-00759-9

Ghaderzadeh, M., Asadi, F., Jafari, R., Bashash, D., Abolghasemi, H., & Aria, M. (2021). Deep convolutional neural network–based computer-aided detection system for covid-19 using multiple lung scans: Design and implementation study. *Journal of Medical Internet Research*, *23*, e27468. https://doi.org/10.2196/27468

Gpus on compute engine — compute engine documentation. (2024, January). *Google Cloud*. Retrieved January 19, 2024, from https://cloud.google.com/compute/docs/gpus

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition.

Huang, J., & Ling, C. (2005). Using auc and accuracy in evaluating learning algorithms. *IEEE Trans. Knowl. Data Eng.*, *17*.

IBM. (2023). Data preparation overview. *www.ibm.com*. https://www.ibm.com/docs/en/spss-modeler/saas?topic=preparation-data-overview

Kerner, H., Campbell, J., & Strickland, M. (2022). Chapter 1 - introduction to machine learning. In J. Helbert, M. D'Amore, M. Aye, & H. Kerner (Eds.), *Machine learning for planetary science* (pp. 1–24). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-818721-0.00007-0

Latif, R. K., Clifford, S. P., Baker, J. A., Lenhardt, R., Haq, M. Z., Huang, J., Farah, I., & Businger, J. R. (2023). Traumatic hemorrhage and chain of survival. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *31*. https://doi.org/10.1186/s13049-023-01088-8

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2020). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *42*(2), 318–327. https://doi.org/10.1109/TPAMI.2018.2858826

Murphy, H. (2022, November). 'tremendous pressure' to keep up with radiologist workloads results in another missed diagnosis. *healthimaging.com*. Retrieved October 22, 2023, from https://healthimaging.com/topics/healthcare-management/healthcare-staffing/radiologist-workloads-result-missed-diagnosis

O'Hara, R., S. Haylon, L., & M. Boyle, D. (2023). A data analytics mindset with crisp-dm — ima. *IMA*. Retrieved October 22, 2023, from https://www.sfmagazine.com/articles/2023/february/a-data-analytics-mindset-with-crisp-dm#:~:text=PHASE%201%3A%20BUSINESS%20UNDERSTANDING

Pawlowski, N., Ktena, S. I., Lee, M. C. H., Kainz, B., Rueckert, D., Glocker, B., & Rajchl, M. (2017). DLTK: state of the art reference implementations for deep learning on medical images. *CoRR*, *abs/1711.06853*. http://arxiv.org/abs/1711.06853

Petmezas, G., Cheimariotis, G.-A., Stefanopoulos, L., Rocha, B., Paiva, R. P., Katsaggelos, A. K., & Maglaveras, N. (2022). Automated lung sound classification using a hybrid cnn-lstm network and focal loss function. *Sensors*, *22*, 1232. https://doi.org/10.3390/s22031232

Ramírez, S. (2023). Fastapi. *fastapi.tiangolo.com*. Retrieved January 19, 2024, from https://fastapi.tiangolo.com/

RedHat. (2020, May). What is a rest api? *www.redhat.com*. Retrieved January 19, 2024, from https://www.redhat.com/en/topics/api/what-is-a-rest-api

Stefanovskyi, O. (2023, March). The most important phase of crisp-dm you need to get right: Business understanding. *Medium*. Retrieved October 22, 2023, from https://medium.com/@stefanovskyi/business-understanding-crisp-dm-1111bfbc7b8d#:~:text=The%20Business%20Understanding%20phase%20is%20critical%20to%20the%20success%20of

Streamlit documentation. (2024). *docs.streamlit.io*. Retrieved January 19, 2024, from https://docs.streamlit.io/

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the inception architecture for computer vision.

Tan, M., & Le, Q. V. (2021). Efficientnetv2: Smaller models and faster training. *CoRR*, *abs/2104.00298*. https://arxiv.org/abs/2104.00298

Theo, V. (2024, January). Theo viel — grandmaster. *www.kaggle.com*. Retrieved January 19, 2024, from https://www.kaggle.com/theoviel

Tu, Z., Talebi, H., Zhang, H., Yang, F., Milanfar, P., Bovik, A., & Li, Y. (2022). Maxvit: Multi-axis vision transformer.

Wirth, R., & Hipp, J. (2000). Crisp-dm: Towards a standard process model for data mining. *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, *1*, 29–39.

Wu, X., Sahoo, D., & Hoi, S. C. H. (2019). Recent advances in deep learning for object detection.

Xie, L., Yu, Q., Zhou, Y., Wang, Y., Fishman, E. K., & Yuille, A. L. (2020). Recurrent saliency transformation network for tiny target segmentation in abdominal ct scans. *IEEE Transactions on Medical Imaging*, *39*, 514–525. https://doi.org/10.1109/tmi.2019.2930679

Xue, Z., Antani, S., Long, L. R., Demner-Fushman, D., & Thoma, G. R. (2012). Window classification of brain ct images in biomedical articles. *AMIA Annual Symposium Proceedings*, *2012*, 1023–1029. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540547/

Zhang, Y., & Yang, Q. (2022). A survey on multi-task learning. *IEEE Transactions on Knowledge and Data Engineering*, *34*(12), 5586–5609. https://doi.org/10.1109/TKDE.2021.3070203

7. Appendix

7.1. User Manual

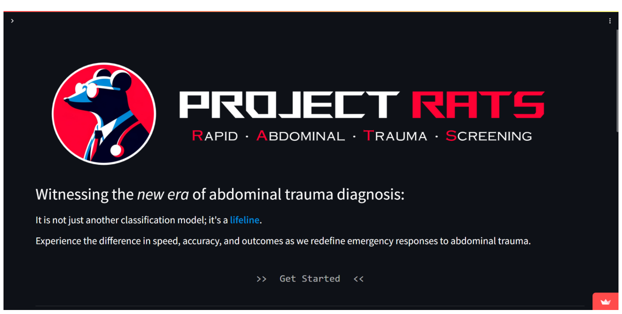
This user manual provides detailed steps and explanations for correctly setting up and using the Project RATS dashboard. Project RATS is an AI powered abdominal trauma classification tool, aiming to provide rapid screening of abdominal CT scans for trauma assessment. Follow these steps to ensure an effective utilization of the dashboard.

7.1.1. Content Outline

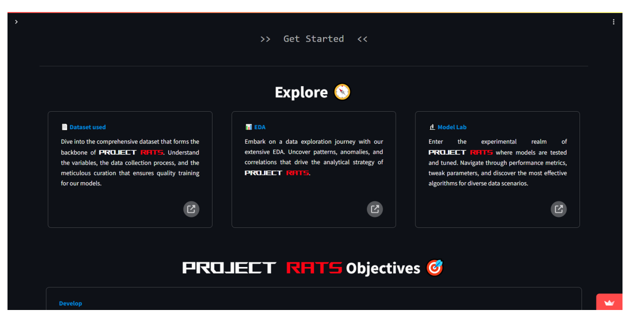
- Using the Project RATS Dashboard
  - Navigation and Features
    - Datasets Page
    - EDA Page
    - Model Lab Page
  - Abdominal Trauma Classification System

7.1.2. Using the Project RATS Dashboard

Navigation and Features.
Upon entering https://project-rats.streamlit.app, you can scroll down to **Explore** section to navigate to available pages in Project RATS.



(a) Landing Page



(b) Explore section

Alternatively, you can open the side menu bar and navigate to desired page. The side menu button is situated at the top left corner of the web page.

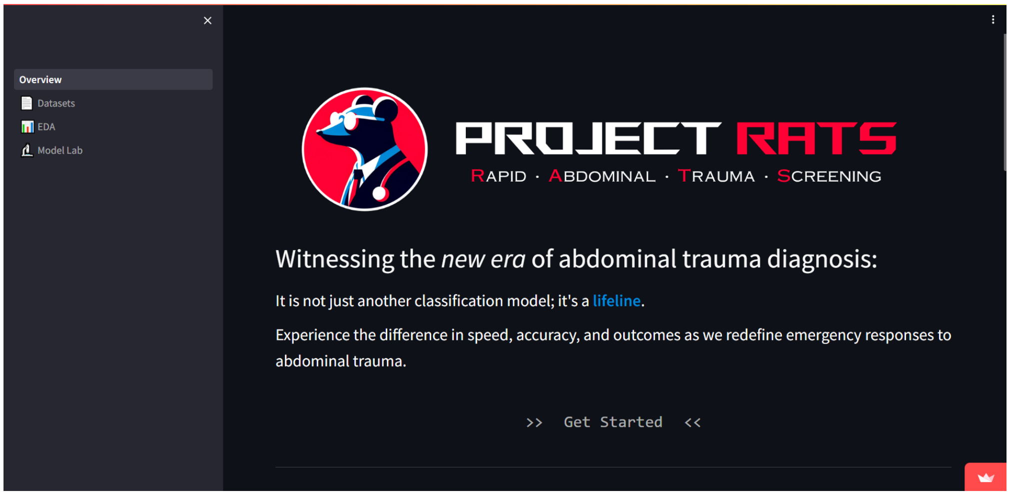


Figure A1: Side menu bar opened

Datasets Page.
This page shows the information of the dataset used to train the model. Here, users are able to understand the nature and composition of the utilized dataset that powered Project RATS AI inference system. Interactive previews of the dataset is provided.

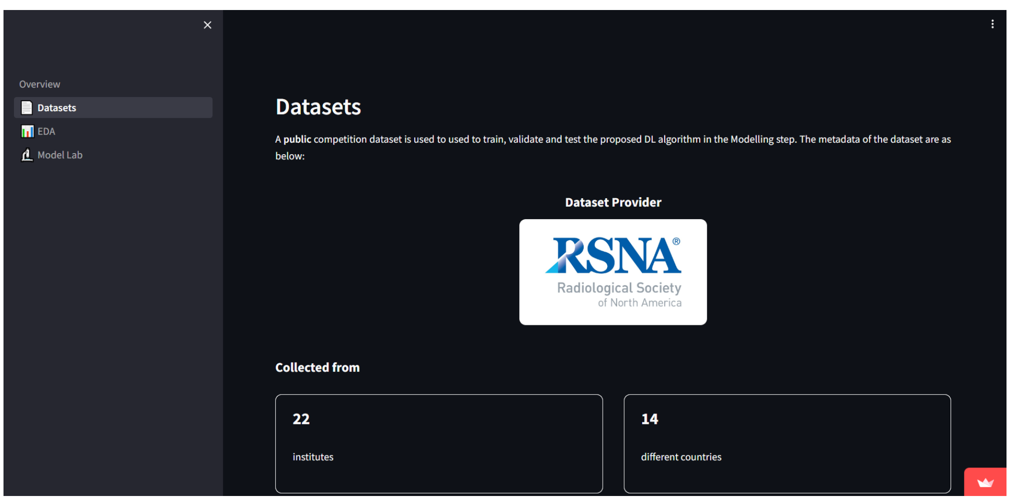


Figure A2: Datasets Page

EDA Page.
This page covers the exploratory data analysis conducted on the dataset. Here, users can better understand the underlying patterns and relationships exist within the dataset. Charts and explanation of findings are provided

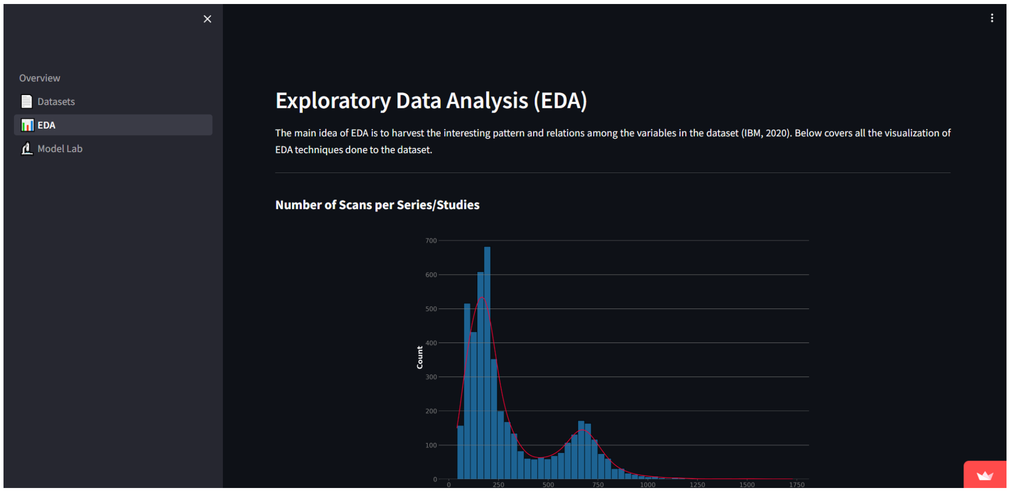


Figure A3: EDA Page

Model Lab Page.
This page hosts the interface to the inference service. Users are allowed to select preferred model and upload the CT scans series for injury inference. The guideline to use the integrated abdominal trauma classification system will be covered in the next section.

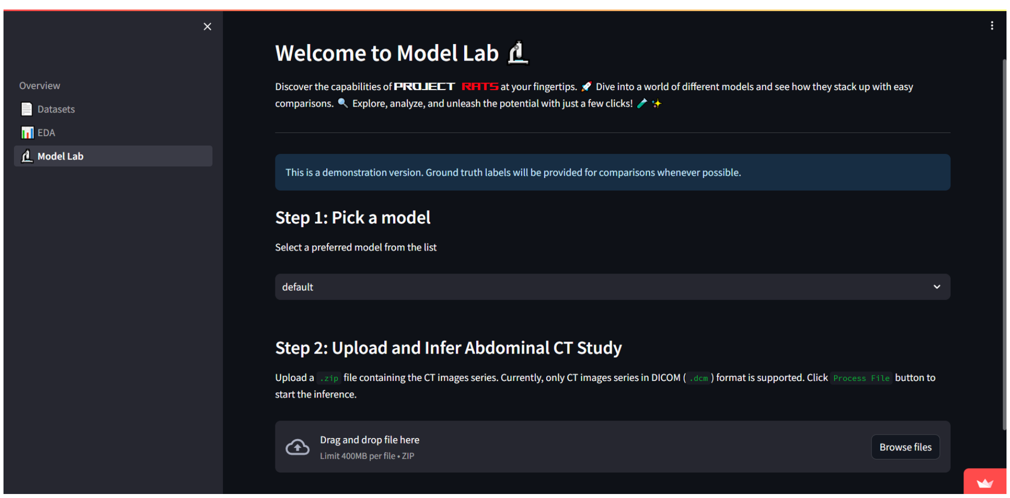


Figure A4: Model Lab Page

Abdominal Trauma Classification System.
This section covers the steps to use Project RATS. Only a few simple steps are required:
Step 1: Pick a model
Select the model of choice using the drop down selector.

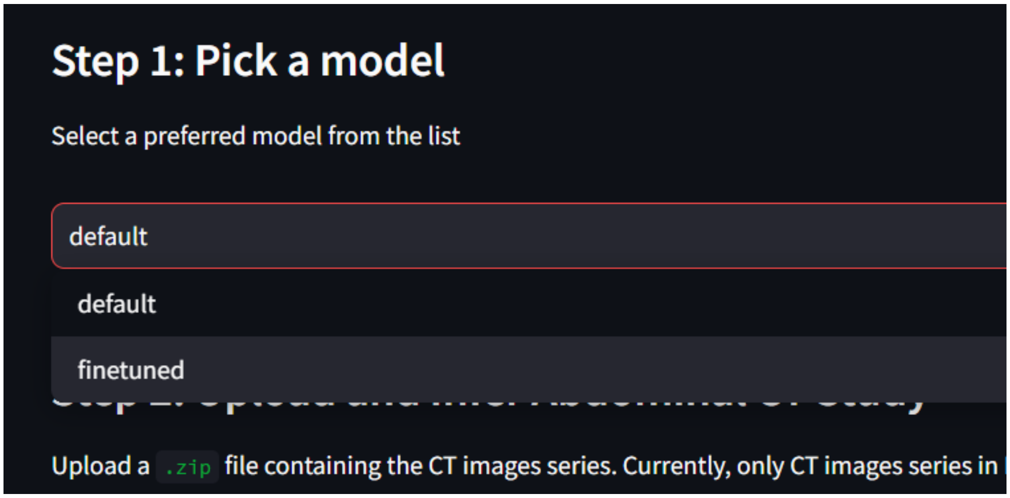


Figure A5: Select model of choice

Step 2: Upload the CT scans series zip
Upload the file zip containing all the CT series to the dashboard. Currently, the backend only recognizes the CT images **must** in DICOM format (.dcm). The name of the zip file is advised to be in the format of [patient ID].[series ID].zip. .

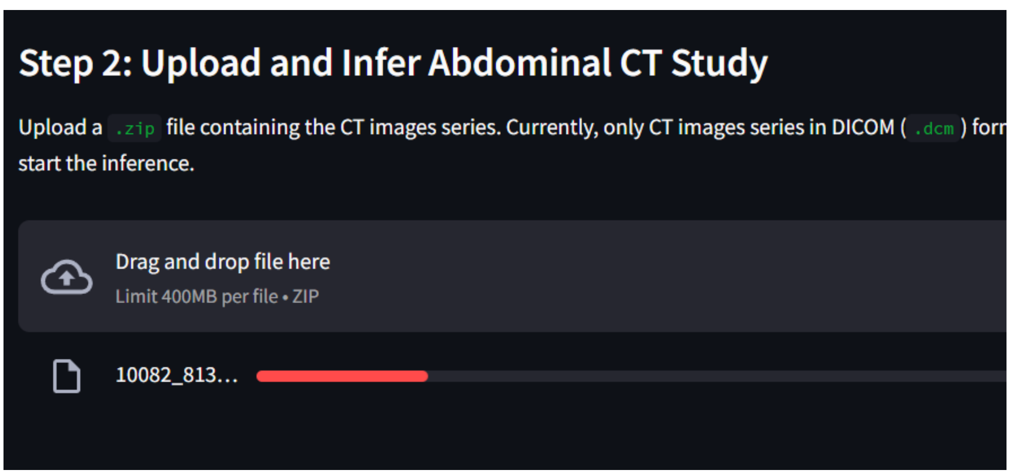
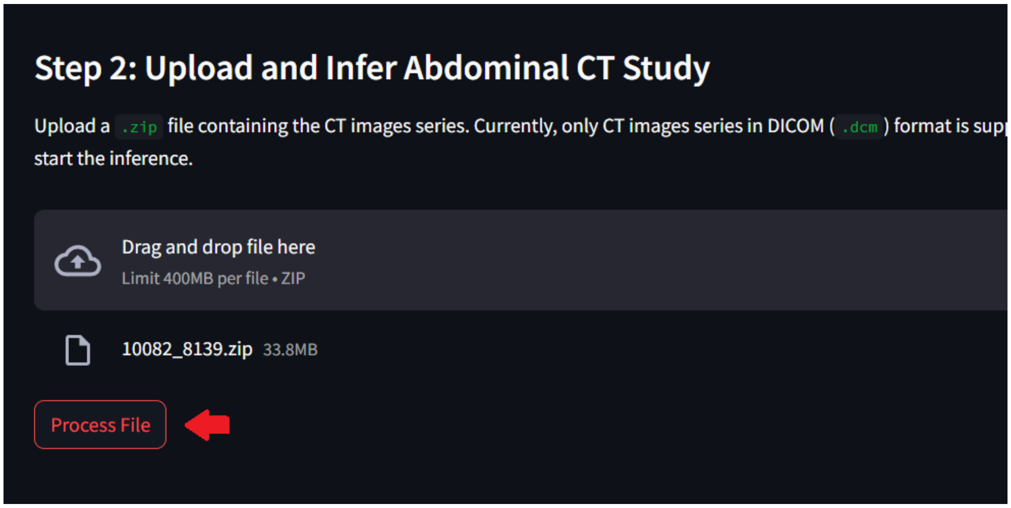


Figure A6: Upload the CT zip file

Step 3: Start inferencing
Click the "Process File" button to start the inference task. The status of the inference will be logged to the screen in real time.



(a) Click the Process File button

(b) Inference in progress

Step 4: Analyze prediction output
The inference process usually takes around 30 seconds to complete. Once completed, the result of the prediction will be shown on shown on the screen. The results are summarized in a table of what injury level of each organ is along with the model confidence.

Results

Number of Image Processed: 163
Elapsed Time: 26.21 seconds
Patient_ID: 10082
Series_ID: 8139

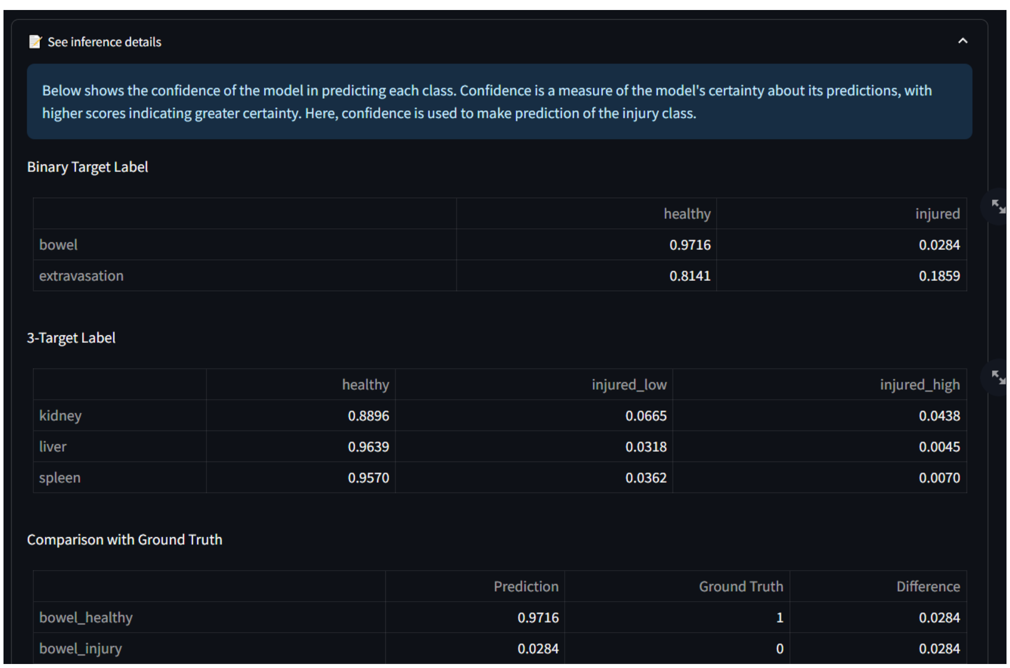
	Organ	Predicted	Confidence (%)
0	Bowel	Healthy	97.1600
1	Extravasation	Healthy	81.4100
2	Kidney	Healthy	88.9600
3	Liver	Healthy	96.3900
4	Spleen	Healthy	95.7000

See inference details

Process Next

Figure A8: Results of Inference

Users can also choose to view the detail of the inference by expanding the ”See Inference Details” section. This section reveals the probability of each organ injury class predicted by the model.



See inference details

Below shows the confidence of the model in predicting each class. Confidence is a measure of the model's certainty about its predictions, with higher scores indicating greater certainty. Here, confidence is used to make prediction of the injury class.

Binary Target Label

	healthy	injured
bowel	0.9716	0.0284
extravasation	0.8141	0.1859

3-Target Label

	healthy	injured_low	injured_high
kidney	0.8896	0.0665	0.0438
liver	0.9639	0.0318	0.0045
spleen	0.9570	0.0362	0.0070

Comparison with Ground Truth

	Prediction	Ground Truth	Difference
bowel_healthy	0.9716	1	0.0284
bowel_injury	0.0284	0	0.0284

Figure A9: Detailed view

Ground truth labels are provided for comparison whenever possible.

Comparison with Ground Truth

	Prediction	Ground Truth	Difference
bowel_healthy	0.9716	1	0.0284
bowel_injury	0.0284	0	0.0284
extravasation_healthy	0.8141	1	0.1859
extravasation_injury	0.1859	0	0.1859
kidney_healthy	0.8896	1	0.1104
kidney_low	0.0665	0	0.0665
kidney_high	0.0438	0	0.0438
liver_healthy	0.9639	1	0.0361
liver_low	0.0318	0	0.0318
liver_high	0.0045	0	0.0045
spleen_healthy	0.9570	1	0.0430
spleen_low	0.0362	0	0.0362
spleen_high	0.0070	0	0.0070

Process Next

Figure A10: Compare with Ground Truth

Users can then click ”Process Next” to proceed with more inference. That’s all for this user manual.