Project Overview

With more than 5 million deaths caused by traumatic injury each year, our project aims to address the critical issue of prompt and accurate diagnosis of abdominal trauma, Prompt and accurate diagnosis of traumatic injuries is crucial for improving patient outcomes and increasing survival rates. Among various diagnostic tools, computed tomography (CT) has emerged as a vital technology for evaluating individuals suspected of having abdominal injuries. CT scans provide detailed cross-sectional images of the abdomen, aiding in the detection and assessment of traumatic injuries.

The need for timely intervention and appropriate treatment underscores the importance of improving the diagnostic process. Among our research questions were; How effective are AI algorithms in automatically detecting traumatic injuries to internal abdominal organs like the liver, kidneys, spleen, and bowel using CT scans? What features and patterns in CT scans are most indicative of different severities of abdominal injuries, and how can they be utilized for automated injury grading? What are the appropriate metrics for evaluating the performance of the developed AI algorithms in terms of both machine learning benchmarks and clinical utility?

Our methodology for the analysis entailed data understanding, exploratory data analysis, image processing and visualization, modelling, conclusions, and recommendations.

Among the challenges that we encountered were:

i. Class imbalance with limited CPU as we had to oversample to balance the class
ii. Limited time when it came to the project completion timeframes

Data understanding

We had multiple datasets: the labels dataset, train dataset, train meta dataset and image dataset. For the labels data set, the data types were appropriate with integer types for identifiers and object (string) type for the injury name, the train data set all columns are of integer type while for the train meta dataset the data types are appropriate with integer and float types.

Exploratory Data Analysis (EDA)

Univariate analysis

From our analysis extravasation was more frequently identified in the provided images as compared to bowel injuries indicating class imbalance. This highlighted the need to address the imbalance but also added to the difficulty in addressing the minority class and avoid model bias towards the majority class.

Further analysis made using the histogram suggested a leptokurtic distribution. Most observations clustered closely around the mean indicating a higher degree of peakedness and less variability in the dataset. The distribution also had a longer right tail indicating positive skeweness

Bivariate analysis

We undertook the following visualizations: distribution of 'aortic_hu' based on 'injury_name', the relationship between 'injury_name' and 'incomplete_organ',. The performance of the model was that

- 'Kidney Status,' 'Liver Status,' and 'Spleen Status' carry substantial weight in predicting injury presence.
- Recognizing that 'Kidney Healthy', 'Liver Healthy' and 'Spleen Healthy' are the most prevalent states in injury cases will help influence the severity levels assigned when an injury is detected in these regions. Such nuanced adjustments based on prevalence patterns enhance the

precision of our severity predictions, supporting effective treatment decisions by medical professionals.

- The predominance of healthy organ statuses in the presence of injury is crucial information. It underscores the importance of swiftly identifying cases where organ status deviates from the healthy norm. By incorporating these patterns into our models, we can enhance the clinical applicability of our predictions, aligning with our project objective of providing medical experts essential tools for accurate injury detection and severity classification.

After visualizations, analysis of the following was done: Binary patient features based on the presence or absence of injuries and Correlation Analysis for the merged dataset,

Preprocessing

For this particular phase, we decided to undertake on the following

- Rescaling- this entailed adjusting the intensity values to a standard scale
- Resizing- this was to make sure all images have the same size, especially if they are being fed into a neural network.
- Histogram Equalization that enhanced the contrast of images.
- Normalization to remove the mean and scaling to unit variance.
- Data Augmentation by using techniques such as rotation, zooming, and flipping to artificially increase the size of the dataset (useful for training deep learning models).
- Smoothing to help improve on the data quality to contribute towards increasing accuracy of the predictive models
- Padding

Following these steps, the project was able to ensure the image output was clearer as compared to the original image.

Modelling

The process involved the following

- Data preparation where the data was split into training and validation sets
- Data Augmentation to artificially increase the size of the training dataset.
- Model Architecture to define the CNN architecture.
- Model Compilation to specify the loss function, optimizer, and metrics.
- Model Training to train the model using the training data.
- Model Evaluation that was to evaluate the model's performance on the validation data.

Baseline Model

Further, we went ahead with checking with checking for the class balance for each individual label in both the training set and the validation set. An upsampling of the minority class was done by identify majority and minority class combinations, upsampling each minority class combination and later combining the majority class with the upsampled minority classes.

To Check class balance in the validation set we undertook, image augmentation and preprocessing. The performance of the model is based on the following; the validation loss started at 7.07, the accuracy on the validation set varied: bowel (43.75%), extravasation (39.75%), liver (82.38%), kidney (79.25%), and spleen (48.37%).

Model two (2)

First, the dataset was split into groups and checked for data leakage. A total of 219 patient_ids shared between training and validation datasets. Also, we checked for class balance in both training set and the validation set.

Augmentation and processing of the images were also undertaken. Among other steps that we conducted were; getting the next batch of images and labels for training, mocking some values for configuration and training the model. We also visualized the training plots and stored the model for inferences.

Model three (3)

For our third model, we defined functions for reading DICOM images, augmentation, and image preprocessing, defined image generator and data loading, defined a function that will build the model and further undertook hyperparameter tuning. We also got the best hyperparameters and trained the final model using the best hyperparameters. Best hyperparameters had the following learning rate: 3.841017210511828e-05, dropout rate: 0.4 and weight decay: 4.597469246271629e-05

Upon evaluating the final model on the validation set the following accuracies and precisions were noted; bowel Accuracy: 0.5275 | Bowel Precision: 0.5304, Extra Accuracy: 0.6927 | Extra Precision: 0.6946, Liver Accuracy: 0.8813 | Liver Precision: 0.8821, Kidney Accuracy: 0.8147 | Kidney Precision: 0.8158, Spleen Accuracy: 0.6830 | Spleen Precision: 0.6850

Model four (4)

The changes made in this model were:

i. Architecture: We utilized DenseNet121, a deeper and more complex architecture than ResNet50 in Model 3.
ii. Hyperparameter Tuning: Further tuning with focus on minimizing false positives.
iii. Training Duration: Model trained for 10 epochs with early stopping enabled after 10 trials to prevent overfitting.

The best hyperparameters included: Learning Rate: 3.841017210511828e-05, Dropout Rate: 0.4, Weight Decay: 4.597469246271629e-05,

Bowel Class: Significant challenges exist in accurately classifying instances from this class, indicated by both low accuracy and precision. Targeted optimizations, more data and advanced techniques will be necessary.

BEST MODEL: Model 3

Model 3 not only exhibited stable accuracy and precision across multiple classes but also showcased consistent performance on both the validation and test datasets. Its ability to maintain accuracy while significantly reducing false positives makes it the ideal choice for our abdominal trauma classification task.

Based on its balanced accuracy, enhanced precision, and consistent performance across various classes and datasets, Model 3 stands as the optimal choice for our abdominal trauma classification and severity prediction task

Conclusion

Our analysis highlighted the prevalence of specific injury in the presence of any abdominal injury. Bowel injury and extravasation injury tend to be more common when any abdominal injury is detected.

This can help medical professionals identify common injury combinations for more targeted assessments.

Understanding these severity patterns will play a crucial pattern in the classification and priority cases for medical treatment.

Incomplete organ instances underscores the importance of developing models that can accurately diagnose injuries even in cases where organ visibility might be compromised due to incomplete imaging data.

Understanding the relationships between different injury types can guide the model's focus on specific injury types and associated organ statuses.

Patterns in CT Scans: The scan images revealed subtle differences in hyperdensity, potentially indicating injury severity. These insights suggest the importance of capturing intricate patterns in abdominal trauma, indicating the need for advanced image processing techniques.

Despite improvements, accurately detecting spleen injuries remained a significant hurdle, indicating a need for advanced approaches specific to this class.

The analysis of the complexity of the Aortic HU distributions provided insights into the variation in CT scan values.

Addressing class imbalance significantly improved model performance, ensuring accurate predictions across various organ injuries.

Precision Over Accuracy: Emphasizing precision, especially in a medical context, led to the reduction of false positives, crucial in avoiding unnecessary interventions and treatments.

The models were complex and thus careful attention to tuning and understanding the trade-offs between accuracy and false positives played a pivotal role in the model's success.

Model 3 emerged as the optimal choice due to its consistency and reliability

Recommendations

Utilization of injury combinations for targeted assessments

Focus on Addressing Spleen Injury Challenges: Allocate research efforts specifically to address challenges in spleen injury classification.

Practitioners can closely work to enhance models that can accurately diagnose injuries even when organ visibility is compromised using additional imaging techniques or supplementary tests.

Understanding the complexity of Aortic Hounsfield Units (HU) its variations and skewed distributions, will aid in the development of future, better performing algorithms.

Consider implementing techniques for explainable AI to provide insights into how the model arrives at its predictions.

Next steps

1. It is important to provide training sessions to medical practitioners who will be utilizing the model.
2. Integrating explainable AI techniques into the model to provide transparent insights into the decision-making process.
3. Continuous Evaluation