SpringBoard

Detecting Abdominal Trauma

By: Eric Wolf

Introduction

Traumatic injuries pose a significant global health threat, accounting for a substantial number of deaths in the first four decades of life. Addressing this challenge is crucial, with over 5 million annual deaths worldwide. Timely and accurate diagnosis plays a pivotal role in improving patient outcomes. Computed Tomography (CT) has become indispensable for evaluating abdominal injuries, providing crucial cross-sectional images for accurate assessments.

However, interpreting CT scans for abdominal trauma, especially in cases involving multiple injuries or subtle bleeding, is a complex task. This report explores the potential of artificial intelligence (AI) and machine learning to revolutionize the diagnostic process, aiming to assist medical professionals in rapidly and precisely detecting injuries and determining their severity.

The use of advanced algorithms has the potential to elevate trauma care and improve patient outcomes globally.

Blunt force abdominal trauma, often resulting from motor vehicle accidents, is a prevalent form of traumatic injury. Such incidents can lead to damage and internal bleeding of vital organs, requiring prompt detection for effective treatment. Traditional clinical diagnostic methods face many challenges, emphasizing the critical role of medical imaging in patient care.

The RSNA Abdominal Trauma Detection AI Challenge, organized in collaboration with the American Society of Emergency Radiology (ASER) and the Society for Abdominal Radiology (SAR), focuses on building models capable of detecting severe injuries to internal abdominal organs. This includes the liver, kidneys, spleen, and bowel, as well as identifying any active internal bleeding. By combining expertise in radiology and AI, the challenge aims to propel the field forward, offering innovative solutions to transform emergency medical care worldwide.

Method

Source

Data sourced from 23 research institutions across 14 countries was utilized to create the challenge dataset, consisting of de-identified abdominal CT studies and related clinical information. The dataset included target labels for organs such as bowel, extravasation, kidney, liver, or spleen and healthy, low, or high to indicate severity of injury. Corresponding data was provided with the CT scan data in DICOM format such as series metadata, image-level labels, segmentations, DICOM tags, and other auxiliary files.

Data Wrangling | Exploratory Data Analysis

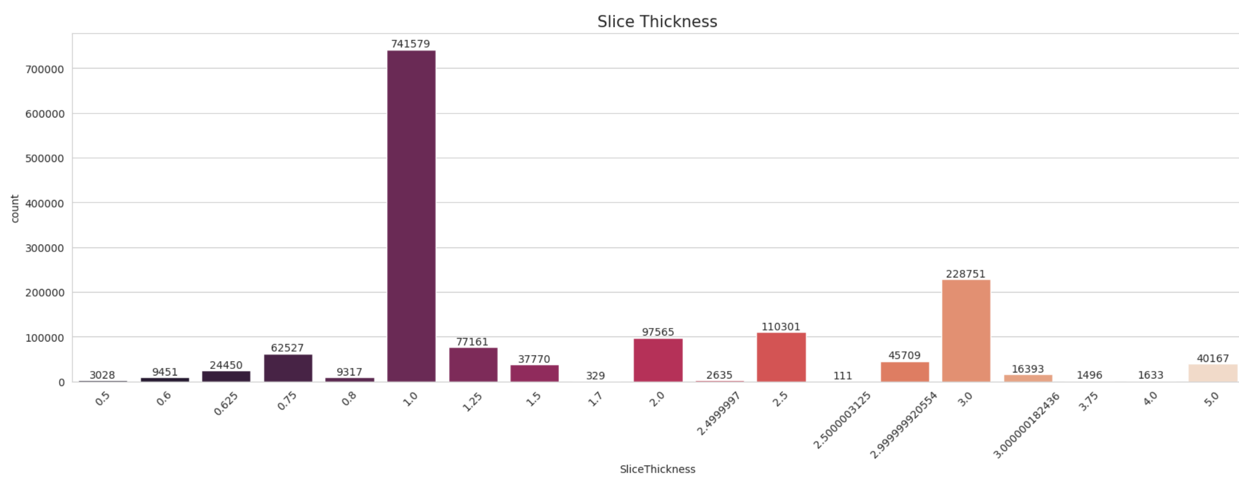
Upon delving into the raw training data, a comprehensive examination revealed a total of 3,147 entries with 15 attributes, portraying a well-structured dataset. Crucially, the data exhibited a one-hot encoded format, eliminating any concerns about null or missing values. The training metadata, consisting of 4 attributes and 4,711 entries, unveiled 3,147 unique patient_ids, suggesting the possibility of repeat or multiple entries per patient, prompting a need for careful consideration during subsequent analyses.

Further exploration of the image data uncovered a total of 12,029 entries, with only 246 unique patient IDs and 330 instance numbers. This observation suggested the existence of multiple images per patient, and indicated that not every patient in the raw training data corresponded to an image. This revelation underscored the importance of ensuring data integrity and avoiding assumptions based solely on patient IDs from the raw data set.

In terms of reported injuries within the training set, approximately 73% of patients had documented injuries. However, the distribution of specific injuries varied, with roughly 2% reporting bowel injuries, 6% extravasation, 6% kidney injuries, 10% liver injuries, and 11% spleen injuries. Notably, only 6.64% of the scans contained identifiable organs, emphasizing the need for precise detection algorithms.

A potentially crucial aspect of the data analysis was the examination of CT scan thickness, revealing a range from 0.5 to 5.0. To visualize this distribution, a bar graph of slice thickness was generated (see Figure 1). This histogram provided insights into the variation in scan thickness across the dataset, aiding in understanding the characteristics of the imaging data.

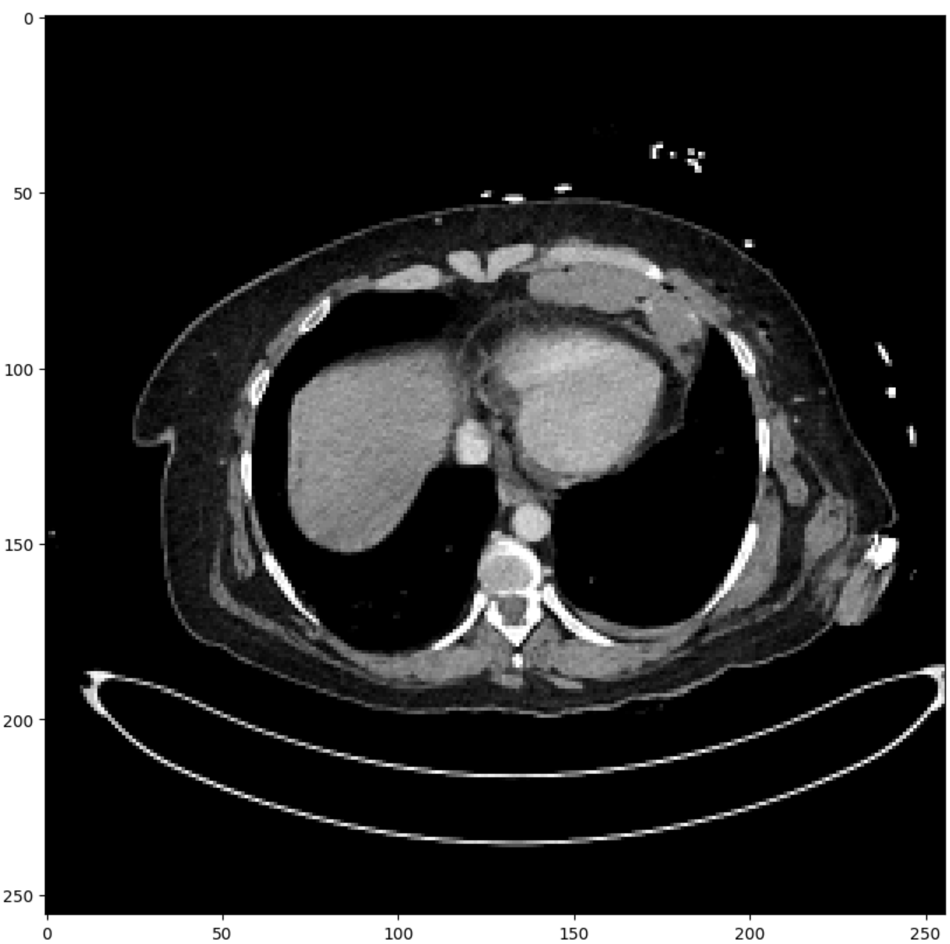
Figure 1: Slice Thickness



The bar graph illustrated the distribution of CT scan thickness and highlighted potential variations that could impact the analysis and interpretation of the medical images. This information is potentially pivotal in designing preprocessing steps to account for differences in scan characteristics during subsequent modeling.

Modeling | Results

Following meticulous data cleaning and preprocessing, the dataset for the RSNA Abdominal Trauma Detection AI Challenge underwent careful preparation for integration into computer vision models. Originally in DICOM format, a standard for medical imaging, the decision was made to convert the dataset into PNG images derived from DICOM. This transformation was driven by a combination of factors, including technical compatibility, ease of preprocessing, robust community support, visualization benefits, and storage considerations tailored to the intricate requirements of our machine learning workflow. An example of the converted image is shown below for illustration purposes.



The subsequent phase involved the selection of four distinct models for evaluation, each bringing unique strengths to the task at hand: a baseline model for established comparison, ResNet-50, EfficientNet B7, and Xception.

ResNet-50: Renowned for its effectiveness in training deep neural networks,emerged as a powerful choice due to its depth and capacity, enabling it to capture intricate features in complex medical images. The introduction of residual blocks mitigates the vanishing gradient problem, a critical advantage for tasks demanding nuanced understanding.

EfficientNet B7: As part of the EfficientNet architecture, EfficientNet B7 stood out for its design tailored for high computational capacity and superior performance in demanding computer vision tasks. Its efficiency across scales, transfer learning capabilities, resource efficiency, scalability, and state-of-the-art performance made it a compelling candidate for abdominal trauma detection.

Xception: Deviating from traditional designs, Xception utilizes depthwise separable convolutions, emphasizing efficiency, global contextual information capture, transfer learning capabilities, and adaptability to different scales. Its innovative architecture and high computational efficiency positioned it as a cutting-edge solution for our task.

With the models selected for testing each notebook was augmented for the appropriate image preprocessing required. Additionally to optimize each model hyperparameter tuning was executed using Bayesian Optimization, an adaptive and data-efficient technique designed for efficient exploration of the hyperparameter space. The results of each model can be found below. Each model was scored on a separated validation set from the preprocessed data, accuracy, recall, and precision shown is an averaged overall score for each image identified in the best epoch determined by lowest validation loss.

Results Summary:

Base Model:

Best Loss: 3.159, Accuracy: 75.5%, Recall: 73.9%, Precision: 76.2%

Organ Metrics: Varied accuracy, recall, and precision for different organs.

ResNet-50:

Best Loss: 0.753, Accuracy: 100%, Recall: 100%, Precision: 100%

Organ Metrics: Achieved the highest accuracy, recall, and precision for all organs.

EfficientNet B7:

Best Loss: 1.050, Accuracy: 97.8%, Recall: 97.7%, Precision: 97.9%

Organ Metrics: Demonstrated high performance for all organs.

Xception:

Best Loss: 1.336, Accuracy: 94.8%, Recall: 94.5%, Precision: 95.0%

Organ Metrics: Showed strong performance for all organs.

In the meticulous process of selecting the final model, ResNet-50's consistent dominance across all three critical metrics—accuracy, recall, and precision—made it the unequivocal recommendation. Scoring the highest in each category, ResNet-50 not only demonstrated technical prowess but also exhibited a capacity to capture subtle abnormalities in diverse abdominal organs.

The results underscore the efficacy of advanced deep learning architectures, particularly ResNet-50, EfficientNet B7, and Xception, in significantly enhancing abdominal trauma detection. These models exhibit not only impressive accuracy, recall, and precision but also a capacity to capture subtle abnormalities in diverse abdominal organs. The choice of Bayesian Optimization for hyperparameter tuning contributed to the fine-tuning of each model, optimizing their performance for the specific requirements of the task. The success of these models marks a significant stride towards leveraging AI for improved patient outcomes in abdominal trauma cases.

Future Improvements

Moving forward, our focus lies on a multifaceted approach to refine and elevate the capabilities of AI in abdominal trauma detection. One pivotal aspect involves the iterative refinement of existing models, with a keen emphasis on enhancing their performance and generalization across diverse datasets. Additionally, we envision exploring ensemble methods that amalgamate the strengths of multiple models, fostering synergistic improvements in accuracy and robustness.

Our commitment to collaboration with the medical community remains unwavering. Continuous engagement with healthcare professionals would allow us to refine and validate our models using authentic clinical data. We recognize the invaluable feedback from practitioners as an essential element in ensuring our models align seamlessly with the practical requirements of clinical settings.

In our pursuit of innovation, we are open to investigating advanced architectures beyond the currently employed models. This exploration aims to uncover novel approaches that may offer superior performance in abdominal trauma detection. Concurrently, we remain attuned to emerging techniques and architectures, acknowledging the dynamic nature of the deep learning landscape.

An integral dimension of our future efforts involves incorporating interpretability tools into our models. By enhancing transparency and providing clear insights into the decision-making process, these tools aim to foster trust among healthcare professionals, ultimately contributing to the seamless integration of AI-assisted diagnostic decisions into clinical workflows.

Addressing emerging challenges in the field remains a steadfast priority. Ongoing research initiatives will be dedicated to staying abreast of the latest advancements and adapting our models to accommodate changes in data distribution, as well as integrating insights derived from cutting-edge medical research.

As we advance, the continuous refinement of performance evaluation metrics is essential. Our goal is to align these metrics more closely with clinical relevance, ensuring that the measured success of our models directly translates into meaningful impacts on patient outcomes.

Lastly, our exploration extends to the realm of hyperparameter tuning strategies. We aim to delve into advanced techniques that further optimize model performance, scrutinizing the impact of hyperparameters on robustness and generalization. This meticulous examination will contribute to the ongoing evolution of our models, fostering their effectiveness in real-world clinical scenarios.