AI-driven Cervical Spine Fracture Detection from CT Scans of the Human Spine

Md Mushfiqur Rahman Chowdhury
School of Computing and Electrical Engineering
Indian Institute of Technology, Mandi
D23134@students.iitmandi.ac.in

Soujatya Sarkar
School of Computing and Electrical Engineering
Indian Institute of Technology, Mandi
S23106@students.iitmandi.ac.in

Kriti Khare
School of Computing and Electrical Engineering
Indian Institute of Technology, Mandi
S23069@students.iitmandi.ac.in

Krishna Kumar Dixit
School of Biosciences and Bioengineering
Indian Institute of Technology, Mandi
B21013@students.iitmandi.ac.in

***Abstract*—For early patient care and medical intervention, the accurate and rapid diagnosis of cervical spine fractures from CT scans is essential. In this study, we suggest using cutting-edge AI methods to improve fracture detection's effectiveness and precision. Our primary goal is to apply state-of-the-art architectures, like Distilled VIT and Vision Transformer with deformable attention, and optimize for quicker computation without sacrificing accuracy or higher accuracy with comparable processing resources. Preprocessing CT scan images, using annotated data to train the AI model, and optimizing the chosen architecture to focus on cervical spine fracture detection comprise our methodology. The dataset used in this experimental project has been collected from (RSNA - RADIOLOGICAL SOCIETY OF NORTH AMERICA, a featured code competition held in 2022 for Cervical Spine Fracture Detection). Through the utilization of deep learning and novel attention mechanisms, our goal is to create a dependable system that can accurately and quickly detect fractures. The project's goal is to develop a highly effective and precise AI-driven cervical spine fracture detection system that can expedite diagnosis and enhance patient outcomes. These developments have the potential to transform diagnostics for medical imaging, optimize clinical workflows, and ultimately save lives. The results of this initiative have the potential to transform medical imaging diagnostics, improve fracture detection precision, and speed up patient treatment decisions.**

***Index Terms*—Fracture detection, CT scans, Vision Transformer, Deformable attention, Distilled VIT, Medical imaging, Diagnostic accuracy, Computational efficiency**

I. Introduction

The spinal cord is one of the most important parts of the human nervous system, it is a critical intermediary for transmitting sensory and motor information between the brain and specific regions of an individual's body. This intricate structure, protected by the vertebral column - or spine as it is most commonly known — consists of a number of bony segments: the vertebrae. Third, the vertebral column is itself divided into independent regions with their anatomical features and made-up functions.

Comprising the highest portion of the vertebral column is The cervical spine which includes seven vertebrae termed C1 to C7. These specific vertebrae serve many valuable functions such as providing support for our head and permitting numerous movements like rotation, extension, and flexion. Figure 1 shows the vital cervical vertebrae that constitute the neck region; these are essential for safeguarding every delicate structure of the spinal cord so that they can function well.

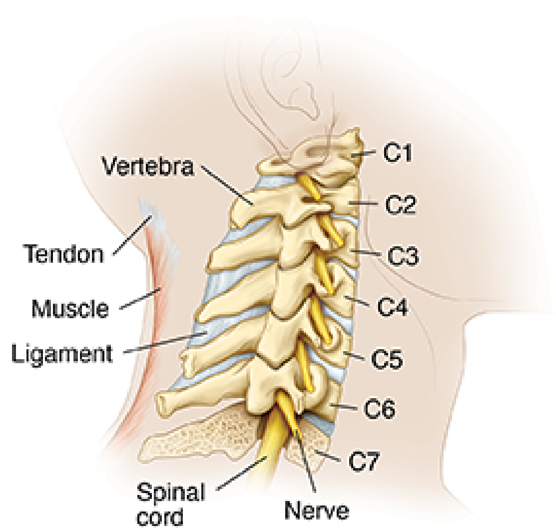

Fig. 1: The Cervical Spine

With major socioeconomic ramifications, the prevalence of spinal cord injuries (SCIs) resulting from vertebral fractures is a matter of great public health concern. Recent data indicates that there are about 17,730 new occurrences of SCI in the United States alone each year, with falls and car accidents causing the majority of these injuries [24]. These injuries not only cause great physical and psychological suffering for those who sustain them, but they also place a significant financial strain on healthcare systems around the globe.

There is an urgent need for reliable and effective detection techniques since prompt and correct diagnosis is crucial to reducing the negative effects of spinal fractures. Innovative cervical spine fracture identification methods have been made possible by recent developments in artificial intelligence (AI) and medical imaging. These methods can potentially change clinical practice and improve patient outcomes completely.

This article's goal is to investigate the use of AI-driven methods for detecting cervical spine fractures from human spine CT scans. We aim to create a scalable and resilient system that can identify cervical spine fractures early and accurately. This will enable medical professionals to treat patients promptly and enhance patient outcomes. We will achieve this by utilizing cutting-edge deep learning algorithms and cloud computing.

II. Literature Review

The diagnosis and management of spinal injuries, particularly cervical spine fractures, represent a complex clinical challenge. Recent scholarly endeavors have capitalized on emerging technologies such as deep learning (DL) and cloud-based computation to augment detection systems, thereby offering promising avenues for enhancing patient care and outcomes [1].

Clinical evidence underscores the pivotal importance of timely surgical intervention in the management of spinal cord injuries [2]. Nevertheless, accurate and efficient detection of spinal fractures remains an ongoing clinical concern. Initiatives such as the RSNA 2022 Cervical Spine Fracture Detection competition have sought to bridge this gap by fostering innovation in fracture detection algorithms [3].

Computational methods have emerged as indispensable tools in advancing spine imaging, furnishing clinicians with robust analytical frameworks for diagnosis and treatment evaluation. Noteworthy publications like "Computational Methods and Clinical Applications for Spine Imaging" provide comprehensive insights into the convergence of computational techniques and clinical practice [4]. Similarly, benchmark datasets like Verse serve as invaluable resources for the development and evaluation of vertebrae labeling and segmentation algorithms [5].

A nuanced comprehension of the pathophysiological underpinnings and associated risk factors of vertebral compression fractures is paramount for refining preventive and therapeutic strategies [6]. Moreover, infections affecting the spinal cord and contiguous structures necessitate precise imaging modalities for accurate diagnosis and effective management [7]. Scholarly contributions such as "Spinal Imaging and Image Analysis" furnish clinicians and researchers alike with indispensable knowledge in this domain [8].

Contemporary reviews underscore the escalating significance of DL in medical imaging, with a particular emphasis on applications pertinent to spinal health [9]. Authoritative works such as the "Handbook of Medical Image Computing and Computer-Assisted Intervention" furnish exhaustive coverage of DL methodologies and their manifold applications in medical imaging [10]. Furthermore, iterative fully convolutional neural networks and other DL architectures hold promise in automating vertebra segmentation and identification tasks, thereby streamlining clinical workflows [11, 12].

Notwithstanding significant strides, challenges persist in achieving precise detection of vertebral fractures. Innovative approaches such as cortical shell unwrapping and model-based segmentation frameworks endeavor to bolster fracture detection sensitivity and specificity [13, 14]. Furthermore, research endeavors in allied fields such as colonography and head-and-neck carcinoma detection offer transferable insights and methodologies to the realm of spinal imaging [15, 16].

Recent advancements in DL, computational methods, and evidence-based clinical practice have substantially reshaped the landscape of spinal imaging and fracture detection. Collaborative synergies among clinicians, researchers, and technologists continue to propel innovation, promising heightened diagnostic accuracy and optimized patient care in the domain of spinal injury management.

III. Objectives, Milestones, and Deliverables

A. *Objectives*

- Develop an AI-driven model specifically designed for the identification of cervical spine fractures using CT scans.
- Utilize a diverse and extensive dataset of annotated CT scans of the human spine to train and evaluate the developed model.
- Implement cutting-edge machine learning algorithms, particularly deep learning approaches, to enhance the accuracy and efficiency of fracture detection.
- Compare the performance of the AI-driven model with current methods to assess improvements in diagnostic efficiency and accuracy.
- Evaluate the potential impact of the developed model on clinical decision-making by measuring the reduction in diagnosis time.
- Investigate the potential enhancement of patient outcomes through early and accurate detection of cervical spine fractures.
- Contribute to the advancement of AI in medical imaging by providing radiologists and doctors with a reliable tool for early and accurate fracture detection.

B. *Milestones and Deliverables*

The project Milestones have been represented in the form of a Gannt Chart in Fig. 2. This timeline shows a sequential progression of tasks, with overlapping periods where multiple tasks are conducted concurrently. This helps in understanding the allocation of time and resources for each phase of the project.

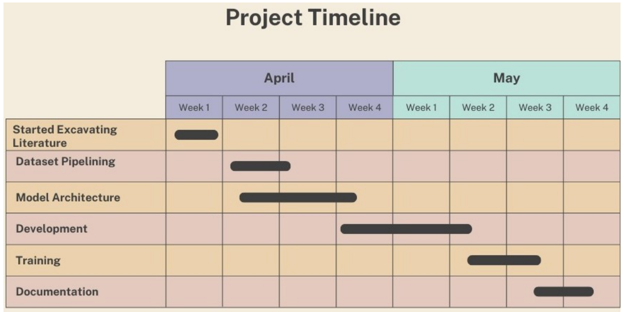

Fig. 2: Project Timeline

1. Started Excavating Literature: This task started in Week 1 of April and continued into Week 2 of April.

2. Dataset Pipelining: We began the work in Week 2 of April and continued through Week 3 of April.

3. Model Architecture: Started in Week 3 of April and extended till Week 1 of May.

4. Development: Commenced in Week 2 of May and continued till Week 3 of May.

5. Training: Started in Week 3 of May and finished just a few days before our project submission into Week 4 of May.

6. Documentation: This task was done from the start of the last week of May and finished on the 25th of May.

IV. Dataset Description

The objective of this competition is to discern fractures in CT scans of the cervical spine (neck) at both the individual vertebrae level and the overall patient level. Timely detection and precise localization of any vertebral fractures are crucial to prevent neurological deterioration and paralysis following trauma.

This competition entails a hidden test dataset. Upon submission, the actual test data, including a comprehensive sample submission, will be provided to the submitted notebook.

A. *Files*

- **train.csv**: Metadata for the training dataset.
  - **StudyInstanceUID**: The unique study ID for each patient scan.
  - **patient_overall**: One of the target columns indicating the patient level outcome, i.e., if any of the vertebrae are fractured.
  - **C[1-7]**: The other target columns indicate whether the given vertebrae is fractured.
- **test.csv**: Metadata for the test set prediction structure. Only the initial rows of the test set are accessible for download.
  - **row_id**: The row ID.
  - **StudyInstanceUID**: The study ID.
  - **prediction_type**: Indicates which one of the eight target columns requires a prediction in this row.

- **[train/test]_images/[StudyInstanceUID]/[slice_number].dcm**: The image data, organized with one folder per scan. Expect approximately 1,500 scans in the hidden test set. Each image is in DICOM file format with ≤ 1 mm slice thickness, axial orientation, and bone kernel. Note that some DICOM files are JPEG compressed.

- **sample_submission.csv**: A valid sample submission.
  - **row_id**: The row ID.
  - **fractured**: The target column.

- **train_bounding_boxes.csv**: Bounding boxes for a subset of the training set.

- **segmentations/**: Pixel-level annotations for a subset of the training set, provided in the NIfTI file format. A portion of the imaging datasets have been segmented automatically using a 3D UNET model, with radiologists modifying and approving the segmentations. The provided segmentation labels range from 1 to 7 for C1 to C7 (seven cervical vertebrae) and 8 to 19 for T1 to T12 (twelve thoracic vertebrae).

  Please note that the NIfTI files contain segmentations in the sagittal plane, while the DICOM files are in the axial plane. Utilize the NIfTI header information to ensure alignment between DICOM images and segmentations.

B. *Our Dataset*

Initially, we utilized the 86 segmentation sample studies provided by the organizers. We redefined the mask labels as follows:

- 0: Background
- 1 to 7: C1 to C7

V. Methodology

A. *Dataset Preparation*

We utilized a dataset consisting of 86 CT scans of the spines from different patients. The segmentations are stored in the NIfTI format (with the extension `.nii` or `.nii.gz` , which is widely used in medical imaging for storing volumetric data. The individual slices of the CT scans are in the DICOM format (`.dcm`), a standard format for medical imaging data that ensures compatibility with various imaging devices and software.

For our experiments, we divided the dataset as follows:

- **Training and Validation**: We allocated 84 CT scan images for training and validation purposes. This subset was further split into training and validation sets in a 70:14 ratio, resulting in 70 images for training and 14 images for validation. This split helps ensure that the model is exposed to a sufficient variety of examples during training while having a distinct validation set to monitor and tune model performance.

- **Testing**: The remaining 2 CT scan images were set aside for testing. This test set is used to evaluate the model's performance on unseen data, providing an unbiased assessment of its generalization capability.

B. *Summary of Data Split*

- **Total CT Scans**: 86
- **Training and Validation**: 84
  - **Training Set**: 70 images
  - **Validation Set**: 14 images
- **Test Set**: 2 images

This approach ensures that our model is trained, validated, and tested on appropriately split data, enabling reliable performance evaluation and potential for generalization to new data.

C. *Data Augmentation*

We applied data augmentation techniques to the CT scan slices to enhance the dataset and improve model robustness. The process for creating augmented data is as follows:

1) Segmentation Data:

- For segmentation files in NIfTI format (e.g., `2.nii`), we created a single `.npy` file with 6 channels. This file is named `2.npy`.

2) Slice Data:

- For the DICOM slices (e.g., `0.dcm`, `1.dcm`, `2.dcm`, `3.dcm`, and so on), we generated `.npy` images for every even-numbered file (`0.dcm`, `2.dcm`, `4.dcm`, `6.dcm`, `8.dcm`, ..., `28.dcm`). This resulted in a total of 15 `.npy` files.

3) Vertebrae and Slices:

- Each vertebra is represented by 30 slices.
- Considering the cervical vertebrae (C1 through C7), we have:
  - **C1 to C7**:

$$30 \text{ slices/vertebra} \times 7 \text{ vertebrae} = 210 \text{ slices}$$

- Only these 210 slices are deemed important for our study; the rest are discarded.
- From these 210 slices, using our data augmentation technique, we generate 105 `.npy` files. This results in one `.npy` file for every pair of slices, effectively giving us data for every second slice.

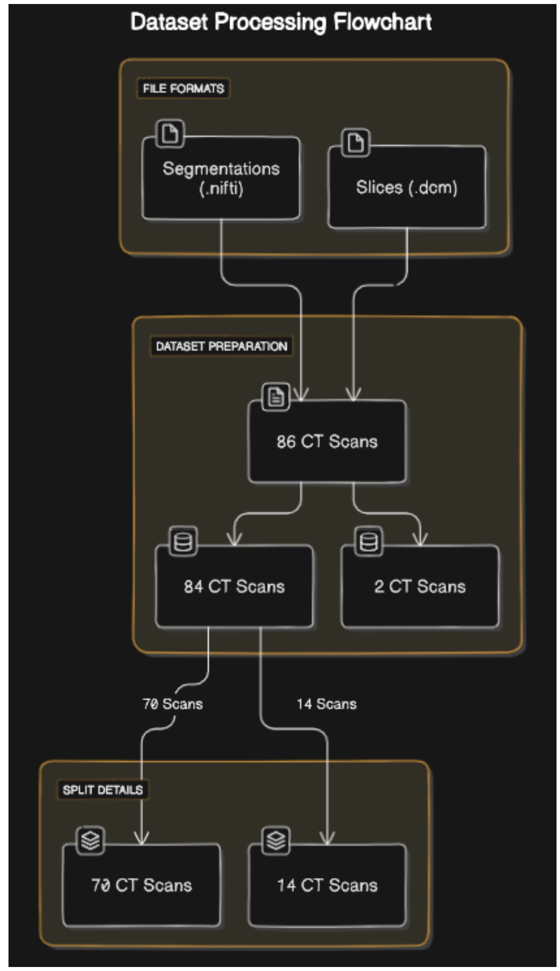

Fig. 3: Flow of Data Preparation

This data augmentation strategy ensures that we maximize the use of available data, thereby enhancing the training process and potentially improving the model's performance on segmentation tasks.

D. *Model Architecture*

For our task, we employed the Vision Transformer (ViT) as in Fig. 4 due to its strong performance in classification tasks. The ViT architecture is particularly suitable for handling image data by utilizing a transformer encoder to process image patches.

1) Transformer Encoder: We utilized a ViT with 12 transformer encoder layers. The key components of our model architecture are as follows:

- **Volumetric Patch Embeddings**: The input volumetric data is divided into smaller patches, which are then flattened and projected linearly to create patch embeddings.

- **Positional Encoding**: Positional encodings are added to the patch embeddings to retain the spatial information within the sequence of patches.
- **Transformer Encoder Layers**: The sequence of patch embeddings, augmented with positional encodings, is fed into the transformer encoder. Each encoder layer consists of multiple layers of multi-head self-attention mechanisms and feed-forward neural networks.
- **Classification Token (CLS Token)**: A special learnable classification token (CLS token) is prepended to the sequence of patch embeddings. The final hidden state corresponding to this CLS token is used as the representation for the entire image, which is then utilized for the classification task.

2) Classification Output: The model outputs an array of length 7, corresponding to the presence or absence of fractures in each of the seven cervical vertebrae (C1-C7). Each element in the array is a binary value (0 or 1), indicating the absence or presence of a fracture in the respective vertebra.

3) Training and Loss Calculation: During training, we calculate the loss between the predicted array and the ground truth labels. This loss is then backpropagated through the network to update the weights, improving the model's performance.

- **Loss Calculation**: The loss function compares the predicted classification array to the ground truth labels. We have used Binary Cross Entropy loss.
- **Backpropagation**: The calculated loss is used to adjust the model's weights through backpropagation, enabling the model to learn and improve over time.

This comprehensive use of the Vision Transformer architecture allows us to effectively address the classification task, providing a robust framework for detecting fractures in cervical vertebrae based on volumetric CT scan data.

Our 3D ViT approach can be summarised in the form of an architectural diagram as shown in Fig. 4.

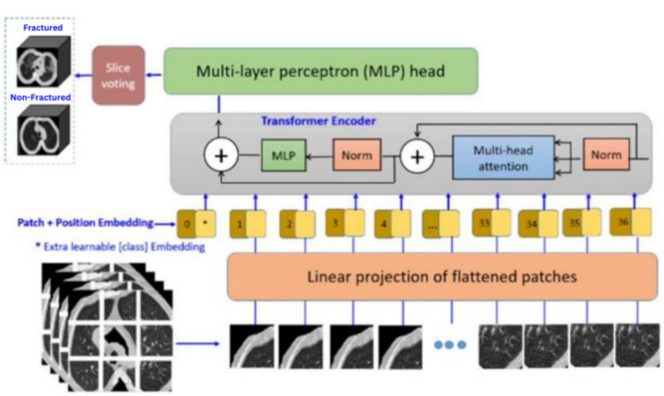

Fig. 4: Architecture of 3D Simple ViT

VI. Implementation Results

We have implemented 3 models with different hyper-parameters for achieving our objective. The ablation study for them are shown in Table. 1.

TABLE I: Ablation study of our 3 trained models

Model	Train Loss	Val Loss	Train Accuracy	Val Accuracy
3D ViT - 1/3e-4/0.0	0.4893	0.4879	90.2899	90.4176
3D ViT - 1/5e-4/0.1	0.7179	0.8186	78.6929	75.069
3D ViT - 1/3e-4/0.1	0.4169	0.4158	93.6419	94.4136

As our objective nearly aligns with a Kaggle competition, we have compared our model with the state-of-the-art architectures that have been publicly showcased in the Kaggle leaderboard of this competition. Our implementation results against the state-of-the-art Kaggle models, are summarized in Table. 2.

TABLE II: Comparative study with State-of-the-Art models

Model	Train Loss	Val Loss
2.5D CNN + LSTM classification	0.2117	0.2047
2.5D CNN + Transformer encoder	0.4049	0.4140
3D ViT B2 [on 1/25 data]	0.4169	0.4158
3D ViT B1 [on 1/25 data]	0.4893	0.4879
EfficientNet-V2-S classifier	0.4812	0.4982

VII. Conclusion

From the above results we can conclude that our model though trained on 1/25 of the total dataset has performed very efficiently compared to the SOTA architectures. We expect this model to perform much better and beat the SOTA trends in a much more efficient and organized manner.

VIII. Acknowledgement

The authors would like to acknowledge the opportunity and guidance provided by Dr. Sneha Singh, Assistant Professor of the School of Computing and Electrical Engineering, IIT Mandi. This project has been carried out as a task for the course: Medical Imaging and Applications (EE-XXX).

References

[1] P.Chład and M.R. Ogiela, "Deep Learning and Cloud-Based Computation for Cervical Spine Fracture Detection System", Electronics 2023, 12, 2056. https://doi.org/10.3390/electronics12092056

[2] M.G. Fehlings and R.G. Perrin, "The Timing of Surgical Intervention in the Treatment of Spinal Cord Injury: A Systematic Review of Recent Clinical Evidence". Spine 2006, 31, S28–S35

[3] A. Flanders, C. Carr, E. Colak, F. Kitamura, H.M. Lin, J.Rudie, J.Mongan, K.Andriole, L.Prevedello and M.Riopel, RSNA 2022 Cervical Spine Fracture Detection. 2022. Available online: https://www.kaggle.com/competitions/rsna-2022-cervicalspine-fracture-detection/overview (accessed on 5 March 2023).

[4] Cai, Y., Wang, L., Audette, M., Zheng, G. and Li, S. "Computational Methods and Clinical Applications for Spine Imaging", (Springer, 2020).

[5] Sekuboyina, A.,"Verse: a vertebrae labelling and segmentation benchmark", arXiv preprint arXiv:2001.09193 (2020).

[6] Faruqi, S.,"Vertebral compression fracture after spine stereotactic body radiation therapy: a review of the pathophysiology and risk factors", Neurosurgery 83, 314–322 (2018).
[7] Balériaux, D. L. and Neugroschl, C. "Spinal and spinal cord infection", Eur. Radiol. Suppl. 14, E72–E83 (2004).
[8] Li, S., Yao, J.,"Spinal Imaging and Image Analysis" (Springer, 2015).
[9] Zhou, S. K.,"A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises", Proc. IEEE (2021).
[10] Zhou, S. K., Rueckert, D. and Fichtinger, "G. Handbook of medical image computing and computer-assisted intervention" (Academic Press, 2019).
[11] Lessmann, N., Van Ginneken, B., De Jong, P. A. and Išgum, I, "Iterative fully convolutional neural networks for automatic vertebra segmentation and identification", Med. image analysis 53, 142–155 (2019).
[12] Payer, C., Stern, D., Bischof, H. and Urschler, M., "Coarse to fine vertebrae localization and segmentation with spatialconfiguration-net and u-net", In VISIGRAPP (5: VISAPP), 124–133 (2020).
[13] Yao, J., Burns, J. E., Munoz, H. and Summers, R. M, "Detection of vertebral body fractures based on cortical shell unwrapping", In International Conference on Medical Image Computing and Computer-Assisted Intervention, 509–516 (Springer, 2012).
[14] Korez, R., Ibragimov, B., Likar, B., Pernuš, F. and Vrtovec, T. "A framework for automated spine and vertebrae interpolationbased detection and model-based segmentation", IEEE transactions on medical imaging 34, 1649–1662 (2015).
[15] Johnson, C. D., "Accuracy of ct colonography for detection of large adenomas and cancers", New Engl. J. Medicine 359, 1207–1217 (2008).
[16] Bejarano, T., De Ornelas Couto, M. and Mihaylov, "I. Head-and-neck squamous cell carcinoma patients with ct taken during pre-treatment, mid-treatment, and post-treatment dataset", the cancer imaging archive; 2018.
[17] Simpson, A. L."A large annotated medical image dataset for the development and evaluation of segmentation algorithms", arXiv preprint arXiv:1902.09063 (2019).
[18] Harmon S. A.,"Artificial intelligence for the detection of covid-19 pneumonia on chest ct using multinational datasets." Nat. communications 11, 1–7 (2020).
[19] Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. and Maier-Hein, "K. H. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation." Nat. Methods 18, 203–211 (2021).
[20] Yushkevich, P. A., "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability", Neuroimage 31, 1116–1128 (2006).
[21] Ronneberger, O., Fischer, P. and Brox, "T. U-net: Convolutional networks for biomedical image segmentation", In International Conference on Medical image computing and computer-assisted intervention, 234–241 (Springer, 2015).
[22] Peng, C., Lin, W.-A., Liao, H., Chellappa, R. and Zhou, S. K. Saint: "spatially aware interpolation network for medical slice synthesis", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7750–7759 (2020).
[23] Egger, J, "Gbm volumetry using the 3d slicer medical image computing platform", Sci. reports 3, 1–7 (2013).
[24] National Spinal Cord Injury Statistical Center, "Spinal Cord Injury Facts and Figures at a Glance," Birmingham, AL, USA, 2021. [Online]. Available: https://www.nscisc.uab.edu/Public/Facts