

# Automatic lesion analysis for increased efficiency in outcome prediction of traumatic brain injury

Margherita Rosnati<sup>1,2,3</sup>, Eyal Soreq<sup>2,3</sup>, Miguel Monteiro<sup>1</sup>, Lucia Li<sup>2,3</sup>,  
Neil S.N. Graham<sup>2,3</sup>, Karl Zimmerman<sup>2,3</sup>, Carlotta Rossi<sup>4</sup>, Greta Carrara<sup>4</sup>,  
Guido Bertolini<sup>4</sup>, David J. Sharp<sup>2,3</sup>, and Ben Glocker<sup>1</sup>

<sup>1</sup> BioMedIA Group, Department of Computing, Imperial College London, UK

<sup>2</sup> UK Dementia Research Institute Care Research and Technology Centre, Imperial College London and the University of Surrey, Guildford

<sup>3</sup> Department of Brain Sciences, Faculty of Medicine, Imperial College London, UK

<sup>4</sup> Laboratorio di Epidemiologia Clinica Dipartimento di Salute Pubblica, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Italy  
margherita.rosnati12@imperial.ac.uk

**Abstract.** The accurate prognosis for traumatic brain injury (TBI) patients is difficult yet essential to inform therapy, patient management, and long-term after-care. Patient characteristics such as age, motor and pupil responsiveness, hypoxia and hypotension, and radiological findings on computed tomography (CT), have been identified as important variables for TBI outcome prediction. CT is the acute imaging modality of choice in clinical practice because of its acquisition speed and widespread availability. However, this modality is mainly used for qualitative and semi-quantitative assessment, such as the Marshall scoring system, which is prone to subjectivity and human errors. This work explores the predictive power of imaging biomarkers extracted from routinely-acquired hospital admission CT scans using a state-of-the-art, deep learning TBI lesion segmentation method. We use lesion volumes and corresponding lesion statistics as inputs for an extended TBI outcome prediction model. We compare the predictive power of our proposed features to the Marshall score, independently and when paired with classic TBI biomarkers. We find that automatically extracted quantitative CT features perform similarly or better than the Marshall score in predicting unfavourable TBI outcomes. Leveraging automatic atlas alignment, we also identify frontal extra-axial lesions as important indicators of poor outcome. Our work may contribute to a better understanding of TBI, and provides new insights into how automated neuroimaging analysis can be used to improve prognostication after TBI.

## 1 Introduction

Traumatic brain injury (TBI) is a leading cause of death in Europe [27]. In the UK alone, 160,000 patients with TBI are admitted to hospitals annually, with an estimated yearly cost of £15 billion [21]. The accurate prediction of TBI outcome is still an unresolved challenge [13,9] whose resolution could improve the therapy

and after-care of patients. Due to its acquisition speed and wide availability, computed tomography (CT) is the imaging modality of choice in clinical practice [14] and a key component in TBI outcome prediction [4]. The Marshall score [17] is one of the most widely used metrics to evaluate TBI injury severity. However, it does not leverage the rich information content of CT imaging [2] and requires a radiologist to assess the CT scan manually, which is time-consuming. Years of acquisition of CT of TBI patients generated rich datasets, opening the possibility of automatic extraction of CT biomarkers. These could enable a deeper and broader use of imaging data, augmenting the skills of radiologists and reducing the workload, allowing them to see more patients. Machine learning for medical imaging is a growing research field with advances in medical imaging segmentation [25] and classification [6]. It can be used for fast and autonomous outcome prediction of TBI using imaging data.

This work explores the predictive power of novel TBI biomarkers computationally extracted from hospital admission CT scans. TBI lesion volumes are automatically extracted from the scans; then, lesion statistics are derived to inform the prediction of TBI outcome. We compare the discriminate power of our proposed features to the Marshall score features independently and when paired with clinical TBI biomarkers.

In particular, we make the following contributions:

- **Novel machine-learning driven imaging biomarkers.** We extract interpretable measurements for TBI lesion quantification;
- **Human-level performance on unfavourable outcome prediction.** We reach comparable, if not superior, performance to manually extracted CT biomarkers when predicting unfavourable outcome, both in isolation and when paired with clinical TBI biomarkers;
- **Imaging biomarker relevance.** We show that features relating to extra-axial haemorrhage in the frontal lobe are important for the prediction of outcome, confirming previous clinical findings in a data-driven manner.

## 2 Related work

The prediction of TBI of outcome has primarily been tackled using clinical features. Jiang et al. [12] and Majdan et al. [15] used clinical features such as age and motor score to predict the patient outcome with a regression model. Pasipanodya et al. [22] focused on predictions for different patient subgroups. Huie et al. [10] provide an extensive review of clinical biomarkers for the prediction of unfavourable outcome in TBI. More recently, Bruschetta et al. [3] and Matsuo et al. [18] investigated using the same predictors with machine learning models, such as support vector machines and neural networks. Researchers also used more complex features to predict TBI outcome, such as electroencephalograms [20] and magnetic resonance imaging (MRI) [8].

A stream of research within TBI focuses on features extracted from CT. Recent work adopted neural networks for TBI lesion segmentation and midline shift quantification [19,11].

Similarly to our work, Plassard et al. [24] and Chaganti et al. [5] used multi-atlas labelling to extract radiomics from different brain regions and predict a variety of TBI end-points, yet excluding unfavourable outcome. Pease et al. [23] trained a convolutional neural network using CT scans and clinical TBI biomarkers to predict the TBI outcome and achieved comparable performance to IMPACT. Unlike the biomarkers we designed, the authors extracted deep learning features, which are not interpretable by humans. In parallel to our work, Yao et al. [28] trained a neural network to segment epidural haematomas and used the resulting segmentation volumes to predict patient mortality. Our work differs because we focused on the more challenging and clinically relevant problem of predicting TBI unfavourable outcome. TBI outcomes are measured by the patient state scale Glasgow Outcome Scale Extended (GOS-E), where a score of 4 or below defines an unfavourable outcome.

### 3 Methods

**Study design** We analysed data from the observational studies The Collaborative REsearch on ACute Traumatic Brain Injury in intensive Care Medicine in Europe (CREACTIVE, NCT02004080 ), and BIOmarkers of AXonal injury after TBI (BIO-AX-TBI, NCT03534154 [7]), the latter being a follow-up study to CREATIVe. From these observational studies, sub-studies collecting CT scans recruited 1986 patients admitted to intensive care with a diagnosis of TBI between 2013 and 2019 from 21 European hospitals. CREATIVe recruited patients admitted to a hospital due to suspected TBI, whereas for BIO-AX-TBI, the criterion was moderate-severe TBI as defined in [16]. The studies did not define a protocol for acquisition or scanner type. Hence, the CT scans collected were a heterogeneous dataset akin to clinical scenarios.

**Data cleaning** We discarded patients for whom trauma time, hospital admission and discharge, admission CT scan, and the patient outcome measure GOS-E were not recorded. In addition, we discarded patients who had surgery before the scan, whose scan could not be automatically aligned to an atlas, and for whom one of the clinical TBI biomarkers - age, Glasgow Coma Scale motor score and pupil responsiveness, hypoxia and hypotension - or Marshall score was missing. A flowchart of the patient selection can be found in the appendix Figure A.I. Of the remaining 646 patients, 389 (60.2%) had an unfavourable outcome, of which 225 (34.8%) died. The median age is 56.6, and 74.9% of the patients are male. Out of the 21 centres, we selected three centres with 59 patients as an independent replication holdout test set, and we used the remaining 18 centres with 587 patients as a training set.

**Feature extraction** All CT scans were standardised via automatic atlas registration to obtain rigid alignment to MNI space. We then used the deep-learning

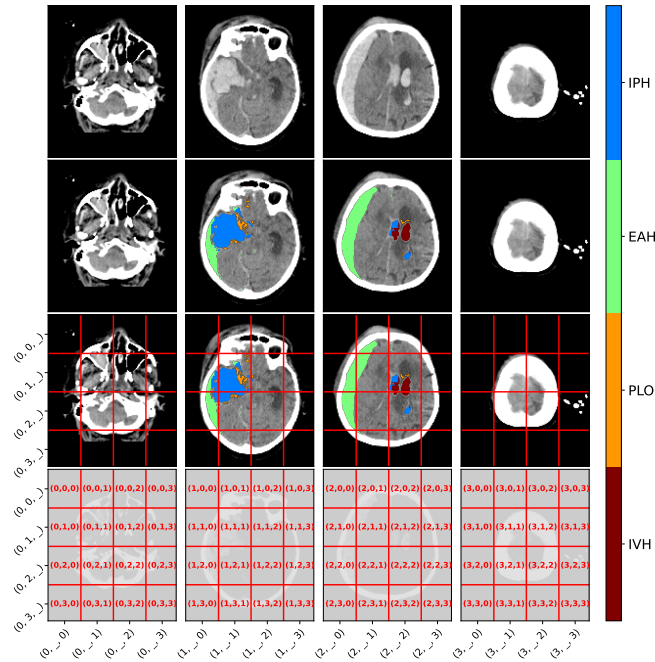


Fig. 1: Example of feature extraction. A CT scan (1st row) is fed through a segmentation neural network, outputting lesion volumes (2nd row + legend). We either aggregated the lesion volumes globally or calculated the volumes over each cuboidal region to provide localised measures (3rd and 4th rows).

lesion segmentation method, BLAST-CT, described in [19]. The method produces a voxel-wise classification of intraparenchymal haemorrhages (IPH), extra-axial haemorrhages (EAH), perilesional oedemas (PLO) and intraventricular haemorrhages (IVH), and reports respective Dice similarity coefficients of 65.2%, 55.3%, 44.8% and 47.3% for lesions above 1mL.

From the 3D segmentations, we extracted two types of statistical features, *global* and *local* lesion biomarkers. We calculated all connected components for each lesion type and discarded any connected region of 0.02mL or less to remove noise. We determined this threshold by visual inspection of the scans. From the separate lesions<sup>5</sup>, we extracted the following *global* lesion biomarkers: the total number of lesions, the median lesion volume, the 25th and 75th lesion volume percentiles, and the mean lesion volume. Next, we partitioned the registered 3D scans four ways for each axis, creating  $4^3 = 64$  equidimensional cuboids. We chose four partitions to balance feature granularity and human interpretability. We extracted *local* lesion biomarkers for each lesion type by calculating the total lesion volume in each cuboidal region. Figure 1 shows an example of brain partitioning and the corresponding indexing.

<sup>5</sup> Each lesion is a connected component.

Table 1: Models performance when predicting unfavourable outcome. Marshall score is described in [17], *global* features refer to disjoint lesion statistics, and *local* features refer to lesion volume per cuboid (see **Feature extraction**). The clinical biomarkers are age, motor and pupil responsiveness, hypoxia and hypotension.

Features	Cross-validation	Hold-out set		
	AUROC	AUROC	Precision	Recall
Marshall score	76.7 +/- 7.7	73.2	69.6	61.5
<i>global</i> lesion features	72.4 +/- 5.8	80.9	70.0	<b>80.8</b>
<i>local</i> lesion features	77.2 +/- 5.5	83.3	65.6	<b>80.8</b>
<b><i>local</i> + <i>global</i> lesion features</b>	<b>77.2 +/- 6.5</b>	<b>84.0</b>	<b>74.1</b>	76.9

Features	Cross-validation	Hold-out set		
	AUROC	AUROC	Precision	Recall
Marshall score + clinical biomarkers	81.6 +/- 3.9	84.7	69.2	69.2
<i>global</i> + clinical biomarkers	82.1 +/- 4.3	87.5	80.8	<b>80.8</b>
<b><i>local</i> + clinical biomarkers</b>	<b>83.0 +/- 4.4</b>	<b>87.7</b>	<b>84.0</b>	<b>80.8</b>
<i>local</i> + <i>global</i> + clinical biomarkers	81.1 +/- 4.8	87.2	77.8	<b>80.8</b>

**Modelling and performance evaluation** We used a Random Forest Classifier with 300 estimators to predict a patient’s favourable or unfavourable outcome. An unfavourable outcome, defined as a GOS-E score of 4 or below, is a typical target in TBI outcome prediction. We compared eight predictive models based on different sets of features. The first four models use imaging features alone: 1) the Marshall score; 2) *global* lesion biomarkers; 3) *local* lesion biomarkers; 4) *global* and *local* lesion biomarkers. The second set of models uses the imaging features above, together with the clinical TBI biomarkers (age, motor and pupil responsiveness, hypoxia and hypotension). Note that the clinical TBI biomarkers and Marshall score are the same as those used in the state-of-the-art IMPACT-CT model [26].

We evaluated model performance using the area under the receiver-operator curve (AUROC), precision, recall, and true positive rate at a 10% false positive rate. In addition to evaluating performance on the holdout test set, we also measured cross-validation performance on the training set. We calculated statistical significance through a permutation test on the holdout set’s metrics and the statistical relevance of each feature using the average Gini importance. In addition, a cross-validation per clinical centre can be found in appendix Table A.II.

## 4 Results

***Local* and *global* lesion biomarkers performed similarly or better than Marshall score.** Using *local* and *global* lesion biomarkers achieved a cross-validation AUROC of  $76.7 \pm 7.7\%$  compared to  $77.0 \pm 6.6\%$  when using the Marshall score (Table 1 top). On the holdout set, the improvement of AUROC was 10.8%, from 73.2% using the Marshall score to 84.0% using *local* and *global* lesion biomarkers. Similarly, the precision improved by 4.5% and the recall by

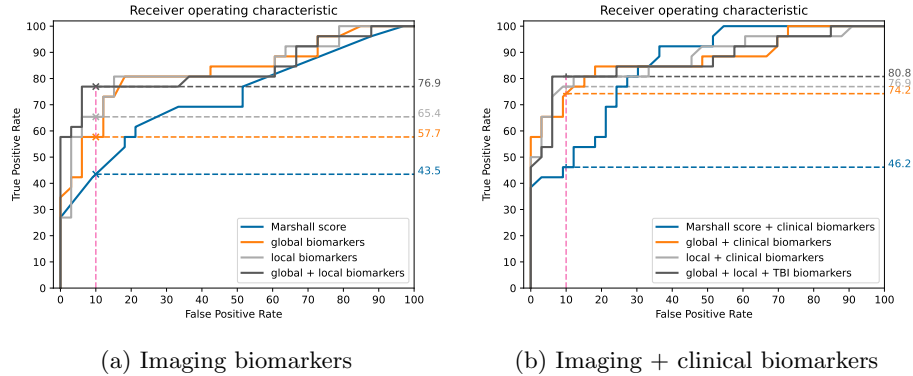
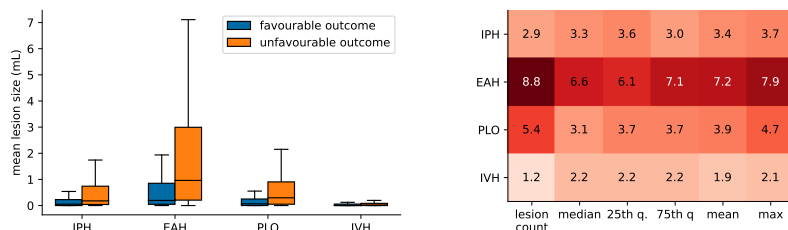


Fig. 2: ROC curves of models predicting unfavourable outcome. The dashed lines indicate the true positive rate for a fixed false positive rate of 10%. *Global* and *local* biomarkers always produce a better or equivalent performance than the Marshall score.

15.4%. For a false positive rate of 10%, the volumetric features yielded a true positive rate of 73.1% compared to 43.5% for the Marshall score (Figure 2a). Testing the statistical significance of these improvements, we found that improvement in AUROC in the holdout set was statistically significant (one-way p-value  $< 0.05$ ), whereas all other metrics on the holdout set were statistically comparable.

**Local lesion biomarkers and clinical features performed similarly or better than features used in IMPACT-CT.** When adding the clinical TBI features to the *local* lesion biomarkers, the AUROC was  $83.0 \pm 4.4\%$  in cross-validation and 87.7% on the holdout set, compared to  $81.6 \pm 3.9\%$  and 84.7%, respectively, for clinical TBI and Marshall score biomarkers ((Table 1) bottom). Similarly, the holdout set’s precision, recall and true-positive rate improved by 3%, 14.8% and 34.6%, respectively (Figure 2b). When tested, the improvement in true positive rate was significant (one-way p-value  $< 0.05$ ), whereas the remaining holdout set metrics for the two experiments were statistically comparable.

**Extra-axial haemorrhage was the most important feature.** When considering feature importance when using *global* lesion biomarkers (Figure 3b), EAH was the statistical feature with the highest importance score, and IVH was the feature with the lowest. The lesion count and maximum lesion size were the most important factors. As per the *global* lesion biomarkers, EAH was the statistical feature with the highest importance scores when using *local* lesion biomarkers to predict unfavourable outcome (Figure 4). In addition IVH in the second-bottom transverse plane, second-bottom coronal plane  $((1, 2, -))$ , EAH in the second-front coronal plane  $((1, 1, -))$  and  $(2, 1, -)$  were important.



(a) Mean lesion size stratified per lesion type and outcome. Unfavourable outcome can be qualified by larger average IPH, EAH and PLO.

(b) Feature importance of *global* lesion biomarkers when predicting unfavourable outcome. The numbers and colour intensity refer to the Gini importance (rescaled by  $10^{-2}$ ).

Fig. 3: *Global* feature statistics and their importance for outcome prediction.

## 5 Discussion

We show that the predictive power of the automatically extracted imaging features is comparable to or superior to that of the Marshall score. Furthermore, the automatically extracted features in conjunction with clinical TBI biomarkers perform at least as well as the features used in IMPACT-CT (Marshall score and clinical TBI biomarkers). The advantage of our approach is that, unlike the Marshall score, automatically extracted imaging features do not require a radiologist to manually review the scan, allowing faster patient care, and reducing workload. Our method generalises well across different scanner types and acquisition protocols as shown from the consistent results on the training set cross-validation and the independent hold-out set. Although other approaches, such as advanced fluid biomarker or magnetic resonance imaging, have also shown promise in improving outcome prediction [8], the described method has the advantage of using data which is currently collected routinely, obviating the need for revised clinical investigation protocols.

The interpretability of the lesion features is an important step in discovering data-driven prognostic biomarkers, contributing to the clinical understanding of TBI. Reinforcing previous results [1], we found that frontal EAH is an important indicator of poor TBI outcomes.

Although no ground truth segmentation was available, the clinicians (LL and DS) qualitatively reviewed a subset of the automatic segmentations. We concluded that the segmentation model tended to produce some errors, such as partially mislabelling lesion types and under-estimating their size. An example can be seen in the appendix Figure A.II. Unfortunately, given the absence of ground truth, we could not quantify the extent of these issues. Nevertheless, there is strong evidence for the soundness of the model through both the predictive performance and the feature importance maps. For example, the feature importance of IVH is high where the ventricles occur.

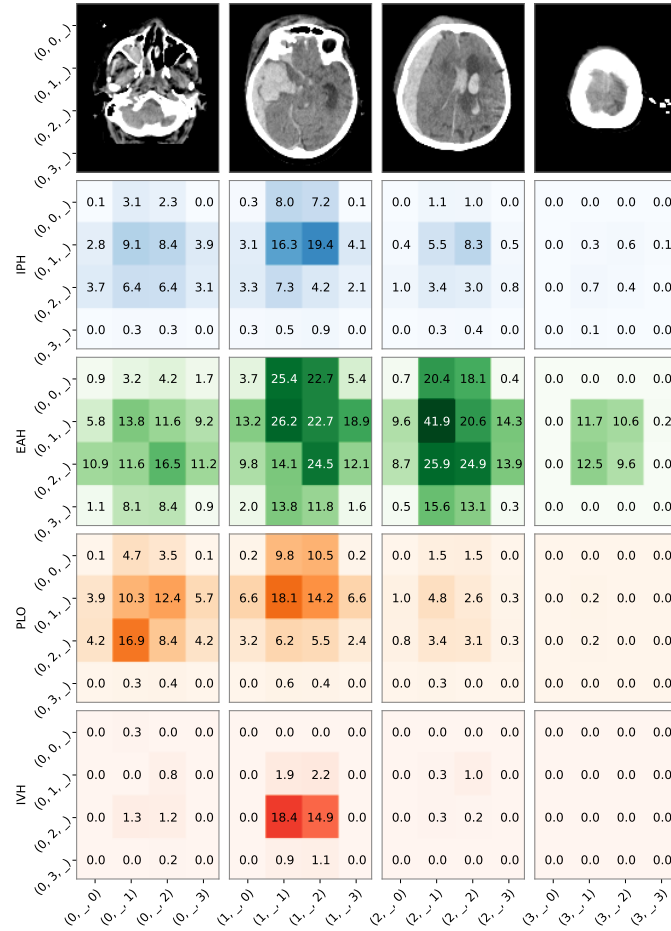


Fig.4: Feature importance of local lesion biomarkers when predicting unfavourable outcome. Each row represents the feature importance of different lesion types, and each column represents a transversal slice, similarly to Figure 1. For each transverse slice, each row represents a coronal slice, whereas each column represents a sagittal slice. The numbers and colour intensity refer to the Gini importance (rescaled by  $10^{-3}$ ).

In summary, our results show that automatically extracted CT features achieve human-level performance in predicting the outcome of TBI patients without requiring manual appraisal of the scan. In future work, the interpretability of the machine learning features may allow for a deeper clinical understanding of TBI, a notably complex condition. However, further work is needed to improve the robustness of lesion segmentation models and their evaluation on new unlabelled datasets.



**Acknowledgements** MR is supported by UK Research and Innovation [UKRI Centre for Doctoral Training in AI for Healthcare grant number EP/S023283/1]. LL is supported by NIHR, Academy of Medical Sciences. NSNG is funded by a National Institute for Health Research (NIHR) Academic Clinical Lectureship and acknowledges infrastructure support from the UK Dementia Research Institute (DRI) Care Research and Technology Centre at Imperial College London and the NIHR Imperial Biomedical Research Centre (BRC). The BIOAX-TBI data was collected thanks to the ERA-NET NEURON Cofund (MR/R004528/1), part of the European Research Projects on External Insults to the Nervous System call, within the Horizon 2020 funding framework. The CREACTIVE project has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under Grant Agreement number 602714.

## References

1. Atzema, C., Mower, W.R., Hoffman, J.R., Holmes, J.F., Killian, A.J., Wolfson, A.B., Group, N.E.X.R.U.S.N.I., et al.: Prevalence and prognosis of traumatic intraventricular hemorrhage in patients with blunt head trauma. *Journal of Trauma and Acute Care Surgery* **60** (2006)
2. Brown, A.W., Pretz, C.R., Bell, K.R., Hammond, F.M., Arciniegas, D.B., Bodien, Y.G., Dams-O'Connor, K., Giacino, J.T., Hart, T., Johnson-Greene, D., et al.: Predictive utility of an adapted marshall head ct classification scheme after traumatic brain injury. *Brain injury* **33** (2019)
3. Bruschetta, R., Tartarisco, G., Lucca, L.F., Leto, E., Ursino, M., Tonin, P., Pioggia, G., Cerasa, A.: Predicting outcome of traumatic brain injury: Is machine learning the best way? *Biomedicines* **10** (2022)
4. Carter, E., Coles, J.P.: Imaging in the diagnosis and prognosis of traumatic brain injury. *Expert opinion on medical diagnostics* **6** (2012)
5. Chaganti, S., Plassard, A.J., Wilson, L., Smith, M.A., Patel, M.B., Landman, B.A.: A bayesian framework for early risk prediction in traumatic brain injury. In: *Medical Imaging 2016: Image Processing*. vol. 9784. International Society for Optics and Photonics (2016)
6. Cireşan, D.C., Giusti, A., Gambardella, L.M., Schmidhuber, J.: Mitosis detection in breast cancer histology images with deep neural networks. In: *International conference on medical image computing and computer-assisted intervention*. pp. 411–418. Springer (2013)
7. Graham, N.S.N., Zimmerman, K.A., Bertolini, G., Magnoni, S., Oddo, M., Zetterberg, H., Moro, F., Novelli, D., Heslegrave, A., Chierigato, A., et al.: Multicentre longitudinal study of fluid and neuroimaging biomarkers of axonal injury after traumatic brain injury: the bio-ax-tbi study protocol. *BMJ open* **10** (2020)
8. Graham, N.S., Zimmerman, K.A., Moro, F., Heslegrave, A., Maillard, S.A., Bernini, A., Miroz, J.P., Donat, C.K., Lopez, M.Y., Bourke, N., et al.: Axonal marker neurofilament light predicts long-term outcomes and progressive neurodegeneration after traumatic brain injury. *Science Translational Medicine* **13** (2021)
9. Helmrich, I.R., van Klaveren, D., Dijkland, S.A., Lingsma, H.F., Polinder, S., Wilson, L., von Steinbuechel, N., van der Naalt, J., Maas, A.I., Steyerberg, E.W.: Development of prognostic models for health-related quality of life following traumatic brain injury. *Quality of life research* (2021)
10. Huie, J.R., Almeida, C.A., Ferguson, A.R.: Neurotrauma as a big-data problem. *Current opinion in neurology* **31** (2018)

11. Jain, S., Vyvere, T.V., Terzopoulos, V., Sima, D.M., Roura, E., Maas, A., Wilms, G., Verheyden, J.: Automatic quantification of computed tomography features in acute traumatic brain injury. *Journal of neurotrauma* **36** (2019)
12. Jiang, J.Y., Gao, G.Y., Li, W.P., Yu, M.K., Zhu, C.: Early indicators of prognosis in 846 cases of severe traumatic brain injury. *Journal of neurotrauma* **19** (2002)
13. Kalanuria, A.A., Geocadin, R.G.: Early prognostication in acute brain damage: where is the evidence? *Current opinion in critical care* **19** (2013)
14. Kim, J.J., Gean, A.D.: Imaging for the diagnosis and management of traumatic brain injury. *Neurotherapeutics* **8** (2011)
15. Majdan, M., Brazinova, A., Rusnak, M., Leitgeb, J.: Outcome prediction after traumatic brain injury: Comparison of the performance of routinely used severity scores and multivariable prognostic models. *Journal of Neurosciences in Rural Practice* **8** (2017)
16. Malec, J.F., Brown, A.W., Leibson, C.L., Flaada, J.T., Mandrekar, J.N., Diehl, N.N., Perkins, P.K.: The mayo classification system for traumatic brain injury severity. *Journal of neurotrauma* **24** (2007)
17. Marshall, L.F., Marshall, S.B., Klauber, M.R., Van Berkum Clark, M., Eisenberg, H., Jane, J.A., Luerssen, T., Marmarou, A., Foulkes, M.: The diagnosis of head injury requires a classification based on computed axial tomography. *J neurotrauma* **9** (1992)
18. Matsuo, K., Aihara, H., Nakai, T., Morishita, A., Tohma, Y., Kohmura, E.: Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury. *Journal of neurotrauma* **37** (2020)
19. Monteiro, M., Newcombe, V.F., Mathieu, F., Adatia, K., Kamnitsas, K., Ferrante, E., Das, T., Whitehouse, D., Rueckert, D., Menon, D.K., et al.: Multiclass semantic segmentation and quantification of traumatic brain injury lesions on head ct using deep learning: an algorithm development and multicentre validation study. *The Lancet Digital Health* **2** (2020)
20. Noor, N.S.E.M., Ibrahim, H.: Machine learning algorithms and quantitative electroencephalography predictors for outcome prediction in traumatic brain injury: A systematic review. *IEEE Access* **8** (2020)
21. Parsonage, M.: Traumatic brain injury and offending. *Centre for Mental health* (2016)
22. Pasipanodya, E.C., Teranishi, R., Dirlikov, B., Duong, T., Huie, H.: Characterizing profiles of tbi severity: predictors of functional outcomes and well-being. *The Journal of Head Trauma Rehabilitation* (2022)
23. Pease, M., Arefan, D., Barber, J., Yuh, E., Puccio, A., Hochberger, K., Nwachuku, E., Roy, S., Casillo, S., Temkin, N., Okonkwo, D.O., Wu, S., , Badjatia, N., Bodien, Y., Duhaime, A.C., Feeser, V.R., Ferguson, A.R., Foreman, B., Gardner, R., Gopinath, S., Keene, C.D., Madden, C., McCrea, M., Mukherjee, P., Ngwenya, L.B., Schnyer, D., Taylor, S., Yue, J.K.: Outcome prediction in patients with severe traumatic brain injury using deep learning from head ct scans. *Radiology* **0** (2022)
24. Plassard, A.J., Kelly, P.D., Asman, A.J., Kang, H., Patel, M.B., Landman, B.A.: Revealing latent value of clinically acquired cts of traumatic brain injury through multi-atlas segmentation in a retrospective study of 1,003 with external cross-validation. In: *Medical Imaging 2015: Image Processing*. vol. 9413. International Society for Optics and Photonics (2015)
25. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer (2015)

26. Steyerberg, E.W., Mushkudiani, N., Perel, P., Butcher, I., Lu, J., McHugh, G.S., Murray, G.D., Marmarou, A., Roberts, I., Habbema, J.D.F., et al.: Predicting outcome after traumatic brain injury: development and international validation of prognostic scores based on admission characteristics. *PLoS Med* **5** (2008)
27. Tagliaferri, F., Compagnone, C., Korsic, M., Servadei, F., Kraus, J.: A systematic review of brain injury epidemiology in europe. *Acta neurochirurgica* **148** (2006)
28. Yao, H., Williamson, C., Gryak, J., Najarian, K.: Automated hematoma segmentation and outcome prediction for patients with traumatic brain injury. *Artificial Intelligence in Medicine* **107** (2020)

## Appendix

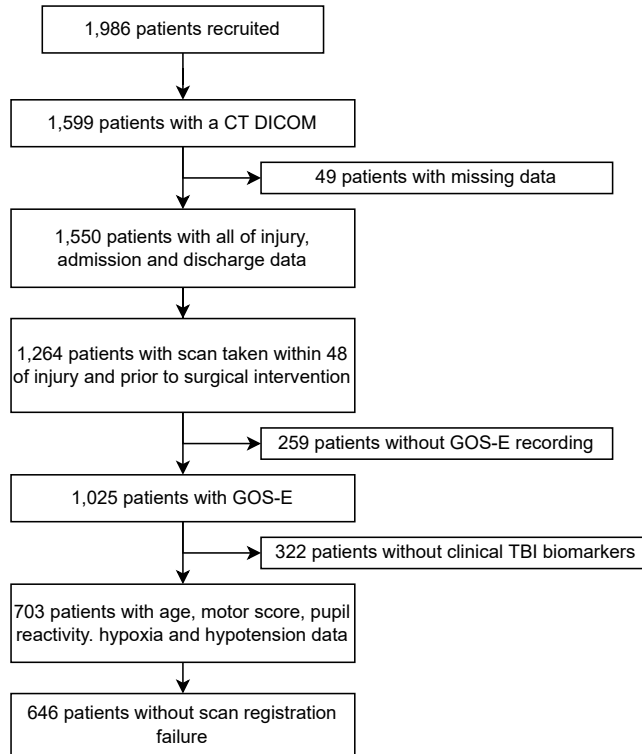


Fig. A.I: Patient selection flow-chart

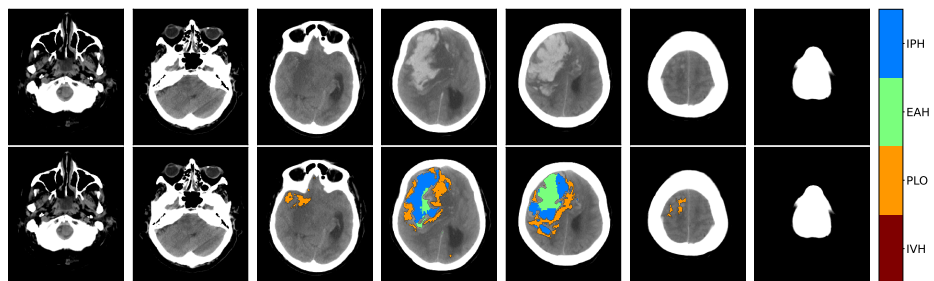


Fig. A.II: Extreme example of mis-segmented CT scan. The lesion labels are intraparenchymal (IPH), extra-axial (EAH) and intraventricular (IVH) haemorrhages, and perilesional oedemas (PLO). The extra-axial label on fourth and fifth slices is erroneous; the opacity on the left side of the brain in the third slice should be labelled as oedema.

