# ICH-DIANet: A Modality-Disentangled and Attention-Expert Interaction Network for Brain Injury Assessment in Intracerebral Hemorrhage

Chengqing Liu[1*], Zicheng Xiong[2*], Yuxuan Rao[3], Guohui Lu[4], Meihua Li[4], Hoileong Lee[5], Jingzhen Niu[6], Shengbo Chen[7✉], and Xujun Shu[8✉]

[1] School of Mathematics and Computer Sciences, Nanchang University, Nanchang, China
[2] School of Computer Science, Nanjing University, Nanjing, China
[3] Ryan Companies US, Inc., Minneapolis, MN, USA
[4] Department of Neurosurgery, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, China
[5] Faculty of Electronic Engineering and Technology, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia
[6] School of Computer and Information Engineering, Henan University, Kaifeng, China
[7] School of Software, Nanchang University, Nanchang, China
[8] Department of Neurosurgery, Jinling Hospital, Affiliated Hospital of Medical School, Nanjing University, Nanjing, China

**Abstract.** Intracerebral hemorrhage (ICH) is a critical condition where rapid, objective brain injury assessment is vital for treatment planning and prognosis. In clinical practice, although the Glasgow Coma Scale (GCS) evaluates brain injury severity, it relies on indirect assessments by eye-opening, verbal, and motor responses, making its accuracy susceptible to the patient's clinical status. Furthermore, this manual evaluation is hindered by significant subjectivity and limited efficiency. Consequently, automated multimodal evaluation via deep learning has become a research priority. However, existing methods struggle with feature redundancy and distributional heterogeneity across modalities, often failing to capture complex inter-patient variations via static fusion strategies. To address this problem, we propose ICH-DIANet, a novel framework integrating 3D Computerized Tomography (CT) images and clinical text for automated brain injury grading. The framework's Modality Disentanglement Module (MDM) projects features into both a modal-consistent subspace and a modality-specific subspace via orthogonality constraints, effectively eliminating redundancy while preserving complementary information. Additionally, a proposed Attention-Based Mixture-of-Experts Module (AMoE) employs a dynamic gating mechanism to adaptively weight expert outputs based on individual patient characteristics, while integrating an attention mechanism to facilitate deep cross-modal semantic alignment. Extensive experiments on both private and public datasets demonstrate that ICH-DIANet significantly outperforms the state-of-the-art methods in accuracy and robustness.

* Equal contribution, ✉ Corresponding Author.

**Keywords:** ICH · Multi-modality · Disentanglement · Attention-Based Experts.

## 1 Introduction

Intracerebral hemorrhage (ICH) represents a major global public health challenge with high mortality rates [1]. Typically involving seizures and neurological deficits, ICH prognosis is closely linked to timely acute-phase intervention [2, 3]. Rapid and objective assessment of brain injury severity is vital for clinicians to formulate precise treatment strategies, directly impacting patient outcomes. The Glasgow Coma Scale (GCS), a cornerstone clinical tool for quantifying consciousness impairment, ranges from 3 to 15 based on assessments of eye-opening, verbal, and motor responses [4]. Clinically, brain injury is categorized as severe (3–8), moderate (9–12), or mild (13–15) based on GCS [5]. However, this indirect evaluation based on external behavioral feedback is highly susceptible to the patient's clinical status. Under consciousness suppression or sedation, the scales reliance on subjective responses often yields underestimated scores, failing to objectively reflect the actual brain injury severity. Furthermore, GCS assessment involves significant inter-rater variability and high human-resource dependency, where evaluations may fluctuate across different observers or even for the same observer over time.

To address the limitations of GCS in evaluating brain injury, advancing Artificial Intelligence (AI) offers standardized, automated, and more direct clinical diagnostic solutions. Although research has transitioned from traditional machine learning to three dimension (3D) deep learning to address early deficiencies in semantic modeling and spatial reliability [6, 7, 8, 9, 10, 11, 12], unimodal data remains inherently limited in reflecting comprehensive patient information. Consequently, some studies attempt multimodal fusion, which only adopts direct input of laboratory results or scale scores, failing to effectively model the latent correlations between imaging and clinical features [10, 13]. To further exploit inter-modal interaction potential, explicit fusion mechanisms have been introduced to mitigate cross-modal heterogeneity through deeper alignment operations [14, 15, 16, 17]. These methods typically employ fixed fusion strategies, overlooking inter-patient pathological heterogeneity. In fact, the diagnostic contributions of imaging features versus clinical text vary significantly across cases. Moreover, unstripped intra-modal noise and redundancy often impede effective cross-modal semantic alignment.

To deal with these aforementioned problems, in this paper, we propose ICH-DIANet, a modality-disentangled and attention-expert interaction network for brain injury assessment by integrating 3D CT images with clinical reports. First, to address the prevalent issues of information redundancy and noise interference in multimodal data, we design a Modal Disentanglement Module (MDM). By employing an orthogonality constraint mechanism, the MDM decouples raw features into modal-shared and modality-specific features, effectively eliminating redundancy while maximizing the preservation of complementary key features.

Second, since inherent inter-individual variability often renders a single fusion strategy inadequate for extracting critical features, we introduce an Attention-Based Mixture-of-Experts Module (AMoE). This allows the module to adaptively focus on the unique feature distributions of different patients, fully exploiting the distinct information within each modality to achieve more precise predictions. Extensive experiments on private and public datasets consistently validate the effectiveness and superiority of our approach.

## 2 Methods

### 2.1 Model Architecture

The overall architecture of the proposed ICH-DIANet is illustrated in Fig. 1. The framework consists of four primary components: a feature extraction module, a modality disentanglement module, an attention-based mixture-of-experts module, and a classifier. In the feature extraction module, a pre-trained 3D ResNet-10 [18] and BioClinicalBERT [19] serve as the vision and text encoders to extract modality representations $f_v$ and $f_t$, respectively, where $v$ and $t$ denote the visual and textual modalities. The classifier comprises two fully connected layers.

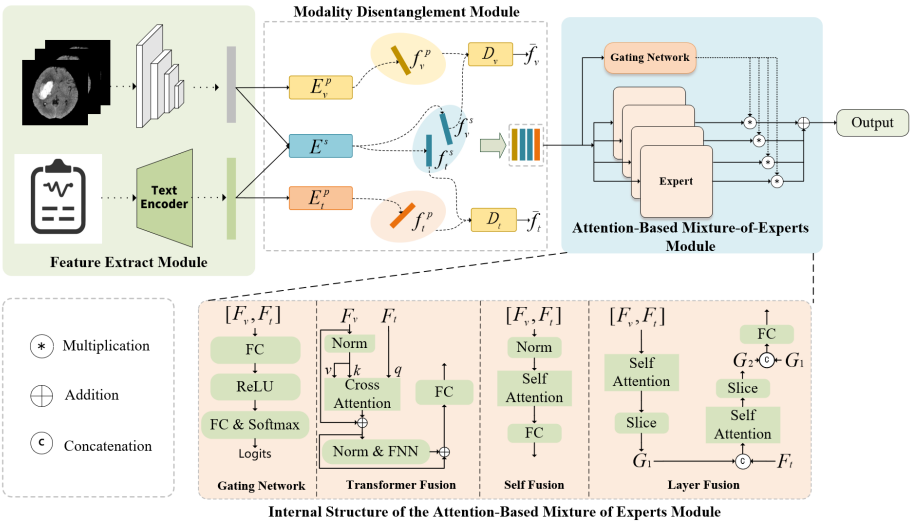


Fig. 1: Architecture of Model.

### 2.2 Modality Disentanglement Module

To decouple intra-modal shared and specific features, we design the Modality Disentanglement Module, which projects input features from each modality $m \in \{v, t\}$ into distinct latent subspaces. Specifically, the shared feature $f_m^s$ captures

cross-modal consistency, while the specific feature $f_m^p$ extracts modality-specific heterogeneity. The decomposition is formalized as:

$$f_m^s = E^s(f_m), \tag{1}$$

$$f_m^p = E_m^p(f_m), \tag{2}$$

where $E$ denotes a two-layer fully-connected encoder, and both $f_m^s$ and $f_m^p$ are $d$-dimensional feature vectors. To prevent semantic distortion during projection, we introduce a self-reconstruction mechanism. The reconstruction encoder $D_m$ encodes the decoupled features to generate reconstructed features $\bar{f}_m$, thereby preserving the integrity of the original input information. The reconstruction process is defined as:

$$\bar{f}_m = D_m(z_m), \tag{3}$$

where $z_m = f_m^s + f_m^p$ and $D_m$ similarly consists of two fully-connected layers.

### 2.3 Attention-Based Mixture-of-Experts Module

Our proposed Attention-Based Mixture-of-Experts Module consists of a gating network and four parallel expert networks. The gating network dynamically assigns weights to experts based on input features, yielding a robust fused representation via weighted summation. Diverging from traditional MoE architectures, our design enables each expert to specialize in a distinct feature subset, ensuring collective contribution to the final output.

**Gating Network.** The gating network generates the expert weight distribution $w$ using $f_m^s$ and $f_m^p$ from the MDM as inputs. The process is formalized as follows:

$$logits = W_1 \times (Relu(W_2 \times [f_v^p, f_v^s, f_t^p, f_t^s])), \tag{4}$$

$$w = softmax(logits), \tag{5}$$

where $W_1$ and $W_2$ represent learnable weight matrices.

**Transformer Fusion.** This expert is designed to capture long-range cross-modal dependencies between visual features $Fv$ and text features $Ft$. By leveraging a cross-attention mechanism, the module enables text semantics to actively align with and select the most relevant visual regions. This process facilitates the generation of text-guided visual representations with deep semantic-level fusion. The specific implementation process is as follows:

$$F_v = Norm([f_v^s, f_v^p]), F_t = [f_t^s, f_t^p], \tag{6}$$

$$Q_j = W_j^Q \times F_t, K_j = W_j^K \times F_v, V_j = W_j^V \times F_v, \tag{7}$$

$$Head_j = Attention(Q_j, K_j, V_j) = softmax\left(\frac{Q_j K_j^T}{\sqrt{d}}\right) V_j, \tag{8}$$

$$F_{v \leftarrow t} = concat(Head_1, Head_2, ..., Head_n), \tag{9}$$

$$F_{add} = F_{v \leftarrow t} + F_v, \tag{10}$$

$$F_{trans} = W_{trans}(F_{add} + FFN(Norm(F_{add}))), \tag{11}$$

where $Q_j$, $K_j$, and $V_j$ denote the query, key, and value vectors in the $j$-th attention head, respectively, and $d$ represents the vector dimension. $W_{trans}$, $W_j^Q$, $W_j^K$, and $W_j^V$ are all learnable weight matrices.

**Self Fusion.** This expert focuses on learning the internal structure of modal features. By emphasizing intra-modal dependencies, it ensures the model preserves modality-specific information while enhancing representation capacity. The process is formally described as follows:

$$F_{self} = W_{self} \times SA(Norm([F_v, F_t])), \tag{12}$$

where $SA(\cdot)$ denotes the self-attention mechanism and $W_{self}$ represents a learnable weight matrix.

**Layer Fusion.** Inspired by MOME [20], this expert is designed to facilitate hierarchical interaction between visual features $F_v$ and text features $F_t$. This mechanism achieves indirect coupling and bidirectional alignment through progressive attentional interactions, effectively avoiding direct inter-modal information exchange. Specifically, the process first utilizes text features as reference information to guide the internal information interaction within $F_v$:

$$G_1 = SA([F_v, F_t])[:2, :], \tag{13}$$

where $G_1$ represents the visual feature slice integrated with text-based reference information. Subsequently, $G_1$ provides semantic guidance back to the text features $F_t$:

$$G_2 = SA([F_t, G_1])[:2, :], \tag{14}$$

where $G_2$ denotes the text feature slice integrated with a visual-based reference information. Finally, $G_1$ and $G_2$ are fused via a fully-connected ($FC$) layer.

**Drop Fusion.** This expert aims to strengthen feature extraction for a single dominant modality. The process is formulated as:

$$F_{Drop} = F_v. \tag{15}$$

### 2.4 Loss Function

To promote inter-modal consistency, we introduce the modal sharing loss $L_{gather}$ to maximize the similarity between $f_v^s$ and $f_t^s$. We employ cosine similarity to

measure feature correspondence and minimize their distance within the modality-consistent subspace:

$$L_{gather} = -\frac{1}{n}\sum_{i=1}^{n}\frac{f_v^s(f_t^s)^T}{\|f_v^s\|\|f_t^s\|}. \tag{16}$$

Where $T$ denotes the transposition and $n$ denotes the batch size. To ensure effective feature disentanglement and minimize redundancy, we introduce the intra-modal decoupling loss $L_{diff}$ to maximize the difference between shared and specific features. This loss guides feature separation by increasing the representational distance in the subspace:

$$L_{diff} = \frac{1}{n}\sum_{i=1}^{n}\|f_v^s(f_v^p)^T\|_F + \|f_t^s(f_t^p)^T\|_F, \tag{17}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Furthermore, to mitigate representation distortion, we employ Mean Squared Error ($MSE$) loss to quantify the information loss between $\bar{f}_m$ and $f_m$:

$$L_{rec} = MSE(\bar{f}_v, f_v) + MSE(\bar{f}_t, f_t). \tag{18}$$

Finally, weighted binary cross-entropy loss serves as the primary task loss $L_{task}$. The total objective function $L_{total}$ for model optimization is defined as:

$$L_{total} = L_{task} + L_{gather} + L_{diff} + L_{rec}. \tag{19}$$

## 3 Experiments

### 3.1 Dataset and Experimental Details

**Dataset.** This study evaluates the proposed method on one private and one public dataset. The private dataset comprises 292 non-traumatic ICH cases. We also incorporate a public dataset [21], where we filter 110 cases for non-traumatic ICH. Both datasets consist of 3D CT images paired with clinical reports, e.g., gender, age, onset time, and so on. The private dataset includes 122 mild, 108 moderate, and 62 severe cases, while the public dataset contains 46 mild, 41 moderate, and 23 severe cases. For CT preprocessing, we first separate the skull using a pre-trained nn-Unet [22] to minimize background interference. All images are subsequently resampled to a uniform resolution of $64 \times 256 \times 256$, followed by normalization and zero-padding to satisfy the model input requirements.

**Experimental details.** Experiments are conducted on a single NVIDIA A40 GPU for 300 epochs with a batch size of 8. We utilize the Adam optimizer with an initial learning rate of 1e-4, adjusted by a cosine annealing scheduler and a weight decay of 5e-4. Performance is evaluated via five-fold cross-validation using Accuracy, Precision, AUC, and F1-score. To mitigate overfitting, training data are augmented with random rotations within $[-20°, 20°]$ across various axes and random translations within [-10, 10] pixels.

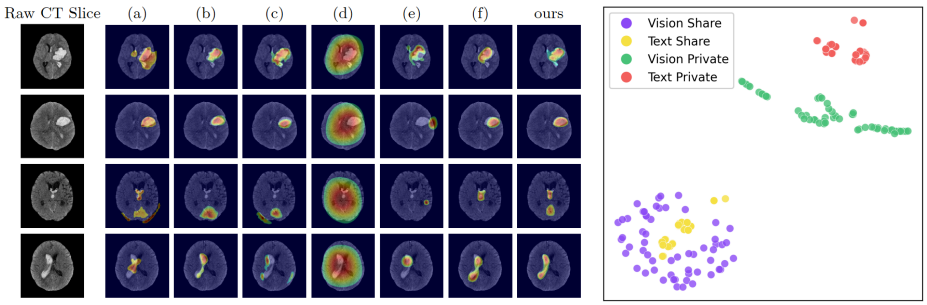


Fig. 2: Model visualization results. (i) Comparison of Grad-CAM heatmaps from baseline methods (a)-(f) (ResNet, DL-Based Method, DAFT, IRENE, CFDL, and ICHPro) against our approach, where color intensity reflects the model's focus on lesion regions. (ii) t-SNE visualization of the 2D feature distribution processed by the MDM.

Table 1: Comparative performance on private and public datasets.

| Method | Private Dataset | | | | Public Dataset | | | |
|---|---|---|---|---|---|---|---|---|
| | Acc(%) | Pre(%) | AUC | F1 | Acc(%) | Pre(%) | AUC | F1 |
| IB-OP [23] | 57.63 | 59.40 | 0.7566 | 0.5735 | 63.63 | 58.79 | 0.7815 | 0.5457 |
| ResNet [24] | 73.57 | 74.58 | 0.8256 | 0.7314 | 62.50 | 59.76 | 0.6942 | 0.5707 |
| DL-PP [13] | 76.43 | 77.31 | 0.8570 | 0.7600 | 71.25 | 74.68 | 0.7279 | 0.7098 |
| IRENR [15] | 58.21 | 55.12 | 0.6731 | 0.5472 | 60.00 | 61.17 | 0.6470 | 0.5541 |
| DAFT [25] | 66.07 | 67.44 | 0.7580 | 0.6599 | 65.00 | 69.14 | 0.7160 | 0.6505 |
| CFDL [26] | 69.14 | 62.57 | 0.8307 | 0.6468 | 61.82 | 53.40 | 0.7088 | 0.5360 |
| ICHPro [16] | 75.00 | 75.96 | 0.8404 | 0.7462 | 70.91 | 74.83 | 0.7443 | 0.6908 |
| **Ours** | **81.50** | **81.77** | **0.8801** | **0.8129** | **79.09** | **82.71** | **0.8289** | **0.7836** |

### 3.2 Comparison Results

To evaluate the proposed method, we compare it against seven state-of-the-art baselines on both datasets, with the results summarized in Table 1 and Fig. 2(i). Compared to the 2D Image-Based Method [23], our approach yields effective accuracy improvements by incorporating comprehensive 3D spatial context. Against the 3D-only ResNet [24], ICH-DIANet achieves accuracy gains of 7.93% and 16.59%, and F1-score increases of 0.0815 and 0.2129 on the private and public datasets, respectively. This substantial improvement is primarily attributed to the effective integration of clinical text. Furthermore, our method outperforms the joint-attention-based ICHPro [16] across all metrics and achieves the highest AUC on both datasets, demonstrating its superiority in feature de-redundancy, deep cross-modal interaction, and class discrimination.

### 3.3 Ablation Experiment

To validate the effectiveness of the proposed MDM and AMoE, we conduct ablation studies on both private and public datasets, with the results summarized in Table 2. The inclusion of MDM alone yields superior performance over the

Table 2: Ablation study results on the private and public datasets.

| MDM | AMoE | Private Dataset | | | | Public Dataset | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Acc(%) | Pre(%) | AUC | F1 | Acc(%) | Pre(%) | AUC | F1 |
| × | × | 76.03 | 77.19 | 0.8539 | 0.7566 | 72.73 | 77.54 | 0.7937 | 0.7061 |
| ✓ | × | 79.13 | <u>80.04</u> | <u>0.8794</u> | <u>0.7913</u> | 74.55 | 76.46 | 0.8045 | 0.7371 |
| × | ✓ | <u>79.44</u> | 79.81 | 0.8630 | 0.7891 | <u>75.45</u> | <u>79.69</u> | <u>0.8147</u> | <u>0.7536</u> |
| ✓ | ✓ | **81.50** | **81.77** | **0.8801** | **0.8192** | **79.09** | **82.71** | **0.8289** | **0.7836** |

baseline across both datasets, indicating that the module enhances feature discriminability through effective disentanglement. We visualize the output features from the modal-shared and modality-specific encoders by projecting them into a 2D plane, as shown in Fig. 2(ii). These intuitive spatial distributions demonstrate that the MDM effectively separates modal-shared and modality-specific features while maintaining cross-modal consistency for shared information. Similarly, integrating only the AMoE significantly improves metrics such as AUC and precision, demonstrating its efficacy in inter-patient feature variance and cross-modal alignment. The full configuration, combining MDM and AMoE, achieves peak performance across all evaluation metrics, substantially outperforming both the baseline and single-module variants. This demonstrates that MDM and AMoE are complementary rather than redundant. Their synergy effectively overcomes the limitations of individual modules.

Table 3: Ablation results of loss functions on private and public datasets.

| $L_{diff}$ | $L_{gather}$ | $L_{rec}$ | Private Dataset | | | | Public Dataset | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Acc(%) | Pre(%) | AUC | F1 | Acc(%) | Pre(%) | AUC | F1 |
| × | × | × | 77.74 | 79.60 | 0.8490 | 0.7751 | 72.73 | 73.97 | 0.7630 | 0.7134 |
| ✓ | × | × | 77.05 | 78.42 | 0.8491 | 0.7667 | 73.64 | 75.88 | 0.7974 | 0.7309 |
| × | ✓ | × | 78.76 | 80.90 | 0.8536 | 0.7836 | 75.45 | 73.31 | 0.7443 | 0.7345 |
| × | × | ✓ | 77.40 | 78.09 | 0.8441 | 0.7727 | 72.73 | 76.28 | 0.7890 | 0.7098 |
| ✓ | ✓ | × | 79.80 | 80.84 | 0.8552 | 0.7914 | 76.36 | 79.51 | 0.8048 | 0.7571 |
| ✓ | × | ✓ | <u>80.81</u> | 81.49 | <u>0.8790</u> | <u>0.8075</u> | 77.27 | <u>80.30</u> | **0.8327** | 0.7670 |
| × | ✓ | ✓ | 80.47 | <u>81.59</u> | 0.8763 | 0.8007 | <u>78.18</u> | 78.43 | 0.8015 | <u>0.7804</u> |
| ✓ | ✓ | ✓ | **81.50** | **81.77** | **0.8801** | **0.8192** | **79.09** | **82.71** | <u>0.8289</u> | **0.7836** |

To evaluate our proposed loss functions, we conduct ablation experiments as summarized in Table 3. Initially, adding only $L_{diff}$ or $L_{rec}$ yields limited improvements over the baseline, suggesting that without synergistic interaction,

these individual losses fail to provide sufficient constraints for the feature space. Conversely, introducing $L_{gather}$ alone improved accuracy across both datasets, demonstrating that enhancing modal-shared subspace consistency directly optimizes cross-modal alignment. Furthermore, combining any two loss functions significantly outperformed both the baseline and individual losses, while the integration of all three losses achieves optimal performance, confirming the efficacy and necessity of each component.

## 4 Conclusion

In this work, we propose ICH-DIANet, a network for assessing brain injury severity in ICH patients by fusing CT imaging and clinical reports. We employ modality disentanglement to minimize intra-modal redundancy and design a dynamic fusion mechanism to adaptively weight expert outputs, effectively capturing global cross-modal dependencies for enhanced alignment and fusion. Experimental results on real-world clinical datasets demonstrate that our method outperforms the state-of-the-art baselines across all performance metrics.

**Acknowledgments.** This work was supported in part by the High-level Overseas Talent Project of Jiangxi Province under Grant No.20232BCJ25026.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References


1. Laurent Puy, Adrian R Parry-Jones, Else Charlotte Sandset, Dar Dowlatshahi, Wendy Ziai, and Charlotte Cordonnier. Intracerebral haemorrhage. *Nature Reviews Disease Primers*, 9(1):14, 2023.
2. David J Seiffge, Simon Fandler-Höfler, Yang Du, Martina B Goeldlin, Wilmar MT Jolink, Catharina JM Klijn, and David J Werring. Intracerebral haemorrhage-mechanisms, diagnosis and prospects for treatment and prevention. *Nature Reviews Neurology*, 20(12):708–723, 2024.
3. Tsong-Hai Lee. Intracerebral hemorrhage. *Cerebrovascular Diseases Extra*, 15(1):1–8, 2025.
4. Geoffrey T Manley and Andrew IR Maas. The glasgow coma scale at 50: looking back and forward. *The Lancet*, 404(10454):734–735, 2024.
5. Paul M Brennan, Charlotte Whittingham, Virendra Deo Sinha, and Graham Teasdale. Assessment of level of consciousness using glasgow coma scale tools. *bmj*, 384, 2024.
6. Lei Pei, Tao Fang, Liang Xu, and Chenfeng Ni. A radiomics model based on ct images combined with multiple machine learning models to predict the prognosis of spontaneous intracerebral hemorrhage. *World Neurosurgery*, 181:e856–e866, 2024.
7. Nacim Yanes, Leila Jamel, Bayan Alabdullah, Mohamed Ezz, Ayman Mohamed Mostafa, and Hossameldeen Shabana. Using machine learning for detection and prediction of chronic diseases. *IEEE Access*, 12:177674–177691, 2024.
8. Siddharth Patel, Zayd Hassan, S Iniyan, and Usha Desai. Multi cancer prediction using deep learning and cnn algorithm. In *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*, pages 214–221. IEEE, 2024.
9. Xuhao Shan, Xinyang Li, Ruiquan Ge, Shibin Wu, Ahmed Elazab, Jichao Zhu, Lingyan Zhang, Gangyong Jia, Qingying Xiao, Xiang Wan, et al. Gcs-ichnet: Assessment of intracerebral hemorrhage prognosis using self-attention with domain knowledge integration. In *2023 IEEE international conference on bioinformatics and biomedicine (BIBM)*, pages 2217–2222. IEEE, 2023.
10. S Suchitra, Lalitha Krishnasamy, and RJ Poovaraghan. A deep learning-based early alzheimers disease detection using magnetic resonance images. *Multimedia tools and applications*, 84(16):16561–16582, 2025.
11. G Praveena and GP Ramesh. Early detection of alzheimer's disease and dementia using deep convolutional neural networks. In *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, pages 1–5. IEEE, 2024.
12. H Abu Owida, Ghayth AlMahadin, Jamal I Al-Nabulsi, Nidal Turab, Suhaila Abuowaida, and Nawaf Alshdaifat. Automated classification of brain tumor-based magnetic resonance imaging using deep learning approach. *International Journal of Electrical and Computer Engineering*, 14(3):3150–3158, 2024.
13. Amaia Pérez del Barrio, Anna Salut Esteve Domínguez, Pablo Menéndez Fernández-Miranda, Pablo Sanz Bellón, David Rodríguez González, Lara Lloret Iglesias, Enrique Marqués Fraguela, Andrés A González Mandly, and José A Vega. A deep learning model for prognosis prediction after intracranial hemorrhage. *Journal of Neuroimaging*, 33(2):218–226, 2023.
14. Wenao Ma, Cheng Chen, Jill Abrigo, Calvin Hoi-Kwan Mak, Yuqi Gong, Nga Yan Chan, Chu Han, Zaiyi Liu, and Qi Dou. Treatment outcome prediction for intracerebral hemorrhage via generative prognostic model with imaging and tabular data. In *International conference on medical image computing and computer-assisted intervention*, pages 715–725. Springer, 2023.

15. Hong-Yu Zhou, Yizhou Yu, Chengdi Wang, Shu Zhang, Yuanxu Gao, Jia Pan, Jun Shao, Guangming Lu, Kang Zhang, and Weimin Li. A transformer-based representation-learning model with unified processing of multimodal input for clinical diagnostics. *Nature biomedical engineering*, 7(6):743–755, 2023.
16. Xinlei Yu, Xinyang Li, Ruiquan Ge, Shibin Wu, Ahmed Elazab, Jichao Zhu, Lingyan Zhang, Gangyong Jia, Taosheng Xu, Xiang Wan, et al. Ichpro: Intracerebral hemorrhage prognosis classification via joint-attention fusion-based 3d cross-modal network. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2024.
17. Zicheng Xiong, Kai Zhao, Like Ji, Xujun Shu, Dazhi Long, Shengbo Chen, and Fuxing Yang. Multi-modality 3d cnn transformer for assisting clinical decision in intracerebral hemorrhage. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 522–531. Springer, 2024.
18. Guan-Hua Huang, Wan-Chen Lai, Tai-Been Chen, Chien-Chin Hsu, Huei-Yung Chen, Yi-Chen Wu, and Li-Ren Yeh. Deep convolutional neural networks on multiclass classification of three-dimensional brain images for parkinsons disease stage prediction. *Journal of Imaging Informatics in Medicine*, pages 1–17, 2025.
19. Tiago Almeida, Plinio Moreno, and Catarina Barata. Prediction of 30-day hospital readmission with clinical notes and ehr information. In *Iberian Conference on Pattern Recognition and Image Analysis*, pages 220–232. Springer, 2025.
20. Conghao Xiong, Hao Chen, Hao Zheng, Dong Wei, Yefeng Zheng, Joseph JY Sung, and Irwin King. Mome: Mixture of multimodal experts for cancer survival prediction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 318–328. Springer, 2024.
21. Amaia Pérez del Barrio, Pablo Menéndez Fernández-Miranda, Pablo Sanz Bellón, Anna Esteve Domínguez, Lara Lloret Iglesias, Enrique Marqués Fraguela, and David Rodríguez González. Head-ct 2d/3d images with and without ich prepared for deep learning. 2022.
22. Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
23. Jawed Nawabi, Helge Kniep, Sarah Elsayed, Constanze Friedrich, Peter Sporns, Thilo Rusche, Maik Böhmer, Andrea Morotti, Frieder Schlunk, Lasse Dührsen, et al. Imaging-based outcome prediction of acute intracerebral hemorrhage. *Translational Stroke Research*, 12(6):958–967, 2021.
24. tefan-Vlad Voinea, Ioana Andreea Gheonea, Rossy Vlădu Teică, Lucian Mihai Florescu, Monica Roman, and Dan Seliteanu. Refined detection and classification of knee ligament injury based on resnet convolutional neural networks. *Life*, 14(4):478, 2024.
25. Sebastian Pölsterl, Tom Nuno Wolf, and Christian Wachinger. Combining 3d image and tabular data via the dynamic affine feature map transform. In *International conference on medical image computing and computer-assisted intervention*, pages 688–698. Springer, 2021.
26. Tianling Liu, Hongying Liu, Fanhua Shang, Lequan Yu, Tong Han, and Liang Wan. Completed feature disentanglement learning for multimodal mris analysis. *IEEE Journal of Biomedical and Health Informatics*, 2025.