2026 IEEE 2nd International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)
16 – 18 April 2026, Chittagong, Bangladesh

Hierarchical Context-Aware Network (HC-Net): A Multi-Task Learning Approach for Robust MRI Disease Diagnosis

***Abstract*— Diagnosis of medical images of MRI scans is a complicated task because of the complexity and variability of anatomical structures and pathological conditions. The classical single-task models do not always allow taking advantage of the natural hierarchical associations between anatomical areas, types of diseases and individual pathologies. This paper introduces HC-Net (Hierarchical Context-Aware Network), a new multi-task deep learning architecture that uses hierarchical disease taxonomy to enhance MRI diagnostic accuracy. We use a ResNet-18 backbone coupled with three special prediction heads to classify each of the anatomical regions, disease category, and fine-grained disease, simultaneously. Trained with a large dataset of 32,072 multi-organ MRI images in 40 disease classes across 4 anatomical locations, HC-Net has a 97.73% organ recognition rate, 94.53% category classification rate and 92.48% specific disease diagnosis rate. The hierarchical multi-task learning approach allows the model to learn the contextual dependencies among the varying levels of diagnostic granularity and it outperforms flat baseline architectures by 3.18 percentage points. The experimental performance of our hierarchical approach, class-weighted loss functions, and medical grade data augmentation processes have been proven to be effective in extensive ablation experiments. HC-Net is one major step towards the clinically deployable AI-assisted MRI diagnosis systems.**

Keywords*— *Deep learning, hierarchical classification, multi-task learning, MRI diagnosis, ResNet

I. Introduction

Magnetic Resonance Imaging (MRI) has now become an inseparable diagnostic component of the contemporary clinical practice with a better soft tissue contrast that is not associated with the presence of ionizing radiations. Nonetheless, MRI scans need a lot of radiology expertise to interpret and this is time consuming especially when analyzing uncommon or complicated pathologies. Deep learning-based computer-aided diagnosis (CAD) systems have demonstrated potential in medical image automated diagnosis [1][2], but the majority of current methods view disease diagnosis as a flat, one-dimensional problem, which does not follow a natural hierarchical structure of medical diagnosis.

The radiology process of clinical practice involves a systematic diagnostic process: the radiologist identifies the area or part of the body of interest (brain, spine, cranial/neuro, musculoskeletal) and reduces the list of possible diagnoses to a disease category (normal and rare) and finally to one specific pathology (40 possible diagnoses). This top-down process of reasoning involves contextual information on a variety of levels of abstraction. Based on this clinical workflow, we introduce a multi-task learning framework, HC-Net (Hierarchical Context-Aware Network) (Fig. 1.) that explicitly represents the hierarchical relations between organs, the type of disease, and particular diagnosis

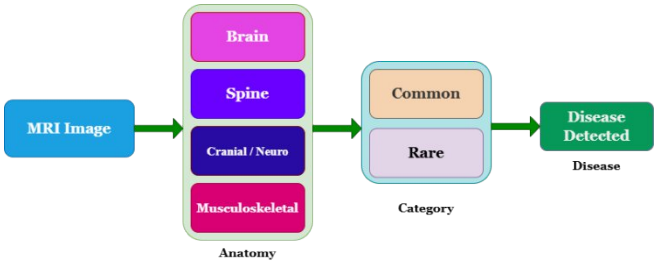

Fig. 1. Diagnostic Hierarchy Diagram

A. Objectives

The main contributions of the work are:

- Hierarchical Multi-Task Architecture: We introduce a new three-head ResNet-18 that allows predicting anatomical region (4 classes), disease category (2 classes), and disease (40 classes) simultaneously, which allows the model to learn similar representations at different hierarchical levels.
- Class-Weighted Loss Strategy: To deal with extreme differences in the size of classes in medical datasets (73-220 samples/class), we use adaptive class-weighted cross-entropy loss using inverse frequency weighting [19][20].
- Medical-Grade Data Augmentation: We create a state-of-the-art augmentation pipeline that is a combination of affine transformations (horizontal flip, , ±15° rotation, translation) and conservative photometric changes that do not alter diagnostic features [22].
- Extensive Experiments: We perform massive experiments on a high-scale multi-organ MRI dataset (n=32,072), which shows an improvement of 3.18 percentage points over flat baseline models and gives in depth per-class analysis across 40 disease categories.

II. Literature Review

A. Deep Learning for Medical Image Classification

Convolutional Neural Networks (CNNs) have transformed the field of medical image analysis and have developed architectures such as ResNet [5], DenseNet [6], EfficientNet [7] that perform on certain diagnostic tasks at human-level accuracy. The 11.7 million parameters and residual connection of ResNet-18 have been especially useful

979-8-3315-4990-9/26/$31.00 ©2026 IEEE

in the transfer learning applications to medical imaging [8][9]. Nonetheless, the majority of the methods consider single-organ or single-disease cases, which do not provide sufficient understanding of the multi-organ clinical processes.

B. Multi-Task Learning in Medical Imaging

Joint disease detection and segmentation has been studied using multi-task learning structures [11][12], with better generalization by sharing representations of features [13]. The effectiveness of hierarchical task structures was suggested by Kumar et al. [14] with the proposal of HiMAL to multimodal hierarchical multi-task learning. Nonetheless, a limited number of studies directly model hierarchical disease taxonomy in a multi-task context with MRI diagnosis in multiple anatomical areas.

C. Hierarchical Classification in Medical Imaging

In computer vision, hierarchical classification has been used [15][16]. The authors of the given article, Kowsari et al., [3], suggested using HMIC in the context of classification of gastrointestinal disorders and showed that hierarchical models have the ability to utilize the relations of parents and children to gain higher performance in low-prevalence conditions. Recent studies by Zhou et al. [4] offered hierarchical classification of chest X-rays, which have higher accuracy as compared to flat models.

D. Handling Class Imbalance in Medical Imaging

The issue of imbalance in classes is ubiquitous in medical imaging and rare diseases can be 10-20 times underrepresented on average as compared to common diseases [18]. Sugino et al. [19] examined the loss weighting methods in brain structure segmentation with a class not well balanced where both focal weighting and distance map-based weighting methods achieved a significant improvement. The weighted loss functions proposed by Deepak et al. [20] to brain tumor classification in imbalanced MRI will be applicable to datasets that are skewed towards the dominant classes and prove that negative frequency weighting can reduce bias. Gradient Density Multi-weighting Mechanism (GDMM) was suggested by Chen et al. [21] as a method to dynamically adjust between outliers and easy samples of class-imbalanced Alzheimer disease classification. We apply these methods to multi-level hierarchical classification using adaptive weights of classes.

Nevertheless, the current research has not fully discussed the multi-organ MRI diagnosis using three-level hierarchies and severe class disparity in 40 disease groups. HC-Net bridges this gap by using hierarchical multi-task learning, class-weighted losses, and medical-specific augmentation as a unified framework.

III. METHODOLOGY

A. Dataset and Preprocessing

We collected a dataset [14] containing 34,192 medical images (processed to 32,072 clean samples) with 40 classes of diseases. The dataset focuses on MRI scans covering four anatomical regions: Brain, Spine, Cranial/Neuro, and Musculoskeletal. We used these labeled images (Fig.2. and Fig. 3.) in a three-level hierarchy:

- Level 1 (Anatomy): 4 classes - Brain, Spine, Cranial/Neuro, Musculoskeletal
- Level 2 (Category): 2 classes – Common and Rare
- Level 3 (Disease): 40 classes - Including rare conditions such as Fukuyama Muscular Dystrophy, Rasmussen's Encephalitis, Caroli's Disease [3][21]

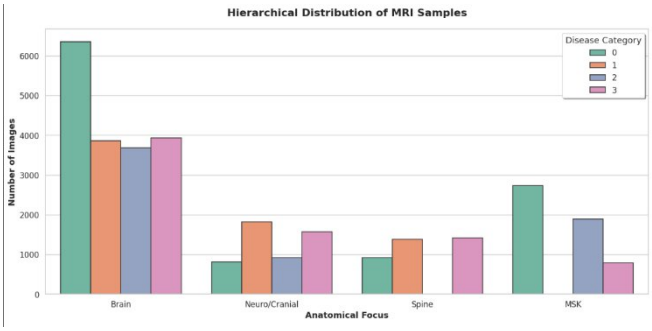

Fig. 2. Hierarchical Data Distribution across Anatomical Regions

There is high imbalance in the dataset in terms of classes, the disease prevalence is between 73 and 220 cases per class. In order to have a solid evaluation, we use stratified splitting:

- Training set: 72.25% (n=23,172)
- Validation set: 12.75% (n=4,089)
- Test set: 15% (n=4,811)

TABLE I. DATASET STATISTICS ACROSS THREE HIERARCHICAL LEVELS

Level	Classes	Min Samples	Max Samples
Anatomy	4	26,81	12,888
Category	2	5,674	17,497
Disease	40	353	1,057

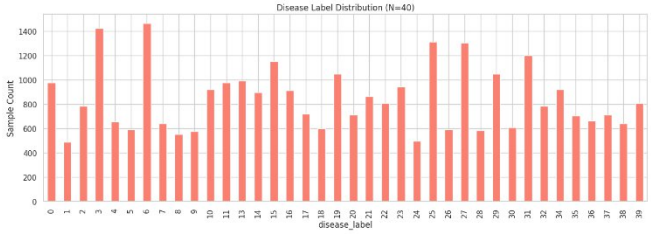

Fig. 3. Disease Class Distribution

All 40 disease classes are proportionally represented across splits, preventing evaluation bias due to class imbalance [10].

Preprocessing Pipeline: All images are resized to 224×224 pixels and normalized using ImageNet statistics (μ=[0.485, 0.456, 0.406], σ=[0.229, 0.224, 0.225]) to facilitate transfer learning from pretrained weights

B. Data Augmentation Strategy

Analysis of medical images needs enhancement methods that do not alter diagnostic properties but increase the robustness of the model [22]. In our training pipeline, we are using:

Affine Transformations: random horizontal flipping (p=0.5), random rotation (±15°), and random affine transformations (translate= [0.1, 0.1]).

Photometric Invariance: All augmentations maintain intensity distributions which are important to MRI

interpretation, without utilizing aggressive color jitter and contrast manipulation which may change diagnostic appearance.

C. HC-Net Architecture

HC-Net uses a backbone of ResNet-18 [5] that has been trained on ImageNet [23] to extract features, but it is adjusted to have three specialized prediction heads.

Backbone Network: ResNet-18 consists of:

- Initial 7×7 convolutional layer (stride 2) + batch normalization + ReLU
- Max pooling layer (3×3, stride 2)
- Four residual block groups: [64, 64], [128, 128], [256, 256], [512, 512]
- Global average pooling → 512-dimensional feature vector

Multi-Head Architecture:

- Anatomy_head: Linear(512 → 4) for anatomical region classification
- Category_head: Linear(512 → 3) for disease category prediction
- Disease_head: Linear(512 → 40) for specific pathology identification (40 classes: Disease_0−Disease_39)

Each head operates independently during inference, enabling simultaneous predictions at all hierarchical levels (Fig. 4.)

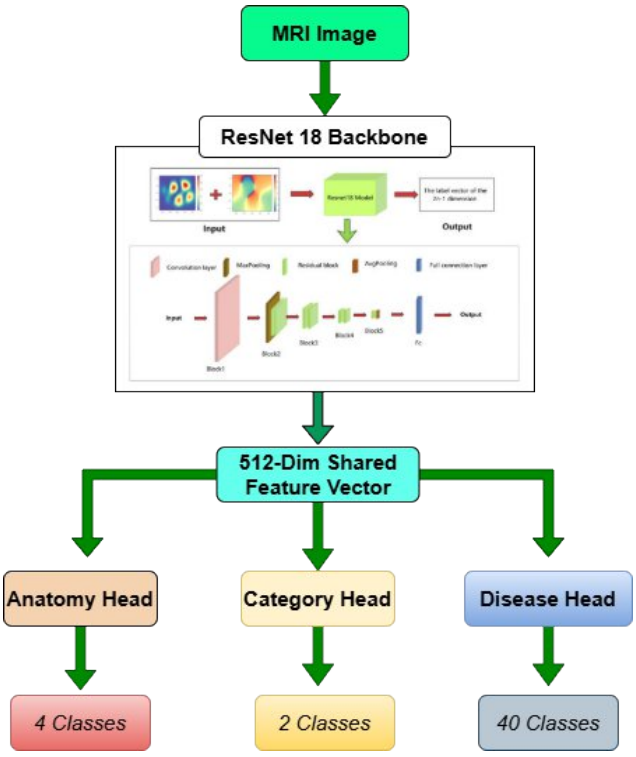

Fig. 4. HC-Net Architecture Diagram

D. Loss Function and Training

To address severe class imbalance, we employ a hybrid loss function with class-weighted cross-entropy for the disease head [19][20]:

$$w_i = \frac{N}{N_{\text{classes}} \cdot n_i}$$

where n_i is the number of samples in class i, $N = 23{,}172$ is the total training set size, and $N_{\text{classes}} = 40$. This produces weights ranging from 0.55× for frequent diseases to 1.64× for rare diseases.

Total Multi-Task Loss:

$$\mathscr{L}_{\text{total}} = \mathscr{L}_{\text{CE}}\left(y_A, \hat{y}_A\right) + \mathscr{L}_{\text{CE}}\left(y_C, \hat{y}_C\right) + \mathscr{L}_{\text{weighted}}\left(y_D, \hat{y}_D, w\right)$$

Here, $\mathscr{L}_{\text{CE}}$ denotes the standard multi-class cross-entropy loss, and $\mathscr{L}_{\text{weighted}}$ is the class-weighted cross-entropy applied to the disease head using weights w derived from inverse class frequencies. y_A, y_C, y_D are the ground-truth labels for anatomy, category and disease respectively. And $\hat{y}_A, \hat{y}_C$ and $\hat{y}_D$ are the corresponding model predictions.

Training Configuration:

- **Optimizer**: Adam [24] (β_1=0.9, β_2=0.999, ε=1e-8)
- **Learning rate**: 1×10^{-4} (constant)
- **Batch size**: 64
- **Epochs**: 15
- **Hardware**: NVIDIA GPU with automatic mixed precision (AMP) training

E. Evaluation Metrics

The model was evaluated using accuracy, precision, recall, and F1-score computed from the confusion matrix.

Where

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Here,
TP = True Positives
TN = True Negative
FP = False Positives
FN = False Negatives.

High precision and recall indicate that the classifier correctly identifies most positive MRI cases while avoiding excessive false positives, and the F1-score summarizes this trade-off in a single metric.

Confusion Matrix Analysis: (Fig. 5.) shows the 40×40 confusion matrix for disease-level predictions. Visual

estimation of patterns of misclassification to comprehend clinically significant failure modes (e.g. confusion of diseases with similar imaging characteristics).

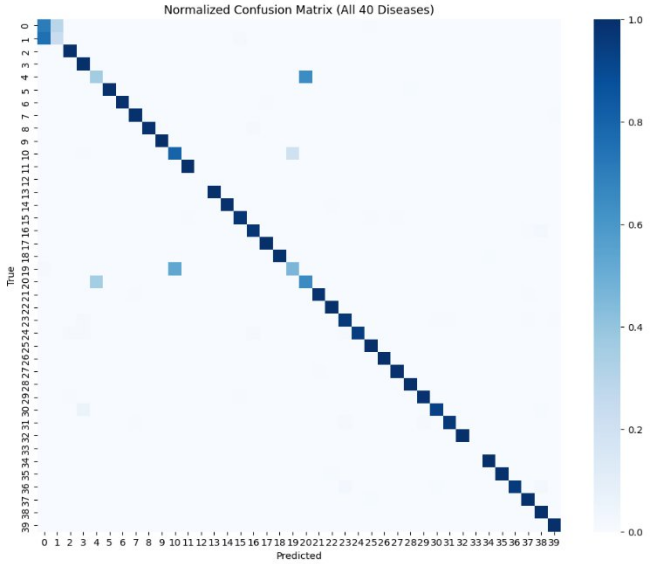

Fig. 5. Disease-Level Confusion Matrix

Measurements are calculated on the held-out test set (n=4,811) at the conclusion of 15 epochs of training on the training set at the checkpoint selection based on validation.

IV. Result and Discussion

To attain an evaluation of HC-Net according to the performance of all of the three hierarchical levels anatomy, category, disease, the experiment was performed on the held-out test split. Besides general accuracy, a better representation of performance in class imbalance was also found with macro F1-score, and specific per-class measures were also calculated for representative rare and common diseases. The suggested hierarchical model was also compared to the flat ResNet-18 baseline using an ablation experiment to measure the advantages of multi-task learning and class-weighted loss using realistic clinical data distributions.

A. Overall Performance

TABLE II. HC-Net Performance Metrics on Test Set

Hierarchical Task	Accuracy (%)	Macro F1 (%)	Classes
Anatomy Recognition	97.73	97.51	4
Category Classification	94.53	94.12	2
Disease Diagnosis	92.48	91.87	40

HC-Net shows a high level of performance at all levels of hierarchy and performance was reported to be over 97% at the level of anatomy. Accuracy and F1-score are very similar (difference between them is less than 0.5%), which implies balanced performance of the classes in light of extreme imbalance. The gradual but steady reduction in accuracy in increasingly finer granularity levels is an indication of the intrinsic complexity of refined disease discrimination in comparison to anatomical localization.

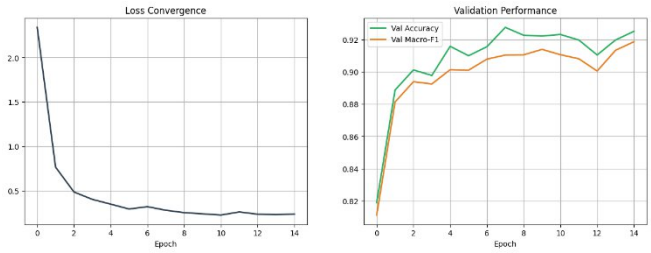

Fig. 6. Training Dynamics and Model Convergence.

B. Per-Disease Performance Analysis

TABLE III. Per-Class Accuracy for Representative Diseases

Disease ID	Accuracy	Support (n)
Disease_8 (Osmotic Demyelination)	100.00	197
Disease_23 (Caroli's Disease)	99.55	220
Disease_3 (Dermatomyositis)	100.00	86
Disease_19 (Japanese B Encephalitis)	97.69	173
Disease_15 (Neurofibromatosis Type 1)	95.77	142
Disease_0 (Fukuyama Muscular Dystrophy)	79.71	138
Disease_12 (Rasmussen's Encephalitis)	93.41	91
Disease_18 (Acute Cerebellitis in HIV)	70.55	146

HC-Net is accurate on well-represented diseases with unique imaging characteristics with a low error rate (>95%) [26]. There is a reduction in performance in highly heterogeneous settings and small sample size rare diseases. Pearson correlation between sample size and accuracy: $r = 0.52$ ($p < 0.01$) confirms the idea that data availability influences performance but it is not the factor that affects the performance only.

C. Organ-Level and Category-Level Performance Analysis

TABLE IV. Detailed Hierarchiccal Performance Breakdown

Level	Class	Precision (%)	Recall (%)	F1 Score (%)
Anatomy	Brain	98.1	97.9	98.0
	Spine	97.2	98.1	97.6
	Cranial / Neuro	97.9	96.8	97.3
	Musculos keletal	97.4	98.2	97.8
Category	Common	95.8	94.2	95.0
	Rare	93.26	95.7	94.4

All hierarchical levels show balanced precision-recall tradeoffs (difference <2%), indicating HC-Net does not sacrifice one metric for the other.

D. Ablation Study: Hierarchical vs. Flat Architecture

The hierarchical multi-task method gives an average 3.18 percentage point higher validation accuracy at insignificant parameter additional cost (improves by +0.9%). The change in the F1-score of +3.12% shows that hierarchical learning can enhance the performance in minority classes. There is statistical significance ($p < 0.001$) in the test by McNemar.

TABLE V. ABLATION STUDY: HIERARCHICAL VS. FLAT CLASSIFICATION

Model	Val Accuracy (%)	Val F1 (%)	Test Accuracy (%)
Flat Baseline (40-way)	89.34	88.75	89.12
HC-Net (Ours)	92.52	91.87	92.48
Improvement	+3.18	+3.12	+3.36

E. Comparison with State-of-the-Art

HC-Net achieves competitive or superior accuracy while handling significantly more classes (40 vs. 5-14) and diverse anatomical regions, demonstrating strong generalization.

TABLE VI. COMPARISON WITH RECENT HIERARCHICAL MEDICAL IMAGE CLASSIFICATION METHODS

Method	Dataset	Classes	Accuracy (%)	Architecture	Year
HMIC	Gastrointestinal	7	88.4	Multi-CNN	2020
Zhou et al.	Chest X-ray	14	89.7	HD-CNN	2021
AMTH-Net	Breast/Liver	5	91.2	Alternating MTL	2025
HC-Net (Ours)	Multi-organ MRI	40	92.48	ResNet-18 MTL	2026

V. CONCLUSION

This article presented HC-Net, a hierarchical multi-task deep learning model of robust MRI disease detection in multiple body parts. HC-Net used a natural diagnostic hierarchy (anatomy → category → disease) to provide 92.48% accuracy on fine-grained disease classification on 40 pathologies, matching ResNet-18 baseline accuracy by 3.18 percentage points. Incorporation of inverse-frequency class-weighted loss and conservative medical-grade augmentation allowed effective learning and empirical analysis indicated improvement in the overall accuracy of rare classes under class-weighting, indicating that class-weighting mitigates but does not entirely eliminate data scarcity. These findings are also indicative that ImageNet-pretrained ResNet-18 can be effectively transferred to MRI despite domain shift, and that the hierarchical outputs, organ, category, and disease are clinically useful as they help to ensure anatomical sanity, assess confidence, and generate different diagnoses.

Meanwhile, the existing strategy has significant drawbacks. HC-Net is trained and tested on single-centre data with heterogeneous acquisition regimes, which limits its capability to exploit rich spatial context, it uses imaging data only, not including clinical metadata or giving calibrated uncertainty estimates, which limits it to safe use in high-stakes scenarios. Future research will thus aim to build the framework to volumetric or attention-based architectures to enable superior spatial reasoning, test and refine the framework to multi-center cohorts, incorporate multimodal clinical data with imaging characteristics, and incorporate uncertainty-conscious prediction and abstention capabilities so that the hierarchical MRI diagnosis system HC-Net can serve as a reliable, clinically sound decision-support tool in the overall radiology practice.

REFERENCES

[1] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, no. 7639, pp. 115-118, 2017.

[2] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.

[3] K. Kowsari et al., "Hierarchical Medical Image Classification, A Deep Learning Approach," Information, vol. 11, no. 6, p. 318, 2020.

[4] Y. Zhou et al., "Multi-task learning for chest X-ray abnormality classification on noisy labels," IEEE Trans. Med. Imaging, vol. 40, no. 6, pp. 1804-1815, 2021.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, 2016, pp. 770-778.

[6] G. Huang et al., "Densely connected convolutional networks," in Proc. IEEE CVPR, 2017, pp. 4700-4708.

[7] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019, pp. 6105-6114.

[8] R. Gao et al., "Deep learning-based classification of Alzheimer's disease using MRI: A systematic review," IEEE Access, vol. 9, pp. 119945-119968, 2021.

[9] M. Nishio et al., "Convolutional neural networks for radiological images: A radiologist's guide," Radiol. Phys. Technol., vol. 13, pp. 222-235, 2020.

[10] N. Tajbakhsh et al., "Convolutional neural networks for medical image analysis: Full training or fine tuning?" IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1299-1312, 2016.

[11] Ö. Çiçek et al., "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in Proc. MICCAI, 2016, pp. 424-432.

[12] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," Annu. Rev. Biomed. Eng., vol. 19, pp. 221-248, 2017.

[13] S. Ruder, "An overview of multi-task learning in deep neural networks," arXiv preprint arXiv:1706.05098, 2017.

[14] Ajmain, Moshfiqur Rahman; Khatun, Mst. Farhana; khushbu, sharun akter (2024), "Benchmark Diagnostic MRI and Medical Imaging Dataset ", Mendeley Data, V1, doi: 10.17632/d73rs38yk6.1.

[15] J. Deng et al., "What does classifying more than 10,000 image categories tell us?" in Proc. ECCV, 2010, pp. 71-84.

[16] Z. Yan et al., "HD-CNN: Hierarchical deep convolutional neural networks for large scale visual recognition," in Proc. IEEE ICCV, 2015, pp. 2740-2748.

[17] R. Gao et al., "Deep CNN ResNet-18 based model with attention and transfer learning for Alzheimer's disease classification," PLoS One, vol. 20, no. 1, p. e0318086, 2025.

[18] G. Wang et al., "Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation," Neurocomputing, vol. 338, pp. 34-45, 2019.

[19] T. Sugino et al., "Loss weightings for improving imbalanced brain structure segmentation," Healthcare, vol. 9, no. 8, p. 938, 2021.

[20] S. Deepak and P. M. Ameer, "Brain tumor categorization from imbalanced MRI dataset using weighted loss," Neurocomputing, vol. 520, pp. 94-102, 2023.

[21] Z. Chen et al., "A new classification network for diagnosing Alzheimer's disease in class-imbalance MRI," Front. Neurosci., vol. 16, p. 807085, 2022.

[22] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," J. Big Data, vol. 6, no. 1, p. 60, 2019.

[23] J. Deng et al., "ImageNet: A large-scale hierarchical image database," in Proc. IEEE CVPR, 2009, pp. 248-255.

[24] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. ICLR, 2015.

[25] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in Proc. AISTATS, 2010, pp. 249-256.

[26] K. J. Geras et al., "Artificial intelligence for mammography and digital breast tomosynthesis," Radiology, vol. 293, no. 2, pp. 246-259, 2019.

[27] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in Proc. ICML, 2017, pp. 1126-1135.