Ahsanullah University of Science & Technology

Department of Computer Science and Engineering

CSE 4138: Soft Computing Lab

Assignment on

"Reviewing Paper"

- *Explainable Cost-Sensitive Deep Neural Networks for Brain Tumor Detection from Brain MRI Images considering Data Imbalance*
- *Transformer-Based Disease Identification for Small-Scale Imbalanced Capsule Endoscopy Dataset*

Submitted To:

Md. Tanvir Rouf Shawon	Lecturer Grade-II
Raihan Tanvir	Lecturer

Submitted By:

Shuvashis Sarkar – 20200104116

LAB SECTION: C1; GROUP NO: 2

Date of Submission: 27th November,2023

Explainable Cost-Sensitive Deep Neural Networks for Brain Tumor Detection from Brain MRI Images considering Data Imbalance

IDEA: This paper introduces a novel approach using CNN, ResNet50, InceptionV3, EfficientNetB0, and NASNet-Mobile models for brain tumor detection. The proposed models, particularly a fine-tuned InceptionV3, achieve high accuracy (99.33%) on a balanced dataset. The study also incorporates Explainable AI and a cost-sensitive neural network approach for imbalanced datasets, outperforming conventional models. The results highlight the potential for practical application in improving brain tumor detection accuracy.

DATASET: Two datasets were utilized for brain tumor detection. The first, Br35H, comprises 1500 instances each of Tumor and No-Tumor categories, ensuring a balanced distribution. The second dataset, Brain MRI Images for Brain Tumor Detection, consists of 155 Tumor and 98 No-Tumor MRI images. Samples from both datasets showcase the presence or absence of tumors in brain MRI images.

METHODOLOGY:

1. Image Input: MRI brain images are loaded for preprocessing, serving as the initial step before model input.

2. Image Preprocessing: The images undergo resizing to 224x224 using bilinear interpolation. Subsequently, the dataset is divided into training (80%), validation (10%), and test (10%) sets to ensure effective model training and evaluation.

3. Check Data Equivalence: A critical assessment is conducted to determine the balance between Tumor and No-Tumor classes in the dataset.

4. Train Classification Models: Standard CNN and pre-trained models, including ResNet50, InceptionV3, EfficientNetB0, and NASNetMobile, are trained. In cases of imbalanced data, a cost-sensitive neural network technique is implemented to emphasize the significance of minority classes.\

5. Performance Evaluation: Model evaluation is conducted, incorporating key metrics such as accuracy, precision, recall, f1-score, and specificity to assess the models' effectiveness.

6. Explainability: The interpretability of models is addressed through various Explainable AI methods. For CNN, explanations include Vanilla Saliency, SmoothGrad, Grad-CAM, Grad-CAM++, Score-CAM, and Faster Score-CAM. Additionally, LIME is employed for explaining pre-trained models, providing insights into the models' predictions.

RESULT ANALYSIS: The performance analysis of the proposed models for brain tumor detection reveals notable findings. On the Br35H dataset, all fine-tuned pre-trained models demonstrate excellent accuracy, with **ResNet50 and InceptionV3** reaching **99.33%**. However, a standard CNN lags behind. The cost-sensitive models on the imbalanced Brain MRI Images for Brain Tumor Detection (BTD) dataset outperform conventional models, with InceptionV3 achieving the highest accuracy of **92.31%.** The ROC analysis confirms the models' efficacy, particularly the cost-sensitive InceptionV3, which achieves an **AUC score** of **0.97**. Explainable AI techniques, such as Grad-CAM and LIME, provide insights into model decision-making, highlighting areas of focus for correct and incorrect classifications. The cost-sensitive models show promising results, emphasizing the importance of addressing class imbalances in medical image classification tasks.

Performance Comparison: The proposed cost-sensitive models, particularly InceptionV3, outperform existing works on brain tumor detection datasets (Br35H and Brain MRI Images). In comparison to Mondal et al.'s deep CNN achieving 99.00% accuracy, our InceptionV3 achieves 99.33% accuracy on Br35H. For the imbalanced Brain MRI Images dataset, our cost-sensitive CNN surpasses Sailunaz et al.'s CNN with MRI features, achieving 90.20% accuracy compared to their 76.47%. Additionally, our approach demonstrates a recall value of 1, outperforming Sruti et al.'s hybrid convolutional SVM with a recall value of 0.98.

Structure and Writing: The paper demonstrates a well-organized structure, presenting a coherent flow of information through its sections. The writing is clear, precise, and technically sound, ensuring accessibility for readers with varying levels of expertise. The inclusion of citations adds credibility to the presented findings, contributing to the overall strength of the paper.

Transformer-Based Disease Identification for Small-Scale Imbalanced Capsule Endoscopy Dataset

IDEA: The paper introduces a novel approach to train Vision Transformer (ViT) on small and imbalanced capsule endoscopic datasets. By combining ViT's self-attention mechanism with spatial pooling, the method achieves high accuracy in disease classification tasks, particularly in wireless capsule endoscopy. Training from scratch on two datasets, Kvasir-Capsule and Red Lesion Endoscopy, demonstrates significant performance improvements over existing baselines, providing a promising solution for diagnosing gastrointestinal diseases through capsule endoscopy.

DATASET: The experiments utilized two datasets: the Kvasir-Capsule dataset, consisting of 11 classes related to capsule endoscopy images, with imbalanced distribution and resolutions of 336 × 336; and the Red Lesion Endoscopy (RLE) dataset, originally for hemorrhage segmentation, rearranged into normal and bleeding classes for binary classification at a resolution of 320 × 320. To enhance robustness and prevent overfitting, data augmentation techniques such as random rotation, flipping, and random noise were applied. The datasets were split into training and testing sets with a ratio of 4:1.

METHODOLOGY: The methodology adopts the Vision Transformer (ViT) architecture, originally designed for image classification tasks. The ViT structure comprises patch embedding, position embedding, transformer encoder, class token, and a classification head. The transformer encoder employs multi-head self-attention to process flattened image patches. To enhance ViT's performance on small and imbalanced capsule endoscopy datasets, a spatial pooling configuration is introduced, inspired by Convolutional Neural Network (CNN) pooling techniques. This involves downsampling ViT's spatial dimensions three times using depthwise convolution and reshaping operations, ultimately improving the model's generalization ability. The class token is aligned with the pooling module, and the final prediction results are obtained through a classifier head after three consecutive downsamplings of the processed data. The methodology integrates data augmentation strategies for robustness and prevents overfitting. The approach is applied and evaluated on the Kvasir-Capsule and Red Lesion Endoscopy datasets, showcasing improved performance over existing baselines.

RESULT ANALYSIS: The model evaluation reveals the superiority of the proposed methodology over both CNN and ViT-based algorithms in the context of wireless capsule endoscopy (WCE) disease classification. Achieving an accuracy of **79.15%** on the **Kvasir-Capsule dataset** and **98.63%** on the RLE dataset, the introduced approach outperforms CaiT and Swin Transformer. Among CNNs, ShuffleNetV1 stands out with excellent results on both datasets. ViT-based models, despite slightly slower inference times, demonstrate superior performance, highlighting their potential for real-time diagnostic applications. Comparison with existing WCE classification solutions underscores the competitive edge of the proposed methodology, emphasizing the efficacy of Vision Transformer in enhancing disease identification accuracy in small and imbalanced datasets.

Structure and Writing: The paper follows a conventional structure, beginning with an Introduction addressing the challenges in wireless capsule endoscopy. The Methodology section details the use of Vision Transformer (ViT) with spatial pooling. Model Evaluation presents experimental results, comparing with CNN and ViT-based algorithms. The Discussion section critically analyzes findings, emphasizing the methodology's unique contributions. The paper concludes with a concise summary and potential future research directions.