Notification No. 575/2025
Date of Award: 28-02-2025
Name of Scholar: Samta Rani
Name of Supervisor: Prof. Tanvir Ahmad
Name of the Department: Computer Engineering
Topic of Research: Hybrid Artificial Intelligence System for Cancer Predictions

Findings

1. **Enhanced Cancer Prediction with Hybrid Pre-Processing –** This study focuses on early prediction of breast and prostate cancer using machine learning models, enhanced by Z-score normalization and feature selection. It addresses class imbalance using five oversampling techniques on gene expression data. Among eight classification models tested, Logistic Regression achieves the highest accuracy (92%) with oversampling, followed by SVM (90%) using SVMSMOTE. Results indicate that while most oversampling techniques improve diagnostic efficacy, no single method consistently outperforms others across all scenarios.
2. **Superior Performance of Hybrid Deep Learning Models –** This chapter explores deep learning techniques for breast cancer image classification using CNN and hybrid CNN-LSTM models on BreakHis and BACH2018 datasets. CNN achieved 95% and 92% accuracy, while CNN-LSTM outperformed with 99.59% (binary) and 95.55% (multiclass). Image preprocessing and augmentation enhanced performance. These models significantly improved classification accuracy compared to previous studies, demonstrating the effectiveness of deep learning in diagnosing breast cancer from medical images.
3. **Multimodal Data Integration for Better Diagnosis –** This study emphasizes the significance of multimodal data fusion for breast cancer detection, integrating imaging, histology, genetics, and electronic medical records. Conducted in three phases, it applies machine learning models to gene expression data, histopathological images, and multimodal data. Results show multimodal models achieve 94% accuracy, outperforming unimodal models (86% and 80.21%). High recall values ensure fewer false negatives, crucial for early detection. The study also improves image processing by slicing large images and filtering noise for better accuracy.

4. **Optimized Learning Settings for Label Efficiency –** This study examines the impact of annotated and non-annotated cancer images on pre-trained models using three learning settings: Supervised (SL), Semi-Supervised (Semi-SL), and Self-Supervised (Self-SL). Using three datasets and seven curated training sets, ResNet50, VGG16, and EfficientNetB0 were trained for cancer prediction. EfficientNetB0 consistently achieved the highest accuracy. Semi-SL results closely matched SL, while Self-SL also performed comparably. Pseudo-labeling was used for Semi-SL, and contrastive learning for Self-SL, demonstrating effective learning from unlabeled data.
5. **Significant Contribution to AI-Driven Cancer Research –** This study explores AI-based cancer prediction using hybrid approaches across multiple dimensions. It applies hybridization in data pre-processing (feature selection, Z-score normalization, and oversampling techniques like SMOTE), modeling (CNN, CNN-LSTM for image classification), dataset modalities (multimodal fusion for improved accuracy), and learning settings (supervised, semi-supervised, and self-supervised learning using ResNet50, VGG16, and EfficientNetB0). Results show CNN-LSTM achieves 99.59% accuracy, multimodal data enhances diagnosis to 94%, and EfficientNetB0 performs best across learning settings. The study highlights AI's potential for early cancer detection and improved treatment outcomes.