An Interpretable Deep Learning Framework for Diabetic Retinopathy Detection Using Patch-Level Variational Autoencoders and Graph Attention Networks

D. Tejaswini
Assistant Professor, Department of Data Science, Anurag University, Hyderabad, India

Bagareddy Gari Chaithanya Reddy
Department of Computer Science and Engineering – Data Science, Anurag University, Hyderabad, India

Malkai Nikhil
Department of Computer Science and Engineering – Data Science, Anurag University, Hyderabad, India

Rajput Gowtham Prasad
Department of Computer Science and Engineering – Data Science, Anurag University, Hyderabad, India

Abstract

Diabetic Retinopathy (DR) is a sight-threatening complication of diabetes that affects millions worldwide. Early detection and accurate grading of DR severity are critical for timely treatment. In this work, we propose an interpretable and modular pipeline that combines patch-level Variational Autoencoders (VAE) for compact feature encoding, spatial graph construction to capture anatomical context, and Graph Attention Networks (GAT) for classification with region-level interpretability. The approach emphasizes clinical relevance by producing attention heatmaps that align with lesion locations. We evaluate the method on the EyePACS dataset and report competitive performance across five DR grades while offering improved interpretability over conventional end-to-end CNNs.

Index Terms

Diabetic Retinopathy, Graph Attention Network, Variational Autoencoder, Explainable AI, Medical Imaging, Fundus