Deep Learning Approach for Early-Stage Diabetes Risk Prediction using Southern Indian Diabetes Dataset

Dr. Sheetalrani R Kawale
Department of Computer Science, Karnataka State Akkamahadevi Women University, Vijayapura, Karnataka, India

PoojaKallappagol
Department of Computer Science, Karnataka State Akkamahadevi Women University, Vijayapura, Karnataka, India

Dr. Sharankumar Holaychi
Department of Community Medicine, Koppal Institute of Medical Sciences, Koppal, Karnataka, India

Abstract

This proposed work investigates the efficiency of a hybrid data augmentation approach (SMOTE-ENN + GAN), SHAP-based feature selection, and deep learning models for classifying the 2001 dataset, which exhibits significant class imbalance. Five deep learning models Radial Basis Function Neural Networks (RBF-NN), Long Short-Term Memory (LSTM), Simple Recurrent Neural Networks (RNN), Convolutional Neural Networks (CNN), and Deep Neural Networks (DNN) were evaluated for their performance. The proposed hybrid augmentation technique significantly improved the model's ability to simplify by balancing the class distribution, resulting in a notable increase in model accuracy. The DNN and CNN models achieved a test accuracy of 98%, outperforming other approaches in the literature. SHAP-based feature selection further enhanced model interpretability, identifying critical features such as age, obesity, and visual blurring. The comparative results confirm that the proposed methodology, combining hybrid data augmentation and SHAP feature selection, provides substantial improvements in classification tasks, showcasing its potential for real-world applications where high accuracy is crucial.

Keywords

Deep Learning, Feature Importance, Class Imbalance, Generative Adversarial Networks.