

Dissertation Title Enhancing Early Detection of Dementia Using
Inter-Relation-Based Features and Oversampling Technique
Author Yanawut Chaiyo
Degree Doctor of Philosophy (Computer Engineering)
Advisor Associate Professor Punnarumol Temdee, Ph. D.

ABSTRACT

Dementia affects both individuals and society, making early detection essential for effective management. However, reliance on advanced laboratory tests and specialized expertise limits accessibility, hindering timely diagnosis. To address this challenge, this study pioneers a novel approach by employing readily available biochemical and physiological features from electronic health records to develop a machine learning-based binary classification model, enhancing accessibility and early detection. This study utilizes a dataset from Phachanukroh Hospital in Chiang Rai, Thailand, for model construction. A hybrid data enrichment framework using feature augmentation and data balancing was proposed to increase data dimensionality. Inter-relation-based Features (IRFs) were suggested as a means to enhance data diversity and promote explainability by making features more informative through the application of medical domain knowledge. To balance the data, K-Means Synthetic Minority Oversampling Technique (K-Means SMOTE) was applied to generate synthetic samples in underrepresented regions of the feature space, improving class imbalance handling. Extra Trees (ET) was proposed for model construction because of its noise resilience and ability to manage multicollinearity. The performances were compared with Support Vector Machine (SVM), K-nearest Neighbors (KNN), Artificial Neural Networks (ANN), Random Forest (RF), and Gradient Boosting GB. Results revealed that the ET model significantly outperformed other models for the combined dataset with four Inter-Relation-Based Features (IRFs) and K-Means SMOTE across key metrics, including accuracy (96.47 %), precision (94.79 %), recall (97.86 %), F1-score (96.30%), and area under the curve of the Receiver Operating Characteristic (99.51 %).

Keywords: Dementia, Classification, Machine learning, Oversampling Technique