Dengue Fever Prognosis Study

Kshitij Kadam
Texas A&M University
Department of Computer Science
and Engineering
kkadam3@tamu.edu

Adekola Okunola
Texas A&M University
Department of Electrical and
Computer Engineering
phirlly@tamu.edu

1 Dataset Reference

The dengue fever prognosis dataset contains gene expression data from peripheral blood mononuclear cells (PBMCs) collected from patients in the early stages of fever. The dataset includes gene expression profiles for 1981 genes and clinical outcomes categorized into classical dengue fever (DF), dengue hemorrhagic fever (DHF), and febrile non-dengue cases. [1]

Proposal for Dataset Analysis

Data Cleaning and Preprocessing

We will use dimensionality reduction techniques to handle the high-dimensional gene expression data. Principal Component Analysis (PCA) or t-SNE will be employed to reduce dimensionality while preserving variance and structure.

Feature Selection and Extraction

We will apply univariate feature selection methods, such as ANOVA F-tests or mutual information scores, to identify the most relevant genes for predicting DHF early in the disease's progression. For multivariate feature selection, Recursive Feature Elimination (RFE) combined with cross-validation will be implemented to iteratively select the best subset of features.

Classification Methods

To predict clinical outcomes based on gene expression profiles, we will use classifiers such as Support Vector Machine (SVM) and Random Forest. Linear Discriminant Analysis (LDA) will also be considered for its interpretability. Model performance will be evaluated using a k-fold cross-validation scheme to ensure robustness against overfitting. [2]

References

[1] Nascimento, E., Abath, F., Calzavara, C., Gomes, A., Acioli, B., Brito, C., Cordeiro, M., Silva, A., Andrade, C. M. R., Gil, L., and Junior, U. B.-N. E. M. (2009). Gene expression profiling during early acute febrile stage of dengue infection can predict the disease outcome. PLoS ONE, 4(11):e7892. doi:10.1371/journal.pone.0007892.

[2] Ulisses Braga-Neto, Fundamentals of Pattern Recognition and Machine Learning, Springer Nature Switzerland AG, 2020. ISBN 978-3-030-27655-3. DOI: 10.1007/978-3-030-27656-0