Robust and Interpretable Feature Selection Pipeline

Student: Nithya Krishnan   Supervisor: A.S. Madhukumar

Variation 1

| Normalized Dataset | → | Causality-based method, A | → | Redundancy-minimization method, W | → | Selected Features |
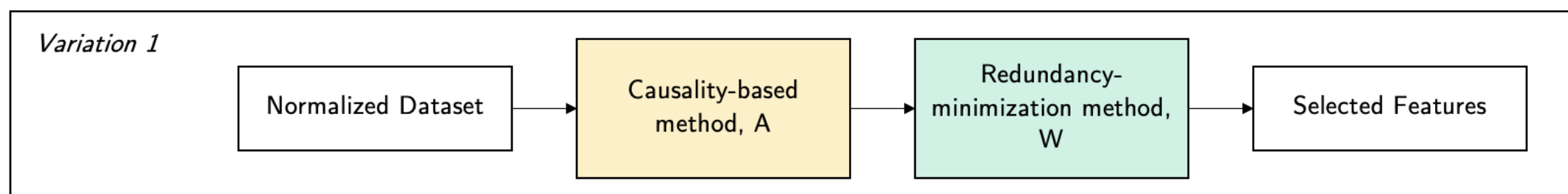
Project Objectives:

This project aims to design a novel feature selection pipeline that reduces the dimensionality of datasets without a significant drop in performance. This is done through the removal of redundant features while retaining relevant features. The pipeline proposes a stacked architecture using established feature selection methods.
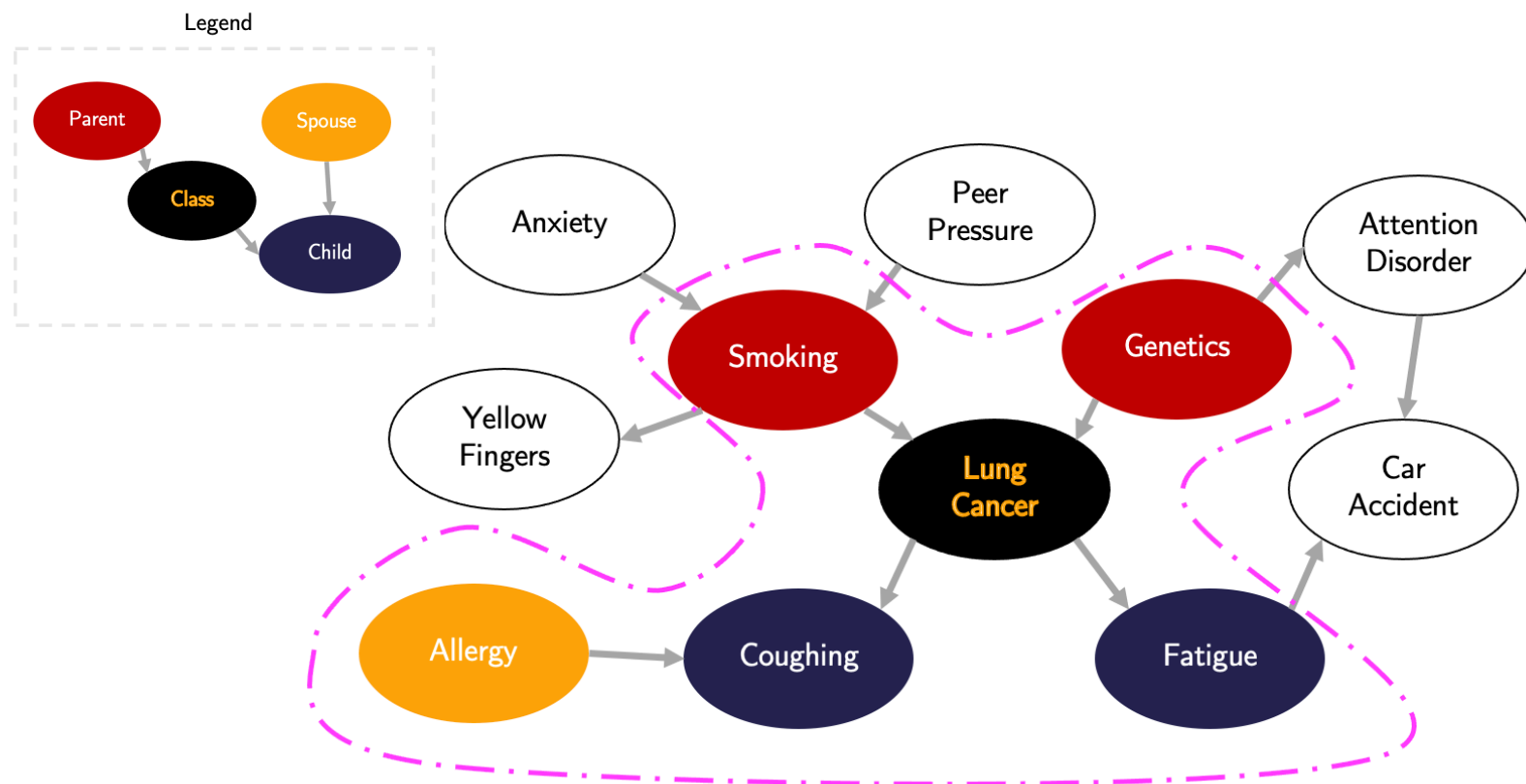
Causality-Based:

We first select relevant features through causality-based feature selection. We look for features that have a causal relationship with the class variable instead of mere co-occurrences

Principal Feature Analysis:

The latter part of the architecture consists of redundancy minimization through the Principal Feature Analysis algorithm, where we use Singular Vector Decomposition to obtain eigenvector components and perform clustering on the components to group features based on correlation.

| Standardised Dataset | → | Eigenvector Ranking (PCA based) | → | Clustering | → | Non-redundant Features |