

An Integrated Multi-Clustering Approach for Patient Stratification and Disease Prediction Using Electronic Medical Records in Vector-Borne Diseases

Dr. S. Muthukumaran

Assistant Professor

Department of Advanced
Computing and Analytics

Vels Institute of Science and Advanced Studies, Chennai.

Ms. V. Nivathra

Assistant Professor

Department of Pharmaceutical Chemistry and Analytics,
Vels Institute of Science and Advanced Studies, Chennai.

Dr. P. Arivazhagan

Assistant Professor

Department of Advanced
Computing and Analytics

Vels Institute of Science and Advanced Studies, Chennai.

Abstract

An electronic medical record comprises both structured and unstructured pieces of patient information, including but not limited to personal identifiers, vital signs, lab test results, physician's notes for diagnosis, and prescriptions. Clustering techniques are one means to EMR data to cluster patients having similar symptoms based on their laboratory results and to identify anomalies, for example, due to faulty data entry. **Objective:** However, the very large numbers of records and the extremely high dimensionality of EMR datasets make the computational burden really substantial in developing machine learning models for predicting diseases. **Methods:** In this regard, this research anticipates the communication of an **Integrated Multi-Clustering Algorithm (IMCA)** that combines and integrates all the three clustering methods: K-Means, Agglomerative, and DBSCAN clustering algorithms. A dataset on Vector borne Diseases for the study consists of 64 manifestations in relation to 11 different types of fever. The output of an individual clustering algorithm in the IMCA was then evaluated on a personal basis to classify patients with like disease conditions or symptoms and to assist a clinician in identifying subgroups benefiting from specific effective treatments. **Results:** Calinski-Harabasz Index (CHI), Davies-Bouldin Index (DBI), and Silhouette Score are various measures for evaluating the clustering algorithms' performance. The results claimed that clustering by K-means was producing more balanced clusters with a more favourable CHI value but slightly lower compactness compared with Agglomerative and DBSCAN methods.

Keywords: *Vector borne Disease, K-Means Clustering, Agglomerative Clustering, DBSCAN Clustering.*