

K-means Clustering and Hierarchical Cluster Analysis Coupled with Linear Discriminant Analysis to Classify Signals in Osmotic Fragility Test for Thalassemia Screening

**Karuna Jainontee^{1,4}, Vannajan Sanghiran Lee^{2*},
Sukon Prasitwattanaseree^{3,4*} and Kate Grudpan^{1,4}**

¹*Department of Chemistry, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*

²*Department of Chemistry, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia*

³*Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand*

⁴*Center of Excellence in Innovation for Analytical Science and Technology, Chiang Mai University, Chiang Mai 50200, Thailand*

**Corresponding authors: E-mail: vannajan@gmail.com,
sprasitwattanaseree@gmail.com*

ABSTRACT

Investigation has been made in applying chemometric treatment of unbiased approaches to classify signals in osmotic fragility test for thalassemia screening. K-means clustering and hierarchical clustering analysis coupled with linear discriminant analysis were the chemometric techniques employed in this work. A knowledge determined from conventional approaches in osmotic fragility test together with multivariate analysis provides a complementary tool and novel approach for disease diagnosis.

Keywords: K-means clustering, Hierarchical clustering analysis, Linear discriminant analysis, Thalassemia screening, Osmotic fragility test, Flow injection

INTRODUCTION

The osmotic fragility test (OFT) is a popular method for thalassemia screening because of its high efficiency and low cost (Hartwell et al., 2005). OFT is a test that measures the ability of red blood cells to retain their integrity in hypotonic saline solution. The surface areas of red blood cells of a thalassemia patient (positive tests) are larger than in a normal person (negative tests), so the blood cells are more likely to breakdown faster than in negative tests. OFT, coupled with stopped flow injection, was developed as an automated thalassemia screening tool that reduce the risk of contamination in the system (Khonyoung et al., 2009). The signals of stopped flow OFT are used to discriminate patients. The method can classify patients by using the optimized slope from the optimization based on hospital records (Khonyoung et al., 2009). However, no evidence exists guaranteeing its accuracy in predicting unknown samples.