



---

## **MICROARRAY BASED ACUTE LYMPHOCYTIC LEUKEMIA (ALL) AND ACUTE MYELOGENOUS LEUKEMIA (AML) CLASSIFICATION USING DATAMINING**

Edirimanna C.L.<sup>1</sup>, Jinasena T.M.K.K.<sup>2</sup> and Edirisuriya E.A.T.A.<sup>2</sup>

<sup>1</sup>Faculty of Information Technology, University of Moratuwa, Sri Lanka

<sup>2</sup>Department of Computer Science, University of Sri Jayawardanapura, Sri Lanka

lakminiedirimanna@gmail.com

---

### **ABSTRACT**

Acute leukaemia is a type of blood cancer exists in bone marrow which needs immediate treatment than the others. Conventional lab methods take more time to differentiate these types and the invention of the micro array technology made advancement in cancer diagnosis and prognosis. However, Gene expression data indicates curse of dimensionality problem which makes it difficult to find associations and patterns across multiple dimensions. Thus, reduce the dimensionality is required prior to the microarray data classification. Extracting disease related genes is a research challenge problem in microarray data. In here, used benchmark microarray data set consisting of 72 samples with 7000 ttributes which represent curse of dimensionality problem. Filter is one main strategy in extracting relevant genes which basically concern on instric property of the genes regardless of the class predictor. There exists number of filter criteria and gene lists which, when selected by different filter methods vary from each other; same gene obtained a different ranks by different filter methods .Thus; it is a problem to determine which gene list is most suitable for further analysis. Therefore, this research has proposed a method that integrates filter methods through consensus score and multi filtration based method in parallel and sequence approache. Finally, ensemble classifiers used to predict the types of luekamia. Performance of the classification models has been evaluated in terms of accuracy; sensitivity and specificity .In the evaluation, parallel approach perform well other than in terms of accuracy, sensitivity and specificity with average values respectively 98.56, 98.87 and 99.1.

**Keywords:** Bio Informatics, Leukaemia, Multi filters, Data Mining