

# Deep learning-based classification of diffuse large B-cell lymphoma cell-of-origin by gene expression profiles

R. Kumar<sup>1</sup>, **K. Kundal**<sup>1</sup>, A. Viswanathan<sup>2</sup>, A. Sengupta<sup>1</sup>, A. Kumar<sup>1</sup>, K. V. Kumar<sup>1</sup>

<sup>1</sup> *Indian Institute of Technology Hyderabad, Biotechnology, Sangareddy, India*

<sup>2</sup> *Amrita Vishwa Vidyapeetham, School of Biotechnology, Amritapuri, India*

Diffuse large B cell lymphoma (DLBCL) is a fast growing and the most aggressive form of non-Hodgkin's lymphoma. DLBCL cases were categorized into two subtypes i.e., ABC and GCB based on gene expression and of these, ABC subtype is associated with poor outcomes. This classification is also known as cell-of-origin (COO) and found to be an efficient method of predicting response to R-CHOP therapy. Limited number of samples and inclusion of large number of genes were the major drawback of previous COO classifiers. These limitations undermine the robustness of the classifier models and also make it difficult to implement them in clinics. To overcome these limitations, we have developed a robust deep learning-based classifier model using expression data of 20 genes in 381 DLBCL cases (Figure 1). We implemented a multi-layer perceptron (MLP) to train the classifier model called MLP-COO. MLP-COO achieved 99.70% and 94.70% accuracy on the training and test datasets, respectively, with 10-fold cross-validation. We also evaluated its performance using an independent dataset of 294 DLBCL cases and achieved accuracy of 95.90% with an MCC of 0.917. We employed this classifier model to predict the clinical outcome using survival data from two large cohorts of DLBCL cases to demonstrate its wider applicability. Based on their COO in both cohorts, MLP-COO recapitulates the survival outcomes of DLBCL cases (Figure 2). We anticipate that MLP-COO will be useful in accurately predicting the COO of DLBCL cases and their clinical outcomes.