

## Identification of Macrophage-Related Gene Signature of Hepatocellular Carcinoma by using Knowledge-based and Forward Feature Selection with TCGA Database

### **Shih-Yen Weng**

Smart Healthcare Interdisciplinary College, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

### **Chao-Yang Kuo**

Smart Healthcare Interdisciplinary College, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

### **Yen-Chang Huang**

Department of Applied Mathematics, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

### **Shu-Fang Vivienne Wu\***

School of Nursing, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

### **Abstract**

Hepatocellular carcinoma (HCC) is one of the leading causes of cancer-related death. Studies have revealed that changes in immunity within the liver microenvironment may play a crucial role in the development of HCC. Macrophages, the largest group of innate immune cells in the liver, have been linked through their polarization to liver cancer progression. Thus, we believe that macrophage markers, which can be detected in circulation, may serve as ideal non-invasive targets for HCC detection. Based on previous studies, we organized macrophage gene-expression datasets of HCC from the TCGA and GTEx databases for machine-learning analysis. Forward feature selection using Regression Trees, Random Forests, and XGBoost was applied in a benchmark model evaluation pipeline. XGBoost showed the best performance (ACC = 0.983) in prediction, and common genes identified across the three iterations of feature selection—including TLR2, SOCS1, MARCO, STAT6, NOS2, IL6, and C1QA—were selected for model interpretation. Model interpretations were conducted using explainable AI methods, including feature importance and SHAP analysis. Feature importance indicated that the most influential genes were C1QA, NOS2, and MARCO. SHAP analysis further demonstrated that TLR2, SOCS1, and MARCO were significantly increased in HCC patients. These genes warrant more detailed investigation for their potential in predicting HCC. In summary, we have identified a novel set of macrophage biomarkers that may reflect the development of HCC.

### **Index Terms**

Liver Cancer, HCC, Macrophage, Machine Learning, Feature Selection, SHAP, Shapley Value