2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Cascade Flexible Neural Forest Model for Classification of Cancer Subtypes Based on Gene Expression Data

Lianxin Zhong, Qingfang Meng & Yuehui Chen

Abstract—**The correct classification of cancer subtypes is of great significance for the in-depth study of cancer pathogenesis and the realization of accurate treatment for cancer patients. In this paper, the cascade flexible neural forest (CFNForest) model was proposed to accomplish cancer subtype classification.**

Clinical Relevance—**Experiments on RNA-seq gene expression data showed that CFNForest effectively improved the accuracy of cancer subtype classification.**

I. Introduction

Nowadays, cancer has become one of the major causes of human death. There are many molecular subtypes in cancer tissue. Cancer patients with the same symptoms can show significant prognostic differences under the same treatment regimens. The occurrence, development and metastasis of cancer is a complex and multifactorial process. Different cancer subtypes differ significantly in multiple gene expression data. Therefore, cancer research at the genetic level is of great importance for cancer treatment and diagnosis.

II. Methods

In this paper, the cascade flexible neural forest (CFNForest) was proposed for cancer subtype classification. This model is a deep neural network ensemble model based on FNT Group Forest. Different from traditional deep neural network models, CFNForest can automatically optimize the structure and parameters of the internal neural network during the training process. Compared with deep models built on non-differentiable models, the proposed model can stratify genetic features without discrete input features. CFNForest integrates multiple classifier ensemble strategies and improves the overall classification accuracy of the model through feature conversion between levels. Through the introduction of enhanced sample and feature optimization mechanism, the model still shows good classification performance on small sample datasets.

We took a FNT group as a whole and used the bagging ensemble strategy to form a FNT Group Forest and put K FNTs groups on each node. The structure of CFNForest Model was shown in Fig.1. we generated three forests by different grammars. The red forest used the function set {+2, +3, +4}; The blue forest used the function set {+2, +3, +5}; The purple forest used the function set {+2, +4, +5}.

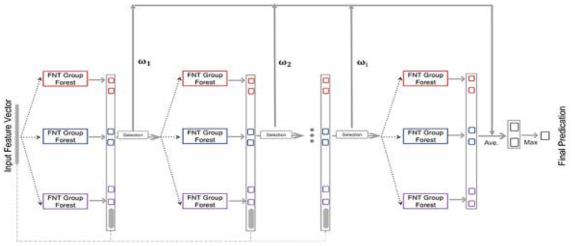

Fig.1 A Cascade Flexible Neural Forest Model.

III. Results

Three types of cancers were used in this paper: Breast Invasive Carcinoma (BRCA), Glioblastoma Multiforme (GBM) and Lung Cancer (LUNG). We compared the proposed model with k-nearest neighbor (KNN), support vector machine (SVM), multi-layer perception (MLP), random forest (RF) and multi-grained cascade forest (gcForest), respectively. The classification results were shown in Fig.2.

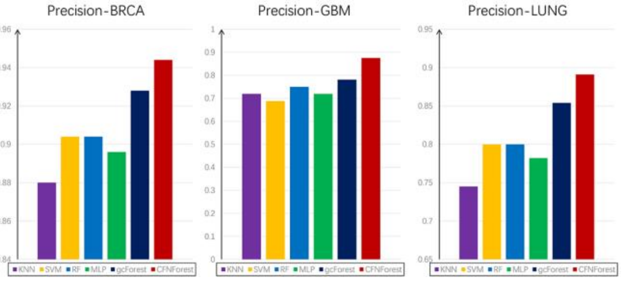

Fig.2 The average precision comparison of classifiers on three datasets.

IV. Discussion & Conclusion

Objective and accurate classification of cancer subtypes enables doctors to correctly understand the pathogenesis and primary location of cancer, which is of great significance to the study of cancer genesis. In this paper, a new cancer subtype classification model was proposed, which is capable of hierarchical processing of gene features without discrete input features compared with deep models built on non-differentiable models. Experimental results showed that the proposed model consistently outperforms the state-of-the-art methods in classifying cancer subtypes by using RNA-seq gene expression data.

*Research supported by the National Natural Science Foundation of China (Grant No.61671220), University Innovation Team Project of Jinan (2019GXRC015), Key Science and Technology Innovation Project of Shandong Province (2019JZZY010324).

LX Zhong, QF Meng and YH Chen are with the School of Information Science and Engineering, University of Jinan, China (corresponding author to provide e-mail: ise_mengqf@ujn.edu.cn).

978-1-7281-1178-0/21/$31.00 ©2021 IEEE