МИНОБРНАУКИ РОССИИ

Федеральный исследовательский центр «Информатика и управление» РАН
Национальный комитет РАН по распознаванию образов
и анализу изображений
Институт информационных технологий Министерства науки и образования
Азербайджанской Республики
Институт проблем передачи информации им. А. А. Харкевича РАН
Белорусский государственный университет
Национальный исследовательский Томский государственный университет
Федеральный исследовательский центр «Карельский научный центр РАН»
Ошский государственный университет
Юго-Западный государственный университет

ОПТИКО-ЭЛЕКТРОННЫЕ ПРИБОРЫ И УСТРОЙСТВА В СИСТЕМАХ РАСПОЗНАВАНИЯ ОБРАЗОВ И ОБРАБОТКИ ИЗОБРАЖЕНИЙ

Распознавание – 2023

Сборник материалов XVII Международной
научно-технической конференции

12–15 сентября 2023 года

Редакционная коллегия:
С. Г. Емельянов, В. С. Титов (отв. ред.),
Т. А. Ширабакина, Э. И. Ватутин, Е. А. Коломиец

Курск
ЮЗГУ
2023

2. The health digital twin to tackle cardiovascular disease—a review of an emerging interdisciplinary field / G. Coorey, G. A. Figtree, D. F. Fletcher [et al.] // NPJ Digit. Med. 2022. Vol. 5(1). P. 126. DOI 10.1038/s41746-022-00640-7.

3. Imran Ahmed, Misbah Ahmad, Gwanggil Jeon. Integrating digital twins and deep learning for medical image analysis in the era of COVID-19 // Virtual Reality & Intelligent Hardware. 2022. Vol. 4, is. 4. P. 292–305. DOI 10.1016/j.vrih.2022.03.002.

UDC 004.93

M. H. Mammadova[1], Z. G. Jabrayilova[1], L. A. Garayeva[1]
e-mail: mmg51@mail.ru; djabrailova_z@mail.ru; karayevalala.01@gmail.com

[1]Institute of Information Technology, Baku, Azerbaijan

ALGORITHM FOR EARLY DIAGNOSIS OF HEPATOCELLULAR CARCINOMA BASED ON GENE PAIR SIMILARITY

The article presents an algorithm for early diagnosing hepatocellular carcinoma, also known as liver cancer. The algorithm determines the similarity of HCC tissues to similar CwoHCC non-cancerous tissues by the minimum deviation and maximum matching of gene pairs. It identifies the characteristics of optimal gene pairs selected based on machine learning methods.

Hepatocellular carcinoma (HCC), which accounts for about 90% of all liver cancer cases, is often diagnosed in the late stages of the disease and therefore causes a high risk of death. Consequently, early diagnosis of HCC is very important in the disease prevention and increases the patient’s survival probability. Currently, the diagnosis of HCC is based on laboratory studies and computed tomography (CT), and X-ray examination. Liver biopsy is estimated to be a good diagnostic option in the clinical condition when CT and X-ray examination cannot provide accurate identification of HCC [1]. Sometimes, non-cancerous tissues (cirrhotic tissues and normal tissues) containing some common molecular features of cancerous tissues are recognized as cancerous tissues. In such cases, gene analysis signatures are included among the available diagnostic signatures to eliminate the risk factor. Gene analysis signatures have a batch effect and are difficult to determine in clinical conditions.

The present article the question of determining the similarity of HCC cancer tissues with identical CwoHCC non-cancerous tissues for the HCC identification, and proposes a machine learning-based solution technique.

Based on the gene analysis of HCC cancer tissue, for its early diagnosis a solution algorithm in the following stages is proposed.

1. ***Gene Expression of HCC tissues.*** Differentiation of HCC cancerous tissue and similar non-cancerous (CwHCC and NwHCC) tissues according to selected genes is used for early HCC diagnosis. [2] performs the solution to this problem

on the basis of 10 genes included in the range of risk factors. The association of each gene with cancerous tissue and similar non-cancerous tissue is determined.

2. ***Gene Expression profiling***. The Gene Expression Omnibus (GEO) and The Cancer Genome Atlas (TCGA) databases can be used for a gene expression profiling. Initially, a database storing relevant HCC biopsy samples (D1), HCC surgical samples (D2), CwoHCC biopsy samples (D3), and CwoHCC surgical samples (D4) is generated. To ensure the objectivity of the created model, the samples of each mentioned type are divided into two data subsets: training data set (80 % samples from each type) and test data set (20 % samples from each type).

3. ***Determination of the Within-Sample Relative Expression Ordering.*** Relative Expression Orderings (REO) technique is used for feature extraction. According to the REO technique, if the gene *a* has a higher analysis level than the gene *b* (or vice versa) in a given sample, they are analyzed as *Ea* >*Eb* (or *Ea* <*Eb*). If at least 95% of the samples for a gene pair have the same gene pair ordering, then that gene is considered to be stable according to the REO technique.

4. ***Feature selection with minimum Redundancy Maximum Relevance (mRMR) and Incremental feature selection (IFS) methods.*** Based on the new profiles, the mRMR method is applied to rank the HCC cancer and non-cancer gene pairs within minimum Redundancy Maximum Relevance conditions. Interaction between genes *I* is defined as:

$$I(g_i,T)=\int p(g_i,T)\ln\left(\frac{p(g_i,T)}{p(g_i)p(T)}\right)dg_i dT. \qquad (1)$$

Here, $I(g_i, T)$ denotes the interaction between the gene pair g_i and type disease *T*. The following formula is used to determine the relevance by all gene pairs:

$$mRMR=\frac{1}{|\Omega|}\sum\nolimits_{g_i\in\Omega} I(g_i,T)-\frac{1}{|\Omega|^2}\sum\nolimits_{g_i,g_j\in\Omega} I(g_i,g_j). \qquad (2)$$

Here Ω represents all given gene pairs, g_i – one of given gene pairs, $I(g_i, g_j)$ – interaction between the genes g_i and g_j. In the next step, optimal gene pairs are selected from the mRMR gene pair in the sample given as a candidate signature, and the IFS method is used for this purpose.

5. Classification of feature. Machine learning methods Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF)) can be used for feature classification.

6. Determination of classification and confusion matrix criteria. Evaluating the classifiers' detection performance is of great importance in machine learning. The criteria of precision, recall, false positive rate (FPR), true positive rate (TP), f-measure and accuracy are used in the evaluation of detection performance.

The proposed algorithm was aimed at improving the HCC diagnosis [3]. In the further studies, it is planned to perform experiments with the application of the SVM, RL, LM classification methods based on the similarity of gene pairs and to select the method with better performance according to the performance evaluation.

REFERENCES

1. Quantitative or qualitative transcriptional diagnostic signatures? A case study for colorectal cancer / Q. Guan, H. Yan, Y. Chen [et al.] // BMC Genomics. 2018. Vol. 19. P. 99. DOI 10.1186/s12864-018-4446-y.

2. Zi-Mei Zhang, Jiu-Xin Tan, Fang Wang. Early diagnosis of hepatocellular carcinoma using machine learning method // Frontiers in Bioengineering and Biotechnology. 2020. Vol. 8 (254). DOI 10.3389/fbioe.2020.00254.

3. Prediction of hepatocellular carcinoma using a machine learning / M. H. Mammadova, Z. G. Jabrayilova, L. A. Garayeva, A. A. Ahmadova // The 16th IEEE International Conference Application of Information and Communication Technologies. Washington DC, 2022. DOI 10.1109/AICT55583.2022.10013575.

UDC 004.85

M. H. Mammadova[1], Z. G. Jabrayilova[1], N. R. Shikhaliyeva[1]
e-mail: mmg51@mail.ru; djabrailova_z@mail.ru; shikhaliyeva.nara@gmail.com

[1]Institute of Information Technology, Baku, Republik of Azerbaijan

CLASSIFICATION OF PATIENT REVIEWS BASED ON INFORMATION IN MEDICAL MEDIA RESOURCES

The article proposes an approach to the use of information collected in the medical social media environment for medical decision-making. Sentiment analysis of the information collected in the attitude segments formed in media resources and review classification algorithm are presented.

Currently, a large number of professional medical social communities have emerged in the Internet environment. The information collected in this environment becomes a valuable resource for decision-making in relevant fields, and intelligent technologies are the utmost required tool for achieving successful results in this field.

One of the important points here is the content analysis of the information related to the physicians, patient, medical institution, which are media subjects, and determining the opinion about media subjects in applications.

The expansion of medical social media, the activity of medical specialists, physicians, patients, and medical clinics on social media has led to the formation of