JOURNAL OF DECISION MAKING AND HEALTHCARE
Volume 1 (2024), No. 1, 30–44
https://doi.org/10.69829/jdmh-024-0101-ta03
Tulipa Opera Scholarum

GENE REDUCTION FOR CANCER DETECTION USING LAYER-WISE RELEVANCE PROPAGATION

SHENG-YI HSU[1], MAU-HSIANG SHIH[2], WU-HSIUNG WU[2], HAO-REN YAO[3], AND FENG-SHENG TSAI[2,4,*]

[1]*Ever Fortune.AI Co., Ltd., Taichung 40360, Taiwan*
[2]*Research Center for Interneural Computing, China Medical University Hospital, China Medical University, Taichung 40402, Taiwan*
[3]*Rehabilitation Medicine Department, Clinical Center, National Institutes of Health, Bethesda, MD 20892, USA*
[4]*Department of Biomedical Engineering, China Medical University, Taichung 40402, Taiwan*

Abstract. Precise detection of cancer types and normal tissues is crucial for cancer diagnosis. Specifically, cancer classification using gene expression data is key to identify genes whose expression patterns are tumor-specific. Here we aim to search for a minimal set of genes that may reduce the expression complexity and retain a qualified classification accuracy accordingly. We applied neural network models with layer-wise relevance propagation (LRP) to find genes that significantly contribute to classification. Two algorithms for the LRP-candidate gene selection and the cycle of gene reduction were proposed. By implementing the two algorithms for gene reduction, our model retained 95.32% validation accuracy to make classification of six cancer types and normal with a minimal set of seven genes. Furthermore, a cross-evaluation process was performed on the minimal set of seven genes, indicating that the selected marker genes in five out of six cancer types are biologically relevant to cancer annotated by the COSMIC Cancer Gene Census.

Keywords. Cancer detection, fully connected networks, gene expression, layer-wise relevance propagation, oncogene.

1. Introduction

Gene expression data increase massively and become available publicly. That raises lots of efforts to unravel the secrets within, especially the mutated genes implicated in oncogenesis [9, 26]. Indeed, researchers have demonstrated that cancer is caused primarily by mutations in specific genes [5, 20, 23]. Those mutations may disable or overact gene expression that transforms normal cells into cancer. So the measure of gene expression can generate big digital data learned to distinguish different disease phenotypes. Identifying those changes in gene expression may yield potentially valuable biomarkers, providing significant insights into the diagnosis and treatment of cancer [10].

The Cancer Genome Atlas (TCGA) catalogs the genomic changes involved in cancer. It is the most comprehensive resource for exploring the impact of genes. Computational modeling techniques are required to analyze TCGA data, improving cancer care from data-driving insights [3, 8, 14, 19]. A central issue of interest to us is cancer detection. That aims at the classification of cancer types and identification of biomarkers for multiclass cancer [15, 20]. Artificial neural network models have been used recently for gene expression-based classification of cancer [1, 6, 16, 18]. Those artificial neural network models outperform the support vector machines (SVM) or the genetic algorithm/k-nearest neighbors algorithm (GA/KNN) in pan-cancer and normal classification. Indeed, Ramaswamy et al. attained 78%

*Corresponding author.
E-mail addresses: shengyi.hsu@everfortune.ai (S.-Y. Hsu), mhshih@math.ntnu.edu.tw (M.-H. Shih), wwh@cs.ccu.edu.tw (W.-H. Wu), hao-ren.yao@nih.gov (H.-R. Yao), and fstsai@mail.cmu.edu.tw (F.-S. Tsai)
Accepted: May 30, 2024.

TABLE 1. FPKM Data for Six Tumor Types and Normal Tissues

Tumor Types	Cancer samples	Normal samples
BRCA	1097	113
KIRC	538	72
LUAD	524	59
PRAD	498	52
THCA	502	58
UCEC	551	35

classification accuracy of 14 cancer types based on SVM [20]. Li et al. used GA/KNN to identify many sets of 20 genes that attained on average 87.6% classification accuracy of 31 cancer types and normal tissues [13]. By contrast, concerning the artificial neural network models, Mostavi et al. achieved 93.9–95% prediction accuracy among classes of 33 cancers and normal [18], far exceeding the accuracy performance of SVM and GA/KNN methods. A remarkable feature of artificial neural network models is the interpretability developed for understanding how the networks work. The interpretable tools, such as the saliency (SA) visualization method [25], the gradient-weighted class activation mapping (Grad-CAM) [24], and the layer-wise relevance propagation (LRP)[2, 17], highlight the input features the networks use to support their prediction. They successfully unravel hot zones in data of 2D or 3D geometric structures, e.g., images in ImageNet [7, 12], that fit well with human intuition. However, for data of 1D structure, it is hard to have an intuitive, geometric explanation of the extracted hot zones. Recently, for data of non-geometric 1D structure, e.g., gene expression profiles, studies have shown that the interpretable tools can potentially extract meaningful patterns from them. Indeed, Lyu and Haque validated that the top 400 genes with high intensities in Grad-CAM of a convolutional neural network are related to tumor-specific pathways [16]. Mostavi et al. identified 2090 marker genes with SA scores greater than 0.5, showing some of them are well-known oncogenes in breast cancer [18].

We will show in this study that LRP serving as a backward propagation technique in networks can be employed to identify marker genes of non-geometric structures. Further, we concern about the question: can the marker genes derived from one LRP be used as a reduced training pool to construct a refined neural network model and then apply LRP again to get marker genes of smaller sizes? The answer to this question in data of non-geometric gene expression profiles is positive. Indeed, we develop two algorithms, namely the LRP-candidate gene selection and the cycle of gene reduction, to minimize the number of marker genes to probe the entwined interactions related to cancer. We propose two hypotheses to explore such relations. The first hypothesis is to identify marker genes in the minimal gene set that can be replaced by other genes in cancer classification, whereas the second hypothesis is to identify marker genes in the minimal gene set that are stably involved in cancer classification. On this basis, we can deduce two or three marker genes that may be tumor-specific for a cancer type. We show that the inferences are concordant with well-known marker genes implicated in cancer.

2. Data Acquisition and Augmentation

The cancer-driving genes were selected from the COSMIC Cancer Gene Census (https://cancer.sanger.ac.uk/census) [9, 26]. A list of 723 cancer-driving genes (version 92, August 2020) was available for download. It allowed us to concentrate on a group of cancer-related genes, ruling out many possibly cancer-irrelevant attributes in cancer classification. Of those 723 genes, 576 in Tier 1 possessed a documented activity and evidence of mutations relevant to cancer, whereas 147 in Tier 2 showed supportive indications but less available evidence in cancer. We downloaded RNA-seq FPKM data from The Cancer Genome Atlas (TCGA, https://www.cancer.gov/tcga) database for six tumor types and normal tissues, namely, breast invasive carcinoma (BRCA), kidney renal clear cell

TABLE 2. Classes for Model Training

Class	Data type	Training samples	Validation samples
I	Cancer (BRCA)	400	150
II	Cancer (KIRC)	400	138
III	Cancer (LUAD)	400	124
IV	Cancer (PRAD)	400	98
V	Cancer (THCA)	400	102
VI	Cancer (UCEC)	400	151
VII	Normal (BRCA)	$67 \rightarrow 67$	46
	Normal (KIRC)	$42 \rightarrow 67$	30
	Normal (LUAD)	$35 \rightarrow 67$	24
	Normal (PRAD)	$30 \rightarrow 67$	22
	Normal (THCA)	$34 \rightarrow 67$	24
	Normal (UCEC)	$21 \rightarrow 67$	14

$\rightarrow$: borderline-SMOTE

carcinoma (KIRC), lung adenocarcinoma (LUAD), prostate adenocarcinoma (PRAD), thyroid carcinoma (THCA), and uterine corpus endometrial carcinoma (UCEC). The selection of those six cancer types relied on the numbers of samples in both cancer and normal tissues sufficient for model training. The gene expression data of 716 out of 723 cancer-driving genes were used for model training and validation, excluding genes IGH, IGK, IGL, TRA, TRB, and TRD that lacked Ensembl stable IDs and gene DUX4L1 that had no expression data. The detailed composition of the data set is provided in Table 1.

For model training, samples of six cancer types were defined as six classes, respectively, and normal samples were all merged as one class. To improve the class imbalance, we used the borderline synthetic minority over-sampling technique (borderline-SMOTE) [4, 11] to increase the number of normal samples. For each class, about 400 samples were selected randomly as the training set and the remaining 100–150 samples were put as the validation set.

3. METHODS

We developed two algorithms, namely the LRP-candidate gene selection and the cycle of gene reduction, to tackle the cancer-and-normal classification. These two algorithms reduced the number of critical genes and evaluated their classification accuracy based on the architecture of multi-layer fully connected networks (FCNs) [21, 22] and LRP [2]. We introduced the average relevance score in the first algorithm to extract LRP-candidate genes and the reduction cycle in the second algorithm to check the significance of extracted genes in classification.

Specifically, we used a simple four-layer FCN to classify input samples into seven classes (six cancer types and normal). The input layer of the FCN contained neurons receiving the expression profiles of selected cancer-driving genes and passing to the neurons in the next layer. The second and third layers contained 300 and 100 neurons, respectively, for processing signals from the previous layer and forwards to the next layer. The output layer of the FCN contained 7 neurons, equipped with the softmax function for classification of samples into seven classes. The training of the FCN was stopped after 1000 epochs and LRP was utilized to determine which genes can make the FCN effective in cancer-and-normal classification. When a sample was fed into the FCN which returned a correct output, LRP redistributed this correct output back to its predecessors and, layer by layer, to the input layer according to the rule:

$$R_i = \sum_{j \in L_k, y_j \neq 0} \frac{w_{ij} y_i}{y_j} R_j \tag{3.1}$$

Algorithm 1. LRP-candidate gene selection

Input: Samples of gene expression data are divided into different classes according to cancer types and normal; Train a multi-layer FCN by one portion of those samples with n genes and test by the other portion;

Output: LRP-candidate genes selected from n genes;

for all samples in the validation portion **do**

Input a sample s to the FCN that yields a network output y indicating the correct class;

Initialize the R-score of neuron j in the final layer of the FCN by the output y if j corresponds to the correct class, otherwise 0;

for consequent layers L_{k-1} and L_k of the FCN **do**

Calculate the R-score of neuron i in the prior layer L_{k-1} by

$$R_i \leftarrow \sum_{j \in L_k, y_j \neq 0} \frac{w_{ij} y_i}{y_j} R_j,$$

where R_j is the R-score of neuron j in the posterior layer L_k, w_{ij} is the coupling weight from neuron i to neuron j, and y_i and y_j are the outputs of neurons i and j, respectively, when processing the sample s with the coupling weights w_{ij} of the FCN;

end for

Denote by $R_1^s, R_2^s, \ldots, R_n^s$ the R-scores calculated in the input layer L_0 for the sample s;

Average the R-scores $R_1^s, R_2^s, \ldots, R_n^s$ and get

$$\bar{R}^s \leftarrow \left(\sum_{m=1}^{n} R_m^s \right) / n;$$

Let $B_m^s = 1$ if the R-score of the mth gene is greater than the average R-score for the sample s, i.e., $R_m^s > \bar{R}^s$, otherwise $B_m^s = 0$;

end for

Calculate the rate of samples taken from a class C such that $B_m^s = 1$ for each given $m = 1, 2, \ldots, n$, i.e.,

$$P_m^C \leftarrow \left(\sum_{s \in C} B_m^s \right) / \sharp C,$$

where $\sharp C$ denotes the number of samples in the class C and P_m^C denotes the SR-sample rate of the mth gene for the class C;

Set the threshold ρ for the SR-sample rates by

$$\rho \leftarrow \min_C \max_m P_m^C;$$

return all of the LRP-candidate genes $m \in \{1, 2, \ldots, n\}$ satisfying $P_m^C \geq \rho$ for at least one class C;

for each neuron i in the prior layer L_{k-1}, where R_j is the relevance score (namely, R-score) of neuron j in the posterior layer L_k, w_{ij} is the coupling weight from neuron i to neuron j, and y_i and y_j are the outputs of neurons i and j, respectively, when processing the sample by the coupling weights w_{ij} of the FCN. As a result, each neuron in the input layer of the FCN was assigned an R-score indicating the effect of the corresponding gene's expression on cancer-and-normal classification (see Fig. 1(a) and (b)

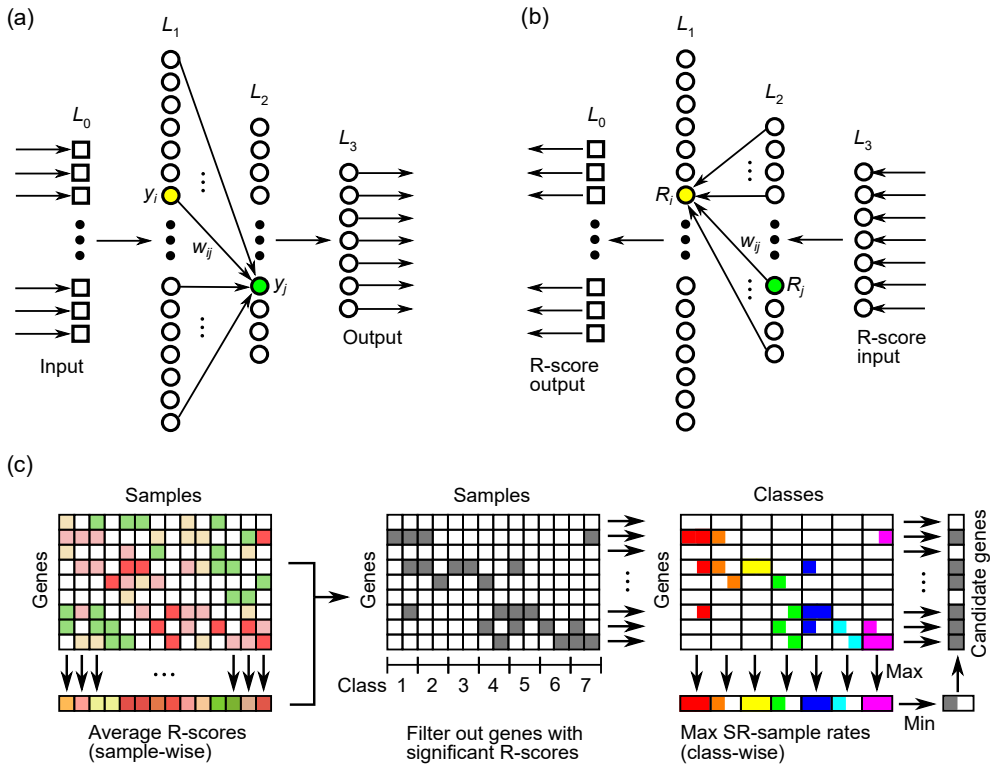

FIGURE 1. Selection of LRP-candidate genes. (a) A simple four-layer FCN implements a feedforward process (layer L_0 to L_3) to make classification of six cancer types and normal based on gene expression. Each neuron in the posterior layer (e.g., green node) generates an output y_j by rectifying incoming signals y_i from all neurons in the prior layer (e.g., yellow node) with the coupling weights w_{ij}. (b) LRP implements a backpropagation process (L_3 to L_0) to redistribute layer-wise relevance from the output of the FCN to the input. A more effective contribution in the feedforward process (e.g., arrow from yellow node to green node in (a)) earns more feedback on the R-score R_i in the backpropagation process (e.g., arrow from green node to yellow node in (b)). An R-score calculated in the input layer of the FCN (R-score output) stands for the effect of a gene's expression on cancer-and-normal classification. (c) A schematic illustration shows the R-scores of genes for each sample (blocks in the left panel; red: high R-scores; green: low R-scores). The R-scores are averaged sample-wise in order to filter out genes (gray blocks in the middle panel) that are significant in a sample. For a specific gene, we count the rate of samples taken from a class, in which the significance of the specific gene is identified (transform gray blocks in the middle panel into color blocks in the right; long color blocks: high SR-sample rates; short color blocks: low SR-sample rates). The maximal SR-sample rate is taken over all genes in a class (down arrows in the right panel). The minimum of the maximal SR-sample rates is taken across all classes (gray block in the lower right corner) and used as a threshold to extract candidate genes (right arrows in the right panel) whose SR-sample rates in at least one class are higher than the threshold. SR: significant relevance.

for an illustration). An R-score greater than 0 (resp., less than 0) means positive (resp., negative) effect, and 0 indicates no effect.

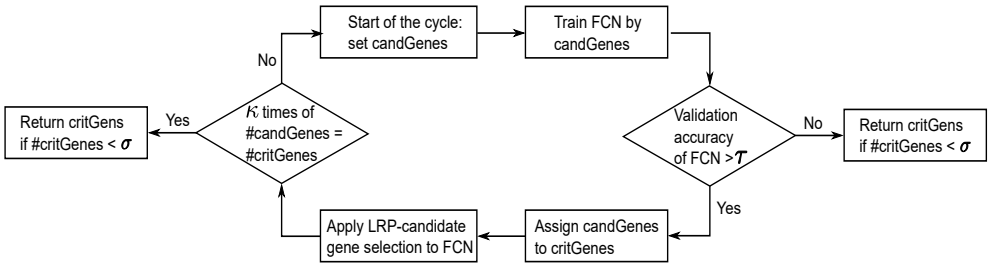

FIGURE 2. The flow chart for the gene reduction cycle. We start a cycle of gene reduction by initiating the candidate genes as all cancer-driving genes. It needs to pass two examinations to support that cycle. The first examination is the training of FCNs by the gene profiles of the candidate genes, which should attain a qualified validation accuracy in cancer-and-normal classification (right diamond). The second examination is the exact reduction of the number of candidate genes implemented by the LRP-candidate gene selection on the trained FCNs (left diamond). The cycle will terminate and return a critical gene set of size less than σ. Here τ, κ, and σ denote the allowable validation accuracy, the allowable number of repeats, and the number of critical genes, respectively.

Instead of selecting genes with prescribed high R-scores (e.g., R-score > 0.5) as marker genes, we used an R-score averaged across all genes in a sample to determine candidate genes having high gene effects (see Algorithm 1 for details). Significant genes with higher R-score (greater than the average R-score) in a sample were selected. So, genes in a sample were divided into significant and insignificant groups. We applied this bisection strategy to all validation samples that were predicted correctly by FCNs. To investigate the criticality of a gene for a class, the rate of validation samples in the class, in which the gene is significant w.r.t. the average R-score, was counted. The higher the rate of samples with significant relevance (SR-sample rate) to a gene, the more likely the gene is critical for the cancer classification. We took the minimum of these maximal SR-sample rates in different classes as a threshold and extracted candidate genes whose SR-sample rates are higher than the threshold (see Fig. 1(c) and (d) for an illustration). This completed the selection of LRP-candidate genes.

Further, to match the importance of LRP-candidate genes to their contribution to classification, we started the cycle of gene reduction that continued to reconstruct an FCN from LRP-candidate genes and reselect LRP-candidate genes from the FCN (see Fig. 2 for an illustration). The gene reduction cycle was terminated under two conditions. The first condition is the validation accuracy of the FCN less than a threshold τ. The second condition happens when applying LRP to the reduction of LRP-candidate genes. Here the allowable number of repeats to try for the same set of candidate genes was set to five ($\kappa = 5$). After five trials with the same candidate gene set, we assumed that it is hard to reduce the number of candidate genes by LRP. In short, the gene reduction cycle continues when LRP can reduce the number of genes and keep a high validation accuracy of the FCN; otherwise, it returns a reduced set of critical genes in the final step (see Algorithm 2 for details).

Note that, in Algorithm 1, the bisection strategy utilizes the average R-score to select LRP-candidate genes across all genes in a sample. So the number of selected genes may be reduced roughly by half in each running cycle. Furthermore, in contrast to the manual choice of a high R-score as a prescribed threshold, the use of the average R-score provides a dynamic way to select candidate genes when the number of genes decreases. Hence, in Algorithm 2, the gene reduction cycle can take this advantage to represent a consolidation process searching a solution of the optimization problem: minimize the number of critical genes that attain qualified cancer-and-normal classification.

Algorithm 2. The cycle of gene reduction

Input: Samples of gene expression data are divided into different classes according to cancer types and normal; An allowable classification accuracy τ; An allowable number κ of repeats; An allowable number σ of critical genes;

Output: A critical gene set of size less than σ and the associated FCN's validation accuracy higher than τ;

Initialize the candidate genes, denoted $\mathrm{candGenes}$, to be all the cancer-driving genes; Initialize the critical genes, denoted $\mathrm{critGenes}$, to be Null; Initialize $\mathrm{nRepeat} \leftarrow 0$;
while $\mathrm{candGenes}$ is not empty **do**
  Train an FCN to classify samples into cancer types and normal based on the expression of $\mathrm{candGenes}$;
  if the FCN's validation accuracy $\geq \tau$ **then**
    Assign the candidate genes to the critical genes, i.e., $\mathrm{critGenes} \leftarrow \mathrm{candGenes}$;
    Apply the LRP-candidate gene selection to the FCN and get a new set of candidate genes $\mathrm{candGenes}$;
    if $\sharp\mathrm{candGenes} < \sharp\mathrm{critGenes}$ **then**
      $\mathrm{nRepeat} \leftarrow 0$;
    else if $\mathrm{nRepeat} < \kappa$ **then**
      $\mathrm{nRepeat}$ += 1;
    else
      $\mathrm{candGenes} \leftarrow \mathrm{Null}$;
      if $\sharp\mathrm{critGenes} > \sigma$ **then**
        $\mathrm{critGenes} \leftarrow \mathrm{Null}$;
      end if
    end if
  else
    $\mathrm{candGenes} \leftarrow \mathrm{Null}$;
    if $\sharp\mathrm{critGenes} > \sigma$ **then**
      $\mathrm{critGenes} \leftarrow \mathrm{Null}$;
    end if
  end if
end while
return $\mathrm{critGenes}$;

4. Results

This study required FCNs' classification accuracies higher than $\tau = 0.935$ and the critical gene set of size less than $\sigma = 30$ in the cycle of gene reduction. We completed 125 trials of reduction simulations. The numbers of critical genes obtained in those trials were slightly different (24 genes on average). Fig. 3 shows the heat map of the occurrence of critical genes in those trials. Of 716 cancer-driving genes, 14 were involved in more than 90% trials of simulations (>112 trials). It raised an interesting question concerning the performance of genes highly involved in those trials. For simplicity, we named the genes arranged from the top (involved in all those trials) to the bottom (involved in only one of those trials) in Fig. 3 by the notations G1, G2, ..., G51, respectively. In this order, we sequentially added a gene to the training pool, constructed and trained a new FCN, and evaluated its validation accuracy.

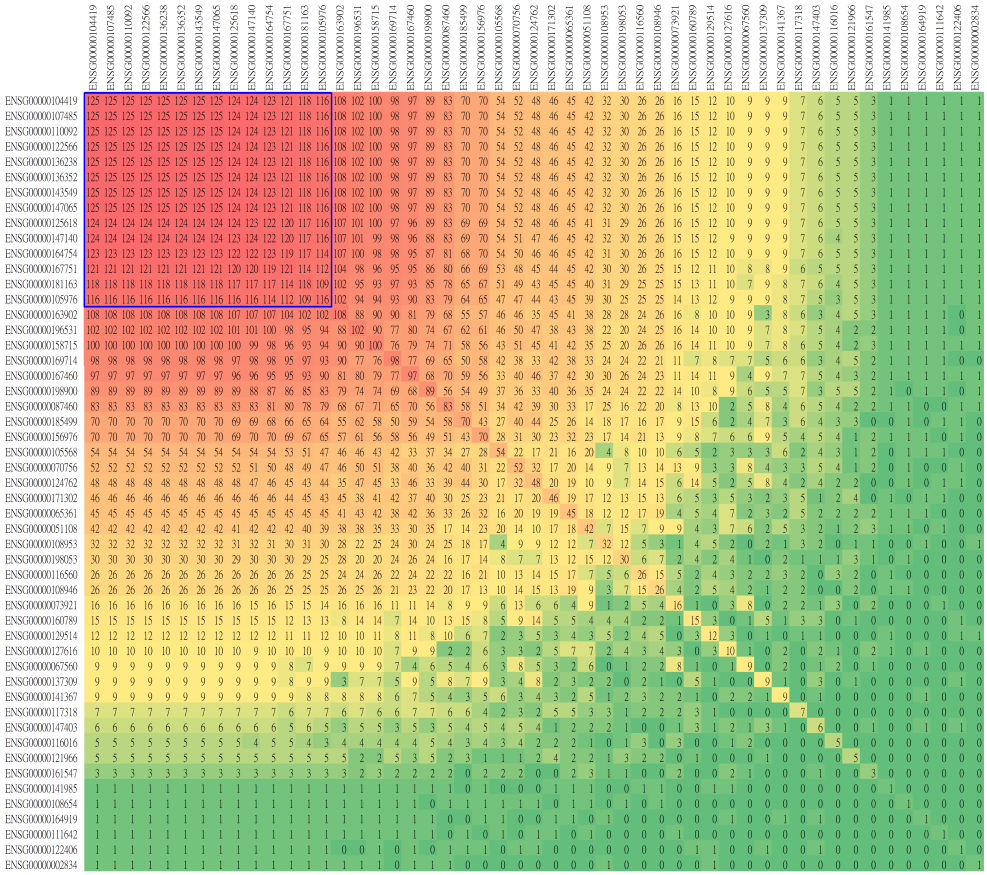

FIGURE 3. The heat map of the occurrence of critical genes in 125 trials of reduction simulations. A colorful block of row i and column j represents the occurrence of genes i and j in critical gene sets obtained from different trials (red: high occurrence; green: low occurrence). Total 51 out of 716 cancer-driving genes are involved in at least one trial (arranged in order of the occurrence times presented on the diagonal) and 14 are involved in more than 90% trials of simulations (>112 trials, blue rectangle in the upper left corner).

We followed those validation accuracies and found a gap within them (see Fig. 4 for an illustration). We observed that FCNs' validation accuracies were lower than 75% until we added G12 to the training pool. Since then, the validation accuracies of FCNs were at the level of 93% no matter how we added other genes to the training pool. Accordingly, we referred to genes G1, G2, ..., G11 as the primary set and genes G12, G13, ..., G19 (for clarity, we chose only the first eight genes) as the complementary set. Combinations of all genes in the primary set and at least one gene in the complementary set revealed an alternative solution for qualified cancer-and-normal classification.

By dividing genes into two groups, we investigated which genes in the primary set can be further removed or replaced by genes in the complementary set. The validation accuracies of FCNs trained by all genes in the primary set P and one gene g in the complementary set were referred to as the baseline performance. Then we removed a gene in the primary set (denoted the removed primary set as P') and evaluated the validation accuracies of FCNs trained by P' plus g. By comparing the validation accuracies of FCNs induced by P' plus g to the baseline performance induced by P plus g for all g taken from the complementary set, we recognized that the removal of one gene from P to P' can increase, slightly decrease, or seriously decrease the baseline performance of FCNs. These changes guided us to

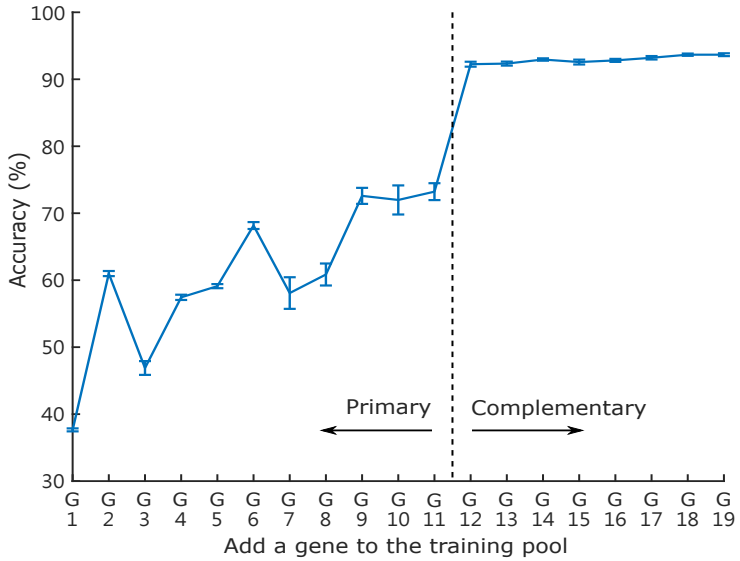

FIGURE 4. When sequentially adding a gene (from G1 to G19) to the training pool of FCNs according to the order of its occurrence times listed in Fig. 3, we find the gap (dashed line) following the validation accuracies of FCNs, which enforces us to separate genes into the primary set and the complementary set.

a removal principle: a gene in the primary set can be removed if it causes an increase or only a slight decrease in the baseline performance. To test this principle, we show in Table 3 that the removal of G3 increases the baseline performance by 15.67% on average. That guided us to set a refined primary set by removing G3. Table 4 shows that the removal of G1 from the refined primary set slightly decreases the baseline performance by 0.15% on average. Again, it guided us to remove G1 from that primary set. We continued removing genes G5, G4, and G11, one by one, which caused a less than 0.3% decrease in the baseline performance. Consequently, we got the final primary set containing only genes G2, G6, G7, G8, G9, and G10. And the removal of any genes from that primary set caused a more than 1.6% decrease in the baseline performance (see Table 5). In general, the removal of a gene q from the primary set P can be formulated by $P' = P \setminus \{q\}$, where q is a minimizer of

$$\varphi(x) = \frac{1}{\sharp\mathcal{C}} \sum_{g \in \mathcal{C}} \left(\mathrm{Acc}((P \setminus \{x\}) \cup \{g\}) - \mathrm{Acc}(P \cup \{g\})\right) \tag{4.1}$$

such that $\min_{x \in P} \varphi(x) > \varepsilon$, where ε equals to -0.003 in this study, $\mathcal{C}$ is the complementary set, $\sharp\mathcal{C}$ denotes the number of genes in $\mathcal{C}$, and $\mathrm{Acc}(\cdot)$ is the validation accuracy of FCNs trained by the specified genes.

So far, we removed genes from the primary set subject to an increase or a slight decrease in the baseline performance of FCNs. We got the minimal primary set that can stably interact with a gene in the complementary set for cancer-and-normal classification, resulting in the validation accuracy of 93.91% on average. In particular, when interacting with the gene G13, we obtained the optimal validation accuracy of 95.32%.

5. Discussion

Algorithms 1 and 2 allowed us to search groups of critical genes that had qualified performances in cancer-and-normal classification. Those groups varied in size (24 genes on average) and the components within. We ranked the appearance frequency of genes in those groups and constructed a series of FCNs

TABLE 3. Removal of G3 from the Primary Set P11 = {G1, . . . , G11}

Genes	Baseline performance	Remove G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
P11∪{G12}	92.48%	92.24%	90.60%	92.17%	92.55%	92.67%	90.77%	90.87%	89.72%	90.92%	91.46%	92.48%
P11∪{G13}	73.69%	67.71%	74.65%	95.28%	70.67%	76.14%	71.87%	70.76%	77.39%	61.28%	76.69%	71.12%
P11∪{G14}	74.56%	66.77%	74.27%	94.97%	73.98%	78.19%	70.29%	73.30%	75.42%	62.84%	77.11%	69.93%
P11∪{G15}	70.35%	68.74%	69.02%	94.61%	71.07%	73.50%	66.76%	69.69%	69.97%	61.55%	70.92%	66.72%
P11∪{G16}	76.57%	68.39%	76.38%	94.73%	75.50%	78.60%	73.97%	75.29%	78.31%	69.95%	75.83%	75.25%
P11∪{G17}	90.44%	90.57%	89.47%	94.46%	89.90%	90.69%	87.80%	88.66%	87.87%	87.41%	88.83%	90.42%
P11∪{G18}	79.93%	74.33%	75.40%	95.28%	79.50%	83.45%	75.70%	75.30%	77.44%	70.81%	79.36%	76.39%
P11∪{G19}	73.13%	65.34%	73.87%	95.03%	73.24%	74.48%	68.52%	71.15%	73.65%	61.75%	73.00%	68.40%
Average removal effect:		-4.63%	-0.94%	**15.67%**	-0.59%	2.07%	-3.18%	-2.02%	-0.17%	-8.08%	0.26%	-2.56%

TABLE 4. Removal of G1 from the Primary Set P10 = P11 \ {G3}

Genes	Baseline performance	Remove G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
P10∪{G12}	92.17%	91.28%	89.81%	–	92.05%	92.25%	90.66%	91.03%	89.41%	90.34%	91.02%	92.30%
P10∪{G13}	95.28%	95.75%	91.70%	–	94.69%	95.20%	93.64%	94.12%	93.49%	92.45%	94.17%	94.98%
P10∪{G14}	94.97%	94.65%	88.38%	–	94.54%	94.59%	94.79%	94.65%	94.19%	93.01%	94.19%	94.49%
P10∪{G15}	94.61%	94.53%	85.56%	–	93.97%	94.14%	93.00%	94.05%	93.30%	91.94%	92.90%	94.82%
P10∪{G16}	94.73%	94.97%	84.63%	–	94.52%	95.05%	93.49%	94.40%	92.95%	91.62%	93.32%	94.52%
P10∪{G17}	94.46%	94.39%	92.05%	–	93.78%	94.61%	92.19%	93.49%	92.54%	92.43%	93.48%	94.22%
P10∪{G18}	95.28%	95.09%	92.22%	–	94.39%	94.48%	92.98%	95.05%	93.14%	92.70%	94.40%	94.46%
P10∪{G19}	95.03%	95.19%	87.42%	–	94.84%	94.62%	93.80%	94.99%	93.00%	92.09%	93.45%	94.74%
Average removal effect:		**-0.15%**	-5.66%	–	-0.53%	-0.26%	-1.56%	-0.66%	-1.88%	-2.56%	-1.27%	-0.32%

TABLE 5. Non-Removal of Genes from the Primary Set P6 = P11 \ {G3,G1,G5,G4,G11}

Genes	Baseline performance	Remove G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11
P6∪{G12}	91.12%	–	88.84%	–	–	–	86.80%	88.74%	87.40%	86.12%	88.71%	–
P6∪{G13}	**95.32%**	–	91.29%	–	–	–	92.50%	92.98%	90.27%	90.47%	92.13%	–
P6∪{G14}	94.29%	–	80.52%	–	–	–	92.20%	93.50%	91.73%	90.68%	92.55%	–
P6∪{G15}	93.69%	–	81.63%	–	–	–	90.92%	92.70%	89.77%	89.81%	90.76%	–
P6∪{G16}	94.77%	–	87.41%	–	–	–	92.00%	92.89%	89.84%	90.20%	91.93%	–
P6∪{G17}	93.47%	–	90.77%	–	–	–	90.67%	90.60%	88.58%	90.51%	89.93%	–
P6∪{G18}	94.17%	–	86.71%	–	–	–	91.53%	93.52%	90.20%	90.88%	91.60%	–
P6∪{G19}	94.47%	–	87.19%	–	–	–	91.79%	93.56%	90.36%	90.76%	92.18%	–
Average removal effect:		–	-7.12%	–	–	–	-2.86%	**-1.60%**	-4.14%	-3.98%	-2.69%	–

according to this rank. We showed that the construction of FCNs by only a few top genes (in this study, six in the primary set and one in the complementary set) retained a high validation accuracy. Table 6 shows the comparison of the number of extracted marker genes and the contributed accuracy in other interpretable models.

The minimization of the number of critical genes helped unlock their entwined interactions in cancer-and-normal classification. A cross-evaluation process was performed on the minimal set of seven genes, in which we may add or delete a gene from the minimal gene set to see its straight effects on other genes. To see this, we first considered FCNs trained by the expression profiles of six genes G2 (GATA3), G6

TABLE 6. Marker Genes and Their Contributed Accuracy

	Interpretable model	Marker genes	Contributed accuracy
[13]	GA/KNN	20	87.6% (31C+1N)
[16]	CNN/Grad-CAM	400	95.59% (33C)
[18]	CNN/SA	2090	93.9–95% (33C+1N)
Our method	FCN/LRP	7	95.32% (6C+1N)

C: cancer class; N: normal class

(NKX2-1), G7 (TPM3), G8 (MSN), G9 (PAX8), and G10 (NONO) in the minimal primary set. As illustrated in Fig. 5 (the first row), only two or three genes were highly related to one cancer type when applying average R-scores to account for their significance (see Fig. 1 for the recall of the SR-sample rates). After that, we constructed FCNs by adding one gene in the complementary set to the training pool. Also, we computed their SR-sample rates and compared them to the results obtained previously from the minimal primary set. Our focus was on how the addition of one gene to the training pool of FCNs affects the SR-sample rates of other genes (all rows except the first row in Fig. 5). We hypothesized that if the increase of the SR-sample rate caused by one additional gene α in the complementary set (e.g., SR-sample rate $> 50\%$) dramatically decreases the SR-sample rate of another gene β (e.g., a more than 20% decrease) in the minimal primary set, then α may play a more dominant role than β in classification. Or, in other words, the gene β may compensate for the knockout of α. Instead, we hypothesized that if a gene β in the minimal primary set maintains high SR-sample rates regardless of additional effects of any genes α in the complementary set (e.g., SR-sample rate $> 50\%$ and has no more than 20% decreases), then the gene β has a true-positive influence on the classification of cancer.

We summarized the genes marked by the above two hypotheses. For BRCA, the genes G2 (GATA3) and G7 (TPM3) attained more than 85% SR-sample rates and satisfied the second hypothesis. For KIRC, the increases of SR-sample rates of additional genes G13 (NPM1), G14 (MET), and G16 (NACA) (resp., 81.17%, 91.90%, and 89.55%) dramatically decreased SR-sample rates of G7 (TPM3) and G10 (NONO), respectively, satisfying the first hypothesis. For LUAD, G6 (NKX2-1) and G7 (TPM3) maintained high SR-sample rates (resp., 82.02% and 54.33% with no more than 5% decreases) regardless of additional effects of any other genes, satisfying the second hypothesis. For PRAD, the increases of SR-sample rates of G12 (KLK2) and G17 (SLC45A3) (resp., 69.54% and 78.73%) dramatically decreased SR-sample rates of other genes, satisfying the first hypothesis. For THCA, G6 (NKX2-1), G8 (MSN), and G9 (PAX8) maintained high SR-sample rates (resp., 77.53%, 56.51%, and 57.95% with no more than 20% decreases) regardless of additional effects of any other genes, satisfying the second hypothesis. For UCEC, G9 (PAX8) and G10 (NONO) maintained high SR-sample rates (resp., 81.43% and 69.47% with no more than 18% decreases) regardless of additional effects of any other genes, satisfying the second hypothesis. In comparison, Table 7 shows the genes marked by the above two hypotheses (denoted by H1 and H2, respectively), the ranks (in decreasing order) of average gene expression over all samples in a cancer type, and the annotations in COSMIC related to those six cancer types. The genes marked by H1 and H2 had their ranking higher than 47 (G7) in BRCA, 55 (G14) in KIRC, 45 (G7) in LUAD, 2 (G17) in PRAD, 25 (G6) in THCA, and 39 (G9) in UCEC, respectively. All of them were the top 10% genes in the ranks, indicating their potential influence implicated in cancer. Notably, hypothesis H1 unraveled the exchange of genes G13 (NPM1), G14 (MET), and G16 (NACA) by genes G7 (TPM3) and G10 (NONO) in the classification of KIRC. It implied that some pairs of genes in the two exchangeable groups may have similar effects on KIRC. Indeed, the annotations from COSMIC indicate the significant relations of G10 (NONO) and G14 (MET) to KIRC. Also, hypothesis H1 unraveled the exchange of G12 (KLK2) and G17 (SLC45A3) by other genes in the classification of PRAD. Here G12 (KLK2) and G17 (SLC45A3) were the

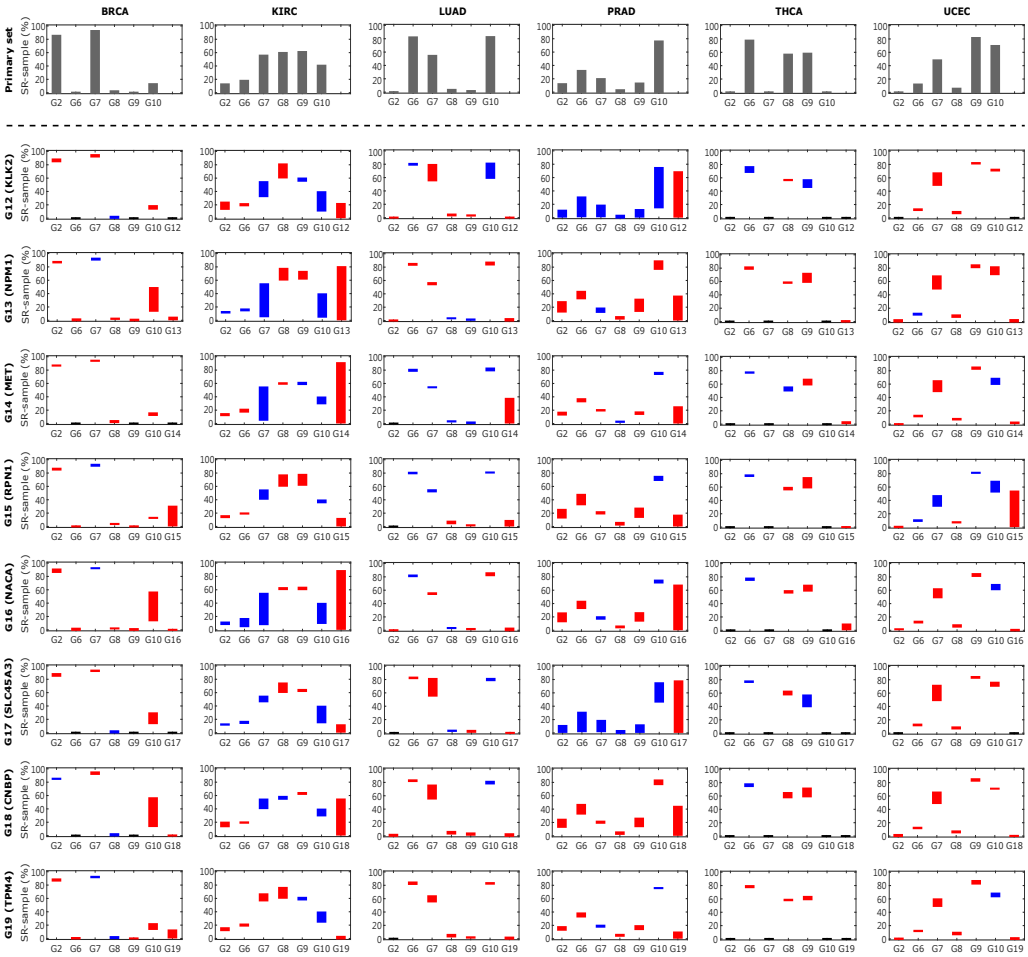

FIGURE 5. Consider FCNs trained by the expression profiles of the six genes in the minimal primary set. We show in the first row the SR-sample rates of those six genes related to six cancer types. Each cancer type has at most three genes of high SR-sample rates (SR-sample rate $> 50\%$). With the same six genes plus one additional gene selected from the complementary set, we construct FCNs and count the related SR-sample rates of those seven genes in six cancer types accordingly. We compare those SR-sample rates and show the changes in the remaining rows. Here a jump from the bottom (resp., the top) to the top (resp., the bottom) of a red bar (resp., a blue bar) indicates an increase (resp., a decrease) in a gene's SR-sample rate corresponding to the change of FCNs trained by the expression profiles of six genes to seven genes. SR: significant relevance.

top 2 genes in the expression rank of PRAD, whose significant relations to PRAD had been annotated by the COSMIC Cancer Gene Census. On the other hand, hypothesis H2 unraveled the reliability of genes in the classification of cancer. At least one of the genes marked by H2 in BRCA was annotated with related breast cancer by the COSMIC Cancer Gene Census. Similar results were found in LUAD and THCA, respectively, except for UCEC. Notice that the Cancer Gene Census annotated only 4.97% of 723 genes with breast cancer, 4.56% with renal carcinoma, 9.12% with lung cancer, 3.87% with prostate cancer, and 4.14% with thyroid cancer. The successful inference of the hypotheses H1 and H2 in cancer BRCA, KIRC, LUAD, PRAD, and THCA primes us the importance of G7 (TPM3), G9 (PAX8), and G10 (NONO) in UCEC in future work.

TABLE 7. Genes Marked by the Hypotheses H1 and H2

IDs	Ensembl IDs	Gene symbol	Tier	BRCA	KIRC	LUAD	PRAD	THCA	UCEC	Annotations in COSMIC
G2	ENSG00000107485	GATA3	1	**H2*** (**14**)	– (499)	– (571)	– (457)	– (417)	– (559)	Breast*
G6	ENSG00000136352	NKX2-1	1	– (689)	– (710)	**H2*** (**44**)	– (633)	**H2** (**25**)	– (567)	NSCLC*
G7	ENSG00000143549	TPM3	1	**H2** (**47**)	– (74)	**H2*** (**45**)	– (111)	– (83)	– (62)	Thyroid, NSCLC*
G8	ENSG00000147065	MSN	1	– (49)	– (16)	– (26)	– (101)	**H2** (**14**)	– (49)	–
G9	ENSG00000125618	PAX8	1	– (631)	– (120)	– (611)	– (628)	**H2*** (**20**)	**H2** (**39**)	Thyroid*
G10	ENSG00000147140	NONO	1	– (41)	– (47)	– (35)	– (42)	– (47)	**H2** (**30**)	Renal
G12	ENSG00000167751	KLK2	1	– (686)	– (698)	– (703)	**H1*** (**1**)	– (627)	– (672)	Prostate*
G13	ENSG00000181163	NPM1	1	– (18)	**H1** (**11**)	– (21)	– (18)	– (18)	– (18)	–
G14	ENSG00000105976	MET	1	– (415)	**H1*** (**55**)	– (59)	– (481)	– (42)	– (213)	Renal*
G15	ENSG00000163902	RPN1	1	– (30)	– (31)	– (27)	– (24)	– (29)	– (19)	–
G16	ENSG00000196531	NACA	2	– (45)	**H1** (**38**)	– (46)	– (29)	– (41)	– (33)	–
G17	ENSG00000158715	SLC45A3	1	– (537)	– (529)	– (457)	**H1*** (**2**)	– (605)	– (537)	Prostate*
G18	ENSG00000169714	CNBP	1	– (15)	– (14)	– (18)	– (11)	– (13)	– (13)	–
G19	ENSG00000167460	TPM4	1	– (31)	– (54)	– (40)	– (57)	– (105)	– (38)	–

H1: the first hypothesis; H2: the second hypothesis; *: match annotations in COSMIC; Brackets: the expression rank of genes in cancer

6. Conclusion

FCNs trained by gene expression profiles were used to make classification of six cancer types and normal. The expression profiles of multiple genes were essential to achieve qualified cancer-and-normal classification. We proposed the LRP-candidate gene selection and the cycle of gene reduction to extract gene sets of size as small as possible. The minimal gene set simplified allowable interactions among genes and enabled us to explore their impact directly on cancer classification. Accordingly, we marked genes that can be exchanged by other genes or have a reliable influence on classification of cancer. Interactions of those marked genes in different types of cancer were recognized by FCNs theoretically, which may help chart a new course to probe their biological information implicated in cancer.

Statements and Declarations

The authors declare that they have no conflict of interest, and the manuscript has no associated data.

Acknowledgments

This research was based on data generated by the COSMIC Cancer Gene Census (https://cancer.sanger.ac.uk/census) and the TCGA Research Network (https://www.cancer.gov/

tcga). Thanks to the National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) in Taiwan for providing computational and storage resources.

References

[1] T. Ahn, T. Goo, C.-h. Lee, S. Kim, K. Han, S. Park, and T. Park. Deep learning-based identification of cancer or normal tissue using gene expression data. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1748–1752, 2018.

[2] S. Bach, A. Binder, G. Montavon, F. Klauschen, K.-R. Müller, and W. Samek. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS ONE*, 10(7):e0130140, 2015.

[3] D. Bertsimas and H. Wiberg. Machine learning in oncology: methods, applications, and challenges. *JCO Clinical Cancer Informatics*, 4:885–894, 2020.

[4] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic minority over-sampling technique. *Journal Of Artificial Intelligence Research*, 16(1):321–357, 2002.

[5] F. S. Collins and A. D. Barker. Mapping the cancer genome. *Scientific American*, 296(3):50–57, 2007.

[6] P. Danaee, R. Ghaeini, and D. A. Hendrix. A deep learning approach for cancer detection and relevant gene identification. *Pacific Symposium on Biocomputing*, 22:219–229, 2017.

[7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[8] G. Eraslan, Ž. Avsec, J. Gagneur, and F. J. Theis. Deep learning: new computational modelling techniques for genomics. *Nature Reviews Genetics*, 20:389–403, 2019.

[9] P. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, and M. R. Stratton. A census of human cancer genes. *Nature Reviews Cancer*, 4:177–183, 2004.

[10] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286(5439):531–537, 1999.

[11] H. Han, W.-Y. Wang, and B.-H. Mao. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning. In D.-S. Huang, X.-P. Zhang, and G.-B. Huang, editors, *Advances in Intelligent Computing*, pages 878–887, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[13] Y. Li, K. Kang, J. M. Krahn, N. Croutwater, K. Lee, U. D. M., and L. Li. A comprehensive genomic pan-cancer classification using The Cancer Genome Atlas gene expression data. *BMC Genomics*, 18(1):e508, 2017.

[14] M. Libbrecht and W. Noble. Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16:321–332, 2015.

[15] Y. Lu and J. Han. Cancer classification using gene expression data. *Information Systems*, 28(4):243–268, 2003.

[16] B. Lyu and A. Haque. Deep learning based tumor type classification using gene expression data. In *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 89–96, New York, NY, USA, 2018. Association for Computing Machinery.

[17] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K. R. Müller. Layer-wise relevance propagation: An overview. In W. Samek, G. Montavon, A. Vedaldi, L. Hansen, and K. R. Müller, editors, *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, volume 11700, pages 193–209. Lecture Notes in Computer Science, Springer, 2019.

[18] M. Mostavi, Y.-C. Chiu, Y. Huang, and Y. Chen. Convolutional neural network models for cancer type prediction based on gene expression. *BMC Medical Genomics*, 13:e44, 2020.

[19] N. Murali, A. Kucukkaya, A. Petukhova, J. Onofrey, and J. Chapiro. Supervised machine learning in oncology: a clinician's guide. *Digestive Disease Interventions*, 4(1):73–81, 2020.

[20] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C.-H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences*, 98(26):15149–15154, 2001.

[21] F. Rosenblatt. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington, DC, 1961.

[22] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, and the PDP research group, editors, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundation*. MIT Press, Cambridge, MA, 1986.

[23] E. E. Schadt, J. Lamb, X. Yang, and et al. An integrative genomics approach to infer causal associations between gene expression and disease. *Nature Genetics*, 37(7):710–717, 2005.

[24] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.

[25] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. *CoRR*, abs/1312.6034, 2014.

[26] Z. Sondka, S. Bamford, C. G. Cole, S. A. Ward, I. Dunham, and S. A. Forbes. The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nature Reviews Cancer*, 18:696–705, 2018.