Learning & Visualizing Genomic Signatures of Cancer Tumors using Deep Neural Networks

Tarek Khorshed
Department of Computer Science & Engineering
The American University in Cairo
Cairo, Egypt
tarek_khorshed@aucegypt.edu

Mohamed N. Moustafa
Department of Computer Science & Engineering
The American University in Cairo
Cairo, Egypt
m.moustafa@aucegypt.edu

Ahmed Rafea
Department of Computer Science & Engineering
The American University in Cairo
Cairo, Egypt
rafea@aucegypt.edu

***Abstract*— Deep learning for medical diagnosis using genomics is extremely challenging given the high dimensionality of the data and lack of sufficient patient samples. Another challenge is that deep models are conceived as black boxes without much interpretation on how these complex models make predictions. We propose a deep transfer learning framework for cancer diagnosis with the capability of learning the sequence of DNA and RNA in cancer cells and identifying genetic changes that alter cell behavior and cause uncontrollable growth and malignancy. We design a new Convolutional Neural Network architecture with capabilities of learning the genomic signatures of whole-transcriptome gene expressions collected from multiple tumor types covering multiple organ sites. We demonstrate how our trained model can function as a comprehensive multi-tissue cancer classifier by using transfer learning to build classifiers for tumors lacking sufficient human samples to be trained independently. We introduce visualization procedures to provide more biological insight on how our model is learning genomic signatures and accurately making predictions across multiple cancer tissue types.**

Keywords— Cancer classification, Convolutional neural networks, Deep learning, Gene expressions, RNA sequencing

I. INTRODUCTION

Advances in structural genomics have allowed studying the full set of DNAs in the human genome [2], [19]. Next generation sequencing (NGS) methods such as whole-genome DNA sequencing and Total RNA sequencing are considered revolutionary technologies for studying genetic changes in cancer [16], [21]. They provide great potential for cancer classification and better understanding of tumor progression given their ability to sequence thousands of genes at one time and detect multiple types of genomic alterations [14], [15], [19]. Gene expression analysis using Total RNA sequencing provides a snapshot of the whole transcriptome rather than a predetermined subset of genes and can detect both coding plus multiple forms of noncoding RNA [16]. These methods have eliminated many limitations involved in microarray based experiments that were traditionally used for gene expression analysis [16], [19], [21].

Despite all these potential capabilities, cancer diagnosis using gene expressions produced from Total RNA sequencing is extremely challenging given the complexity and massive amount of genetic data that is produced [14], [15], [19], [20], [28]. The magnitude of variants obtained from RNA Sequencing is exponential which makes it difficult for traditional machine learning approaches to evaluate genetic variants for disease prediction [2], [16], [17]. Gene expression data is characterized by being very high in dimensionality in terms of having a very large number of features representing the genes, and a very small number of training data representing the patient samples [6], [25]. Complexity is also due to the fact that only a small subset of genes might be driving the tumor progression [1], [2].

Current cancer classification methods avoid processing the full set of genes to overcome these complexities and are mainly based on performing a process of gene feature selection as a prerequisite to the classifier learning process [22]-[25]. These methods are not optimal for early cancer diagnosis as they fall short in taking the full advantage of DNA and RNA sequencing technologies to discover the correlated patterns between genes across the full set of DNAs in the human genome. Current classification methods which are based on gene feature selection will not have the opportunity to detect multiple types of genetic alterations that may be driving the growth of a tumor across the whole transcriptome rather than a predetermined subset of genes [3]. Another limitation of current methods is that they typically rely on gene expressions collected from a single cancer tissue type based on the same anatomical site of origin. This approach does not utilize the full potential of whole-genome sequencing technologies and data produced by large-scale genomic projects that produce molecular characterizations of thousands of tumors using genome-wide platforms [28].

This has motivated our research for early diagnosis of cancer by using deep learning to develop a comprehensive multi-tissue cancer classifier based on molecular signatures of whole-transcriptome gene expressions. Our approach using deep transfer learning eliminates the dependency on huge amounts of training data and the prerequisite gene feature selection by learning the genetic signatures collected from human samples representing multiple cancer tissue types covering multiple organ sites. We propose a new Convolutional Neural Network architecture called "Gene eXpression Network" (GeneXNet) which is specifically designed to address the complex nature of whole-transcriptome gene expressions and with capabilities of learning molecular signatures of DNA and RNA in cancer cells to identify genetic changes that may be driving cancer progression. We contribute in solving one of the biggest challenges in cancer classification which is lack of patient samples. We demonstrate how our trained model can function as a generic multi-tissue cancer classifier by using transfer learning to build classifiers for other tumor types that are lacking sufficient patient samples to be trained independently.

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

We also contribute in providing more confidence in using deep learning for medical diagnosis and solving the problem that deep networks are conceived as "black boxes" without much interpretation on how these complex models are making their decisions [40]. We introduce visualization procedures to provide more insight on how our model is learning genomic signatures of whole-transcriptome gene expressions and accurately making predictions across multiple cancer tumors. We visualize gene localization maps highlighting the important regions in the gene expressions influencing the tumor class prediction. This helps in revealing the genomic relationships of gene expressions that are influential in the cancer tumor progression.

II. Related work

Gene expressions have been extensively used in cancer classification [5]-[11]. By collecting mRNA samples for tumors of known classes, supervised learning can be used to build discriminative models which can learn the gene patterns of the underlying disease and then be used to predict the tumor class of patient samples which were not previously diagnosed [1]. Many microarray experiments have demonstrated how it was possible to distinguish between cancer types using data classification even though they are clinically indistinguishable [1].

Current methods for cancer classification follow the approach of feature engineering and are based on applying innovative gene feature selection techniques as a prerequisite to the classifier learning process to discover a small subset of informative genes which are discriminative among the tumor being analyzed [22], [24]. Gene selection can be generally classified into filtering, wrapping and embedded methods [25]. The accuracy of such a classifier depends heavily on the successful identification of these discriminative features [26], [27]. Substantial work has been done for cancer classification by performing gene feature selection and building on traditional machine learning methods such as Support Vector Machines [11], [13], [23], Random Forests [10], Decision Trees [12], AdaBoost [8], K-Nearest Neighbor [10] and Genetic algorithms [6], [8]. Many other techniques which combine gene feature selection and classification have also been proposed [5]-[9].

III. Proposed Approach

A. Deep Transfer Learning using Genomic Signatures Of Multiple Cancer Tumor Types

Our approach using Deep Transfer Learning provides an alternative solution to feature engineering and eliminates the dependency on huge amounts of training data and the prerequisite gene feature selection. This is achieved by combining the process of gene feature selection and classification into one end-to-end learning system. We develop a comprehensive multi-tissue cancer classifier by designing a Convolutional Neural Network (CNN) with the capability of learning the genomic signatures of whole-transcriptome gene expressions shared across multiple cancer tumor types. By training the model with samples from multiple tissue types collected from multiple organ sites, the classifier is able to learn and extract complex patterns from the gene expressions that represent genomic and transcriptomic alterations. This allows the classifier to more accurately classify cancer tumors which are resulting from DNA or RNA changes that alter cell behavior across multiple tissues and cause uncontrollable growth.

A major advantage is that we can reuse the genomic signatures learned by the trained model to perform very efficient transfer learning to solve one of the biggest challenges in cancer classification which is lack of patient samples. We demonstrate how transfer learning can be used to build and finetune classifiers for other different types of cancer tumors not included in the underlying dataset, which might be lacking sufficient patient samples to be trained independently. By reusing the weights of the pretrained model, we demonstrate how to build a classifier for a different cancer tissue type. The intuition behind transfer learning comes from recent studies which have performed an integrated multiplatform analysis across multiple cancer types that have revealed similar molecular classification within and across tissues of origin [3], [4]. This means that the discriminative molecular features for one cancer classifier will most likely be relevant for other cancer types. Our pretrained model will have already learned the complex types of genetic alterations and genomic signatures collected from multiple cancer tissue types originating from different organs and can effectively function as a generic model for cancer classification. Another advantage is that the classifier will not be limited to learning the molecular characterization of a single tissue type but will have the capability of detecting more complex types of genomic alterations by learning the genetic signatures collected from multiple tumors and across multiple cancer tissue types.

B. Deep Transfer Learning System Architecture

Our end-to-end learning system depends on collecting human samples representing multiple types of cancer tumors collected from multiple tissues spanning different organs across the body. Total RNA sequencing is used for gene expression quantification across the whole-transcriptome, then the data is normalized and converted into a representation suitable for input to our deep learning model. We build and train a deep CNN to automatically learn the molecular signatures of the full set of whole-transcriptome gene expressions and produce a trained model which can be used for classification. Our model, which we refer to as "Gene eXpression Network" (GeneXNet), relies on building an architecture with multiple layers of non-linear functions which transform the gene expressions into feature maps to increase the level of accuracy and invariance of the selected gene features. The genetic signatures learned by the feature maps in the deep layers, eliminate the need for the traditional prerequisite process of gene feature selection because they are insensitive to any insignificant genes or irrelevant variations in the gene expressions [45], [51]. We train the model using supervised learning by feeding the collected samples as input and we use stochastic gradient descent optimization and backpropagation [46] to adjust the network weights.

C. Convolutional Neural Networks (CNNs)

CNNs have contributed to many record breaking results especially in the areas of computer vision [29]-[34]. The development of new CNN architectures continues to be an active research area such as AlexNet [50], VGGNet [48], GoogLeNet [47], InceptionNet [43], ResNet [34], [44], DenseNet [31], MobileNet [29], [33], SENet [32] and NasNet [30].

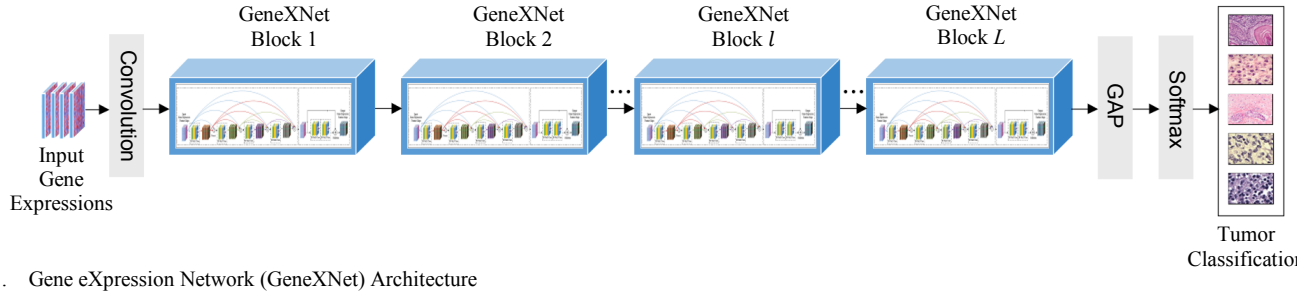

Fig. 1. Gene eXpression Network (GeneXNet) Architecture

The motivation in using CNNs for learning gene expressions is that the convolution operation is very suitable for the high dimensional and sparse nature of the data. It is not practical to use traditional kernel learning methods and fully connected networks since the resulting models will have a huge number of parameters to be learned [45].

D. Gene eXpression Network (GeneXNet) Architecture

In this section we describe the detailed architecture of our proposed CNN model. Recent benchmark results obtained by deep CNNs for image recognition tasks have demonstrated that network depth is of great importance for feature extraction and have managed to achieve outstanding results by designing deeper networks [34], [44]. These models were able to exploit deep architectures because of the availability of large training datasets, such as ImageNet [50], to avoid problems such as overfitting, vanishing gradients and degradation of accuracy [34], [44]. Applying the same CNN architectures for classification of gene expressions is not an evident task. One of the biggest challenges is to build a multi-tissue cancer classifier that can benefit from deep networks to efficiently extract the molecular signatures of a large number of genes, without facing severe overfitting or degradation in performance due to the lack of sufficient training samples. We attempted to build an end-to-end learning system for classification without performing the prerequisite process of gene feature selection by using some of the available state-of-the-art CNN models which have been specifically designed for computer vision tasks. Our experiments have shown that training these deep models suffered from overfitting when presented with the underlying dataset that includes the full set of transcriptome gene expressions. The dataset did not have sufficient samples to train these deep models and achieve the required accuracy.

To solve these problems, we propose a new CNN architecture which we refer to as Gene eXpression Network (GeneXNet) shown in Fig. 1. Our network is designed to specifically address the complex nature of gene expressions and the lack of training samples by incorporating multiple layers of building blocks which we refer to as GeneXNet blocks. These blocks are motivated from both deep residual learning networks [34], [44] and densely connected convolutional networks [31] and are formed by merging together two different types of learning sub-blocks. To formulate our building block, we define our network to have L block layers and define the non-linear transformation of gene expressions denoted by G_l using Residual Learning [44] as:

$$x_{l+1} = f_l[\, G_l(x_l, W_l) \; + \; M(x_l)\,] \tag{1}$$

$$W_l = \{w_{l,i} | \, 1 \le i \le K_l \} \tag{2}$$

where l is the block index, W_l represents the set of weights and biases, $w_{l,i}$ represents the weights of the i^{th} convolutional layer in the l^{th} block, K_l is the number of convolution layers, f_l and $M(x_l)$ are mapping functions and x_l, x_{l+1} represent the input and output features of the l^{th} block. We apply "pre-activation" of weight layers as in [34] by defining the transformation at each layer as a sequence of multiple operations which are Batch Normalization (BN) [42], Rectified Linear Unit (ReLU) [36] and Convolution. Empirical results have shown that residual learning helps to avoid degradation in performance accuracy as the depth of the network increases [44]. We implement both $M(x_l)$ and f_l as identity mappings as this has shown to improve accuracy by creating a more direct path for data propagation in the forward and backward passes [34]. The resulting non-linear transformation of gene expressions and the gradient of the loss function can then be expressed recursively as:

$$x_L = x_l + \sum_{i=l}^{L-1} G_i(x_i, W_i) \tag{3}$$

$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L}\frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L}\left[1 + \frac{\partial}{\partial x_l}\sum_{i=l}^{L-1} G_i(x_i, W_i)\right] \tag{4}$$

where x_L represents the output features, ε is the loss function and $\partial\varepsilon/\partial x_l$ is the gradient obtained by applying the chain rule and backpropagation [34]. We refer to the resulting sub-block as the *Residual Learning block*.

Despite the strong advantages of residual learning, there has been other different intuitions that there might be a great amount of redundancy in residual networks and that not all the layers are required [31]. Densely connected convolutional networks [31] exploit the potential of the network through feature reuse as an alternative to deep or wide architectures by connecting all layers with matching feature-map sizes directly with each other. This design consideration is very important for our task to efficiently extract the molecular signatures of large number of genes, without facing severe overfitting or degradation in performance due to the lack of sufficient human training samples. This has inspired us to further reformulate the design of our GeneXNet building block and augment its learning capability by introducing additional dense layers that precede the residual learning layers. The dense layers follow a similar approach as in DenseNets [31].

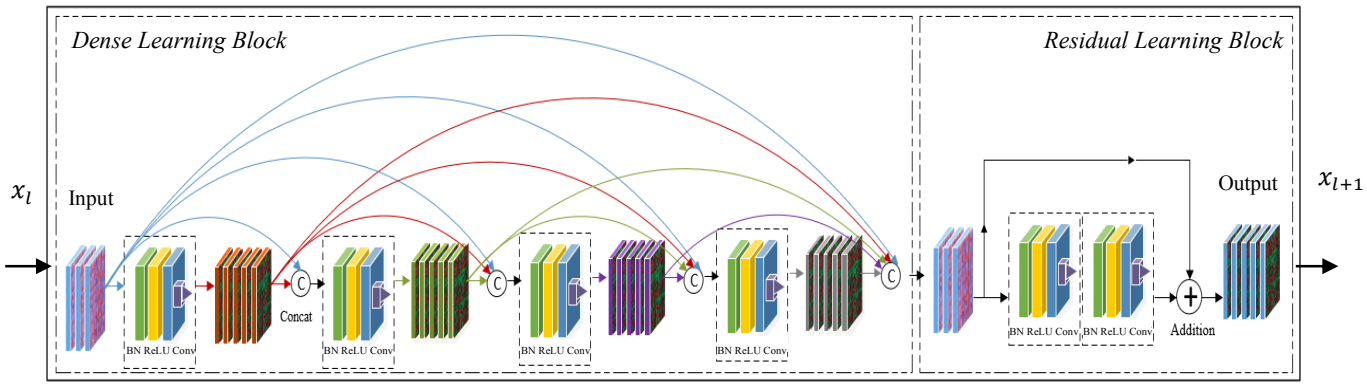

Fig. 2. Gene eXpression Network (GeneXNet) Block

We implement our dense layers by passing additional inputs into each layer from all preceding layers and passing the feature maps of each layer to all subsequent layers. Our aim from this design is to provide each layer with direct access to the gradients from the loss functions and the original input signal which can potentially improve flow of information throughout the network. Our dense layers are formulated as:

$$x_{l+1} = G_l(\, Concat[x_1, x_2, x_3, \dots, x_l]\,) \qquad (5)$$

where $Concat[x_1, \dots, x_l]$ represents the concatenation of the gene expression feature maps resulting from all preceding layers and G_l represents the same transformation as in (1). We refer to the resulting sub-block as the *Dense Learning block*.

Our proposed GeneXNet block is finally formed by merging together these two sub-blocks as shown in Fig. 2. The full Gene eXpression Network (GeneXNet) architecture is shown in Fig. 1. It is implemented by feeding the gene expressions to multiple layers of GeneXNet blocks each containing a combination of dense and residual learning layers as described above. The network ends with a global average pooling [41] after the last block and a fully connected softmax layer for classification.

Our results have demonstrated that our proposed network has allowed training deeper network architectures with complex data such as gene expressions, despite the large number of genes. The dense layers allow the network to efficiently extract the genetic signatures across multiple cancer types by re-using the feature maps learned at different layers, which increases the variation of input signals fed to subsequent layers since it represents the collective knowledge of the network [31]. The residual layers with identity mappings contribute to providing a direct path for information propagation in the forward and backward passes [44] while the connectivity of the dense layers provide each layer with more direct access to the gradients from the loss function and the original input signal [31].

IV. Visualizing Class-Discriminative Genomic Signatures Across Multiple Cancer Tumor Types

One of the challenges in using deep learning for disease diagnosis, is that deep networks are conceived as "black boxes" without much interpretation on how these complex models make their decisions [40]. Extensive work has been done to introduce novel visualization techniques for deep networks to help understand and interpret their record breaking performance in computer vision tasks [39], [40], [49]. The output from these techniques can be interpreted by non-experts when studied in conjunction with image or video datasets because they are visually comprehensible. Unfortunately, these methods are not directly applicable to genomic datasets such as gene expressions, since they cannot be visually rendered in a human-friendly form that allows easy interpretations. Our learning system architecture can contribute in solving this problem, since it is designed to address the complex nature of gene expressions. We introduce a visualization procedure to provide more insight on how our proposed deep network is performing cancer classification across multiple tumor types. Our method is inspired from the work used to visualize intermediate activations for CNNs used in image classification [49]. We also build on the methods for Class Activation Maps (CAM) [39], [40] which visualize heatmaps of class activations for deep networks used in image classification and captioning.

We introduce a visualization method shown in Fig. 3 which uses the gradient information flowing into the last convolutional layer of the GeneXNet model to produce gene localization maps highlighting the important regions in the gene expressions which influenced the resulting tumor class prediction. The gene expression data used to train the network is sparse and very high in dimensionality since it represents a snapshot of the whole transcriptome rather than a predetermined subset of genes. By identifying a class-discriminative localization map in the gene expressions, we can identify the subset of genes driving cancer progression and resulted in the model's prediction. We refer to this localization map as a Gene-Class-Activation-Map (Gene-CAM). For each tumor type, the Gene-CAM is a representation of the discriminative genes used by the network to correctly classify the tumor. The procedure is summarized as follows:

For a GeneXNet with L blocks, the network will produce a set of intermediate activation feature maps as the output of each block. Let F_l represent the output feature maps of the l^{th} block with dimensions (width: X_l, height: Y_l, depth: D_l). This volume represents the molecular features learned by the network that will be activated when matched with similar patterns in the input gene expressions of a given tumor sample. Let $f_L^k(i,j)$ represent the k^{th} feature map for the last block at special location (i,j). Since the network uses Global Average Pooling (GAP) [41] before the final Softmax layer to calculate the spatial average of the feature maps, then the classification score s^c for tumor type c which is used as input to the softmax can be written as:

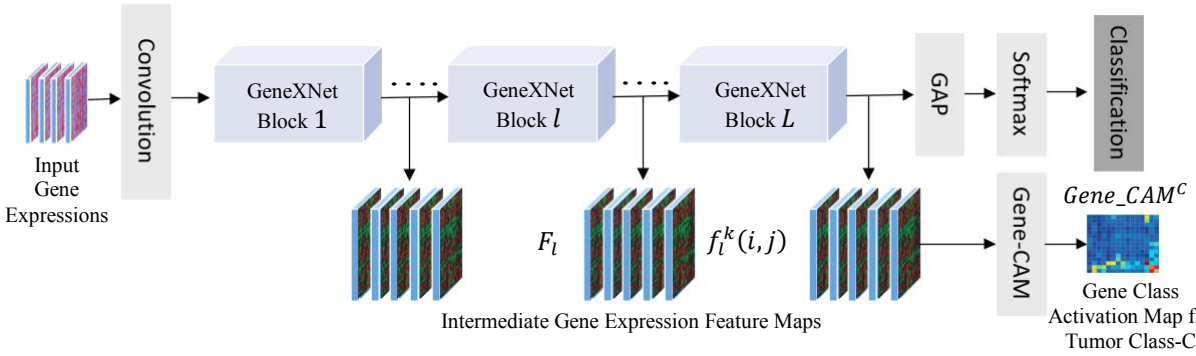

Fig. 3. Visualizing Class-Discriminative Localization Maps of Gene Expressions

$$s^c = \sum_k^{D_L} w_k^c \sum_i^{X_L} \sum_j^{Y_L} f_L^k(i,j) \qquad (6)$$

where c is the tumor class, w_k^c represents the weights for class c with respective to feature map k and D_L is the number of feature maps in the last block before the GAP layer each with width X_L and height Y_L.

We redefine the weights of each feature map with respect to a class as α_k^c by computing the gradient of the score of each class with respect to each feature map as follows:

$$\alpha_k^c = \frac{1}{X_L . Y_L} \sum_i^{X_L} \sum_j^{Y_L} \frac{\partial s^c}{\partial f_{ij}^k} \qquad (7)$$

where the new weights α_k^c represent the importance of each feature map for class discrimination. The Gene-Class-Activation-Map (Gene-CAM) is then calculated as:

$$Gene\_CAM^c(i,j) = ReLU\left[\sum_k^{D_L} \alpha_k^c . f^k(i,j)\right] \qquad (8)$$

The resulting map with dimensions (X_L, Y_L) represents a gene localization for the given tumor sample that captures the discriminative regions in the gene expression input matrix which influenced the prediction of the tumor class. The ReLU [36] is applied to obtain only the features that have a positive contribution to the correct class [40]. Finally, to visualize the Gene-CAM we resize it using up-sampling and overlay it against the input gene expressions. The resulting heatmap highlights the important regions in the gene expressions which in turn helps identify the subset of genes that are possibly influencing the cancer tumor and resulted in the model's prediction.

V. EXPERIMENTS

A. Datasets

Our objective was to design a comprehensive multi-tissue cancer classifier capable of detecting complex types of genetic alterations by learning the genomic signatures of whole-transcriptome gene expressions across multiple cancer tissue types. To achieve this objective, the datasets we selected for our experiments included a total of 11,093 human samples for mRNA gene expression quantification, which were collected from 26 different human anatomical organ sites and covering 33 different cancer tumor types. The datasets were obtained from "The Cancer Genome Atlas" (TCGA) [28] and generated by means of Total RNA sequencing [3]. Each individual human sample represents the whole transcriptome and includes a total of 60,483 genes annotated against a reference genome. One of the biggest challenges in this dataset is the very small number of samples in each of the tumor types, compared to the very large number of genes. Most of the tumors only have several hundred samples and some even have less than a hundred samples.

B. Cancer Classification & Trasnfer Learning

We perform several different multi-class and binary classification tasks. For binary classification we predict whether the given sample represents a tumor versus a normal tissue. For multi-class classification we predict for a given sample the type of cancer tumor within each organ site. We build a multi-tissue multi-class classifier by training our model using ALL the data which includes the genomic signatures from 26 organ sites covering 33 tumors. We also build a multi-tumor classifier for individual organ sites that relatively had the greatest number of samples which included 11 sites as shown in table I. We then repeat the same experiment, but this time by performing transfer learning using the pre-trained model trained with all the tumor data. The objective was to compare the performance between transfer learning using a pre-trained model and Full training.

C. Training, Optimization and Evaluation

We use random sampling to divide our datasets into 85% for training/validation and 15% for final testing. We train all models using stratified k-fold cross-validation experimenting with different fold sizes. We use the validation data to optimize the hyperparameters of our models while the test data is strictly used only to evaluate the final performance. Training a deep multi-layer CNN architecture like GeneXNet is a very complex optimization problem as it involves non-convex loss functions [51]. Among the challenges we faced in model optimization is the very high dimensional landscape of the network weight space resulting from training the network with the whole-transcriptome wide gene expressions for every tumor sample. To overcome these problems, we train our model using mini-batch Stochastic Gradient Descent (SDG) with an adaptive learning rate optimization algorithm [46]. We experiment with Adam [37], AdaGrad [38] and RMSprop [46]. We start with a learning rate of 1e-4 and divide it by half when the validation loss plateaus for more than 50 epochs.

TABLE I. RESULTS OF MULTI-TUMOR CLASSIFICATION & TRANSFER LEARNING FOR 11 INDIVIDUAL ORGAN SITES

Organ Site	Cancer Tumor Type(s)	Total Samples	Full Training		Transfer Learning & Finetuning	
			Accuracy (%)	ROC AUC	Accuracy (%)	ROC AUC
Bladder	Bladder Urothelial Carcinoma (BLCA)	433	96.92	1.0	95.38	0.99
Breast	Breast invasive carcinoma (BRCA)	1222	98.37	0.998	98.37	1.0
Colorectal	Colon adenocarcinoma (COAD), Rectum adenocarcinoma (READ)	698	100	1.0	100	1.0
Head & Neck	Head and Neck squamous cell carcinoma (HNSC)	546	98.78	0.985	92.68	1.0
Kidney	Kidney Chromophobe (KICH), Kidney renal clear cell carcinoma (KIRC), Kidney renal papillary cell carcinoma (KIRP)	1021	100	1.0	100	0.97
Liver	Liver hepatocellular carcinoma (LIHC)	424	100	1.0	98.44	1.0
Lung	Lung adenocarcinoma (LUAD), Lung squamous cell carcinoma (LUSC)	1145	99.42	1.0	99.42	0.94
Prostate	Prostate adenocarcinoma (PRAD)	551	97.59	0.961	97.59	0.94
Stomach	Stomach adenocarcinoma (STAD)	407	96.77	0.979	96.77	0.88
Thyroid	Thyroid carcinoma (THCA)	568	95.35	0.981	93.02	1.0
Uterus	Uterine Corpus Endometrial Carcinoma (UCEC), Uterine Carcinosarcoma (UCS)	643	100	1.0	100	0.89
Bile Duct*	Cholangiocarcinoma (CHOL)	45	-	-	85.71	0.89
Esophagus*	Esophageal carcinoma (ESCA)	173	-	-	92.31	0.99

TABLE II. CLASSIFICATION PERFORMANCE OF GENEXNET IN COMPARISON WITH OTHER CNN MODELS

Network Model	**Accuracy (%)**	**ROC AUC**	**Cross Entropy Loss**
GeneXNet	**98.93**	**0.99 ±0.00**	**0.06**
ResNet-50 v2 [34]	36.96	0.86 ±0.08	4.9
DenseNet-121 [31]	22.33	0.79 ±0.12	6.09
NasNetMobile [30]	21.61	0.84 ±0.07	2.58
MobileNet v2 [29]	24.96	0.80 ±0.12	5.99

We evaluate the performance of our GeneXNet model with different architectures and sizes by tuning the parameters θ_D, θ_R with values (0, 0.25, 0.5 and 1) and θ_k with values (32, 64). These parameters define the percentage of dense and residual sub-blocks in the network and the number of filters used in the convolution layers. We evaluate the classification performance of our GeneXNet models using the receiver operating characteristics (ROC) curves. For all experiments, we report the average classification accuracy and ROC Area Under the Curve (AUC) on the Test dataset. We also evaluate the performance of our model in comparison with some of the current state-of-the-art CNN models specifically designed for computer vision tasks. We perform the same multi-class classification task using all the data but replacing our model with the publicly available implementations of ResNet [34], [44], DenseNet [31], NasNet [30] and MobileNet [29], [33].

D. Results

The results of binary classification for 11 selected individual organ sites are shown in table I. Our GeneXNet model was able to achieve 100% accuracy for 8 different tumor types and between 95.35% to 99.42% accuracy for the remaining tumors. Table I also shows a comparison between the results of Transfer Learning and Full training which demonstrate that Transfer learning achieved excellent results which are comparable to those achieved using Full training. Transfer learning was able to solve the problem for tumor sites such as Bile Duct and Esophagus which did not have sufficient data to be trained independently. By finetuning the pre-trained model, we were able to achieve 92.31% accuracy for Esophagus and 85.71% accuracy for Bile Duct despite that these sites only had 147 and 45 samples respectively. These results have demonstrated how our pretrained model was able to effectively function as a generic model for cancer classification. Our network was able to learn the complex types of genetic alterations and genomic signatures collected from multiple cancer tissue types originating from different organs. We demonstrated how we were able to reuse the genomic signatures learned by the trained model to perform very efficient transfer learning to solve one of the biggest challenges in cancer classification which is lack of patient samples. We demonstrated how transfer learning can be used to build classifiers for cancer tumors which are lacking sufficient patient samples to be trained independently.

The results of performing multi-class classification using ALL the data including 26 organ sites covering 33 tumor types are shown in table II. Our GeneXNet model was able to achieve excellent results with an overall classification accuracy of 98.93% and a ROC AUC of 0.99 on the test dataset. The results for evaluating the performance of our GeneXNet model in comparison with state-of-the-art CNN models are also shown in table II. These results demonstrate that our GeneXNet model consistently outperformed other CNN models by a large margin. The classification accuracy achieved by our model is 98.93% which is significantly higher than the other models which achieve an accuracy below 37%. These results have demonstrated how our model can be used for classification across multiple cancer tissue types without performing the prerequisite gene feature selection. Our model has allowed training deeper network architectures with complex data like whole-transcriptome gene expressions, despite the large number of genes and managed to address the lack of training samples, without suffering from severe overfitting in comparison to state-of-the-art CNN models.

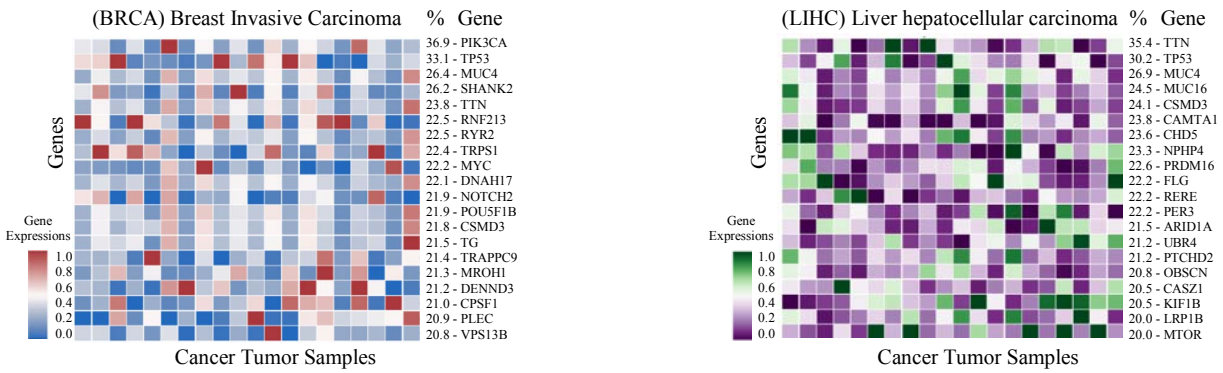

Fig. 4. Visualizing class-discriminative localization maps highlighting the important regions in the gene expressions which influenced the tumor class prediction. Each map shows the top 20 genes across 20 random samples and is a visual representation of the discriminative genes used by our network to correctly classify the tumor. The rows represent genes, columns represent samples and the values are the gene expression levels.

E. Results For Visualizing Class-Discriminative Maps

We apply the visualization procedure to identify a class discriminative Gene-Class-Activation-Map (Gene-CAM) to the underlying dataset to produce a Gene-CAM for each of the 33 individual tumors and then visualize them using heatmaps. Fig. 4 shows the resulting heatmaps of two selected tumors Breast and Liver. By mapping the resulting Gene-CAM to each input sample the network was able to identify a subset of 75 discriminative genes. For visualization, we apply a threshold where each heatmap shows the top 20 genes influencing the underlying tumor across 20 random samples. The rows represent genes, columns represent samples and the values are the gene expression levels. The gene symbols are displayed on the right of each row with the percentage of samples that have also identified this gene in their Gene-CAM. Each map is a visual representation of the discriminative genes used by the network to correctly classify the tumor. The strength of our method is that the network automatically identified a small subset of class-discriminative genes out of the total 60,483 genes originally included in each individual sample. The network automatically identified the TP53 gene as one of the top features common across all tumor types. This result implicitly validates our procedure since TP53 is considered the most commonly mutated gene in cancers that produces a protein to suppress the growth of tumors [2]. We also observed from our experiments that some of the identified discriminative genes were also common in other samples across different tumor types even though the tissues belonged to different organ sites. This subset includes: TP53, TTN, MUC16, LRP1B, CSMD3, PIK3CA, MUC4, RYR2, USH2A, FLG, PTPRD, CSMD1.

VI. CONCLUSIONS

We proposed a deep transfer learning framework for cancer diagnosis with the capability of learning the genomic signatures of whole-transcriptome gene expressions collected from multiple tumor types covering multiple organ sites. We introduced a new CNN architecture specifically designed to address the complex nature of whole-transcriptome gene expressions. We demonstrated how our trained model can be used for transfer learning to build classifiers for tumors which are lacking sufficient patient samples to be trained independently. We introduced visualization procedures to provide more insight on how our model is accurately making predictions. We believe there is great potential for further research to expand on our work for cancer diagnosis. Our work focused on designing a cancer classifier based on Total RNA sequencing using gene expressions from coding mRNA. Future work can explore learning more complex genomic signatures by including data using other multiple forms of Next Generation Sequencing platforms and also enhance prediction accuracy by building ensemble models.

REFERENCES

[1] International Agency for Research on Cancer and World Health Organization, "World Cancer Report 2019," 2019.

[2] US National Cancer Institute (NCI), "Cancer Research." [Online]. Available: https://www.cancer.gov/research.

[3] K. A. Hoadley et al., "Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer," Cell, vol. 173, no. 2, pp. 291-304.e6, Apr. 2018.

[4] K. A. Hoadley et al., "Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin," Cell, vol. 158, no. 4, pp. 929–944, Aug. 2014.

[5] P. Wu and D. Wang, "Classification of a DNA Microarray for Diagnosing Cancer Using a Complex Network Based Method," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 3, pp. 801–808, May 2019.

[6] C. Peng, X. Wu, W. Yuan, X. Zhang, and Y. Li, "MGRFE: multilayer recursive feature elimination based on an embedded genetic algorithm for cancer classification," IEEE/ACM Transactions on Computational Biology and Bioinformatics, pp. 1–1, 2019.

[7] F. Hu, Y. Zhou, Q. Wang, Z. Yang, Y. Shi, and Q. Chi, "Gene expression classification of lung adenocarcinoma into molecular subtypes," IEEE/ACM Transactions on Computational Biology and Bioinformatics, pp. 1–1, 2019.

[8] H. Lu, H. Gao, M. Ye, and X. Wang, "A Hybrid Ensemble Algorithm Combining AdaBoost and Genetic Algorithm for Cancer Classification with Gene Expression Data," IEEE/ACM Transactions on Computational Biology and Bioinformatics, pp. 1–1, 2019.

[9] C.-Q. Xia, K. Han, Y. Qi, Y. Zhang, and D.-J. Yu, "A Self-Training Subspace Clustering Algorithm under Low-Rank Representation for Cancer Classification on Gene Expression Data," IEEE/ACM Trans Comput Biol Bioinform, vol. 15, no. 4, pp. 1315–1324, Aug. 2018.

[10] E. Razak, F. Yusof, and R. A. Raus, "Classification of miRNA Expression Data Using Random Forests for Cancer Diagnosis," in 2016 International Conference on Computer and Communication Engineering (ICCCE), 2016, pp. 187–190.

[11] S. H. Bouazza, N. Hamdi, A. Zeroual, and K. Auhmani, "Gene-expression-based cancer classification through feature selection with

KNN and SVM classifiers," in 2015 Intelligent Systems and Computer Vision (ISCV), 2015, pp. 1–6.

[12] S. A. Ludwig, D. Jakobovic, and S. Picek, "Analyzing gene expression data: Fuzzy decision tree algorithm applied to the classification of cancer data," in 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2015, pp. 1–8.

[13] J.-X. Liu, Y. Xu, C.-H. Zheng, H. Kong, and Z.-H. Lai, "RPCA-Based Tumor Classification Using Gene Expression Data," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 12, no. 4, pp. 964–970, Jul. 2015.

[14] K. Liu, J. Ye, Y. Yang, L. Shen, and H. Jiang, "A Unified Model for Joint Normalization and Differential Gene Expression Detection in RNA-Seq Data," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 2, pp. 442–454, Mar. 2019.

[15] J. M. Knight, I. Ivanov, K. Triff, R. S. Chapkin, and E. R. Dougherty, "Detecting Multivariate Gene Interactions in RNA-Seq Data Using Optimal Bayesian Classification," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 2, pp. 484–493, Mar. 2018.

[16] K. R. Kukurba and S. B. Montgomery, "RNA Sequencing and Analysis," Cold Spring Harb Protoc, vol. 2015, no. 11, p. pdb.top084970, Nov. 2015.

[17] J. E. Dancey, P. L. Bedard, N. Onetto, and T. J. Hudson, "The Genetic Basis for Cancer Treatment Decisions," Cell, vol. 148, no. 3, pp. 409–420, Feb. 2012.

[18] S. Sleijfer, J. Bogaerts, and L. L. Siu, "Designing Transformative Clinical Trials in the Cancer Genome Era," JCO, vol. 31, no. 15, pp. 1834–1841, Apr. 2013.

[19] L. E. MacConaill, "Existing and Emerging Technologies for Tumor Genomic Profiling," JCO, vol. 31, no. 15, pp. 1815–1824, Apr. 2013.

[20] J. M. Rizzo and M. J. Buck, "Key Principles and Clinical Applications of 'Next-Generation' DNA Sequencing," Cancer Prev Res, vol. 5, no. 7, pp. 887–900, Jul. 2012.

[21] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," Nature Reviews Genetics, vol. 10, no. 1, pp. 57–63, Jan. 2009.

[22] C. Liu and H. S. Wong, "Structured Penalized Logistic Regression for Gene Selection in Gene Expression Data Analysis," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 16, no. 1, pp. 312–321, Jan. 2019.

[23] K. R. Kavitha, A. V. Ram, S. Anandu, S. Karthik, S. Kailas, and N. M. Arjun, "PCA-based gene selection for cancer classification," in 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 2018, pp. 1–4.

[24] S. An, J. Wang, and J. Wei, "Local-Nearest-Neighbors-Based Feature Weighting for Gene Selection," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 15, no. 5, pp. 1538–1548, Sep. 2018.

[25] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, Unsupervised, and Semi-Supervised Feature Selection: A Review on Gene Selection," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 13, no. 5, pp. 971–989, Sep. 2016.

[26] J. Tang and S. Zhou, "A New Approach for Feature Selection from Microarray Data Based on Mutual Information," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 13, no. 6, pp. 1004–1015, Nov. 2016.

[27] J.-X. Liu, Y. Xu, Y.-L. Gao, C.-H. Zheng, D. Wang, and Q. Zhu, "A Class-Information-Based Sparse Component Analysis Method to Identify Differentially Expressed Genes on RNA-Seq Data," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 13, no. 2, pp. 392–398, Mar. 2016.

[28] "The Cancer Genome Atlas (TCGA) Research Network." [Online]. Available: https://www.cancer.gov/tcga.

[29] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.

[30] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 8697–8710.

[31] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2261–2269.

[32] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[33] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," CoRR, vol. abs/1704.04861, 2017.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Identity Mappings in Deep Residual Networks," in Computer Vision – ECCV 2016, Cham, 2016, pp. 630–645.

[35] G. Huang, Y. Sun, Z. Liu, D. Sedra, and K. Q. Weinberger, "Deep Networks with Stochastic Depth," in Computer Vision – ECCV 2016, Cham, 2016, pp. 646–661.

[36] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in Proceedings of the 27th International Conference on International Conference on Machine Learning, USA, 2010, pp. 807–814.

[37] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

[38] J. Duchi, E. Hazan, and Y. Singer, "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization," Journal of Machine Learning Research, vol. 12, no. Jul, pp. 2121–2159, 2011.

[39] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning Deep Features for Discriminative Localization," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2921–2929.

[40] 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), 2017, pp. 618–626.

[41] M. Lin, Q. Chen, and S. Yan, "Network In Network," in 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Proceedings, 2014.

[42] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," in Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37, Lille, France, 2015, pp. 448–456.

[43] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-ResNet and the Impact of Residual Connections on Learning," in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, California, USA, 2017, pp. 4278–4284.

[44] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[45] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[46] S. Ruder, "An overview of gradient descent optimization algorithms," CoRR, vol. abs/1609.04747, 2016.

[47] C. Szegedy et al., "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.

[48] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," CoRR, vol. abs/1409.1556, 2014.

[49] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in Computer Vision – ECCV 2014, Cham, 2014, pp. 818–833.

[50] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, p. 2012.

[51] Y. LeCun, "Learning Invariant Feature Hierarchies," in Proceedings of the 12th International Conference on Computer Vision - Volume Part I, Berlin, Heidelberg, 2012, pp. 496–505.