Fine-Tuning Pre-Trained Convolutional Neural Networks for Women Common Cancer Classification using RNA-Seq Gene Expression

Fadi Alharbi[1]
College of Computer and Information Sciences,
Jouf University, Sakaka, KSA

Murtada K. Elbashir[2]
College of Computer and Information Sciences,
Jouf University, Sakaka, KSA
Faculty of Mathematical and Computer Sciences,
University of Gezira, Wad Madani; Sudan

Mohanad Mohammed[3]
School of Mathematics, Statistics and Computer Science,
University of KwaZulu-Natal, Pietermaritzburg,
South Africa
Faculty of Mathematical and Computer Sciences,
University of Gezira, Wad Madani; Sudan

Mohamed Elhafiz Mustafa[4]
College of Computer and Information Sciences,
Jouf University, Sakaka, KSA
College of Computer Science and Information Technology,
Sudan University of Science and Technology,
Khartoum, Sudan

Abstract—Most of the recent cancer classification methods use gene expression profile as features because it can provide very important information regarding tumor characteristics. Motivated by their success in the computer vision area now deep learning has been successfully applied to medical data because it can read non-linear patterns in a complex feature and can allow the leverage of information from unlabeled data of problems that do not belong to the problem being handled. In this paper, we implement transfer learning, which refers to the use of a model trained on one task to perform classification on another task to classify five cancer types that most commonly affect women. We used VGG16, Xception, DenseNet, and ResNet50 as base models and then added a dense layer to reflect our five-class classification problem. To avoid training over-fitting that can result in a very high training accuracy and a low cross-validation accuracy we used L2-regularization. We retrained (fine-tuned) these models using a five-fold cross-validation approach on RNA-Seq gene expression data after transforming it into 2D-image like data. We used the softmax activation function with the prediction dense layer and adam as optimizer in the model fit for all four architectures. The highest performance is obtained when fine-tuning Xception architecture, which achieved classification accuracy = 98.6%, precision = 98.6%, recall = 97.8%, and F1-score = 98% on five-fold cross-validation training and testing approach.

Keywords—*Fine-tuning; RNA-Seq; gene expression*

I. INTRODUCTION

Every cell in multicellular organisms has the same genes and every gene is not transcriptionally active in every cell, therefore the patterns of gene expression differ from cell top another. These variations may play a major role in the difference between disease and health [1]. Therefore, different types of tissues or cells' transcriptomes comparison can reveal an understanding of what constitutes different cells and how changes in transcriptional activity may contribute to diseases. In humans, a small percentage of genetic code i.e. less than 5% of the genome is transcribed from the genome's DNA code into RNA molecules or just a messenger RNA molecule [2], [3]. RNA-Seq or DNA microarray can be used to measure the transcriptome of an organism [4]. The transcription of specific genes is measured by RNA-Seq, which converts long RNAs into a library of complementary DNA (cDNA) fragments, which generate the expression profile. The expression profile can provide very important information regarding tumor characteristics, which offer deep insight into cancer detection problem [5]–[8]. Finding the highly expressed genes in tumor cells but not in normal ones based on gene expression data is considered a problem that needs to be solved using computational techniques. The high dimensionality of the gene expression data that is associated with a small number of samples revealed other challenges to the use of computational techniques. The used computational techniques include the deep learning methods which are popularly used in computer vision problems [9], [10].

Recently deep learning has emerged and succeeded in machine learning applications because it can read non-linear patterns in a complex feature and can allow the leverage of information from unlabeled data of problems that do not belong to the problem being handled. Motivated by their success, now deep learning has been successfully applied to medical data [11], [12]. Transfer learning, which refers to the use of a model trained on one task to perform classification on another task has been successfully implemented in medical data classification and analysis after the introduction of the state-of-the-art deeper learning neural network models that improve the ability of deep learning substantially [13]. There are many state-of-the-art and on-the-shelf pre-trained models that can be used as a transfer learning approach. These state-of-the-art methods include VGG16 [13], Xception [14], DenseNet [15], and ResNets [16], which are convolutional neural networks (CNN) architectures that are trained on a

very large images dataset. Fine-tuning these architectures when applied to medical data is found to be one of the successful approaches because the characteristics of the medical data are not the same as the data in which these pre-trained models are trained on. In this paper, we compared the classification performances of VGG16, Xception, DenseNet, and ResNets after fine-tuning them to classify the common women cancer using RNA-Seq gene expression data. We first converted the gene expression data into 2D-image like data and then we fed the input convolutional layer of these architectures with these 2D-images like data. The results show that the proposed approach achieved high performance as measure by the accuracy, precision, recall, and F1-score using five-fold cross-validation training and testing approach.

II. RELATED WORK

The methods that used deep learning approach for cancer classification based on gene expression data include the work of Rasool et al [17], Chen et al [18], Liao et al. [19], Kong and Yu [20], Lyu and Haque [21], Sevakula et al. [22], Danaee et al. [23]. Rasool et al. used deep learning and unsupervised features learning to detect cancer and analyses cancer types based on gene expression data. They learned a concise feature representation from unlabeled data using a sparse autoencoder. Chen et al presented a method based on deep learning known as D-GEX, which uses a multi-task multi-layer feedforward neural network to infer the expression of target genes from the expression of landmark genes. In their work, the performances of the deep learning method, Linear regression (LR), and k-nearest neighbor (KNN) regression are evaluated on microarray expression and RNA-Seq profile where they found that their deep learning methods outperform the other methods in terms of accuracy. Liao et al proposed a multi-task deep learning method to solve the few data problem of gene expression by leveraging the gene expression data of multi cancer and learn more representation for cancer that has a small number of cases. This way they enhanced the performance of diagnosing all types of cancer. Kong and Yu integrated external relational features information extracted from RNA-seq gene expression of the breast cancer into a deep neural network architecture using Graph-Embedded Deep Feedforward Networks, which enables the network layers to achieve spares connection and avoid over-fitting. They tuned their model's parameters using a grid search approach. Lyu and Haque converted the rows of the RNA-Seq gene expression data into 2D-images like data and then they trained a convolutional neural network using the obtained images like data for classifying multiple cancer types. Sevakula et al. used sparse autoencoders in combination with feature selection and normalization techniques on gene expression data and then they used a transfer learning procedure on their obtained features. They used the data of some tumor types to improve the features representation when classifying other tumor types. Danaee et al. extracted functional features from the gene expression profile using Stacked Denoising Autoencoder (SDAE) and then they used supervised classification to evaluate the performance of the obtained features to be used for cancer detection and identification. Also, they analyzed the SDAE connectivity matrices to identify a set of highly interactive genes.

III. MATERIAL AND METHODS

A. Dataset

Five RNASeq gene expression profile for different types of women cancers were downloaded from the genomic data commons (GDC) data portal. These types of cancers include breast (BRCA), ovarian (OV), colon adenocarcinoma (COAD), lung adenocarcinoma (LUAD), and thyroid (THCA) cancer. We used TCGAbiolinks package in R to download these RNASeq gene expressions profile [24]. TCGAbiolinks has GDCquery function which uses GDC API to search and download the data and it has many arguments such as project, legacy, data.category, platform, data.type, experimental.strategy, sample.type, and workflow.type. These arguments are normally passed to the GDCquery to filter and determine the type of data that should be downloaded. The project argument determines a valid TCGA project data list that should be downloaded. Five project codes corresponding to our five types of cancer, which are TCGA-BRCA, TCGA-OV, TCGA-COAD, TCGA-LUAD, and TCGA-THCA were used as project argument. The legacy parameter is adjusted to "true", to get the unmodified data in the GDC data portal that is stored in the legacy repository. Consequently, to quantifying the gene expression data and to filter the data to be downloaded we adjusted data.type variable to "Gene expression quantification" and data.category has been set to "Gene expression". We used the data produced using the "Illumina HiSeq" platform. The file.type argument is set to "results" to filtering the legacy database, and since we are looking for counts data "RNA-Seq" protocol that was used to perform the laboratory analysis was chosen as experimental.strategy parameter to obtain the expression profiles. In this work, we are interested in the tumor samples only thus, "Primary solid Tumor" adjusted as sample.type argument to filter out the normal samples. The downloaded data is in a form of a matrix, where the columns represent the samples and the rows contain the genes, i.e. features (equivalently covariates). The five types of cancers have 2166 samples, along with 19947 common genes. To reduce the number of the genes, we constructed a symmetric square matrix of Spearman correlation known as Array-Array Intensity Correlation (AAIC) between samples to determine the highly correlated genes. The visualization of this matrix is shown in Fig. 1, where high correlated genes are depicted in dark color. A correlation cut off equal to 0.6 is used to remove the highly correlated genes. To ensure that we can infer the level of expression correctly without biases, we applied a normalization process on the obtained gene expression profile using TCGAanalyze_Normalization function [25]–[28]. Finally, the gene expression profile is filtered by selecting mean values higher than 0.25 across all samples. The final obtained gene expression profile after applying these preprocessing steps has 2166 samples with 14899 genes. The number of samples in each cancer type is as follows BRCA (1082), COAD (135), LUAD(275), OV (304), and THCA (370). These samples are transformed into 2D-images like data to be suitable for the convolutional layer of CNN architecture. The motivation to convert the data into 2D-images comes from many researches works e.g [3], [29].

To capture the linear and non-linear dependencies, we visualized our final obtained data in two-dimensional space using Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Entities (t-SNE). PCA and t-SNE are
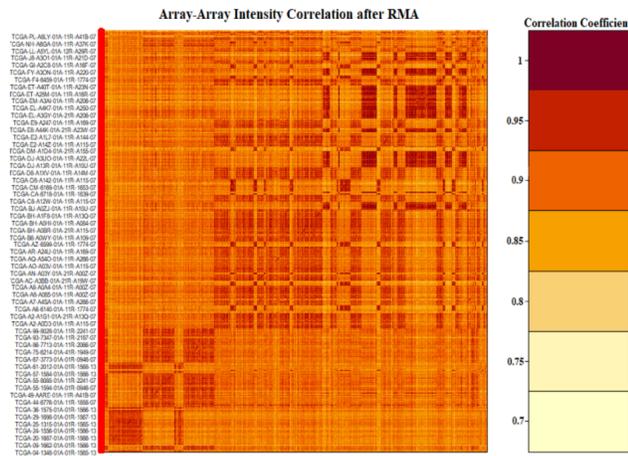
Fig. 1. Symmetric Square Matrix of Spearman Correlation or AAIC Visualization.

Fig. 3. Visualizing the Gene Expression Data using t_SNE.

linear and nonlinear projection methods, respectively. These two methods are used to capture the linear and non-linear dependencies. The obtained projection is depicted in Fig. 2 and Fig. 3. From the two figures, it is clear that the cancers types are overlapped in both the linear and non-linear projection.

Fig. 2. Visualizing the Gene Expression Data using PCA.

B. Problem Formulation

In this paper, we cast common women cancers classification based on gene expression data as a multiclass classification problem. The gene expressions for all the cases are transformed into 2D-images $X = (x_1, x_2, \ldots . x_N)$ that are associated with a ground truth class label $Y = (y_1, y_2, \ldots y_N)$. We are intending to develop a classification function $X \rightarrow Y$. The developed classification function should minimize a loss function using n training samples. We encoded the labels as a vector $y \in \{0, 1\}^M$, where $M = 5$ (the number of the woman common cancer types). We did an investigation using different loss functions and come up with a conclusion that the loss function that gives the highest performance is the categorical_crossentropy, which can be formulated mathematically as given in equation 1.

$$Loss = - \sum_{i=1}^{outputsize} y_i.log\hat{y_i} \qquad (1)$$

Where $\hat{y_i}$, y_i, and outpusize represent i^{th} scaler value, the corresponding target value, and the number of scalar values in the model output, respectively.

Since our data is not large enough to train a CNN model from scratch, we used transfer learning because of its outstanding performance in the computer vision domain in general and in the medical data domain in specific [30]–[35]. We fine-tuned different models as a base model and then added a dense layer to reflect our five-class classification problem. To avoid training over-fitting that can result in a very high training accuracy and a low cross-validation accuracy we used L2-regularization. We compared the classification performance of the following models: ResNet50, DenseNet, Xception, and VGG16.

C. Obtaining the 2D-Images from the Gene Expression Data

We transformed our gene expression data into 2D-images by reshaping them into a square matrix of (123 By 123) to fit the convolutional layer of the used CNN methods. Transforming the gene expression data into 2D-images inspired by the work in [3], [29]. The number of the columns or features in the dataset (14899 genes) is not sufficient to be transformed into 123 by 123 matrix, therefore we appended columns of zeroes to the gene expression data. This kind of modification is normally applied to make the size of the data adjustable to the requirement.

D. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) are deep learning models mostly used for image classification. The connectivity of the neurons in the CNN is similar to that of the animal visual cortex and they have special filters to capture the temporal dependencies in an image features and reduce them into an easier arrangement that can be processed without dropping important features to obtain high classification performance.

A sequence of layers makes the architecture of CNN and each layer can transform one volume of activation to another using a differentiable function. Normally, convolutional, pooling and fully-connected layers are stacked to build a CNN model. CNN takes an image that has n rows, m columns, and 3 color channels ((R, G, and B) as input while considering the special structure of an image into account. CNN uses a convolutional layer with three features map to represent the color channels and f × f local receptors or filter size. The features will be read using a stride. If a stride of 1 is used, the result will be a layer of 3× (m-f+1) × (n-f+1) hidden feature neurons. The convolutional operation, which multiplies the elements of the filter by the element of the image matrix element-wise is used to generate the features map. Sliding the filters across the input image matrix will generate the rest of the features. The mathematical formula of the convolutional operation is given in equation 2.

$$O(i,j) = \sum_{k=1}^{f}(\sum_{l=1}^{f} input(i+k-1, j+l-1)kernel(k,l)) \quad (2)$$

Where i runs from 1 to $m - f + 1$ and j runs from 1 to $n - f + 1$.

E. Transfer Learning

It is very challenging and expensive to acquire medical data and for gene expression datasets the small number of cases and a large number of dimensions can hamper the performance of deep learning significantly. On the other hand, deep learning models require a very large number of data for training to give good classification performance. To overcome this problem, we can use transfer learning to leverage information from other data to understand the distribution of our gene expression data. There are many state-of-the-art and on-the-shelf pre-trained models that can be used as a transfer learning approach. These state-of-the-art methods include VGG16 [13], Xception [14], DenseNet [15], and ResNets [16], which are convolutional neural networks (CNN) architectures that are trained on a very large images dataset. Fine-tuning these architectures, which means re-training them when applied to medical data is found to be one of the successful approaches because the characteristics of the medical data are not the same as the data in which these pre-trained model are trained on.

F. Experimental Setup

After trying many state-of-the-art CNN pre-trained architectures we selected the following models: ResNet50, VGG16, DenseNet, and Xception . These models are considered to be a breakthrough for CNN's progress as they have applied unique deep learning architecture. ResNet50 has 50 layers and is the first to introduce a residue model in CNNs architectures to ease the deeper architectures training and solve the degradation problem, which means that not all architectures are similarly easy to optimize [16]. In ResNet50, instead of learning unreferenced functions, the layers are formulated as learning residual functions with reference to the input layer. VGG16 uses a very small convolutional filter with a very deep architecture. DenseNet is one of the new on-the-shelf pre-trained CNNs

for visual object recognition that has a similar architecture to ResNet with some essential differences. In DenseNet, each layer is connected to every other layer in a feed-forward fashion. With its structure, DenseNet reduced the problem of vanishing-gradient, make the feature propagation strong and promote its reuse, and uses a small number of features map, which makes it parameters efficient [36]. Xception is inspired by inception [37], where it replaces inception modules with separable convolutions. In all the architectures, We used softmax activation with the prediction dense layer and adam as optimizer when fitting the models. Also, we used L2 kernel and bias regularization for all the architectures. The categorical cross-entropy error function is used to perform the training where we used a five-fold cross-validation approach. We used 100 epochs for training in each architecture. To randomize the whole learning producers and ovoid over-fitting, we shuffled the training data in each epoch.

G. Performance Measures

Four measures are used to evaluate the different transfer learning architectures. These measures are the classification accuracy, precision, recall, and F1 -score. They are considered among the most frequent measures that are used to evaluate the performance of computational methods on medical data. The accuracy and F1-score are used to evaluate the comprehensive classification performance while precision and recall are used to evaluate the rate of recognition and sensitivity respectively. The mathematical formulas for these measures are as follows:

$$accuracy = \frac{\sum_i m_{ii}}{\sum_{i,j} m_{ij}} \quad (3)$$

$$recall_j = \frac{m_{jj}}{\sum_i m_{ji}} \quad (4)$$

$$precision_j = \frac{m_{jj}}{\sum_j m_{ji}} \quad (5)$$

$$F1 - score_j = \frac{2 \times recall_j \times precision_j}{recall_j + precision_j} \quad (6)$$

i and j stand for the different classes

IV. RESULTS AND DISCUSSION

In this study, experiments are conducted to classify the five common women cancers: breast, ovarian, colon adenocarcinoma, lung adenocarcinoma, and thyroid cancer. As stated in the methodology, we used five-fold cross-validation, which is a very useful and rigorous validation method for estimating the performance of the classification model, especially with a small dataset. In the five-fold cross-validation approach, the training dataset is divided into five equal sets, four of these sets are used as training and the fifth one is used as a testing set. This process is repeated five times by removing one set to represent the testing set. We used a fine-tunned transfer learning approach in which we tried different architectures as a base model. We tried different activation functions in the prediction dense layer and different optimizer when fitting the model. From the results, we found that the softmax activation

function with the adam optimizer obtained the best results in all the used architectures. The architectures that obtained the best results are ResNet50, DenseNet, Xception, and VGG16. The performances of ResNet50, DenseNet, Xception, and VGG16 have been evaluated for each fold. The final classification performance is calculated as the average of the results of the five testing sets. Table I shows that Xception model has the highest performance in terms of precision, recall, F1-Score, and accuracy compared to DenseNet, ResNet50 , and VGG16 models. The three architectures are trained for 100 epochs. Fig. 4, Fig. 5, and Fig. 6 show the validation accuracy, the validation Loss, and the F-measure graphs respectively for the first fold of the Xception architecture.
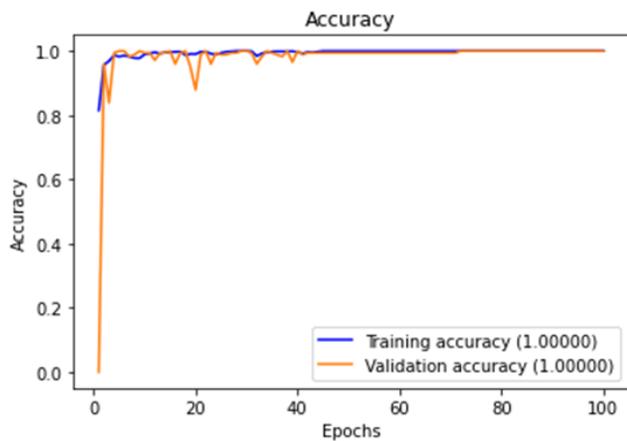

Fig. 6. F1-measure Curve of Xception Architecture.

Fig. 4. Training Accuracy Curve of Xception Architecture.

TABLE I. CLASSIFICATION PERFORMANCES OF THE FINE-TUNED XCEPTION, DENSENET, RESNET50, AND VGG16 ARCHITECTURES.

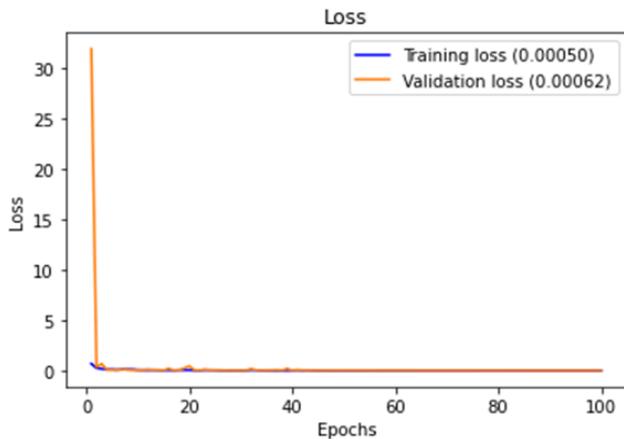| Classification Method	Folds	Performance Metrics (%)			
		Precision	Recall	F1-Score	Accuracy
Xception	Fold1	98	97	97	99
	Fold2	98	96	97	97
	Fold3	99	99	99	99
	Fold4	99	98	98	99
	Fold5	99	99	99	99
	Average	**98.6**	**97.8**	**98**	**98.6**
DenseNet	Fold1	98	97	98	98
	Fold2	98	97	98	98
	Fold3	98	98	98	98
	Fold4	98	97	97	98
	Fold5	97	97	97	98
	Average	**97.8**	**97.2**	**97.6**	**98.0**
ResNet50	Fold1	95	96	96	97
	Fold2	98	97	98	98
	Fold3	99	99	99	99
	Fold4	98	96	97	97
	Fold5	94	94	94	96
	Average	**96.8**	**96.4**	**96.8**	**97.4**
VGG16	Fold1	94	93	93	95
	Fold2	90	85	86	90
	Fold3	96	95	96	97
	Fold4	94	92	93	95
	Fold5	97	96	96	97
	Average	**94.2**	**92.2**	**92.8**	**94.8**

Fig. 5. Training Loss Curve of Xception Architecture.

Schematic Xception architecture diagram for cancer multi-class classification using transfer learning is shown in Fig. 7, where an overview of the layers that comprise the architecture of the base Xception architecture and the layers that we added are depicted. Different colors are used to depict the different layers.

Table I shows that fine-tuned Xception architecture achieved classification accuracy = 98.6%, precision = 98.6%,
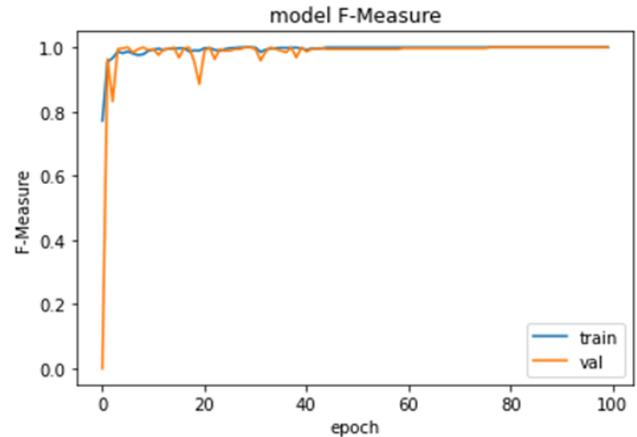
recall = 97.8%, and F1-score = 98% on five-fold cross validation training and testing approach. The table shows that Xception achieved the highest performance in all the performance measures. DenseNet architecture follows Xception in terms of performance. It achieved classification accuracy = 98%, precision = 97.8%, recall = 97.2%, and F1-score = 97.6% on a test set. ResNet50 achieved classification accuracy =97.4%, precision = 96.8%, recall =96.4%, and F1-score = 96.8% on a test set. The lowest performance is achieved when using VGG16, which obtained classification accuracy = 94.8%, precision = 94.2%, recall = 92.2%, and F1-score = 92.8%.

The confusion matrices for the five folds of Xception are shown in Figure 8. Figure 8 also shows the overlapped confusion matrix, which is calculated as the summation of the five folds convolution matrices to reflect the general performance of the Xception model. The overlapped confusion matrix shows that Xception model classified THCA, OV, BRCA, and COAD better than LUAD cancer type in the multi-class classification
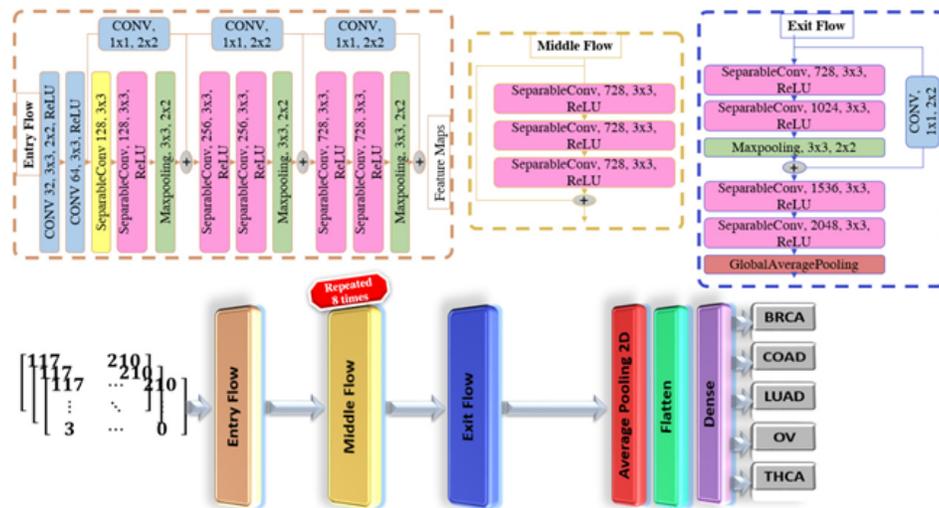
Fig. 7. Schematic Xception Architecture Diagram for Cancer Multiclass Classification using Transfer Learning.

task. This is because the dataset is imbalanced and the classifier does not have an equal number of instances for all the classes during training time.

V. CONCLUSIONS

In this paper, we used the fine-tuning transfer learning approach on RNA-Seq gene expression data to classify five cancer types that mostly affect women. These five types are breast (BRCA), ovarian (OV), colon adenocarcinoma (COAD), lung adenocarcinoma (LUAD), and thyroid cancer (THCA). The RNA-Seq gene expression data for the five cancer types is downloaded from genomic data commons (GDC) data portal using TCGAbiolinks package in R. The downloaded data is in a form of a matrix, where the columns represent the samples and the rows contain the genes. The five types of cancers have 2166 samples, along with 19947 common genes. We used Spearman correlation to reduce the number of the genes by removing the highly correlated genes using correlation cut off equal to 0.6. To ensure that we can infer the level of expression correctly without biases, we applied a normalization process on the obtained gene expression profile using TCGAanalyze Normalization function. Finally, the gene expression profile is filtered by selecting mean values higher than 0.25 across all samples. The final obtained gene expression profile after applying these preprocessing steps has 2166 samples with 14899 genes. These samples are transformed into 2D-images like data to be suitable for the convolutional layer of CNN architecture. We fine-tuned four pre-trained models on the RNA-Seq gene expressing data, namely, ResNet50, DenseNet, Xception, and VGG16. Xception architecture shows the highest performance where it achieved classification accuracy = 98.6%, precision = 98.6%, recall = 97.8%, and F1-score = 98% on five-fold cross-validation training and testing approach.

ACKNOWLEDGMENT

REFERENCES

[1] J. Adams, "Transcriptome: connecting the genome to gene function," *Nat Educ*, vol. 1, no. 1, p. 195, 2008.

[2] M. C. Frith, M. Pheasant, and J. S. Mattick, "The amazing complexity of the human transcriptome.," *European journal of human genetics: EJHG*, vol. 13, no. 8, p. 894, 2005.

[3] M. K. Elbashir, M. Ezz, M. Mohammed, and S. S. Saloum, "Lightweight convolutional neural network for breast cancer classification using rna-seq gene expression data," *IEEE Access*, vol. 7, pp. 185338–185348, 2019.

[4] Z. Wang, M. Gerstein, and M. Snyder, "Rna-seq: a revolutionary tool for transcriptomics," *Nature reviews genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[5] M. Maienschein-Cline, J. Zhou, K. P. White, R. Sciammas, and A. R. Dinner, "Discovering transcription factor regulatory targets using gene expression and binding data," *Bioinformatics*, vol. 28, no. 2, pp. 206–213, 2012.

[6] E. E. Schadt, J. Lamb, X. Yang, J. Zhu, S. Edwards, D. GuhaThakurta, S. K. Sieberts, S. Monks, M. Reitman, C. Zhang, *et al.*, "An integrative genomics approach to infer causal associations between gene expression and disease," *Nature genetics*, vol. 37, no. 7, pp. 710–717, 2005.

[7] J. Krammer, K. Pinker-Domenig, M. E. Robson, M. Gönen, B. Bernard-Davila, E. A. Morris, D. A. Mangino, and M. S. Jochelson, "Breast cancer detection and tumor characteristics in brca1 and brca2 mutation carriers," *Breast cancer research and treatment*, vol. 163, no. 3, pp. 565–571, 2017.

[8] M. Mohammed, H. Mwambi, B. Omolo, and M. K. Elbashir, "Using stacking ensemble for microarray-based cancer classification," in *2018 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE)*, pp. 1–8, IEEE, 2018.

[9] H. Pan, B. Wang, and H. Jiang, "Deep learning for object saliency detection and image segmentation," *arXiv preprint arXiv:1505.01173*, 2015.

[10] J. Han, D. Zhang, G. Cheng, N. Liu, and D. Xu, "Advanced deep-learning techniques for salient and category-specific object detection: a survey," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 84–100, 2018.

[11] H. Jia, Y. Xia, Y. Song, W. Cai, M. Fulham, and D. D. Feng, "Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging," *Neurocomputing*, vol. 275, pp. 1358–1369, 2018.

[12] J. Zhang, Y. Xia, Y. Xie, M. Fulham, and D. D. Feng, "Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1521–1530, 2017.
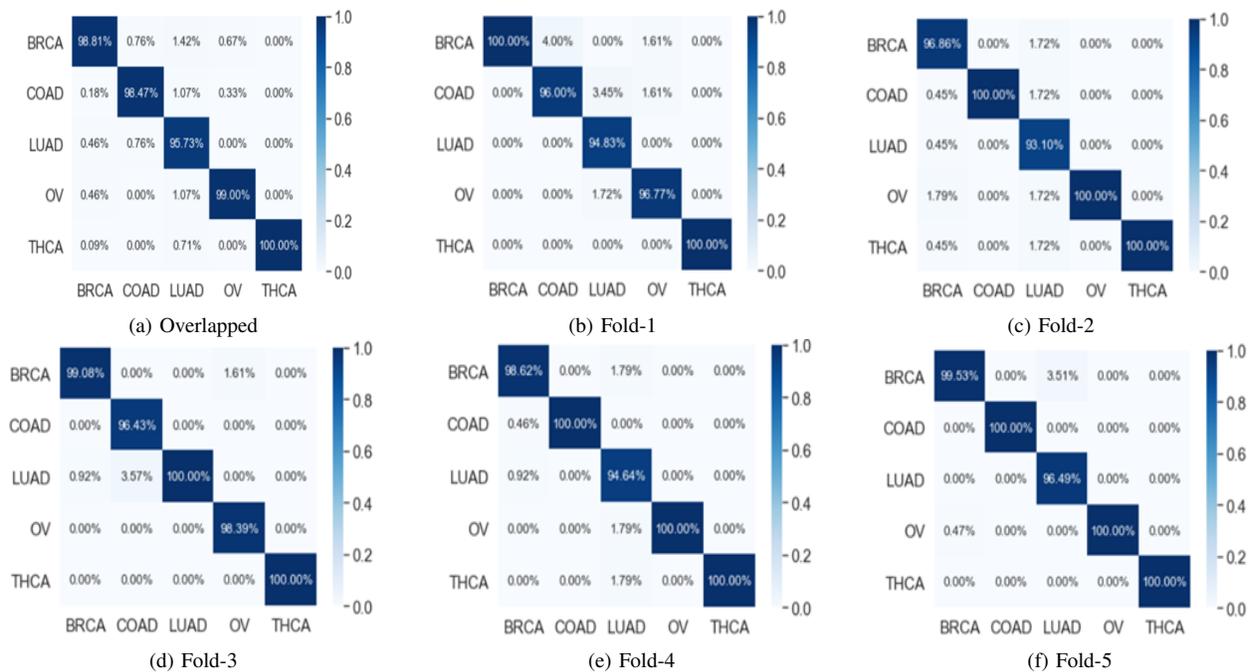
Fig. 8. The Overlapped and 5-fold Confusion Matrix Results of the Multiclass Classification Task: (a) Overlapped Confusion Matrix, (b) Fold-1 CM, (c) Fold-2 CM, (d) Fold-3 CM, (e) Fold-4 CM, and (f) Fold-5 CM.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[14] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.

[15] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[16] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[17] R. Fakoor, F. Ladhak, A. Nazi, and M. Huber, "Using deep learning to enhance cancer diagnosis and classification," in *Proceedings of the international conference on machine learning*, vol. 28, ACM New York, USA, 2013.

[18] Y. Chen, Y. Li, R. Narayan, A. Subramanian, and X. Xie, "Gene expression inference with deep learning," *Bioinformatics*, vol. 32, no. 12, pp. 1832–1839, 2016.

[19] Q. Liao, Y. Ding, Z. L. Jiang, X. Wang, C. Zhang, and Q. Zhang, "Multi-task deep convolutional neural network for cancer diagnosis," *Neurocomputing*, vol. 348, pp. 66–73, 2019.

[20] Y. Kong and T. Yu, "A graph-embedded deep feedforward network for disease outcome classification and feature selection using gene expression data," *Bioinformatics*, vol. 34, no. 21, pp. 3727–3737, 2018.

[21] 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*, pp. 89–96, 2018.

[22] R. K. Sevakula, V. Singh, N. K. Verma, C. Kumar, and Y. Cui, "Transfer learning for molecular cancer classification using deep neural networks," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 16, no. 6, pp. 2089–2100, 2018.

[23] P. Danaee, R. Ghaeini, and D. A. Hendrix, "A deep learning approach for cancer detection and relevant gene identification," in *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2017*, pp. 219–229, World Scientific, 2017.

[24] A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, *et al.*,

"Tcgabiolinks: an r/bioconductor package for integrative analysis of tcga data," *Nucleic acids research*, vol. 44, no. 8, pp. e71–e71, 2016.

[25] D. Risso, K. Schwartz, G. Sherlock, and S. Dudoit, "Gc-content normalization for rna-seq data," *BMC bioinformatics*, vol. 12, no. 1, p. 480, 2011.

[26] J. H. Bullard, E. Purdom, K. D. Hansen, and S. Dudoit, "Evaluation of statistical methods for normalization and differential expression in mrna-seq experiments," *BMC bioinformatics*, vol. 11, no. 1, pp. 1–13, 2010.

[27] K. D. Hansen, R. A. Irizarry, and Z. Wu, "Removing technical variability in rna-seq data using conditional quantile normalization," *Biostatistics*, vol. 13, no. 2, pp. 204–216, 2012.

[28] W. Zheng, L. M. Chung, and H. Zhao, "Bias detection and correction in rna-sequencing data," *BMC bioinformatics*, vol. 12, no. 1, p. 290, 2011.

[29] 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*, pp. 89–96, 2018.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[31] X. Liu, C. Wang, J. Bai, and G. Liao, "Fine-tuning pre-trained convolutional neural networks for gastric precancerous disease classification on magnification narrow-band imaging images," *Neurocomputing*, vol. 392, pp. 253–267, 2020.

[32] T. Tan, Z. Li, H. Liu, F. G. Zanjani, Q. Ouyang, Y. Tang, Z. Hu, and Q. Li, "Optimize transfer learning for lung diseases in bronchoscopy using a new concept: sequential fine-tuning," *IEEE journal of translational engineering in health and medicine*, vol. 6, pp. 1–8, 2018.

[33] G. Wimmer, A. Vécsei, and A. Uhl, "Cnn transfer learning for the automated diagnosis of celiac disease," in *2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6, IEEE, 2016.

[34] H. G. Kim, Y. Choi, and Y. M. Ro, "Modality-bridge transfer learning for medical image classification," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–5, IEEE, 2017.

[35] E. Ribeiro, A. Uhl, G. Wimmer, and M. Häfner, "Transfer learning for colonic polyp classification using off-the-shelf cnn features," in *International Workshop on Computer-Assisted and Robotic Endoscopy*, pp. 1–13, Springer, 2016.

[36] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE confer-* ence on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9, 2015.