**COM865: Research Project 2022-23**

**Research Paper**

**School of Computing, Engineering & Intelligent Systems**

**Anthony Omowumi**

**B00892305**

**Cancer Gene Expression Classification Using Transfer Learning and Explainable AI**

**Supervisor: Dr. Pratheepan Yogarajah**

**Second Marker: Prof. Kevin Curran**

**15th September 2023**



**Cancer Gene Expression Classification Using Transfer Learning and Explainable AI**

**Anthony Omowumi (B00892305)**

***MSc Data Science***

***Ulster University Magee Campus***

***Northern Ireland.***

[***anthony-o@ulster.ac.uk***](mailto:anthony-o@ulster.ac.uk)

***Abstract***- ***Existing traditional methods mostly use feature selection combined with simple classifiers for gene expression analysis showing average classification performance.*** ***In recent years, deep learning algorithms have been applied in different applications and achieved better performance compared to these traditional methods, but deep learning algorithms require a huge amount of data for training purposes***. ***However, Gene expression analysis is a complex process due to relatively few samples available with high dimensional data. That is why the transfer learning technique was employed in this paper to tackle this issue and RNA-Seq gene expression data from Pan-Cancer Atlas was used in the classification of 33 prevalent tumor types. The high dimensional RNA-Seq data was embedded into 2-D images and four pre-trained base models which include Xception, DensetNet169, ResNet50, and VGG19 were used. Some layers of the last convolutional block in the base model were trained to allow the models to learn some patterns/forms in the medical data since the models were trained on ImageNet data that does not have any class label belonging to medical data/gene expression data. ResNet50 performed better with an accuracy of 92% on the 10% test set with a precision of 90%, recall of 88%, and F1 score of 89%. This performance on the unseen data (test data) shows that there is a good generalization performance of the model. Furthermore, Explainable AI - IntegratedGradients was used to check the attribution of the image predicting a specific class label. This is to understand the part of the images leading to the prediction of the outcome. Overall, the study was good in achieving its aim and could be applied to other genomics data.***

***Keywords- RNA-Seq, Transfer learning, Fine-tuning, Explainable AI, IntegratedGradients.***

1. **INTRODUCTION**

Cancer is one of the leading causes of death worldwide according to the World Health Organization (WHO), accounting for close to 10 million deaths in 2020, or close to 1 in 6 deaths [1]. According to Cancer Research United Kingdom [2], projected 18.1 million new cases of cancer worldwide in 2020 and 28 million new cases annually by 2040. The UK's incidence of cancer is higher than that of the third of Europe and higher than that of 90% of the world. Cancer results from abnormal growth of cells because of the complex interaction between genes (deregulation due to mutation and epigenetic modification) and environment (That is, the release of Carcinogens substances) [3].

Gene Expression Analysis provides information about the roles played by different genes in the development and progression of cancer and can serve as a marker for early cancer detection. Recently, gene expression analysis has become an essential approach for overcoming difficulties in cancer detection and medication discovery. In multicellular organs, all the genes are identical, but because some genes are transcriptionally inactive in some cells, different patterns of gene expression exist in different cells. These variations may have a substantial effect on how sickness and health are distinguished. This means that comparing the transcriptomes of various tissues or cell types may assist in understanding how different cell types are made up and how differences in transcriptional activity may result in disease and this is what is known as Gene Expression [4], [6]. The transcriptome of an organism can be assessed using DNA microarray or RNA-Seq [5]. The conversion of RNA molecules into complementary DNA (cDNA) and determination of the nucleotide sequence in the cDNA are necessary steps in the Next-Generation Sequencing (NGS) method known as RNA-sequencing (RNA-Seq), which is used to analyze and quantify gene expression. RNA-Seq data is taken into consideration for this investigation since it has better selectivity, resolution, sensitivity to differential expression, and dynamic range than DNA microarrays [6], [7]. The existing traditional methods mostly use feature selection combined with simple classifiers for gene expression analysis showing average classification performance [4], [8]. In recent years, deep learning algorithms have been applied in different applications and achieved better performance compared to these traditional methods, but deep learning algorithms require a huge amount of data for training purposes. However, Gene expression analysis is a complex process due to relatively few samples available with high dimensional data.

Even though the paper [8] used a Convolutional Neural Network to achieve an accuracy of 95.59%, this paper opted for a transfer learning approach to deal with the complexity of the data and the limited number of samples that were available, as well as to save time and computational resources. The paper [4] also used this methodology on five common women's cancers, but this paper used 33 common tumor types obtained from the Pan-Cancer Atlas.

First,to fit the pre-trained base models using the transfer learning approach with some pre-trained base models – VGG19, Xception, DenseNet169, and ResNet50 [9] the genes with small variance were first filtered out across the samples to reduce the number of genes from 20531 to 10363. Next, the high dimensional data (10363x1) was embedded into 2-D images (102x102) by reshaping the data from a 10363x1 array into 102x102 images by adding zeros to the end. The pre-trained base models mentioned above underwent fine-tuning, which involved freezing some model parts, retraining some model blocks/layers, and retraining the Fully Connected layer, which serves as the model's classification layer. Since none of these qualities—that is, the characteristics of medical data—are present in the objects or images used to train the models, this approach was carried out to enable some layers of the models to learn patterns or forms in the medical data. Furthermore, importantly to enable the end user to understand, interpret, and trust the outcome of the classification results, an explainable AI approach ***-*** IntegratedGradients was used to check the attribution of the image predicting a specific class label. This is to understand the part of the images leading to the prediction of the outcome. The motivation for this research work is based on the paper [4], [8]

The remaining part of the paper is structured as follows. Section II analyzed relevant prior research on Explainable AI, transfer learning, and tumor type classification. The dataset used, the preprocessing, additional methods used in the study, and the experimental setting are all described in Section III. The experimental results and discussion and comparison with the research work [8] are in Section IV, while Section V entails the Conclusion of the paper and Section VI talks about Future Works.

1. **EXISTING WORK**

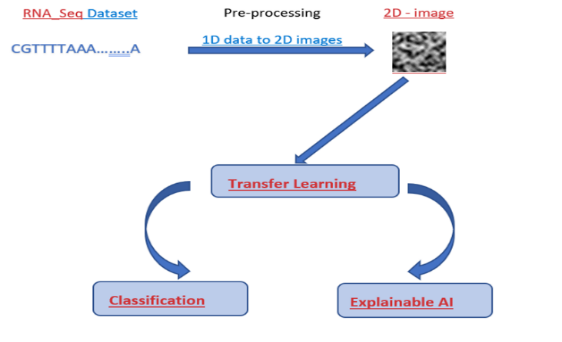
Alharbi et al. [4] used a transfer learning approach for the classification of the five cancers that are most common in women: thyroid, breast, ovarian, colon, and lung adenocarcinomas. VGG19, Xception, DenseNet, and ResNet50 were the four pre-trained models utilized in the study as the base models for the research to assess their performance in the categorization of tumors. The methodology involved converting the gene expression data into 2D image-like data, which was then supplied to the input convolutional layer of the networks. Five-fold cross-validation was used to fine-tune/retrain the models, and when the average of each model was calculated, Xception emerged as the model with the highest accuracy, at 98.6% among the compared models [4]. Another Research team classified 33 common tumor types using the Convolutional Neural Network by embedding the high dimensional RNA-Seq data in 2D pictures, with an accuracy of 95.59%. Additionally, Guided Grad Cam was employed to pinpoint the area of the input image that had the greatest influence on the classification score [8].

Also, another researcher used a transfer learning approach to identify tomato plant disease based on leaf images with 6 different pre-trained models- VGG16, AlexNet, DenseNet121, ResNet50, InceptionV3, and MobileNetV1 were used with a fine-tuning method of some layers been use/retrained and some layers were removed to enhance the accuracy. MobileNetV1 23 last layers were retrained to give an accuracy of 99% and the models built have accuracies greater than 97% except one with an accuracy of 95%. The work shows a good generalization performance of the models without data augmentation. This technique will be utilized in this study as well [10].

In a recent study, Meena and Hasija used XAI-SHAP on the Boost ML model to find promising diagnostic biomarkers in Squamous cell carcinoma (SCC) for non-melanoma skin cancer. There are 23 major genes that have been determined to be linked to the development of SCC. In particular, the work in the field of bioinformatics for the identification of biomarkers helpful for prognostic and predictive purposes [11] gives evidence in support of the application of XAI on ML models to quantify and carefully analyze the outcomes of predictions.

1. **METHODS AND MATERIALS**

Publicly available RNA-Seq Expression data (Secondary data) are used in this work [10]. The Cancer Genome Atlas (TCGA) has sequenced and handled a sizable quantity of tumor tissues. From these samples, TCGA further evaluated over 11,000 tumors from the 33 most common types of cancer, which enabled the creation of the Pan-Cancer Atlas. The National Cancer Institute (NCI) and the National Human Genome Research Institute jointly founded the TCGA project in 2006 with the goal of customizing and identifying the genetic mutations responsible for various cancer types using genome sequencing and bioinformatics [12]. The methodology approach that would be used in this research is a synthesis of Lyu and Haque [8], and Alharbi et al [4]. The Python programming language and its packages are used in this investigation. For data mining, data modeling, validation, testing, and explainability/interpretability, libraries such as TensorFlow, Keras libraries, and Alibi Explain were installed and used [13]. The overall workflow in this paper is represented in the diagram “Figure 1” which shows the sequence in which this paper will take in achieving its proposed work.



***Figure 1: Basic Proposed Work Architecture***

1. ***The Identified Dataset***

33 tumor types of RNA-Seq gene expression data, normalized to level 3 were used in this study (Available at Pan Cancer Atlas) [12]. The information includes a cancer-type classification column, 20531 genes were reduced to 10363 by filtering out the genes using a threshold variance of 1.19 to filter out genes whose samples remain almost unchanged across the expression levels, and 10446 tumor samples (Rows). The details of the 33 tumor types, cohort names, and the number of samples are shown in “Table 1”.

***TABLE 1: Tumor Types, Cohort, and their Samples Count.***

|  |  |  |
| --- | --- | --- |
| **Tumor Type** | **Cohort Names** | **Number of Samples** |
| Adrenocortical Carcinoma | ACC | 79 |
| Bladder urothelial carcinoma | BLCA | 427 |
| Breast invasive carcinoma | BRCA | 1212 |
| Cervical and Endocervical cancers | CESC | 309 |
| Cholangiocarcinoma | CHOL | 45 |
| Colon Adenocarcinoma | COAD | 328 |
| Lymphoid Neoplasm Diffuse Large B-cell Lymphoma | DLBC | 48 |
| Esophageal Carcinoma | ESCA | 196 |
| Glioblastoma Multiforme | GBM | 171 |
| Head and Neck Squamous Cell Carcinoma | HNSC | 566 |
| Kidney Chromophobe | KICH | 91 |
| Kidney Renal Clear Cell Carcinoma | KIRC | 606 |
| Kidney Renal Papillary Cell Carcinoma | KIRP | 323 |
| Acute Myeloid Leukemia | LAML | 173 |
| Brain Lower Grade Glioma | LGG | 530 |
| Liver Hepatocellular Carcinoma | LIHC | 423 |
| Lung Adenocarcinoma | LUAD | 576 |
| Lung Squamous Cell Carcinoma | LUSC | 552 |
| Mesothelioma | MESO | 87 |
| Ovarian Serous Cystadenocarcinoma | OV | 307 |
| Pancreatic Adenocarcinoma | PAAD | 183 |
| Pheochromocytoma and Paraganglioma | PCPG | 187 |
| Prostrate Adenocarcinoma | PRAD | 550 |
| Rectum Adenocarcinoma | READ | 105 |
| Sarcoma | SARC | 265 |
| Skin Cutaneous Melanoma | SKCM | 473 |
| Stomach Adenocarcinoma | STAD | 450 |
| Testicular Germ Cell Tumors | TGCT | 156 |
| Thyroid Carcinoma | THCA | 568 |
| Thymoma | THYM | 122 |
| Uterine Corpus Endometrial Carcinoma | UCEC | 201 |
| Uterine Carcinosarcoma | UCS | 57 |
| Uveal Melanoma | UVM | 80 |

1. ***Data Preprocessing and Transformation***

In this procedure, the raw data was cleaned and transformed into an effective format that is suited for developing models using data mining techniques.

***Feature Selection***: According to the approach described by Lyu and Haque [8], feature columns (representing gene names) were chosen by matching genes with an annotation file derived from NCBI. A variance threshold of 1.19, which served to exclude genes with little change in expression across samples, was a crucial requirement for inclusion. The variance threshold approach, which is frequently used to select features, automatically disqualifies zero variance features or those that keep the same values throughout all samples. The decision to use 1.19 as the threshold variance score was influenced by the justification put forward in reference [8], [14] which is that the cutoff optimizes classification accuracy while maintaining the integrity of the gene set. A lower criterion would keep more genes, ensuring that true gene-specific biomarkers were not overlooked, whereas a higher threshold would exclude more "weak features," potentially improving classification accuracy.

After this preliminary filtering, a thorough manual review was performed on the dataset to find and remove any discrepancies and redundant data, such as duplicate entries, in an Excel worksheet. At this point, a feature name file that had already undergone preprocessing was used to cross-reference and compare with the dataset in the Excel spreadsheet. Following a filtering process to remove any gene or feature names not included in the reference file, the number of genes was reduced to 10363 in total.

***Log Transformation and Image Synthesis***: The ranges of the expression read count for the gene are enormous while some values are smaller than 1, therefore, a log transform was used; y = logbase2(x + 1) to reduce the scale of the expression count of the genes. Because most pre-trained models are Convolutional Neural Network architecture, the input data must be images suitable for the convolutional layer of CNN. The data was reshaped from a 10363x1 array into a 102x102 image by adding zeros at the last line of the image. This motivation to convert the data into 2D images comes from much research work [6], [8]. After this, the images were normalized to make sure the range was between [0, 255]. The images were then converted to a greyscale image with a single channel which was converted to having 3 channels through a channel duplication mechanism- it was achieved using the ImageGenerator (Image\_dataset\_from\_directory) function in Keras to allow the images to fit the pre-trained models [15]. This was done before passing the images through the pre-trained models because most medical imaging comes in black and white [16]. The generated greyscale images are shown in “Figure 2”.

A screenshot of a computer screen

Description automatically generated

***Figure 2: An example of 2D images generated for Class (ACC)0 to Class (UVM) 32 (33 Embedded Image Classes)***

1. ***Transfer Learning:*** Transfer learning (TL) frequently makes use of a model that has already undergone extensive training on a substantial amount of training data for a challenging image classification problem. The information and characteristics of this pre-trained model can act as a useful base, effectively acting as a general model of the visual environment. This strategy can be used to resolve a variety of computer vision issues because collecting medical data can be expensive, particularly for gene expression data with a small number of cases and different characteristics that can affect DL performance [17]. Importantly, this is the reason the TL is used in this study because, in the field of computer vision, the first/initial layer is used for edge detection, the second/middle layer is used for forms/shapes/patterns detection in the images, and the final layers are used to identify characteristics specific to the task [10], [18]. The final layers are retrained primarily through ***"FINE-TUNING"*** whereas the first and second layers are used in the setting of transfer learning. Shorter training periods and improved neural network performance are the conclusions of this study, and they are especially helpful in the lack of sufficient data.
2. ***Explainable AI:*** Researchers are increasingly focusing on enhancing both the interpretability and performance of models to better understand how models make decisions, particularly from a human perspective. However, the absence of data on unusual clinical conditions is a recurring factor in the limitations of new data-driven deep diagnostic techniques. Additionally, traditional deep transfer learning techniques utilized in medical image processing alone are occasionally insufficient for capturing features, resulting in limited interpretability. In addition, a major critique of deep neural networks is that they are difficult to interpret. This is because there is an obscure relationship between the first and last layers (the input and output layer) in a DNN, making it nearly impossible to understand the hidden layer (also known as the "black box") in neural networks. Tracing a data set through a network in a model with 100 neurons to determine how the model works and which aspects are significant is already impractical. This is why in this study, a transfer learning approach plus an XAI tool is employed – Alibi Explain in Python library was used for this purpose – In this, Integrated Gradients was imported. Alibi Explain is a free and open-source Python toolkit designed for analyzing and interpreting machine learning models. The library's main goal is to offer top-notch implementations of local, global, black-box, and white-box explanation methods for regression and classification models while Integrated Gradients allow for the possibility to check the input of a deep learning model on their importance for the output. Integrated gradients generate an attribution value for each feature (in this case, for each pixel and channel in the image) by integrating the model's gradients about the input along a straight path from a baseline x' instance to the input instance x [13].
3. ***Performance Metrics:*** Four metrics were used in this study to assess the model performance: accuracy, precision, recall, and F1-score.

* Accuracy = (TP+TN) / (TP+FP+TN+FN) (1)
* Precision = TP/TP + FP (2)
* Recall = TP/TP + FN (3)
* F1-Score = 2 \* Precision \*Recall/ Precision + Recall (4)

i. The number of positive cases that the model predicts will be positive is known as the True Positive or TP.

ii. The number of false positives (FP), or cases of negative data that the model interpreted for positive data.

iii. The number of cases of negativity that the model correctly classified as negativity (True Negative, TN).

iv. False Negative (FN) indicates the number of cases where a positive outcome was incorrectly predicted as a negative outcome.

1. ***Experimental Setup***

After the final step of the data preprocessing processes, the 33 image class data were randomly split using the split-folder package in Python in the ratio of 80:10:10 (8342: 1030: 1074) for the pre-trained model training, validation, and testing of the model's performance. The classes were represented with numerical values ranging from 0 to 32, where 0 stands for ACC and 32 is the last class UVM. Also, the images were modified into 3-channel images through a channel duplication mechanism allowing the pre-trained model to process the images as RGB channel images. The grayscale intensity is essentially used as the value for all three color channels in the model, which handles all three channels equally by the models [15], [16].

Four models—VGG19, ResNet50, DenseNet169, and Xception—were chosen after experimenting with a variety of base CNN pre-trained model architectures, including EfficientNetB0, MobileNet, and others. VGG19 is a deep architecture CNN network of 19 layers, including 1 SoftMax layer, 5 max-pooling layers, 3 fully connected layers, and 16 convolutional layers.

A variant of densely linked convolutional networks called DenseNet169 was created to maximize accuracy by avoiding the vanishing gradient problem and feature mappings that reduce parameter efficiency. It is like ResNet but takes all previous output as an input for future layers while ResNet uses an additive method, which means it takes a previous output as an input for future layers [19]. It turns a conventional network into a residual network by "skipping over" some levels using the idea of shortcut connections. ResNet50 builds on a bottleneck design of 1x1 convolution, employs a stack of three layers rather than two levels, and decreases the number of parameters and matrix multiplication to speed up the training of each layer [20]. A deep convolutional neural network called Xception, which is an extreme variant of Inception, uses depth-wise Separable convolutional.

This study's motivation for training the final convolutional block alongside the fully linked layer comes from [10], [16], [18] since the models were all trained on ImageNet, which does not have a class label for any medical images, a prediction-dense layer with a SoftMax activation, an Adam optimizer, and a learning rate of 10e-4 was used when fitting the models, this is to at least ensure that the models learn some patterns in the medical data. To enhance generalization performance and avoid overfitting, Dropout, and L2 kernel regularization were applied. For multiclass classification problems utilized in the training of the models, the loss function is categorical cross-entropy, and a training epoch for all the models was 25 except for Xception trained at 15 epochs. A second test dataset that the model had never seen was used to assess the model's performance and generality. The training and validation dataset was used to train the model and validate its performance. The IntegratedGradients were used to demonstrate the explainability of the models built, heatmap visualizations were made, each of which showed the regions on the images that determined the rationale for its forecasts, to show how easily the models could be explained.

1. **EXPERIMENTAL RESULTS & DISCUSSION**

The Accuracy results of the fine-tuning of the pre-trained base model are shown in “Table 2” and “Table 3.”

**TABLE 2: ACCURACY PERFORMANCE FOR THE TRAIN, VALID, AND TEST DATA FOR THE 4 PRE-TRAINED BASE MODEL**

|  |
| --- |
| **Fine-Tuned Classification Model** |
| **Train Accuracy (80%)** | **Valid Accuracy (10%)** | **Test**  **Accuracy**  **(10%)** |
| **Xception** | 99.69% | 91.75% | 90.00% |
| **DenseNet169** | 98.20% | 92.33% | 91.00% |
| **VGG19** | 95.26% | 91.76% | 91.00% |
| **ResNet50** | 1.00% | 94.08% | 92.00% |

**TABLE 3: XCEPTION, DENSENET169, RESNET50, AND VGG19 CLASSIFICATION PERFORMANCE METRICS BASED ON THE 10% TEST DATA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Fine-Tuned Classification Model** | **PERFORMANCE METRICS ON THE 10% TEST DATA** | | | |
| **Precision** | **Recall** | **F1-Score** | **Accuracy** |
| **Xception** | 88% | 86% | 86% | 90% |
| **DenseNet169** | 89% | 86% | 87% | 91% |
| **VGG19** | 87% | 86% | 86% | 91% |
| **ResNet50** | 90% | 88% | 89% | 92% |

The macro-average score for the test data was chosen for precision, recall, and f1-score because the macro-average considered all classes equally contributing to the final averaged metric which is important for the generality of the model irrespective of the weighted size of the classes' contribution.

Inspecting the performance metrics, accuracy was less considered in this study because the data contains an imbalance class distribution. Therefore, taking the performance metrics of the ResNet50 architecture, the precision of 90% explains that when the model predicts a tumor presence, there is a 90% chance that the model is accurate while Recall/Sensitivity of 88% indicates that the model correctly predicted 88% of the actual tumor instances. In other words, the model has an 88% chance of discovering and predicting a tumor when one is present. Figure 3, and “Figure 4” show the ResNet50 model accuracy and model loss for the train (8342 samples in total) and validation test (1030 samples in total) for 25 epochs.

A graph of a graph

Description automatically generated

***Figure 3: The Model Accuracy Curve for ResNet50 Architecture***

***A graph of a model loss

Description automatically generated***

***Figure 4: The Model loss Curve for ResNet50 Architecture***

The generated confusion matrix of ResNet50 on the 10% test data of 1074 (unseen during training) is shown in “Figure 5”. The purpose of the test data is to evaluate the performance of the model. Looking at the confusion matrix shows some misclassification of some of the classes.

1. 7 of class 16 (LIHC) samples were misclassified as class 28 (THCA) while 5 samples of THCA were misclassified as LIHC.
2. 5 of class 19 (OV) samples were misclassified as class 30 (UCEC) while 3 of UCEC samples were misclassified as OV samples.
3. 4 class 1 (BLCA) samples were misclassified as class 23 (READ) while 1 of READ was misclassified as BLCA.

According to paper [8], READ class and LIHC were also misclassified. This misclassification could be due to small samples in the test data since it only contains 10% of the total dataset in this study and the class distribution is not evenly distributed. It is an imbalanced dataset.

A diagram of a graph

Description automatically generated***Figure 5: The Confusion Matrix generated for ResNet50.***

To get the interpretability of the model prediction, IntegratedGradients was imported from Alibi Explainers. An image from class 1 (BLCA) was used to inspect this. The image was preprocessed and transformed using the ResNet50 preprocess input library in the TensorFlow keras library. *“Figure 6”* shows the explainability of the model prediction and why it is predicting the class label “BLCA/class1” in the code. By overlaying the attribution values for each pixel over the original image on the heatmap with the black image baseline, the attributions are shown. The pixel's attribution value is obtained by adding the three-color channels' individual attribution values. Green pixels represent positive attributions, whereas red pixels represent negative attributions. The attributions are scaled in a range and show which genes or portions of the image are responsible for the observed results.

Finally, compared with the paper [8], CNN architecture was built, and it was trained and tested using a 10-fold cross-validation. It gives an average accuracy of 95.59%. Meanwhile, the transfer learning approach used in this paper gave a 92% accuracy for ResNet50, using 10% test data (Unseen dataset set apart from training and validating of the model). This shows that the approach used in this paper has a good performance and good generalizability performance.

A close-up of a screen

Description automatically generated

***Figure 6: Shows the red and green attribution of part of the images causing the observed outcomes/prediction of the class label in ResNet50.***

1. **CONCLUSION**

In this study, based on the Pan-Cancer Atlas provided 33 most prevalent tumor types, a transfer learning methodology to classify genomic data (Gene Expression data) was employed. The experimental result in this study shows that the pre-trained model- VGG19, Xception, DenseNet169, and ResNet50 with a fine-tuning procedure like retraining of some part of the convolutional layer/block to effectively learn some patterns/forms in the gene expression/medical data and the common fully connected layer retraining method of the models is good in classifying gene expression data. ResNet50 performed better in this study achieving an accuracy of 92% on the 10% test set with a precision of 90%, recall of 88%, and F1 score of 89%. This performance on the unseen data (test data) shows that there is a good generalization performance of the model.The proposed methodology in this paper, transfer learning and base models’ architectures fine-tuning can work well to identify tumor types based on their gene expression and will work well with other tumor types not included in this study as well as other genomics data analysis. It is faster and less computationally expensive. Also, the use of IntegratedGradients on the images gives a clearer view of which part of the image is attributing the cause of the prediction of the various classes. It is shown in this work that a good generalization performance is achieved. The accuracy of the methodology used in this paper shows good performance of the pre-trained models on tumor type classification as well as the interpretability of the results given using Explainable AI.

1. **FUTURE WORKS**

In this study, the dataset class distribution is imbalance, Efforts can be made for future studies to balance the class distribution for each class using some machine learning balancing data techniques like Random Over-Sampling, Under-Sampling, or Synthetic Minority Oversampling Technique (SMOTE), and much more can be considered to balance the class distribution which might improve the performance of the DNN pre-trained model classifiers. Also, the training time could be increased and some hyperparameters can be further finetuned to find the most optimum parameters to improve the model performance. The concept of training some of the convolutional block/layers in the pre-trained models can further be investigated into further to know the number of layers that should be retrained to learn the pattern and forms in the medical data as the pre-trained models do not contain any classes pointing to medical data. The black image baseline was used in this study, however, other available baselines like random image baseline can be used for a more efficient view of the explainability of the study. Also, it is proposed that for further study, the gene attribution shown on the heatmap be investigated, to be printed out in their names to allow for more clearer understanding of the genes causing the alterations and leading to a particular tumor. The methodology used in this study can be employed and suited well for medical imaging data analysis using Transfer learning. It also shows that Transfer learning will perform well when used for medical imaging analysis.

1. **ACKNOWLEDGEMENT**

This study relies on the RNA-Seq Expression data sequenced by The Cancer Genome Atlas (TCGA) cohort, a project by the joint effort of the National Cancer Institute (NCI) and the National Human Genome Research Institute and further analyzed the tumor types to 33 from 11,000 leading to the Pan-Cancer Atlas accomplished, seeks to acknowledge the team for making the dataset available. The Deep learning methods- transfer learning and techniques adopted in this research work were also adapted from the classroom learnings and most especially appreciate and acknowledge the Guidance from Supervisors, Dr. Pratheepan Yogarajah and Professor Kevin Curran.

1. **REFERENCE**

[1] Cancer, “World Health Organization,” [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer. [Accessed 19 July 2023].

[2] World Cancer Statistic, Cancer Research UK, [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer. [Accessed 19 July 2023].

[3] Khalsan, M., Machado, L.R., Al-Shamery, E.S., Ajit, S., Anthony, K., Mu, M., and Agyeman, M.O., "A survey of machine learning approaches applied to gene expression analysis for cancer prediction," vol. 10, no. IEEE Access, pp. pp.27522-27534.

[4] Alharbi, F., Elbashir, M.K., Mohammed, M. and Mustafa, M.E. "Fine-Tuning Pre-Trained Convolutional Neural Networks for Women Common Cancer Classification using RNA-Seq Gene Expression," International Journal of Advanced Computer Science and Applications, vol. 11, p. 11, 2020.

[5] Wang, Z., Gerstein, M. and Snyder, M. "RNA-Seq: a revolutionary tool for transcriptomics.," Nature reviews genetics, vol. 10, no. 1, pp. pp.57-63, 2009.

[6] Alharbi, F. and Vakanski, A., "Machine learning methods for cancer classification using gene expression data: A review.," Bioengineering, vol. 10, no. 2, p. p.173, 2023.

[7] 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, p. pp. 1358–1369, 2018.

[8] Boyu Lyu and Anamul Haque, "Deep Learning Based Tumor Type Classification Using Gene Expression Data.," in ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, New York, USA, 2018.

[9] Rukhsar, L., Bangyal, W.H., Ali Khan, M.S., Ag Ibrahim, A.A., Nisar, K. and Rawat, D.B., 2022. Analyzing RNA-seq gene expression data using deep learning approaches for cancer classification. *Applied Sciences*, *12*(4), p.1850.

[10] L. A. Chamli Deshan, M. K. Hans Thisanke and D. Herath, "Transfer Learning for Accurate and Efficient Tomato Plant Disease Classification Using Leaf Images," 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS), Kandy, Sri Lanka, 2021, pp. 168-173.

[11] Meena, J. and Hasija, Y., "Application of explainable artificial intelligence in the identification of Squamous Cell Carcinoma biomarkers," Computers in Biology and Medicine, vol. 146, pp.105505, 2022.

[12] Welcome to the Pan-Cancer Atlas, "Pan-Cancer Atlas," [Online]. Available: https://www.cell.com/consortium/pancanceratlas. [Accessed 21 August 2023].

[13] Explainable AI: Integrated Gradients|, "Data Basecamp," 17 December 2021. [Online]. Available: https://www.databasecamp.de/en/ml/integrated-gradients-nlp/. [Accessed 29 August 2023].

[14] A. Gupta, "Analytics Vidhya," 26 April 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/10/feature-selection-techniques-in-machine-learning/. [Accessed 20 July 2023].

[15] Keras “Image Loading” [Online]. https://keras.io/api/data\_loading/image/ [Accessed on 15 September 2023].

[16] Towards Data Science “Transfer Learning on Greyscale images: How to Fine-Tune Pretrained Models on Black-and-White Datasets”. 24 February 2022 [Online]. https://towardsdatascience.com/transfer-learning-on-greyscale-images-how-to-fine-tune-pretrained-models-on-black-and-white/ [Accessed 15 September 2023].

[17] Zhang, X., Chen, Z., Gao, J., Huang, W., Li, P. and Zhang, J., "A two-stage deep transfer learning model and its application for medical image processing in Traditional Chinese Medicine.," Knowledge-Based Systems, vol. 239, p. p.108060., 2022.

[18] Siddiqi, R., "Effectiveness of transfer learning and fine-tuning in automated fruit image classification.," In the proceeding of the 3rd International Conference on Deep Learning Technologies (pp. 91-100)., Pakistan, July 2019.

[19] Shivam Baldha, "Introduction to dense nets (Dense CNN)," Analytics Vidhya, 7 April 2022. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/03/introduction-to-densenets-dense-cnn/. [Accessed 28 August 2023].

[20] " ResNet 50- The basic and a quick tutorial Datagen," Datagen, [Online]. Available: https://datagen.tech/guides/computer-vision/resnet-50/. [Accessed 29 August 2023].