

---

# Cancer Types Classification using Deep Convolutional Neural Network with Transfer Learning Model

---

Burçin Bozkurt Günay, bgunay21@ku.edu.tr  
Ece Alptekin, ealptekin21@ku.edu.tr

Due date: January 26, 2023 – 23:59 PM

## Abstract

Cancer is one of the major health concerns for people from all around the world. In each year, cancer causes millions of deaths. Besides, the cost for the treatment is very high. Therefore, early detection of cancer, in terms of cost and accuracy, plays a critical role. The motivation for the project is to reduce the workload of the pathologists to perform manual detection by providing a novel approach on the prediction of different cancer types using whole slide images obtained from the Cancer Genome Atlas (TCGA) dataset. We proposed a stacking ensemble of a target deep learning model on top of a pre-trained DenseNet201 architecture with transfer learning. Our model tends to provide two different outputs including a prediction for tissue classification of SKCM, LUSC and COAD cancer types and the prediction of the age of the patient owning the TCGA tissue sample. The performance of the proposed model is compared with other CNN architectures built on top of the pre-trained VGG16 and ResNet152V2 architectures. After a wide hyperparameter search, it is concluded that DenseNet201 performed the best with having overall accuracy 84.60%.

## 1. Introduction

Diagnosis of the cancer types and categories pose new computational challenges for developing analytical solutions. The improvements in Deep Learning methodologies and proposals for the pre-trained Convolutional Networks leverage the fast and accurate detection of the tissues having tumour. Computational tools have become more important than ever in assisting oncologists in cancer diagnosis and prognosis, ensuring that patients receive the identification results in advance. So, the purpose of this work is to classify tumor samples based on the tissue samples taken for pathological surgery and diseases. Apart from the classification, our proposed model predicts the initial pathological

diagnosis ages of patients. To achieve our goal, we first convert tumor samples into tiles. Then, the tiles are fed into the model. Here, transfer learning is performed with a base model and a target model. DenseNet201 [16] CNN architecture that was pre-trained on approximately 1.28 million images (1,000 object categories) from the 2014 ImageNet Challenge has been utilized and a target model has been proposed. Our dataset is trained on the target model using transfer learning. DenseNet201 model is used to provide three-class classification for tumour type and patient's age regression output layers separately. The VGG16 [17] and ResNet152V2 [27] are solely used to provide three-class classification. The target CNN models are trained using 4000 tumour pathological image belonging to three classes (skin cancer, lung cancer and colon cancer).

## 2. Related Work

The classification of cancer tumours is a major research area for deep learning. We examine previous works which make significant contributions to this field, to give us insights. In work by Mohammed, M., they consider the classification of the most common cancer types among women. The RNAseq gene expression of women is used as the input to their ensemble deep learning model and 1D-CNN model.

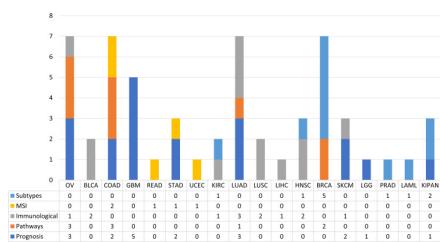
The ensemble deep learning model uses five 1D-CNN models as the base model, and then, the final prediction is made by a meta-model. Hence, the meta-model is able to generalize the predictions. By generalizing, the meta-model performs better than the base model. In their experiments, the ensemble deep learning model outperforms other models by classifying all positive cancer cases correctly.

They argue that the classification based on form and structure has limitations. Hence, a multi-class classification based on RNASeq gene expression is preferred to achieve more accuracy in distinguishing cancer types, but its performance is decreased due to small sample sizes and the huge number of uninformative genes. To overcome this, they perform feature selection to discriminate informative genes and uninformative genes before working on the model.

In data preprocessing, normalization, transformation and filtration among genes are applied to tumour samples. Then, in the phase of regularization and feature selection, Lasso regression provides the elimination of uninformative genes and decreases the dimensionality of the data sets. To prevent over-fitting, they have to reduce the number of parameters in their model or increase the number of input samples. Since the images have high memory usage, a solution for the number of parameters is required. To achieve this, an early stopping mechanism is proposed to avoid overfitting and underfitting.

In the work by Hanczar, B., Multilayer perceptron is selected for the classification of gene expression data. Then, supervised and unsupervised transfer learning is performed due to its ability to reduce the dimension of the data set. In the experiment, the prediction accuracy of target model is considered. According to the results, the accuracy depends on the type of cancer. In addition, using less data increases the performance of transfer learning. Their work shows that transfer learning is highly affected by the training data set size. In their results, they observe that for the RNA-seq data set of The Cancer Genome Atlas (TCGA) [7], batch normalization improves the performance in the prediction of cancer type. In addition, in the optimization process, SGD gives the best result compared with the RMSprop and ADAM.

An interesting survey is conducted by Liñares-Blanco J., where studies including ML techniques have been explored for the analysis of different types of cancer using TCGA data. The summary of studies is given according to most frequent cancer types.

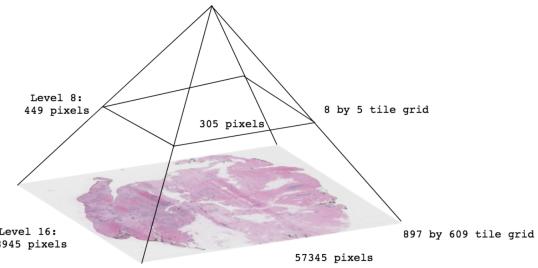


As mentioned in the survey, the era of individual analysis has passed, and the era of data integration studies at the clinical-genomic level and medical patient data is emerging. For regression, patient clinical data, including the age, gender, region of birth, etc. of the patients, could be used for diagnosis or the prognosis of the patient by means of the Karnofsky Performance Status Scale [15]. Death of the patient, regional differences among the cancer types, and age and gender analysis could serve as further research topics.

### 3. Data Preparation

All data sets are collected from tumour images available through the Cancer Genome Atlas (TCGA) data portal (<https://tcga-data.nci.nih.gov/tcga>) with .SVS extension. An SVS file is a Tiled TIFF image that consists of additional pages including the overview image, slide label, and a few smaller, scaled copies of the scanned slide.

A tumour image stores the information in tiles in a pyramid shape. The dimensions differ in each level of the picture. For example, at level 0, the image has the maximum resolution and dimensions. The structure of an image is shown as an illustration.

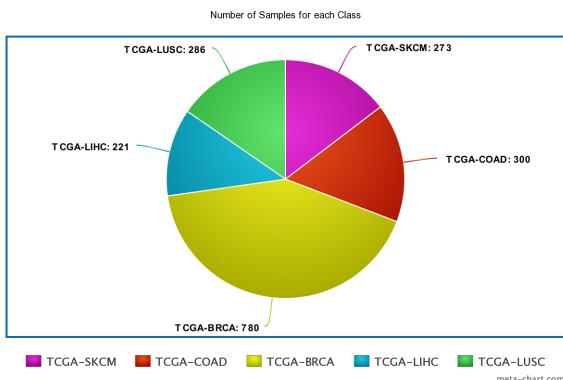


GDC Application Programming Interface (API) [18] is used to download the categories of tumour tissue types. The GDC API also provides metadata and pathology report of each image in the upper side of the web page. Through the data, the information of the patient, the condition of the tumor and the diagnosis of the doctor is accessible. The healthy tissue slides are also available in the TCGA portal. In the portal, clinical manifestations are categorized into data collections based on the human body parts. Hence, each data collection contains tumour images related to the body categories.

In the proposal, TCGA-BRCA(Breast), TCGA-LUAD(Bronchus and lung), TCGA-LUSC(Bronchus and lung), TCGA-COAD(Colon Adenocarcinoma), TCGA-SKCM(Skin), TCGA-STAD(Stomach), TCGA-TGCT(Testis), TCGA-KIRP(Kidney), TCGA-UVM(Eye and adnexa), TCGA-LIHC(Liver and intrahepatic bile ducts) categories had been selected to be used in the proposed architecture. But, further analysis of the data has shown that .SVS files are more than one GB each, and the cost of processing and storage requirements of each file exceeded our limitations. Furthermore, the age information belonging to the patient may not be available in the patient's records.

Due to these concerns, tumour samples of TCGA-COAD(Colon Adenocarcinoma), TCGA-LUSC(Bronchus and lung) and TCGA-SKCM(Skin) are taken into consideration. The data set consisted of 967 tumour sample in total. Each patient is assumed to have one sample from any of TCGA-COAD, TCGA-LUSC and TCGA-SKCM

classes. The distributions of samples can be visualized in the following chart.



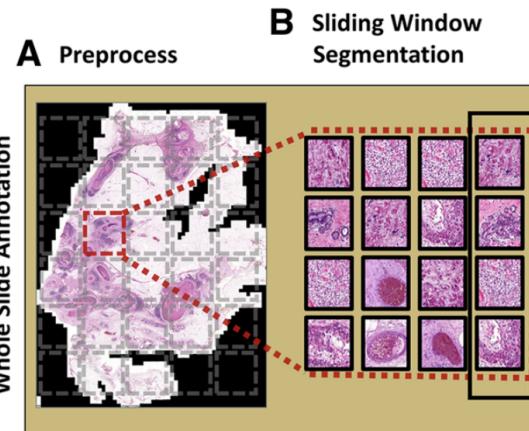
### 3.1. Labeling Data

The data labeling pipeline involves four main steps:

1. Download of SVS image file and patient metadata from TCGA portal via GDC API
2. Transformation of a SVS image into TIFF files using OpenSlide library
3. Splitting of 70 tiles from TIFF file each having size around 200 KB on average with 224x224x3 dimension
4. Assessment of each tile to check if it meets the minimum tissue coverage threshold to be considered as foreground

In our work, we use a source code [4] that reads slide images with a C library which is called OpenSlide. In transformation of a SVS image into TIFF files, it provides the general information about the slide such as dimensions and properties. To read individual tiles, DeepZoomGenerator is applied on a slide image with a tile size. It returns the amount of levels and level dimensions in a tile. After that, saving each individual tile in tiff format is possible.

Here, each tile file dimension is 224x224x3 since the pre-defined DenseNet201 and VGG16 models are compatible with 224x224x3 input shape. Moreover, tiles of an image is obtained by a specific level of the image. Level 12 is chosen in our work, since these dimensions are suitable for extracting 70 tiles with dimensions 224x224x3. The figure demonstrates the visualizations of obtaining tiles from an image.



After the 3rd step in the pipeline, having 70 tiles from each sample, an assessment has been performed on the tiles, and tiles with a majority of blank backgrounds have been eliminated from the tile set. The following formulation is applied on each tile:

---

```

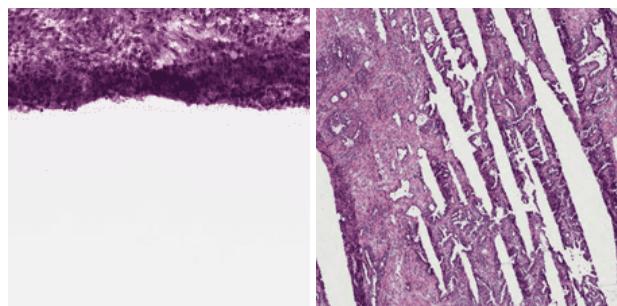
mean ← MEAN OF IMAGE
std ← STANDARD DEVIATION OF IMAGE
if mean < 230 and std > 15 then
    | image ← KEPT
else
    | image ← REMOVED
end

```

---

For the assessment, the standard deviation and mean of the tiles, after converting them into arrays. When these properties of the tiles are examined, it is observed that the images with a standard deviation of more than 200 and a mean of less than 15 are blank images. Therefore, they are removed from the data set. Here, as an illustration, a blank image and a valid tile image are attached.

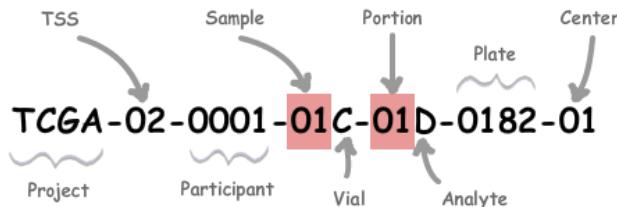
These tiles are then assessed to see if they meet a minimum foreground (tissue) threshold concerning the total area of the tile, in which case they are kept (e) and otherwise discarded.



In support of the aforementioned reasons for the reduction of class types in the model, we have to reduce the number of categories since from some classes, such as TCGA-BRCA,

TCGA-LIHC and TCGA-LUAD, we cannot obtain enough tiles after looking at their standard deviation and mean values.

A tumour sample with SVS extension is stored with a unique title. The following figure presents the basic collection of identifiers to generate a label with the code entries.

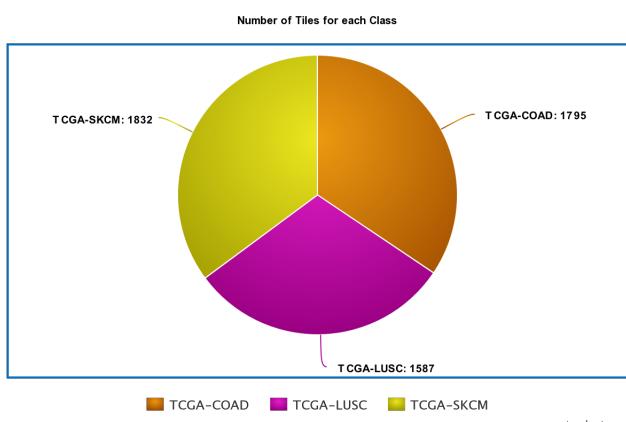


Tiles are also stored with their barcodes generated in compliance with the conventions. Also, each tile image title is parsed to gather the project name, sample type and participant information. The tiles are uniquely identified with the first 12 characters of their barcodes. In the last part, These unique characters are kept to retrieve patient records and obtain images by adding them to an URL.

### 3.2. Data Set Preparation

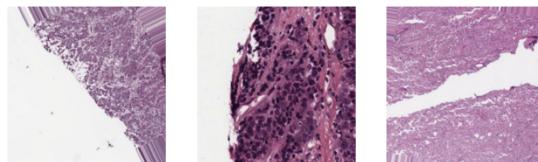
Before constructing the data set, we traverse the tiff files to differentiate them from the sample types. By looking at the sample of the tiff files' barcodes, we separate them into three categories: Tumor types, normal types, and control samples. Then, we only consider the files with tumour types. The patient IDs of these files are received. Thus, ten tiles for each patient are set aside to be added to the data set. In the figure, the number of tiles for each class is given.

The data set is balanced in terms of classes, each having a similar number of tiles.

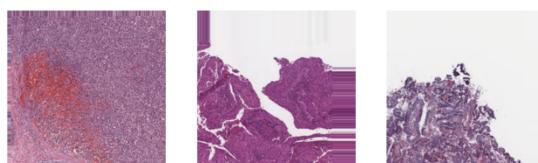


### 3.3. Data Augmentation

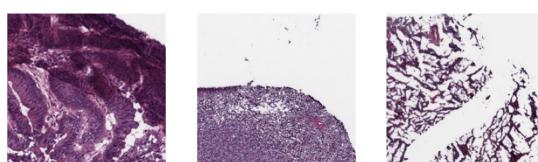
Data augmentation techniques have been applied to generate new tile image samples to improve the model's accuracy. We aim to diversify the data set as much as possible by using various transformations. First, random rotations with a degree range of 30 are performed.



Second, shift transformations with range 0.2 are applied for random horizontal shifts and with range 0.2 for random vertical shifts.



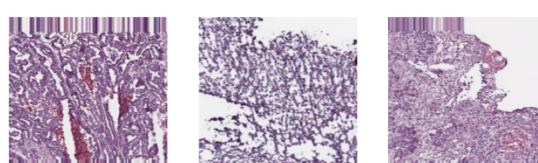
Transformation, including randomly flipping inputs horizontally and vertically, is performed.



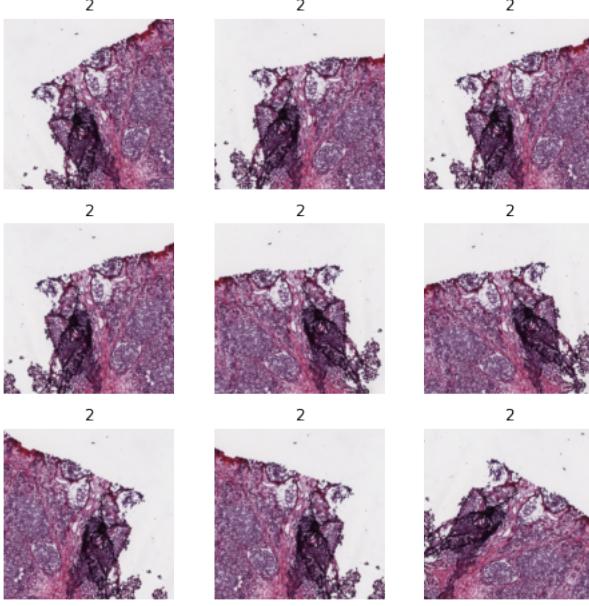
The brightness of images is changed as transformation.



As the last step, zoom transformation is operated on images.



Before giving the data set to the model, different transformations are operated on the data set to see their effect of them on the model performance. We combine rotation and flip operations to apply them to the data set sequentially. The results are given in the following Figure.



### 3.4. Data Splitting

In generating the data set, 200 patients are chosen from each class. These patients are identified by their unique barcode characters. For each unique patient id, 10 tiles are obtained to put them into the data set.

Then, the data set is split into train, validation and test sets in a 2:1:1 ratio. To ensure representative splits, all of the tiles belonging to a patient are located uniquely in one of the splits. In other words, tiles with the same first 12 barcode characters, which include patient information, cannot be elements of different sets.

### 3.5. Age Information

Apart from the classification, the model predicts the initial pathologic diagnosis age of the patients. The correlation between the patients and files can be obtained via GDC API to access age information through patient ids. As it is stated before, each image is stored with a barcode. The first 12 characters of these barcodes are unique for each patient. Therefore, these special characters are derived from their barcodes to find their correlated file ids. Then, these file ids are added to an URL to get the age information of a patient. Here, a list of patients with their file ids' is shown as an example. In the figure, the entity submitter id represents the first 12 characters of a barcode. Associated file ids are located under the entity submitter ids.

```
{
  "data": {
    "hits": [
      {
        "id": "3601d2d7-34a6-4bc3-920c-a6733a415e0c",
        "associated_entities": [
          {
            "entity_submitter_id": "TCGA-37-4133"
          }
        ],
        "file_id": "3601d2d7-34a6-4bc3-920c-a6733a415e0c"
      },
      {
        "id": "5d290588-9af6-42fc-b747-6120e6ce5bd4",
        "associated_entities": [
          {
            "entity_submitter_id": "TCGA-43-6647"
          }
        ],
        "file_id": "5d290588-9af6-42fc-b747-6120e6ce5bd4"
      }
    ]
  }
}
```

Using the file id, it is possible to obtain patient records in an XML file.

```

procurement_status="Completed" restricted="false"
source_system_identifier="1448327">NO</clin_shared:tissue_prospective_collection_indicator
<clin_shared:tissue_retrospective_collection_indicator
preferred_name="retrospective_collection" display_order="3"
cde="3088528" cde_ver="1.000" xsd_ver="2.3" tier="2" owner="TSS"
procurement_status="Completed" restricted="false"
source_system_identifier="1448328">YES</clin_shared:tissue_retrospective_collection_indicator
<clin_shared:days_to_initial_pathologic_diagnosis precision="day"
xsd_ver="1.12" tier="1" cde="3131740" owner="TSS"
procurement_status="Completed"
preferred_name="initial_pathologic_dx_days_to"
display_order="9999"
cde_ver="1
.000"></clin_shared:days_to_initial_pathologic_diagnosis>
<clin_shared:age_at_initial_pathologic_diagnosis precision="day"
xsd_ver="1.12" tier="1" cde="2086657" owner="TSS"
procurement_status="Completed">62</clin_shared:age_at_initial_pathologic_diagnosis>
<clin_shared:year_of_initial_pathologic_diagnosis
preferred_name="initial_pathologic_dx_year" display_order="23"
cde="2896960" cde_ver="1.000" xsd_ver="1.12" tier="1" owner="TSS"
procurement_status="Completed" restricted="false"
source_system_identifier="1448296">1996</clin_shared:year_of_initial_pathologic_diagnosis>
...</clin_shared:patient>

```

In our data set, we use the initial pathologic diagnosis age of the patients who are in the projects of TCGA-COAD(Colon Adenocarcinoma), TCGA-LUSC(Bronchus and lung) and TCGA-SKCM(Skin). In the chart, the counts for age ranges in the train data set are illustrated for each class.

Age Counts of Classes

Class	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100
TCGA-COAD	0	5	15	23	34	40	29	4
TCGA-LUSC	0	0	4	20	52	48	7	0
TCGA-SKCM	4	9	27	31	34	41	9	1

## 4. Methods and Experiments

The primary focus of our work was to train a three-class (COAD, LUSC and SKCM) classifier to predict the cancer type of each sample having a tumour on the pathological image and also predict the age of the patient owning the pathological image based on a pre-trained ConvNet using transfer learning [20] with Keras Applications implementation [21].

The transfer layer pipeline consists of six main activities depicted below.

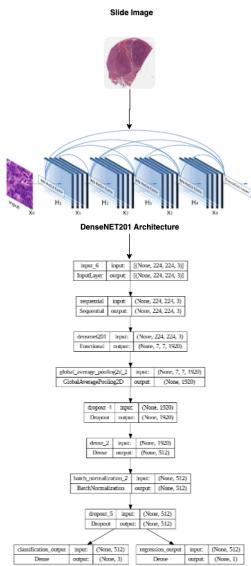


#### 4.1. DenseNet201

Among other ConvNet architectures, DenseNet201 is chosen as the pre-trained network. DenseNet201 is the latest version of DenseNet network with 201 hidden layers that alleviates the vanishing-gradient problem, encourage feature reuse, and substantially reduce the number of parameters. These improvements prove that DenseNet201 can be used for fast and accurate classification of digital pathology images [18]. As DenseNet has been pre-trained on ImageNet, its standard output is for 1000 classes. However, since three classes classification and age regression results are requested, a multi output layer model is designed to diverge in the last layer into two separate layers each satisfying the 3-class classification and 1 output regression layers.

After freezing the top layer, a new model, given in the following figure, is designed and implemented. The new model consists of a Global Average Pooling layer, a Dropout layer, a Dense layer followed by a ReLU activation, a Batch Normalization layer followed by a Dropout layer and in the last part, a Dense Layer for 3 classes with Softmax activation followed by a ReLU activation and in parallel, a dense layer for 1 class Layer with Linear activation.

Two separate output layers are generated for 3-class classification and age regression outputs. The target model should work with 224x224x3 images. Global Average Pooling layer averages the spatial dimensions of a matrix of any size that can take input image of any size.



The loss functions that is used to measure the error from our predicted outputs are mean square error (MSE) loss and Sparse Categorical Cross Entropy loss.

For a training batch of size N, accuracy and mean absolute error (MAE) metrics are used to measure the accuracy

rates of Variations on the model structure were tested including network depth, filter size, epoch size, batch size and inclusion of regularisation.

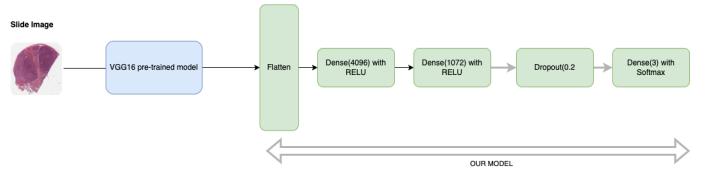
Experiments are conducted initially with the following parameters: batch size 32, 10 epochs, learning rate: 0.00001

#### 4.2. VGG16

Note that VGG16 model is designed solely for the support of multi-class classification output.

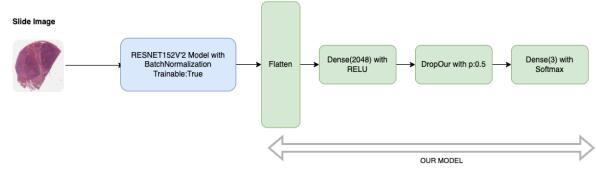
After freezing the top layer, a new model, given in the following figure, is implemented based on VGG16. The new model consists of a Flatten layer, followed by a series of three Dense(4096), Dense(1072) layers with Relu activations and a Dropout(0.2) layer with a final output, a Dense Layer, of size 3 with Softmax activation set.

Loss functions used included Sparse Categorical Cross Entropy.



#### 4.3. ResNet152V2

Note that ResNet152V2 model is designed solely for the support of multi-class classification output. It is designed to mitigate the vanishing gradients problem. The target model is given in the following depict. The new model consists of a Flatten layer, followed by a Dense layer (2048) with Relu activation and a Dropout Layer and finally a Dense Layer, of size 3 with Softmax activation set.



#### 4.4. Experiments

We conducted our experiments on Google colab [22] which provides collaborative coding environment where we can use GPU and high RAM as the runtime environment. Downloading SVS files takes more than 24 hours, each having minimum 1GB size and also increases the necessity of extra hardware storage by more than 350 GB.

We also trained all our models for minimum of 20 epochs (complete passes through the training data) and batch size 32. The loss function we used to measure the error from our

predicted outputs was mean square error (MSE) loss.

Classification accuracy (Acc) and mean absolute error (MAE) are metrics for our model that are used to measure the performances for both two output layers.

Metrics module from *sklearn* library [23] is used to generate a scoring matrix and classification report for target model the DenseNet pre-trained model. Classification Report includes Precision, Recall, F1-score and Support aspects. These aspects are then used to compare performances of pre-trained models.

The source code is available in git repo [24].

## 5. Performance of the Models

### 5.1. Initial Model Testing

Initial training was performed on the entire training dataset using the model built on top of DenseNet201 and VGG16 pre-defined models. DenseNet201 model includes 2 multi output layer whereas VGG16 and ResNet152V2 model includes single output layer for classification.

Learning rates were set in the range of  $1e - 2$  to  $1e - 4$ . Batch size is set to 32 and epoch number is taken as 20.

The Adam optimiser was used with its default settings in VGG16 and DenseNet201. Adamax optimiser is used in ResNet152V2.

### 5.2. Fine Tuning

Fine tuning has been done on the following items:

- Experiment with different learning rates, batch sizes and epochs
- Reduction of the number of neurons in the FC layers
- Increasing/decreasing  $p$  ratio in Dropout Layers
- Including Regularization to Batch Normalization

The learning rate were in the range of  $1e - 2$  to  $1e - 5$  to reduce the extent of over-fitting.

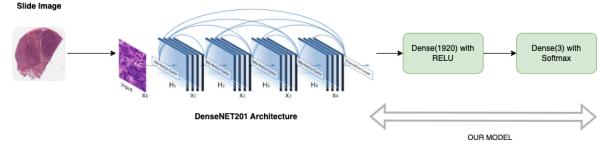
The  $p$  ratio was in the range of 0.2 to 0.5

After many attempts to utilize our model, the regression output layer was removed from proposed model built on top of DenseNet201. The reason of the removal is, MultiOutput layer model could not generalise well to the training set, scoring lower than 40% accuracy.

The updated DenseNet201 architecture initializes pre-trained weights from IMAGENET1K\_V1. It is also implemented in PyTorch. The new architecture includes a

FC(1920) layer with RELU followed by a Softmax classifier layer. It also uses Binary Cross Entropy as the loss function.

The updated model is given in the depict. Instead of Sparse Category Entropy, Binary Cross Entropy is used.

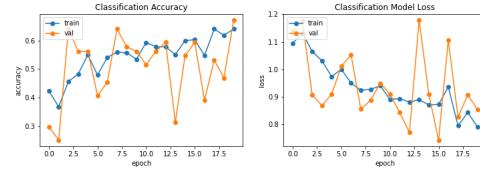


## 6. Results and Comparisons

### 6.1. VGG16 Evaluations

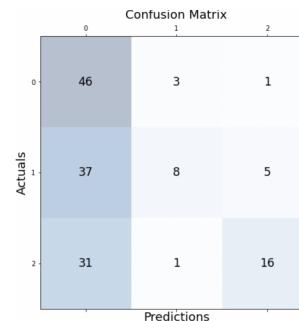
VGG16 model tends to perform lower performance in both training, validation and test.

The loss and accuracy distributions are given in the below figure.



The test accuracy and loss values are also given with the confusion matrix below. Class Labels are given as follows : 0: TCGA-COAD, 1:TCGA-LUSC, 2:TCGA-SKCM

Test Loss	1.0327%
Test Accuracy	0.4729%



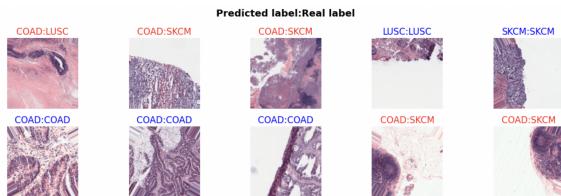
And, finally the classification report including the F1 scores, etc. are given below.

## Final Report

```
F1 Score: 0.42539432767599566
Precision Score: 0.5991493886230729
Recall Score: 0.4711111111111111
precision    recall   f1-score   support
0            0.40     0.92     0.56      50
1            0.67     0.16     0.26      50
2            0.73     0.33     0.46      48

accuracy          0.60
macro avg       0.60     0.47     0.43     148
weighted avg    0.60     0.47     0.42     148
```

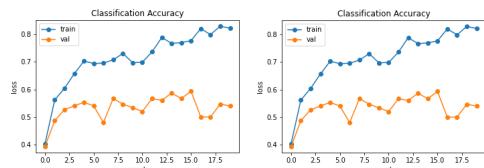
We can randomly select from the test samples and observe the predictions in the model.



### 6.2. ResNet152V2 Evaluations

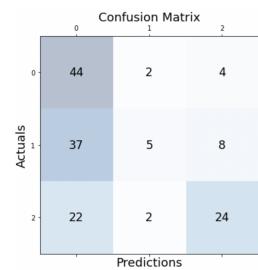
ResNet152V2 model tends to perform better accuracy in both training, validation and test performance.

The loss and accuracy distributions are given in the below figure.



The test accuracy and loss values are also given with the confusion matrix below. Class Labels are given as follows : 0: TCGA-COAD, 1:TCGA-LUSC, 2:TCGA-SKCM

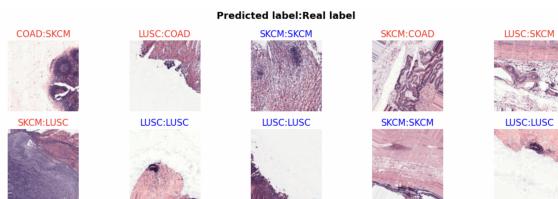
Test Loss	84.60%
Test Accuracy	0.493243%



And, finally the classification report including the F1 scores, etc. are given below.

	precision	recall	f1-score	support
0	0.85	0.86	0.86	179
1	0.67	0.60	0.64	156
2	0.82	0.73	0.77	180
micro avg	0.79	0.74	0.76	515
macro avg	0.78	0.73	0.75	515
weighted avg	0.79	0.74	0.76	515
samples avg	0.74	0.74	0.74	515

We can randomly select from the test samples and observe the predictions in the model.



It is surprising that some samples that has large blank backgrounds are classified correctly whereas the some with full tumour tissue are misclassified.

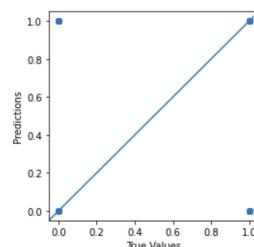
### 6.3. DenseNet201 Evaluations

The DenseNet201 model generalises reasonably well to the training data set after many fine-tuning attempts, scoring 84.60% overall accuracy. Learning rate is set to 0.00001 with 10 epoch and 32 batch size.

The total accuracy results are given in the following table:

Average Test Accuracy	84.60%
Classwise Accuracy	
TCGA-SKCM	89.90%
TCGA-LUSC	79.03%
TCGA-COAD	84.85%

The difference in the distribution of classes is given in the below depict.



The confusion matrix is given below:

TCGA-SKCM	Positive	Negative
Positive	309	27
Negative	25	154
TCGA-LUSC	Positive	Negative
Positive	313	46
Negative	62	94
TGCA-COAD	Positive	Negative
Positive	306	29
Negative	49	131

#### 6.4. Performance Analysis

Nevertheless, across all models, test accuracy percentage is lower than validation and training accuracy. This is most likely due to the models finding unaccounted tumour portions to be suspicious during test, a consequence of training with limited datasets.

DenseNet201 performs better results than VGG16 and ResNet152V2. VGG16 and ResNet152V2 are not enough for extracting features from these tissue images to reach into a satisfied level of classification accuracy. On the contrary, DenseNet201 utilises dense connections between layers, where thorough Dense Blocks, we can generate feature-maps and use them to support classification. Similar to the work done in our proposal, Karaddi, S. H. performed a multi-class disease classification using DenseNet and ResNet [25]. He observed that DenseNet-201, achieved the highest accuracy of 97.2%. Our accuracy percentage is about 84.60%.

Also, in his work, Nillmani proposed multi-classification method for Pneumonia Classification using seven types of ConvNet models like DenseNet, ResNet, VGG16. The best results were using DenseNet201, VGG16, and VGG16, respectively having accuracies of 99.84%, 96.7%, 92.67%. The results are really impressive when compared with our work. VGGs performance is also close to DenseNet in this work [26].

So, performance of the models is directly effected by the input data images. Usage of the features in all these networks may effect the performance.

### 7. Conclusion

In this work, we explore transfer learning based deep learning method implementations to perform multi-class classification for 3 cancer types. The input to our proposed models are tumour images obtained from TCGA Database.

Tumour image slides present a difficult classification task. Each image is in SVS format with minimum 1 GB. the processing of the SVS files took much more than expected.

As expected, DenseNet network performed better results than VGG16 and ResNet. DenseNet201 has an architecture that utilises dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other.

### References

- [1] <https://ysbecca.github.io/programming/2018/05/22/pywsi.html>
- [2] Klimov, S., Miligy, I.M., Gertych, A. et al. A whole slide image-based machine learning approach to predict ductal carcinoma in situ (DCIS) recurrence risk. *Breast Cancer Res* 21, 83 (2019). <https://doi.org/10.1186/s13058-019-1165-5>
- [3] [https://docs.gdc.cancer.gov/Encyclopedia/pages/TCGA\\_Barcodes.html](https://docs.gdc.cancer.gov/Encyclopedia/pages/TCGA_Barcodes.html)
- [4] [https://github.com/bnsreenu/python\\_for\\_microscopists](https://github.com/bnsreenu/python_for_microscopists)
- [5] David A. Gutman, Mohammed Khalilia, Sanghoon Lee, Michael Nalisnik, Zach Mullen, Jonathan Beezley, Deepak R. Chittajallu, David Manthey, Lee A.D. Cooper; The Digital Slide Archive: A Software Platform for Management, Integration, and Analysis of Histology for Cancer Research. *Cancer Res* 1 November 2017; 77 (21): e75–e78. <https://doi.org/10.1158/0008-5472.CAN-17-0629>
- [6] Mohammed, M., Mwambi, H., Mboya, I.B. et al. A stacking ensemble deep learning approach to cancer type classification based on TCGA data. *Sci Rep* 11, 15626 (2021). <https://doi.org/10.1038/s41598-021-95128-x>
- [7] Hanczar, B., Bourgeais, V. Zehraoui, F. Assessment of deep learning and transfer learning for cancer prediction based on gene expression data. *BMC Bioinformatics* 23, 262 (2022). <https://doi.org/10.1186/s12859-022-04807-7>
- [8] Moataz, Laila Salama, Gouda Elazeem, Mohamed. (2021). Skin Cancer Diseases Classification using Deep Convolutional Neural Network with Transfer Learning Model. *Journal of Physics: Conference Series.* 2128. 012013. 10.1088/1742-6596/2128/1/012013
- [9] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Pineros M, et al. Global Cancer Observatory: Cancer Today. Lyon: International Agency for Research on Cancer; 2020 <https://gco.iarc.fr/today, accessed February 2021>
- [10] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248–255)
- [11] <https://cancer.digitalslidearchive.org/>, date of access: 25 Nov 2022
- [12] <https://keras.io/>
- [13] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z (2016). Rethinking the inception architecture for computer vision. In: IEEE conference on computer vision and pattern recognition, pp. 2818–2826
- [14] Chollet, F (2017). Xception: deep learning with depth-

- wise separable convolutions. In: IEEE conference on computer vision and pattern recognition, pp. 1251–1258
- [15] Liñares-Blanco J, Pazos A, Fernandez-Lozano C. Machine learning analysis of TCGA cancer data. PeerJ Comput Sci. 2021 Jul 12;7:e584. doi: 10.7717/peerj-cs.584. PMID: 34322589; PMCID: PMC8293929.
- [16] Huang, G., Liu, Z., van der Maaten, L., Weinberger K., “Densely connected convolutional networks”, Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2261–2269, (2017).
- [17] Simonyan, K., Zisserman, A., “Very deep convolutional networks for large-scale image recognition”, Proceeding of the 2015 International Conference on Learning Representations, (2015).
- [18] <https://gdc.cancer.gov/developers/gdc-application-programming-interface-api>
- [19] Talo, M. Automated classification of histopathology images using transfer learning. Artif. Intell. Med. 101, 101743 (2019).
- [20] Pan, S. J. Yang, Q. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359 (2010).
- [21] <https://keras.io/api/applications/densenet/>
- [22] <https://colab.research.google.com/>
- [23] [https://scikit-learn.org/stable/modules/model\\_evaluation.html](https://scikit-learn.org/stable/modules/model_evaluation.html)
- [24] [https://github.com/burcinbozkurtgunay/my\\_treasure/tree/main/project](https://github.com/burcinbozkurtgunay/my_treasure/tree/main/project)
- [25] Karaddi, S. H., Sharma, L. D. (2023). Automated multi-class classification of lung diseases from CXRimages using pre-trained convolutional neural networks. Expert Systems with Applications, 211, 118650. <https://doi.org/10.1016/j.eswa.2022.118650>
- [26] Nillmani, Jain PK, Sharma N, Kalra MK, Viskovic K, Saba L, Suri JS. Four Types of Multiclass Frameworks for Pneumonia Classification and Its Validation in X-ray Scans Using Seven Types of Deep Learning Artificial Intelligence Models. Diagnostics (Basel). 2022 Mar 7;12(3):652. doi: 10.3390/diagnostics12030652. PMID: 35328205; PMCID: PMC8946935.
- [27] He, K., Zhang, X., Ren, S., Sun, J. (2015). Deep Residual Learning for Image Recognition. ArXiv. <https://doi.org/10.48550/arXiv.1512.03385>