EasyChair Preprint

№ 7392

Transfer Learning Based Classification of MSI and MSS Gastrointestinal Cancer

Zabiha Khan and R Loganathan

January 30, 2022

Transfer Learning Based Classification of MSI and MSS Gastrointestinal Cancer

Zabiha Khan[a]
[a]*Department of Computer Science Engineering, Visvesvaraya Technological University* Belagavi, India
zabi007@gmail.com

Dr. Loganathan R[b]
[b]*Department of Computer Science Engineering, HKBK College of Engineering* Bangalore, India
logu73@yahoo.com

***Abstract*—Gastrointestinal and Colorectal cancers are treated with chemotherapy and its other forms which are not able to provide higher survival rates [1]. Immunotherapy is increasingly becoming popular due to its promising response especially to mutated tumors such as MicroSatellite Instability (MSI) cancers with deficient DNA Mismatch-Repair system (dMMR). Generally, 85% of all the cases related to gastrointestinal and colorectal cancers have proficient DNA Mismatch-Repair system (pMMR) which are also called MicroSatellite Stability (MSS). Only about 15% of the gastrointestinal and colorectal cancer patients have deficient DNA Mismatch-Repair system (dMMR) causing MicroSatellite Instability (MSI) in their tumors. While Immunotherapy responds well to patients with MSI tumors, it is resistant to MSS tumors [2]. Hence, it's important to classify MSI vs. MSS tumors so that appropriate treatment can be given to the patients. Clinically MSI cancers are difficult to be detected after stage III due to their sensitivity to pembrolizumab inhibitors [3][4]. In this work, deep learning based transfer learning approach is detailed that can accurately classify MSI vs. MSS cancers using histological images which are derived from formalin-fixed paraffin-embedded (FFPE).**

***Keywords*: MicroSatellite Instability (MSI), MicroSatellite Stability (MSS), Deep learning, Deep convolutional neural network, Gastrointestinal Cancer, Colorectal cancer, Transfer Learning.**

I. INTRODUCTION

Microsatellites are often termed as short tandem repeats or simple sequence repeats. Microsatellites are short repetitive DNA which are about 6 to 10 base pairs in length.

These short repetitive DNA sequences can be present at thousands of locations in the genome. One of the unique characteristic of these repetitive DNA is that it has higher mutation rate as compared to other regions of the genome. The high mutation rate occurs because of the phenomenon of replication slippage experienced by DNA polymerase during replication. [5]

It has been observed that the repetitive DNA sequences present in the DNA can make the DNA polymerase unstable resulting in addition of wrong nucleotides during replication. The study of microsatellites can give us valuable information for example variation in microsatellite can be used to distinguish between normal cell and an abnormal cell. In case of abnormal cells the microsatellites may show gain or loss in high frequency. Hence the genetic profile of abnormal cell will be different as compared to the normal cell.

Abnormality in microsatellites can be related to either microsatellite stability or microsatellite instability. Most colorectal and gastrointestinal cancers (about 85%) are microsatellite stable (MSS) which have intact DNA and tend to respond well to regular chemotherapy based treatments while conditions related to microsatellite instability needs to be identified for immunotherapy. [6]

Microsatellite instability (MSI) is a molecular tumor phenotype caused by the loss or gain of nucleotides from microsatellites and can be measured by polymerase chain reaction (PCR). These changes can arise from impairments in the mismatch repair (MMR) protein complex. A DNA repair mechanism is critical for maintaining genomic stability. Deficient MMR dMMR causes DNA mutations to accumulate and may lead to high levels of microsatellite instability MSI-H. Immunohistochemistry can be used to determine dMMR.

MSI gastric tumors are a known entity and there are several ways to differentiate or diagnose a tumor being an MSI. One can take a look at the protein level or at the DNA level. Within the protein level one can look at the mismatch repair proteins and check if any of them are deficient. Absence of these protein staining shows that the tumor is likely MSI On the DNA level one can look at this through PCR or so-called fragment analysis which is also very direct and quick test. Also it can also be observed through TCGA data where the field is moving particularly in which we take a look at it in a multiplex fashion by next-generation sequencing. This will allow to get MSI status because of the mutational burden of the tumor especially in gastric cancer where it's not subtle. [7]

MSI tumors regardless of organ of origin are chemotherapy refractory. MSI metastatic gastric cancer progresses on systemic chemotherapy within first four months of treatment and again showing that these tumors are biologically distinct and that immunotherapy should be considered in the earlier lines of therapy. [8]

Microsatellite instability (MSI) is really the main driver for the choice of immune therapies and human associated inflammatory responses. MSI tumors is just a surrogate for hypermutation which tend to have a load of mutations 10 to 100 times increase in the mutational load versus MMR Proficient tumors.

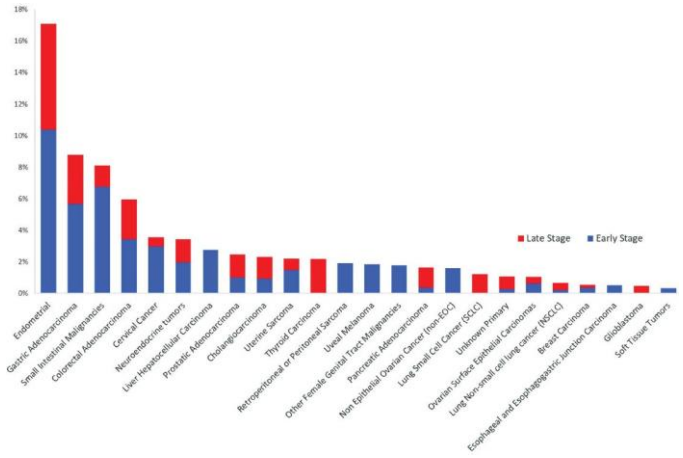

Fig. 1. Mismatch repair deficiency across tumors.

The gastric cancer leads the mismatch repair deficiency across tumor types per [9] which is followed by small intestine, colorectal and then pancreas cancer as shown in Figure 1. Hence MSI high does exist in primarily gastric, then the small intestine and colorectal cancers.

The tumor mutational burden among GI cancers is high on gastric cancer and from clinical standpoint this conveys that immune therapy is suitable.

There is also high correlation relationship between MSI and tumor mutation load for both GI and Colorectal cancers per [10] which essentially means that MSI is a good surrogate for a high mutational load.

The disease control rate via immunotheraphy was about 90% in the case of MMRD cancers and this was about 16% for MMRP cancers as per [5].

In order to identify a given cancer whether it is MSI or MSS, histological study needs to be done on the clinical data such as FFPE images. FFPE means a way of accessing cells or tissues from samples through a formula which is really good for the clinic. The paraffin-embedded process will slow down the cells which are extracted from tissues from dying off and killing themselves and starting the processes that will destroy things within the cells so forming them pauses and slows down a lot of those processes. It also keeps fitta cells and organelles and things the structure can remain the same for doing histological studies. An example of histlogical images derived from FFPE are shown in Figure 2.

II. Related Work

One of the main goals of the computer vision and AI community in collaboration with the pathology community is at the moment to design and develop techniques to learn to represent tissue which is probably the most difficult things. The most difficult task in computational pathology in a modern sense is that we have different type of tissue types and we also have many different diseases.

Generally a human pathologist reviews a glass slide under a microscope that contains a stained specimen which is called the pathology diagnosis.

The challenges are the diversity and variety of this and that many of them are polymorphic. We have mitosis and also the manifestation in body parts which appear differently when we look in the microscope or a digital scanner. So that's what we call in computer science np hardness which is a fancy word for saying that's basically impossible. Since the representation of the tissues has large number of possible combinations, machine learning can be employed to resolve this complexity and classify MSI and MSS cancers.

The whole idea recently is to digitize the details of the stained specimen via a scanner so that it is available in the form of a digital image. [11]. These digital image formats would pave way for exploiting machine learning and other techniques to assist the diagnosis. [12].

Initial application of machine learning algorithms is via the use of feature extraction methods which are manual in nature. The performance of these algorithms purely depend on the manual feature extraction methods using traditional ML algorithms. Efficiency depends on the annotation, filters used and combination of feature sets. [13] Several feature extraction techniques such as texture descriptors, lower-order and higher-order histogram features, local binary patterns, gray-level co-occurrence matrix, gabor filters, perception-like features etc were used. Also several classifiers such as 1- nearest neighbour, Linear and radial basis function support vector machine, Ensemble of decision trees, etc were tried.

There were attempts in creating high-dimensional data by a process called Radiomics via the use of computational power that could transform medical images for prognosis and diagnosis of cancer. [14] highlights that tumor staging can be classified by the use of computed tomography (CT) images from which relevant features could be extracted. Studies were conducted to predict microsatellite instability using computed tomography images in colorectal cancer by extracting Radiomic features. [15] applied Radiomics feature selection in combination with Naïve Bayes classification and have got AUC of 0.598 for the clinical model. The Radiomics Model provided an AUC of 0.688 and the Combined(Radiomics+Clinical) Model achieved AUC of 0.752.

In general the presence of proteins is considered as proficient MMR and if there are any proteins missing, it will be considered as deficient MMR. Predictive biomarkers considered to be more effective as the immune checkpoint inhibitors (ICIs) though received reasonable achievements for gastrointestinal cancers, it didn't create a wider impact due to its low response rate. The high tumor mutational burden (TMB) along with the mismatch repair/microsatellite instability are the categories where the predictive biomarkers can be used to guide.

Formalin-fixed paraffin embedded (FFPE) sections of tumors were leveraged to determine the mismatch repair status where immune checkpoint inhibitors were applied on those images. These were normalized and later standardized so that feature extraction is performed. The extracted features are then scaled using a sigmoid function and later a classifier such as linear support vector machine model was applied. The linear SVM classification model resulted in 0.74 AUC. [16]

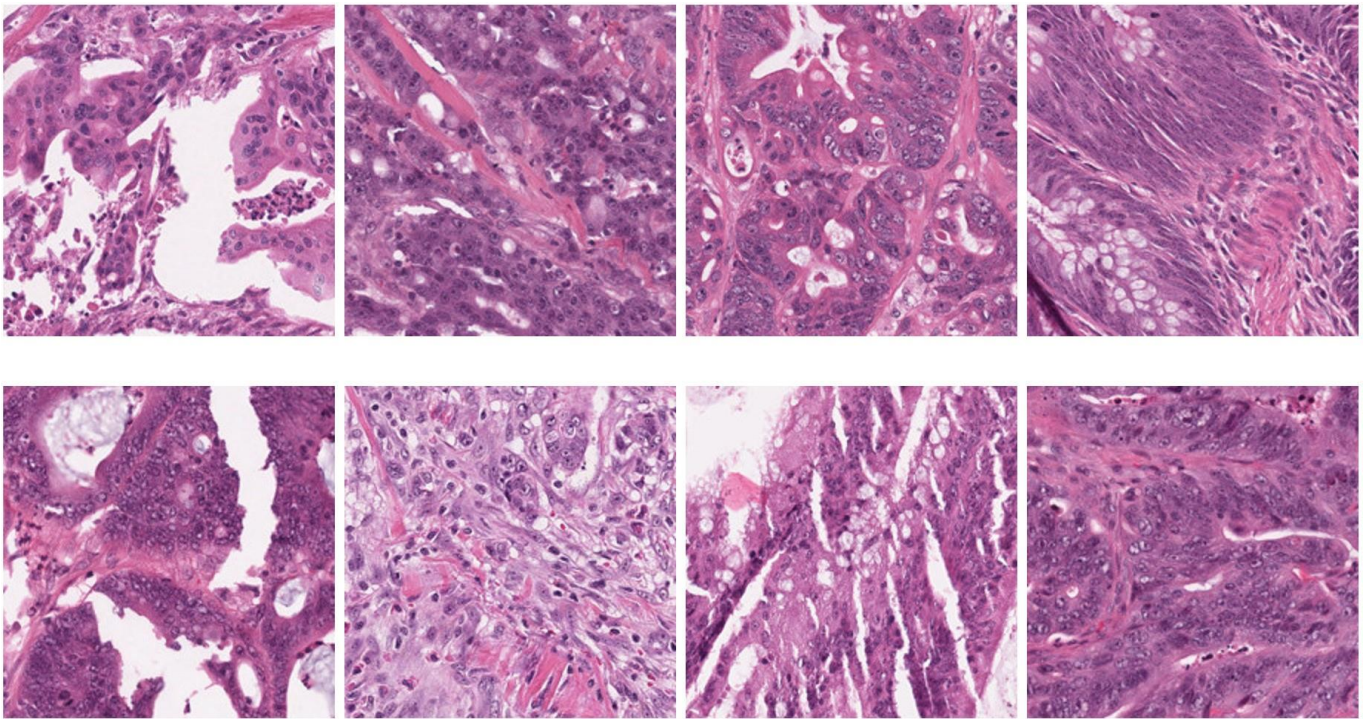

Fig. 2. Histological images of colorectal and gastrointestinal cancer - MSI (Top) and MSS (Bottom)

Similar to [15] but with more sophisticated way, [17] has used domain experts such as radiologists where one of the them segmented the CT images manually around the tumor region using a special software called ITK-SNAP software while the other radiologist double checked the segmented tumor regions to ensure they were accurate enough. Three different approaches were adopted in [17] where clinical features were analyzed using univariate statistical analysis that could highlight MSI in the first approach which gave 0.74 AUC. [18]

The second approach depended on Wilcoxon rank-sum test that could leverage radiomic features and gave an AUC of 0.76. The 3rd model similar to [15] combined the first and second approaches that is, it used both the radiomic features and the clinical variables and unlike previous attempts used a random forest classifier which gave 0.79 as AUC. The third approach gave a better result than the previous approaches.

Regression based methodologies were attempted on the analysis of data related to population in China. For this formalin-fixed paraffin-embedded tissue was used from which the DNA can be extracted and when immunohistochemical (IHC) analysis was applied, it enables to identify the protein expressions thus facilitating to identify the tumor MMR.

Polymerase chain reaction (PCR) was used to analyze the MSI status and this MSI status was more efficiently identified through various machine learning techniques and in this case with the use of Unconditional logistic regression [19] which gave 0.8062 as AUC. So, this unique combination gave a high potential to detect MSI from majority of colorectal cancer populations. One of the primary limitation that was observed to be the inability to clarify the effect of MMR status due to unavailability of therapeutic information on clinical outcome.

The approach of using Radiomic features continued in this study [20] similar to previous studies such as [15] and [17]. Contrast-enhanced computed tomography (CE CT) images were used to extract radiomic features to detect the DNA mismatch repair deficient (MMR-D) and tumor mutational burden-high in patients.

The process adopted was to extract intra-tumoral radiomic features from contrast-enhanced computed tomography (CE CT) images where the tumor contours were used to form Peritumoral-rim after extending.

Several edge detection techniques such as Sobel and Gabor were used from image processing from which appropriate Radiomic features were selected. This is done through generalized linear regression methodology and later the selected features are sent to a recursive feature elimination method and to a random forest classisfier which has achieved an AUC of 0.87. [20] Though the accuracy obtained was reasonable, the overall process is quite complex.

A more recent attempt was done by [21] who has used Ensemble Patch Likelihood Aggregation (EPLA) methodology which leverages a unique hybrid approach that uses both deep learning and traditional machine learning techniques.

This method has got 0.8848 as AUC to predict microsatellite status. This approach is based on histopathology images from asian colorectal cancer data which inturn comes from TCGA-COAD data set. This ensembling of traditional ML methods and deep learning has shown significant improvement in MSI detection to previous methods. [22][23].

III. The Proposed Transfer Learning

A. Digital Pathology

Digital Pathology has become a standard workflow where a patient goes to the hospital and a biopsy/test is taken. The sample with tissue is taken to the lab usually and it gets processed which then goes to some sort of distribution channel and the tissue sample in the end goes to a pathologist. Since the sample is digitized in the process, this opens up a wider set of possibilities.

There are some inherent challenges in the process of digitizing the sample such as blurring of the image and missing/fading of the tissue in the sample. Hence, its essential to have a nice quality digital sample for digital pathology. The validation analysis should see same information from the glass slide as that of the digital sample. There are publicly available data sets where we have the digital images which are then used to apply computer vision or AI related techniques. [24]

B. Artificial Intelligence

Artificial intelligence and machine learning is now commonly getting in vogue in for medicine in general and also in sub-specialty practice of Gastroenterology.

Artificial intelligence is the use of computer-based technology to help solve some of the repetitive tasks that humans have been doing over the last several years and the reason artificial intelligence is used here is that it not only can prevent humans making any errors because of these repetitive tasks but at the same time it improves the efficiency. It also helps in getting a lot of work in a very reasonably short period of time.

The area of artificial intelligence and machine learning have several applications in the field of Gastroenterology. One of the areas that this technology helps is obviously image recognition where the machine can diagnose images based on certain characteristics that we input and these are the algorithms that are actually written and the machines are then able to over the time learn on its own and get better at it.

Machines can also predict better than a human especially when certain parts of the GI tract might get missed out of sight via human eyes when doing endoscopy while artificial intelligence is able to predict certain lesion's.

Traditional Machine Learning methods leverage feature based approaches through algorithms such as Random Trees, Random Forests and Logistic Model Trees as shown in Figure 3. These methods focus on extracting global features from the given data and then train the model using the extracted features. These approaches worked well for smaller datasets where the identification of datasets was relatively simple. As the dataset sizes increases, the level of complexity increases. This is the case for the digital pathology images which are of large size and this in turn requires the involvement of deep neural networks like convolutional neural networks which are popular in image classification and recognition tasks.

A regular neural network has an input layer, hidden layers, and an output layer. The number of inputs depends on the number of dimensions of the input data and the hidden layers perform computations on these inputs. The outcome of the

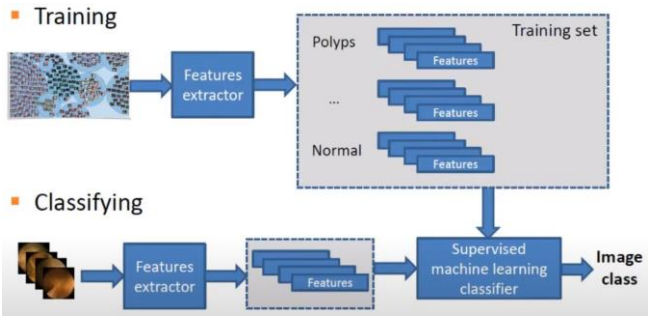

Fig. 3. Machine Learning Feature Based Approaches

computations will pass through the output layer. Each of these layers contain neurons which consist of certain weight and they are in turn connected to neurons in the previous layer. There is absolutely no assumption what so ever about the data being fed into the network. These networks are quite nice with certain type data except data types such as images.

C. Convolutional Neural Networks

Convolutional neural networks introduced by [25] are special type of feed-forward artificial neural networks which is inspired from visual cortex. A small region in the brain called Visual cortex is a region of cells that are sensitive to specific regions of visual field. That is, few neurons in the visual cortex fires when exposed to vertical edges while few fire when exposed to horizontal layers. Few fire and exposed to diagonal edges and that is the motivation behind convolutional neural network.

If an image of 200 x 200 x 3 pixels is fed to a fully connected neural network, around 120 thousand bits are required at the first hidden layer itself which require a lot of parameters. Basically in a convolutional neural network each neuron in one layer is connected with another layer of the network that contains a small region of the layer before it. This topology results in a fewer weights between neurons as the number of connections between layers are low.

CNN considers small segments of the image where these segments/patches are known as features or filters. By finding a matching feature in roughly the same positions in two images, CNN improves on learning the similarity between the whole image matching schemes. In convolution layer, one by one feature is taken and moved it through the entire image. While moving filter CNN's multiplies the pixel value of the image with that of the corresponding pixel value of the filter and can be added and dividing by the total number of pixels to get the output.

Generally a convolutional neural network has three layers. A convolution layer, pooling layer and towards the end a fully connected layer. Convolutional neural networks or CNNS can do some pretty interesting things when they are fed with a bunch of pictures.For instance when the face images are given as an input, the convolution neural networks learns some fo the features such as edges, dots and spots. These multi-layer neural networks learn these edges or gradients in the initial layers and the second layer learns some of the parts of objects

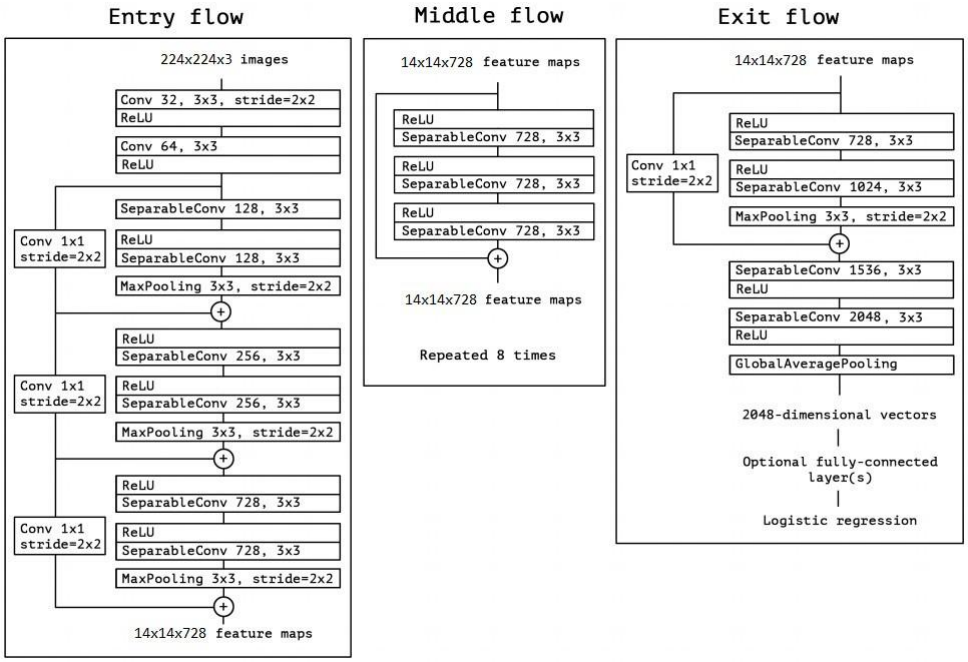

Fig. 4. Xception Network

such as eyes, noses and mouths. The third layer learns objects such as faces.

Convolution is a measure of overlap between two functions as one slides over the other. Mathematically it's a sum of products the standard convolution operation is slow to perform however it can speed up with an alternative method called depth wise separable convolution. A scalar is returned from a regular convolution operation that computes the input's and kernel's sum of products. This operation is continued by sliding the kernel over the input. The concern now is with the cost of this convolution operation which has a number of multiplications required.

D. Depth-wise Separable Convolution

Depth-wise separable convolution which was introduced by [26] applies convolution to a single input channel at a time. This is different from the standard convolution that applies convolution to all channels. The complexity of this convolution can be split into two parts depth-wise convolutions and point-wise convolutions. The number of multiplications thus obtained is the sum of multiplications in the depth-wise convolution stage plus the number of multiplications in the point-wise convolution stage.

Depth-wise separable convolution decreases the computation and number of parameters when compared to standard convolution. Depth-wise separable convolution is a combination of depth wise convolution and point wise convolution. Depth-wise convolution is the filtering step and point wise convolution can be thought of as the combination step. Figure 5 illustrates the working of depth-wise seperable convolution.

E. Xception Network

Xception is a convolution neural network architecture based entirely on depth-wise separable convolution layers. Large datasets such as Google's jft image dataset are applied to Xception network which showed exceptional performance. It's a repository of 350 million images with 17,000 class labels. To put this into perspective the popular ImageNet took 3 days to train, however to train even a subset of this jft dataset it took a month and it didn't even converge. In fact it would have approximately taken about three months to converge, how'd they let it run to its full length. Xception is pushing convolution neural networks to use depth-wise separable convolution as the de facto.

Generally a pre-trained model is used to make predictions so that the results are great. Classifying images into the categories

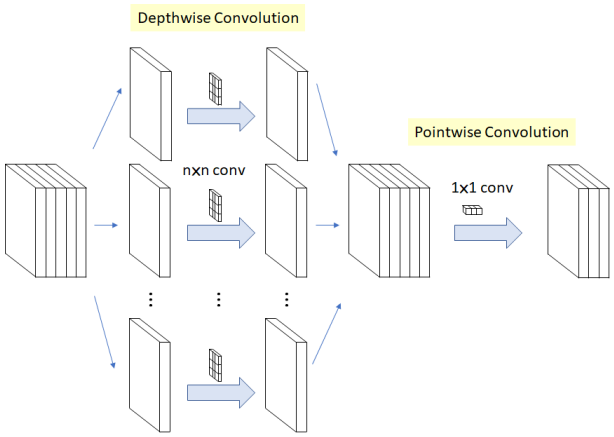

Fig. 5. Depth-wise Seperable Convolution

used by the original models is simple but what if the new use case don't categorize images in exactly the same way as the categories for the pre-trained model are. A new model can be built from scratch for this specific purpose but to get good results it needs thousands of images with labels for which might be not practical in all scenarios. For these requirements transfer learning will give good results with far less data.

F. Transfer Learning

Transfer learning is one of the method of training machine learning algorithms unlike other methodologies like supervised machine learning and unsupervised semi-supervised machine learning. Transfer learning has a special feature where we train for one task and try to actually use that knowledge for another related task. Let's say the objective is to train a convolutional network image classifier to identify Cheeta but we only have a thousand images of cheetah which isn't good enough to train the convolutional neural network model.

We can actually use an existing pre-trained CNN models which is probably trained with millions of animals and already understands how the animal looks like along with all the features of the animals. This pre-trained network is downloadable and can help us simplify training process were there are less number of training images.

G. DataSet

The dataset used in this work comes from [27] which has histological images of gastrointestinal cancer that can be used to classify MSI Vs MSS. The original dataset has 411,890 unique images from original SVS slides of cancer patients of TCGA cohort. In this work a subset of 192,312 images were taken which amounts to approximately 5GB of data.

Formalin-fixed paraffin-embedded (FFPE) slides are the source of these images and these slides were pre-processed prior to formulation of this dataset. The resolution of the images is at 0.5 micro metre/px and the images are tailored to 224 x 224 pixels. All the images were reformatted into the specified resolution and in JPG format and the images were color normalized using Macenko method [28]. The patients were categorized by specialists into MSI and MSS categories so that dataset has labels. The sample image of this dataset can be seen in Figure 2. The dataset is divided into training, validation and testing in the ratio of 80%, 10% and 10% respectively.

IV. Transfer Learning Implementation

A. Architecture

The architecture of the convolutional neural network used in this work comes from the Xception Network. The Xception network uses depth-wise seperable convolutions which are quite simple from compute perspective when compared to regular convolutions. The overall flow in the Xception network is detailed in Figure 4 with input histological image dimensions as 224x224x3.

This network is already pre-trained with larger datasets and hence all the layers in the network are have pre-set weights.

Algorithm 1 Transfer Learning using Xception Network

```
Initialize the Keras Libraries and Colab
Download the Dataset
Get the Train and Test sets with labels
Normalize the images

Data Augmentation
for each sample image do
    Rescale and Rotate image by 45 degree
    Width and Height Shift by 20%
    Flip image Horizontally
    Zoom the image by 50%
end for

Build Neural Network
Download pre-trained Xception Network
The input dimensionality is set to 224x224x3
Create a Keras Sequential Model
Embed pre-trained Xception model as starting block
Pre-trained model o/p to Global Average Pooling
The network ends with a Dense network layers
Compile Transfer Learning model using SGD

Train, Validate and Test
Batch Size of 64 is suitable
Train the network for just 5 epochs and validate
Test the trained model with test data images
Save the weights of the network
```

This indicates that all the initial layers have been already trained to capture edges and gradients in a given image. Also the later layers understand the features such as textures and parts of objects. Once the pre-trained Xception network is imported from the Keras applications, the later layers of the network contain a dense network whose outputs are forwarded to global average pooling layer.

B. Data Augmentation

In order to make the model more robust, the training image data can be augmented. This will address any issue of overfitting and the model gets more generalized. The input training images are rescaled and then rotated by 45 degrees. Then the images are shift by 20% both vertically and horizontally. Later the images are flipped horizontally and zoomed by 50%. This overall process augments the images and improves the validation accuracy.

C. Network Hyperparameters

Some of the most important network hyper parameters are learning rate, epochs and batch-size. It is very important to tune these parameters in such a way that good performance is obtained. The batch size of 64 was observed to be working well for the given dataset and this batch size is also very suitable for GPU architecture to handle 64 images in parallel. For the training Stochastic Gradient Descent optimizer is used with exponentially decaying learning rate and 0.6 momentum with

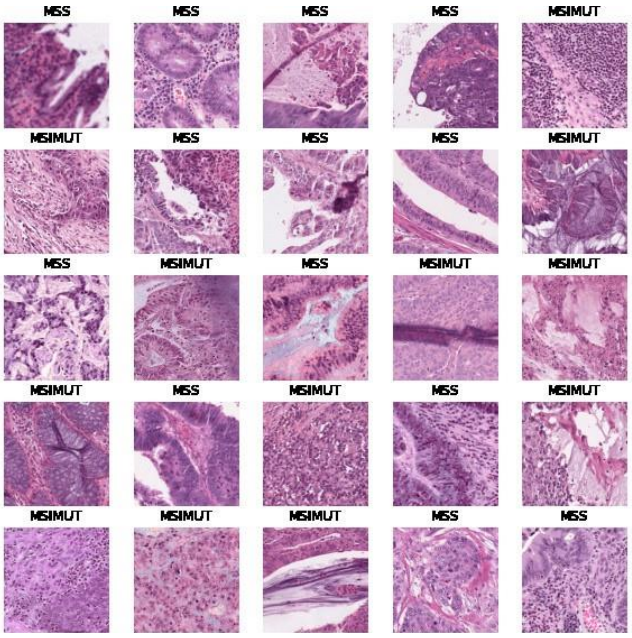

Fig. 6. Labeled Training Images

the decay of 0.8. Algorithm 1 depicts the architecture of neural network used.

V. Results

A. Training and Validation

The training of the pre-trained Xception network is done with histological training images of the dataset. The training dataset has the labeled MSI and MSS images as shown in Figure 6. The overall training images are 153,849 and the overall validation images are about 19230. The overall program execution was done in Google Colab which provides free GPU access.

The training was done for up to 5 epochs which is sufficient as the initial layers of the Xception network are already pre-trained to identify edges, gradients, textures, etc. The training data is augmented as per the specifications mentioned in the algorithm 1. The validation accuracy of 93.18% is obtained as shown in Figure 7 and the overall time taken for training the network took about 4 and half hours.

B. Testing

The trained Xception network was then tested with test histological images of the dataset. The overall test images are about 19230. The resulting testing accuracy is about 90.17% with test AUC of **0.932** which indicates the ability of the pre-trained Xception network on these histological images. The predicted test images labels are shown in Figure 8 and all those images are predicted correct.

VI. Conclusion and Future Work

This paper demonstrates that MSI and MSS gastrointestinal cancer classification can be most accurately done using transfer learning methodology in deep learning. A brief review of literature is performed on the classification accuracies of gastrointestinal cancer datasets and various model behaviors has been investigated. This gives an overall idea of research

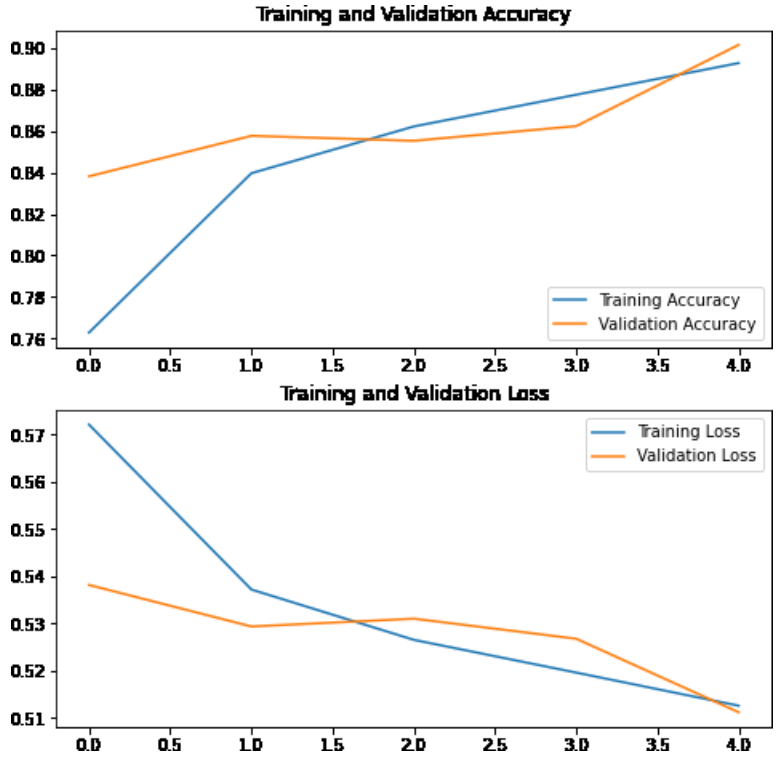

Fig. 7. Training and Validation Accuracy

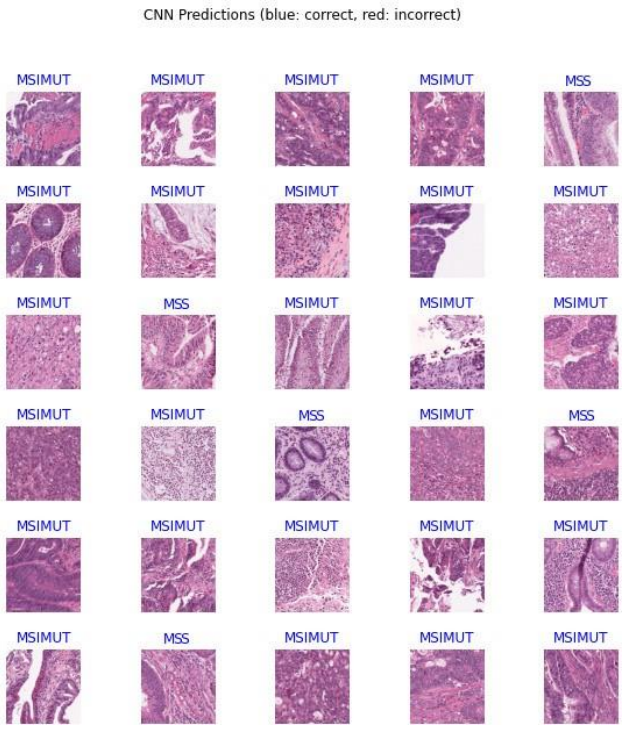

Fig. 8. Predicted Labels for Test Images

focus to this work and in turn facilitates to identify the gaps and direction to undertake.

The required pre-requisites have been captured and design of the deep neural network using pre-trained Xception network has been performed. The formulated pre-trained model is then trained on Google Colab research platform as training these models is compute intensive requiring GPU.

The implementation of the pre-trained model illustrates the usage of GPU on Google cloud platform for training deep learning models. The algorithm has been coded in Python along with Keras API which uses the TensorFlow backend.

As a future work, other gastrointestinal cancer datasets can be trained with the pre-trained Xception model developed in this work. Other pre-trained networks can also be attempted with this dataset and other datasets as well. Also, hybrid combined networks can be tried for improving AUC.

The robustness of the model in this work is improved by implementing suitable data augmentation techniques which in turn enhanced the classification accuracy of the classifier. Thus this work demonstrates that MSI Vs MSS gastrointestinal cancer classification performance is reasonable to be implemented in production and more research in these areas can help patients be diagnosed accurately and especially at the right time.

REFERENCES

[1] G. Golshani and Y. Zhang, "Advances in immunotherapy for colorectal cancer: a review," *Therapeutic Advances in Gastroenterology*, vol. 13, p. 1756284820917527, 2020.

[2] J. N. Kather, A. T. Pearson, N. Halama, D. Jäger, J. Krause, S. H. Loosen, A. Marx, P. Boor, F. Tacke, U. P. Neumann *et al.*, "Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer," *Nature medicine*, vol. 25, no. 7, pp. 1054–1056, 2019.

[3] E. J. Lipson, W. H. Sharfman, C. G. Drake, I. Wollner, J. M. Taube, R. A. Anders, H. Xu, S. Yao, A. Pons, L. Chen *et al.*, "Durable cancer regression off-treatment and effective reinduction therapy with an anti-pd-1 antibody," *Clinical Cancer Research*, vol. 19, no. 2, pp. 462–468, 2013.

[4] N. J. Llosa, M. Cruise, A. Tam, E. C. Wicks, E. M. Hechenbleikner, J. M. Taube, R. L. Blosser, H. Fan, H. Wang, B. S. Luber *et al.*, "The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints," *Cancer discovery*, vol. 5, no. 1, pp. 43–51, 2015.

[5] D. T. Le, J. N. Uram, H. Wang, B. Bartlett, H. Kemberling, A. Eyring, N. S. Azad, D. Laheru, R. C. Donehower, T. S. Crocenzi *et al.*, "Programmed death-1 blockade in mismatch repair deficient colorectal cancer." 2016.

[6] R. Gupta, S. Sinha, and R. N. Paul, "The impact of microsatellite stability status in colorectal cancer," *Current problems in cancer*, vol. 42, no. 6, pp. 548–559, 2018.

[7] S. Ogino and A. Goel, "Molecular classification and correlates in colorectal cancer," *The Journal of Molecular Diagnostics*, vol. 10, no. 1, pp. 13–27, 2008.

[8] M. Kreidieh, D. Mukherji, S. Temraz, and A. Shamseddine, "Expanding the scope of immunotherapy in colorectal cancer: Current clinical approaches and future directions," *BioMed Research International*, vol. 2020, 2020.

[9] D. T. Le, J. N. Durham, K. N. Smith, H. Wang, B. R. Bartlett, L. K. Aulakh, S. Lu, H. Kemberling, C. Wilt, B. S. Luber *et al.*, "Mismatch repair deficiency predicts response of solid tumors to pd-1 blockade," *Science*, vol. 357, no. 6349, pp. 409–413, 2017.

[10] M. E. Salem, J. Xiu, B. A. Weinberg, W. S. El-Deiry, L. M. Weiner, Z. Gatalica, Z. Liu, H. El Ghazaly, N. Xiao, J. J. Hwang *et al.*, "Characterization of tumor mutation burden (tmb) in gastrointestinal (gi) cancers." 2017.

[11] L. Pantanowitz, "Digital images and the future of digital pathology," *Journal of pathology informatics*, vol. 1, 2010.

[12] D. Komura and S. Ishikawa, "Machine learning methods for histopathological image analysis," *Computational and structural biotechnology journal*, vol. 16, pp. 34–42, 2018.

[13] J. N. Kather, C.-A. Weis, F. Bianconi, S. M. Melchers, L. R. Schad, T. Gaiser, A. Marx, and F. G. Zöllner, "Multi-class texture analysis in colorectal cancer histology," *Scientific reports*, vol. 6, p. 27988, 2016.

[14] B. Zhang, J. Tian, D. Dong, D. Gu, Y. Dong, L. Zhang, Z. Lian, J. Liu, X. Luo, S. Pei *et al.*, "Radiomics features of multiparametric mri as novel prognostic factors in advanced nasopharyngeal carcinoma," *Clinical Cancer Research*, vol. 23, no. 15, pp. 4259–4269, 2017.

[15] S. Fan, X. Li, X. Cui, L. Zheng, X. Ren, W. Ma, and Z. Ye, "Computed tomography-based radiomic features could potentially predict microsatellite instability status in stage ii colorectal cancer: a preliminary study," *Academic radiology*, vol. 26, no. 12, pp. 1633–1640, 2019.

[16] Z. Lu, H. Chen, X. Jiao, W. Zhou, W. Han, S. Li, C. Liu, J. Gong, J. Li, X. Zhang *et al.*, "Prediction of immune checkpoint inhibition with immune oncology-related gene expression in gastrointestinal cancer using a machine learning classifier," *Journal for immunotherapy of cancer*, vol. 8, no. 2, 2020.

[17] J. S. G. Pernicka, J. Gagniere, J. Chakraborty, R. Yamashita, L. Nardo, J. M. Creasy, I. Petkovska, R. R. Do, D. D. Bates, V. Paroder *et al.*, "Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation," *Abdominal Radiology*, vol. 44, no. 11, pp. 3755–3763, 2019.

[18] G. Corso, C. Pedrazzani, D. Marrelli, V. Pascale, E. Pinto, and F. Roviello, "Correlation of microsatellite instability at multiple loci with long-term survival in advanced gastric carcinoma," *Archives of Surgery*, vol. 144, no. 8, pp. 722–727, 2009.

[19] W.-Y. Yan, J. Hu, L. Xie, L. Cheng, M. Yang, L. Li, J. Shi, B.-R. Liu, and X.-P. Qian, "Prediction of biological behavior and prognosis of colorectal cancer patients by tumor msi/mmr in the chinese population," *OncoTargets and therapy*, vol. 9, p. 7415, 2016.

[20] H. Veeraraghavan, C. F. Friedman, D. F. DeLair, J. Ninčević, Y. Himoto, S. G. Bruni, G. Cappello, I. Petkovska, S. Nougaret, I. Nikolovski *et al.*, "Machine learning-based prediction of microsatellite instability and high tumor mutation burden from contrast-enhanced computed tomography in endometrial cancers," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.

[21] R. Cao, F. Yang, S.-C. Ma, L. Liu, Y. Zhao, Y. Li, D.-H. Wu, T. Wang, W.-J. Lu, W.-J. Cai *et al.*, "Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in colorectal cancer," *Theranostics*, vol. 10, no. 24, p. 11080, 2020.

[22] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber *et al.*, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLoS medicine*, vol. 16, no. 1, p. e1002730, 2019.

[23] J. Ziegler, J. F. Hechtman, R. Ptashkin, G. Jayakumaran, S. Middha, S. S. Chavan, C. Vanderbilt, D. DeLair, J. Casanova, J. Shia *et al.*, "Mimsi-a deep multiple instance learning framework improves microsatellite instability detection from tumor next-generation sequencing," *bioRxiv*, 2020.

[24] M. Kandemir and F. A. Hamprecht, "Computer-aided diagnosis from weak supervision: A benchmarking study," *Computerized medical imaging and graphics*, vol. 42, pp. 44–50, 2015.

[25] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

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

[27] J. N. Kather, "Histological images for MSI vs. MSS classification in gastrointestinal cancer, FFPE samples," Feb. 2019. [Online]. Available: https://doi.org/10.5281/zenodo.2530835

[28] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt, and N. E. Thomas, "A method for normalizing histology slides for quantitative analysis," in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2009, pp. 1107–1110.