TRM - Research Proposal

Predicting mutational changes of Rhabdomyosarcoma using Deep learning and Convolutional Neural Networks

Vivek Velivela - 12704286

Bachelor of Computing Science (Honours)

supervised by

Prof. Paul Kennedy and A/Prof. Daniel Catchpole

May 29, 2022

Topic

Establishing a Deep Learning framework for predicting key genomic changes in Rhabdomyosarcoma from histological images

Keywords

Deep Neural Networks, RMS Tumour, mutational changes, Rhabdomyosarcoma

Thesis

- Building Deep learning Model on digital microscopic images of tumour tissue to predict mutational changes of RMS Tumour.

Hypothesis

Deep Neural Networks with help of transfer learning can identify and predict the mutational status of Rhabdomyosarcoma tumours

Aims

- Building Deep learning Model on digital image of tumour tissue of transgenic RMS tumours with human genomic changes.

Objectives

- brief knowledge on various histological changes associated with the types of Rhabdomyosarcoma that can describe the mutational status of the tumour.
- Gain In-depth Understanding of Convolutional Neural Networks and acquire knowledge on Building and validating deep neural network models.
- Collect Rhabdomyosarcoma Histology Images and apply certain preprocessing techniques

1 Introduction

Due to pivotal developments in image analysis community, Computer Vision applications have shown huge impact in many industries mainly in healthcare industry. Due to increase in data repositories in various industries, performance of deep learning and convolutional neural network(CNN) models started to develop. classification and segmentation are some of several deep learning applications in healthcare industry[8].

Some classification tasks contains examining whether a tumour is present in a tissue, classifying type of tumour, predicting type of mutations involved in a tumour. early stages of medical image classification models were based on Deep Belief Networks, sparse auto encoders etc. As increased performance in CNN, they have been used widely across various medical diagnosis and prognosis tasks.Usually large amounts of data is needed to train these models but it has proven that transfer learning approach has very big affect on models performance with small amount of data [9]. Therefore there has been a lot of research being conducted over improving the performance of deep learning models using smaller medical image data sets. Recently, Deep learning Networks have been used to classify Rhabdomyosarcoma subtype [3].

Rhabdomyosarcoma(RMS) is an aggressive skeletal muscle-lineage malignancy and most common soft tissue sarcoma in children[1].There are two major types of RMS namely Embryonal RMS(ERMS),Alveolar RMS(ARMS), these are divided according to their distinct clinical features.ERMS is more common subtype whereas ARMS is more aggressive[1].The treatment of varies RMS types varies thus it is important to classify RMS type to determine the type of treatment the patient will undergo.

There are three risk groups of Rhabdomyosarcoma classified by World Health Organisation(WHO)- embryonal as intermediate, and alveolar as unfavorable[2]. Current standard method of tumour identification involves a trained pathologist examining the histology slide and classifying the RMS type. Subtypes of RMS are identified according characteristics of skeletal muscle lineage. ARMS are of two types — classic pattern and solid pattern. Classic ARMS tumor cells contains round and uniform nuclei while Solid ARMS contains cells with round nuclei arranged in sheets. ERMS are of four subtypes- typical, dense, botryoid and spindle cell pattern.Typical pattern are dense cellular regions. Dense pattern characterized by central or angulated nuclei. Botryoid pattern contains linear aggregates of tumor cells. Spindle cells have oval and blunted central nuclei. As Dense pattern ERMS closely resembles ARMS there is a high chance that these types can be often misclassified[2]. It is also important to study the underlying mutational change to understand the behaviour of the tumour.Hence, Some transgenic models are developed to ensure that they are representative to human diseases by histopathology, gene expression and other features which

gives us an opportunity to learn more about the mutational profile and cell of origin [4]. Predicting Mutational status in a disease can be classified as a multi label problem as there will be chance that the tumour can contain more than one mutation.

2 Literature Review

2.1 classification and mutation prediction non-small lung cancer histopathology images using deep learning

Lung cancer is one of the most widely spread cancers in the world caused not only by exposure to toxic chemicals but also due to smoking. Two types of lung cancer are the most common ones namely squamous cell carcinoma(LUSC) and Adenocarcinoma(LUAD) with distinct therepautic protocols. LUAD diagnosis will prompt the search for molecular biomarkers and mutation which will have an impact on treatment options. Most common mutations have a challenging drug targets. Due to quite similar features between LUAD and LUSC it is usually hard to identify the subtype even for a trained pathologist. This literature review will discuss the methods and image processing techniques used to predict the mutational status of the tumour and briefly states methods used for classification of the tumour.

The data contains 1634 whole slide images that were extracted from genomic Data commons database with 1176 tumour tissues(609 LUSC, 567 LUAD) and 459 normal tissues. The data was divided into three sets: training(70%), validation(15%) and testing(15%). As the whole slide images were too large to be sent as a direct input to neural network, each slide was divided into 512*512 pixels tiles obtained from non overlapping patches of whole slide image. tiles with more than 50% of background were omitted and then sent into inception v3 CNN with partial or fully retrained. tiles were classified and the results were finally aggregated per slide to extract the heat maps and AUC statistics.

Three classification models have been developed with same approach: first model was trained on pre trained inception v3 with last layer randomly initialised to classify tumour versus normal which yielded AUC of 0.990 on averaging probabilities of tumour in each slide, same approach has been tested on distinguishing LUAD vs LUSC which yielded AUC of 0.847 and third model which classifies into normal lung or LUAD or LUSC yielded the highest AUC of 0.97.

For predicting gene mutation in LUAD slides they have downloaded gene mutation data for matched patients samples from TCGA. they have only selected images which were mutated at least 10% of available tumours in order

to avoid biasing the network to learn LUAD vs LUSC specific mutations from each slide of LUAD only tiles classified as LUAD by the accurate classification model were used as input. to perform multitask classification the inception v3 was modified to allow multi-output classification results. RMSProp algorithm is used for optimization and fully trained this network for 500,000 iterations with WSI, each one associated with 10 cell vector where cell tells us regarding the presence and absence of the mutation with 1 or 0 respectively. method. 212,000 tiles from 320 slides were used for training and validation and 44,000 tiles from 62 slides were used for testing. AUC values for serine/threonine protein kinase11 (STK11), EGFR, FAT atypical cadherin 1 (FAT1), SET binding protein 1 (SETBP1), KRAS and TP53 were between 0.733 and 0.856.

In conclusion this paper has mentioned methods for two tasks which are multi-class classification to predict the type of cancer and multi label classification to predict the type of mutation involved in particular slide. the workflow and CNN pipeline used was almost the same for both the models except some changes in activation, loss functions to output binary result which specifies the involved mutations. Image pre-processing techniques like dividing slides into tiles and filtering out the background has been done before feeding the data into the model.

In this paper each mutation classification is treated as binary classification task which allow single tile to have multiple mutations. As this is considered as binary classification task against all mutation combinations the last layer of the model will be replaced with sigmoid activation function with [0,1] as limit instead of softmax which supports multi-class classification. This allows each sample with several binary labels. They have also divided each slide into several tiles with magnification to capture various minute features and background has been omitted to avoid model to learn unnecessary features.

2.2 Pan-cancer classifications of tumor histological images using deep learning

Image analysis of histological slides is considered a gold standard in diagnosing cancer type and sub types and plays a crucial role in selection of treatment options. currently, whole slide images are manually examined by trained pathologists which is time consuming[6].Due to major advances in machine learning models for supervised and unsupervised learning tasks over the past few years, tumour histopathology image analysis is well suited for machine learning. This paper explains pan-cancer study of tumour image data by analysing 27815 whole slide hematoxylin and eosin(H&E) images from 23 The cancer genome atlas program(TCGA) cohorts. H&E is a standard type of staining used in tissue biopsy. TCGA is a program which is molecularly characterized over 20,000 primary cancer and matched normal samples spanning 33 cancer types [7]. They

have developed a image processing and neural network pipeline with Inception CNN to classify the status of the tumour whether it is tumour/normal. they report the effectiveness of the pipeline to classify the cancer sub types.They have have compared the ability of the network trained on one cancer type to classify images from another type and stated that network which is able to classify other class by training on different class has some relationship that overlap with tissue biology.Finally, they test the ability of the network to predict the TP53 mutational status among breast, lung and gastric cancers.

Flash frozen Whole Slide Images(WSI) of BRCA,LUAD and STAD cohorts were selected for training the model. BRCA has 985 slides, LUAD has 565 slides and STAD has 437 slides. Pre-processing techniques includes using the result of tumour/normal classifier for removing of the normal tiles from the tumour slides which help tiles with positive mutations labelled as tumour. Under sampling techniques has been used to maintain equal number of images in each class.Each slide has been divided into non-overlapping 512*512 patches. Inception v3 is used and the model was fully trained on the input slides for better results. to predict mutations they have trained two way classifiers, assigning 70% of the images in each tissue to training and 30% for testing. slide level prediction is acquired by averaging all the tile level predictions. cross tissue mutational classification was also performed by training the model with one set of cohorts and performing its prediction on another set of cohorts.

Self-cohort predictions ranges from 0.65-0.80 AUC for the per-slide and AUC 0.63-0.78 for the per-tile evaluations. Fully trained network model with AUC of 0.76 has outperformed a transfer learning simulation with AUC of 0.73 for the same data. Cross-predictions yielded AUC 0.62-0.72 for slides and AUC 0.60-0.70 for tiles.

In conclusion, the methods used in this paper to predict TP53 associated mutations across breast, lung and gastric cancer are almost similar to those methods used in mutation prediction of lung cancer [2] except some minor differences in pre-processing techniques. the improved result for fully trained network suggests that optimizing early layers of CNN had an affect on models performance. For pre-processing under sampling has been performed on the data to avoid class imbalance which might add bias to the model.

The pre-processing techniques used in both the papers aims to reduce bias and improve feature selection in the data either by dividing them into small patches,magnification, under sampling, removing background or removing non-tumour tissues. standard Inception v3 with Imagenet weights has been used for training in both the papers. As the available data was huge compared to lung cancer prediction, the model was fully trained to produce better results for predicting mutations in pan cancer study. The main difference between both the approaches was the final output pan cancer study has one mutation within three disease but in lung cancer study has almost ten mutations within one

disease. The output layer has different size corresponding to number of classes the model is being trained.

3 Challenges

Due to large number of parameters present in CNN models, huge amount of balanced data to be fed to train the model. Although Rhabdomyosarcoma is a most soft common tissue in children and adolescents aged below 20 [1], the data that currently available has high sample variance which arises a class imbalance problem[3]. There are two types of class imbalance problems namely intrinsic imbalance which is a result based on the natural occurrence of the data and on the other hand extrinsic imbalance is a result several external factors like different methods of data storage and collection methods [5].

Varied image intensity might be a potential problem if some of the images vary widely in colour and tone depending on the staining method [10]. Intra-subject variability is one of the source of variation in the data, further variability can be added to the data unintentionally while cutting and preserving the images [11].

4 Methods and Techniques

4.1 Image pre-processing

Most of the Image processing barriers can be addressed through several data level and algorithmic level and hybrid approaches.Data Level approaches include using some techniques like balancing all the classes in the data sets either by data augmenting under represented class or under sampling the over represented class.

Hybrid approaches include using framework like Deep Over Sampling (DOS). DOS is a framework which learns an embedding layer that produces more discriminative features, and then supplements the minority group by over-sampling in deep feature space[5]. In general, hybrid methods are more complex in nature and more difficult to implement than algorithm-level methods and data-level methods [5].

Generative adversarial networks(GAN) to produce more representative data, GAN's can produce high dimensional latent distribution of data which has led to significant performance gains[11]. GAN's have potential to reduce the requirement of manual annotations [11].

Algorithmic approaches includes changing the decision making process of the model that increases the importance to positive class. assigning penalties to each class through cost matrix. Increasing the cost of the minority group is equivalent decreasing the likelihood that the learner will incorrectly classify instance from this group [5].

Stain normalisation is essential pre-processing step to make data robust to various colour intensity variations caused due to manufacturing techniques of stain vendors, difference in colour responses of slide scanners.

4.2 CNN Architectures

4.3 Inception v3

Inception v3 contains 11 stacks of inception modules where each module consist convolutional filters and pooling with rectified linear unit as activation function. Inception v3 contains different filter sizes to produce various feature maps. It avoids representational bottlenecks to define a clear information flow across the network [16]. so, the input dimensions are reduced gradually at various stages of the network. Inception v3 has balanced width and depth which helps in optimal network performance [16].

4.4 Inception-ResNet-v2

This CNN is trained on image-net data set with more than million images. this network is 164 layers deep and can classify more than 1000 general classes. this CNN contains inception modules which can help producing discriminatory features with various convolutions sizes and reduce the number of parameters by using small convolutional layers. Two types of reduction modules are also present that can reduce the image size during training. Default input size is 299*299 [14].

The main difference between Inception-ResNet-v2 and Inception v3 is the prior one has multiple residual connections[15]. differing compositions of two networks inception modules is also a key difference. Each inception module in Inception v3 has different filter with different sizes to enhance networks adaptability to different convolutional kernels. A residual unit is added to each inception module in Inception-ResNet v2 to avoid vanishing gradient problem which is associated with increase in number of layers. residual units acts as skip connections to avoid the vanishing gradient problem [15].

4.5 Transfer learning and fine tuning

Transfer learning has been an integral part of deep learning applications especially in medical imaging [9]. The process of using weights of architectures which were already pre-trained on some natural image data sets like IMAGENET is called transfer learning, adding some custom classification layers and unfreezing some of the last layers of the pre-trained architecture to train the model with specific data set is called fine-tuning.

The normal intuition behind transfer learning is taking advantage of pre-trained feature maps without training a model on large data set from scratch because the model which is trained on large and general data set can effectively serve as generic model for many real world problems.

Fine tuning is used to adapt the pre-trained model to a specific problem by unfreezing some of last layers of pre-trained architecture and adding classification at the bottom for output so that the last layer will be trained to have a problem specific feature maps to produce better results.

According to [9] Usually transfer learning is performed on Imagenet architecture along with its pre-trained weights. However, there are considerable differences among various medical image features and Imagenet data set features(eg: Imagenet dataset contains many classes mostly containing real world objects with very distinct features compared to features of a tumour tissue slide) which suggest that using transfer learning necessarily does not always improve the model performance.

4.6 Activation and Loss Function

In a neural Network model Activation function performs a compound mappings between input and output units. These functions potentially introduce non-linearity into the model by mapping the outputs in a certain range.These functions can also help capture the smallest features of the data [17]. In each layer activation functions are applied after calculating the input which is processed by weights and biases.According to [17], A Neural Network without an activation function will be a linear regression model with limited capabilities and cannot process high dimensional data.As several mutations can be present in one tumour, prediction of mutational status is considered as multi label classification problem. Due to its binary classification nature, Sigmoid activation is most commonly used for multi label classification.

Sigmoid is a straight forward, binary classifier used as an activation function. It decides whether a certain object belongs to a specific class or not, as 0 or 1. unlike other activation functions like Softmax, This function does not consider

the possibilities of an object belonging to another class [17].

$$Sigmoid(x) = \frac{1}{1 + e^{-}x}$$

Cost Function is most important operation to optimise the weights of the neural network which helps in creating a better fitting machine learning model [18]. During training the neural network is run on training data set and the outputs are given in probabilities. These probabilities are compared to the target labels to calculate the loss of each training example[18]. These losses helps to train the model to give good results by adjusting weights in back propagation. Binary cross-entropy(BCE) is one of the standard loss functions

$$BCE = \frac{-1}{\mathrm{M}} \sum_{m=1}^{M} [y_m log(h_\theta(x_m))] + [1 - y_m] log(1 - (h_\theta(x_m))]$$

$$M = number\ of\ training\ examples$$
$$y_m = target\ label\ for\ training\ example\ m$$
$$x_m = input\ for\ training\ example\ m$$
$$h_\theta = model\ with\ input\ training\ weights\ \theta$$

First term $y_m log(h_\theta(x_m))$ is responsible for desensitizing false negatives [18]. If training example has a target 1 and the machine learning model outputs 0.7, it can be said there is a probabilistic false negative of 30%. The loss function penalises the model by 30% by returning the values -log(0.7) = 0.15. In perfect case of output 1, the error will be 0(-log(1) = 0). The same logic applies with the second term to deal with false positives. Binary cross entropy is used in binary classification tasks and it is also a special case of cross-entropy which is used for multi label classification.

4.7 Evaluation Metrics

Many of the current research works model performances are measured against state of the art architectures, with different representative data sets and using various scoring metrics. F1 Score, accuracy, AUC, precision and recall are the frequent metrics that are used in majority of the classification tasks.

Accuracy is one of the basic and most used metric in classification performance [13]. Accuracy is ration between correctly classified samples to total number of samples. Sensitivity, True positive rate (TPR), hit rate, or recall

represent positive correctly classified samples to total number of positive samples. whereas specificity, True negative rate (TNR), or inverse recall is expressed as the ratio of the correctly classified negative samples to the total number of negative samples. F-1 score is harmonic mean of precision and recall, range of F1 score is [0,1] [13]. high f1 score represents high classification performance. AUC stands for area under the ROC curve.AUC gives overall performance of all classes. range of AUC is between 0 and 1 [13].

5 Future Impact/Significance

I will be performing various extensive analysis and experimentation using some combinations or all the techniques that have been mentioned in the above section. based on these mentioned methodolgies I will either improvise or modify some of the techniques to achieve a notable performance improvement in Predicting the mutational status of Rhabdomyosarcoma.

Final goal of my research is to build an efficient methodology that can classify the mutations in RMS histopathological images solving many data related challenges. I am also hoping the proposed model will not only have impact on certain mutations in RMS but on all other tumour which share a similar morphology.

References

[1] Mark E. Hatley, Wei Tang, Matthew R. Garcia, David Finkelstein, Douglas P. Millay, Ning Liu, Jonathan Graff, Rene L. Galindo, and Eric N. Olson1 *A mouse model of rhabdomyosarcoma originating from the adipocyte lineage*,NIH Public access , viewed 19 April 2021 <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3479681/>

[2] Coudray, Nicolas & Ocampo, Paolo & Sakellaropoulos, Theodore & Narula, Navneet & Snuderl, Matija & Fenyö, David & Moreira, Andre & Razavian, Narges & Tsirigos, Aristotelis. (2018). *Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning.* Nature Medicine. 24. 10.1038/s41591-018-0177-5.

[3] Agarwal, Saloni & Eltigani, Mohamedelfatih & Abaker, Osman & Zhang, Xinyi & Daescu, Ovidiu & Barkauskas, Donald & Rudzinski, Erin & Leavey, Patrick. (2020).*Rhabdomyosarcoma Histology Classification using Ensemble of Deep Learning Networks.* 1-10. 10.1145/3388440.3412486.

[4] https://childrenscancertherapydeve.app.box.com/s/f82oib8p2yzmvdco3m9glxvl8qamu6eq

[5] Johnson, Justin & Khoshgoftaar, Taghi. (2019). *Survey on deep learning with class imbalance.* Journal of Big Data. 6. 27. 10.1186/s40537-019-0192-5.

[6] Noorbakhsh, Javad & Farahmand, Saman & Soltanieh-ha, Mohammad & Namburi, Sandeep & Zarringhalam, Kourosh & Chuang, Jeffrey. (2019). *Pan-cancer classifications of tumor histological images using deep learning.* 10.1101/715656.

[7] https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga

[8] Litjens, G., Kooi, T., Bejnordi, B., Adiyoso Setio, A., Ciompi, F., Ghafoorian, M., Van Der Laak, J., Ginneken, B. & Sanchez, C. 2017, 'A survery on deep learning in medical image analysis', *ELSEVIER*, vol. 42, pp. 60-88.

[9] Raghu, Maithra & Zhang, Chiyuan & Kleinberg, Jon & Bengio, Samy. (2019).*Transfusion: Understanding Transfer Learning with Applications to Medical Imaging.*

[10] Kazeminia, Salome & Baur, Christoph & Kuijper, Arjan & Ginneken, Bram & Navab, Nassir & Albarqouni, Shadi & Mukhopadhyay, Anirban. (2020). *GANs for Medical Image Analysis. Artificial Intelligence in Medicine.* 109. 101938. 10.1016/j.artmed.2020.101938.

[11] Yoon, Hongjun & Lee, Joohyung & Oh, Ji Eun & Kim, Hong & Lee, Seonhye & Chang, Hee Jin & Sohn, Dae. (2018). *Tumor Identification in Colorectal Histology Images Using a Convolutional Neural Network. Journal of Digital Imaging.* 32. 10.1007/s10278-018-0112-9.

[12] Tschuchnig, Maximilian Ernst & Oostingh, Gertie & Gadermayr, Michael. (2020).*Generative Adversarial Networks in Digital Pathology: A Survey on Trends and Future Potential.* Patterns. 1. 10.1016/j.patter.2020.100089.

[13] Tharwat, Alaa. (2018). *Classification Assessment Methods: a detailed tutorial.* 10.1016/j.aci.2018.08.003.

[14] Wang, Yunjun & Guan, Qing & Lao, Iweng & Wang, Li & Wu, Yi & Li, Duanshu & Ji, Qinghai & Wang, Yu & Zhu, Yongxue & Lu, Hongtao & Xiang, Jun. (2019). *Using deep convolutional neural networks for multi-classification of thyroid tumor by histopathology: a large-scale pilot study.* Annals of Translational Medicine. 7. 468-468. 10.21037/atm.2019.08.54.

[15] Xie, Juanying & Liu, Ran & Luttrell, Joseph & Zhang, Chaoyang. (2019).*Deep Learning Based Analysis of Histopathological Images of Breast Cancer.* Frontiers in Genetics. 10. 10.3389/fgene.2019.00080.

[16] Szegedy, Christian & Vanhoucke, Vincent & Ioffe, Sergey & Shlens, Jon & Wojna, ZB. (2016). *Rethinking the Inception Architecture for Computer Vision.* 10.1109/CVPR.2016.308.

[17] Lydia, Agnes & Francis, Sagayaraj. (2020). *Multi-Label Classification using Deep Convolutional Neural Network.* 10.1109/ICITIIT49094.2020.9071539.

[18] Ho, Yaoshiang & Wookey, Samuel. (2019). *The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling.* IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2962617.