One Year Confirmation Report

By Peter Donnelly

WORD COUNT 8564
TIME SUBMITTED 05-NOV-2020 11:34AM
PAPER ID 64757536

STUDENT NAME: Peter Donnelly	
MEETING DATE: 8 Nov 2020	**VENUE**: Zoom
COMMITTEE: 4&12	
SUPERVISORS: Dr. Ken Doig and Dr. Thomas Conway	
MENTOR: Dr. Huiling Xu	
LENGTH OF CANDIDATURE TO DATE: 14 months	
PROJECT TITLE: Improving Classification of T-Cell Lymphomas By Applying machine learning to Heterogeneous Data Types	

Name of candidate: Peter Donnelly

Student ID Number: 1059982

Field of Research Codes: 111202, 111203 , 111207

Title of the thesis

"Improving Classification of T-Cell Lymphomas by Applying machine learning to Heterogeneous Data Types"

Abstract

Some kinds of cancers are difficult to sub-type using histopathology images alone.

It is proposed that a deep learning network trained with matched but heterogeneous sample types, namely (1) high resolution scans of histopathology tissue samples, and (2) gene expression data derived from RNA sequencing, may be an effective tool for automatically sub-typing cancers generally, and difficult to classify cancers such as certain T-Cell Lymphoma sub-types in particular.

The proposition will be validated or rejected by conducting experiments on matched WSI and RNA-Seq data using appropriate neural network algorithms and models.

The development and delivery of a suitable experiment system is a key thesis deliverable.

Clinical Motivation

Early and accurate cancer detection and classification is highly consequential: early detection allows treatment to commence before the disease has progressed too far, while accurate classification is the most important determinant of treatment programs and ultimately therefore, patient outcomes.

Perhaps surprisingly - despite the many breakthrough in genomics of the past 20 year or so - cancers are still mainly detected and classified by a pathologist looking at stained biopsy tissue under a microscope. No doubt this is the case because histopathology is effective, relatively cheap, and has a 125 year head start on genomic methods.

Nonetheless, pathologists find some types of cancers particularly difficult to classify from histology images alone: these include, but are not limited to: pancreatic cancer, cancer of unknown primary and lymphoma.

With reference to Figure 1 below, T-Cell Lymphoma has a large number of sub-types, and they're not all easy to differentiate by cell morphology alone. ALK-positive and ALK-negative Anaplastic Large Cell Lymphoma are almost indistinguishable on H&E biopsy slides: indeed the study that Figure is based onfound that inter-pathologist agreement in classifying ALK negative was only 74%. However the two types of Lymphoma have markedly different survival rates and require different treatment regimes, reinforcing that accurate cancer classification is consequential.

Figure 1: example of cancer which is difficult to classify from cell morphology alone

Case for the Proposition of this Thesis

Machine learning has been successful in accomplishing tasks that centre on extracting complex patterns from biological data of arbitrary modality or combinations of modality, including cancer detection and classification.

Many researchers have used machine learning of one kind or another to classify or sub-type cancers using *either* histopathology images or omics data.

It is proposed that a multi-modal machine learning network trained with *matched* Whole Slide Image and RNA-Seq data type is likely to be more effective at classifying cancers than one trained on either alone.

The proposition that a multi-modal approach may out-perform single mode machine learning in the classifying cancers rests chiefly on the assumption that WSI image data and RNA-Seq data each possesses at least some 'classification relevant' information that the other data modality does not. Cell morphology observable in pathology images must embody an integrated but imperfect synthesis of all underlying genomic information; whereas gene expression data should be a 'more perfect' representation of a subset of genomic information: for example methylation data clearly contains information that RNA-Seq data does not, and that information clearly may affect cell morphology.

Hypothesis

A classifier trained on BOTH images AND RNA-Seq will be able to:

(i) classify any cancer; and,

(ii) classify challenging cancers,

better than image-specific or RNA-Seq specific classifiers

Thesis Objectives

Aim 1 (95% complete as at 1 Nov 2020)

- Establish a baseline experiment system using mainstream machine learning software, techniques and models, and use to test the hypothesis on publicly available (TCGA) data, including:
  - ↘ develop an experiment system capable of testing the hypothesis
  - ↘ locate, acquire and pre-process suitable matched experimental data
  - ↘ validate by classifying Whole Slide Image data (and establish a baseline for Aim 2)
  - ↘ validate by classifying gene expression data (and establish a baseline for Aim 2)
  - ↘ test multi-modally with matched Whole Slide Image + gene expression data

Aim 2

- Test the first part of the hypothesis, viz.

  "a classifier trained on BOTH images AND RNA-Seq will classify and subtype any cancer better than existing Image specific or RNA-Seq specific classifiers"

- Method: Using the experiment system and classifiers developed in Aim 1:
  - ↘ classify as many matched TCGA cancer types and cases as is (practically) possible
  - ↘ compare performance to Aim 1 Image-only & RNA-Seq-only classifiers' performance
  - ↘ compare performance to best performing classifiers in the literature
- Metrics: Accuracy, Precision, Recall, F1, Area Under Curve

Aim 3

- Test the second part of the hypothesis, viz.

  "A classifier trained on BOTH images AND RNA-Seq will classify T-Cell Lymphoma better than any existing classifier"

- Method: Using the experiment system and classifiers developed in Aim 1:
  - ↘ Classify 200 matched T-Cell Lymphoma samples provided by PMCC molecular haematology laboratory
  - ↘ Compare performance to best performing classifiers in the literature
- Metrics: Accuracy, Precision, Recall, F1, Area Under Curve

Literature Review

Histopathology and Molecular Science: Worlds Apart

Pathology images and genomics data embody fundamental biological information which is routinely used to detect and diagnose cancer and other diseases. The *interpretation* of pathology images and molecular pathology assays is a knowledge-based, but subjective process conducted by highly trained doctors and scientists, who nonetheless have finite cognitive powers, infallible memories, varying skill and experience levels, and a time-limited career over which they are exposed to only a tiny subset of the infinitude of normal and abnormal tissue types and genomes. Further, despite the fact that each examines a different aspect of the same disease (morphology and genome) for the same purpose (diagnosis), pathologists and molecular biologists live in very different worlds; each with their own training, career paths and culture; and very rarely cross paths operationally.

These observations would be of little moment were it not for the fact that machine learning generally, and deep learning in particular, portend a future in which algorithms can seamlessly bridge these two worlds. To a deep learning algorithm, Whole Slide Images, RNA-Seq FPKM UQ values and Methylation beta values (etc) are all just numbers in memory. There is no theoretical upper limit on the number and kind of biological samples that can be incorporated into a pattern recognition network; the learning process is accretive over an arbitrarily long time span; and the algorithms never get tired or have off days.

Cancer Classification

Cancer classification taxonomies([i]) are the historical product of decades of effort on the part of oncolologists and medical scientists([ii])([iii]). Messy and always contingent([iv]), they are subject to change and refinement as and when new scientific evidence emerges which improves, refines or refines an existing classification([v]).

Until recently, cancer classification was based on the visual evidence of histopathological analysis, but in the past two decades genomics evidence has come to play an important role in cancer classification([vi])([vii])([viii]) ([ix])([x])([xi])([xii]), in a way similar to the way DNA barcoding influences the morphology oriented Linnaean biological classification system([xiii]).

Difficult to Classify Cancers

Pathologists find some kinds of cancers particularly challenging to classify([xiv])([xv])([xvi]). Often only molecular testing can provide the additional information necessary to differentiate morphologically near identical subtypes. Lymphoma provides multiple examples of this: while many B-Cell Lymphomas and Non Hodgkins Lymphomas are straightforward to classify histologically, some of the large number of T-Cell Lymphoma sub-types are notoriously difficult to classify this way. As one example, the ALK-negative([xvii]) and ALK-positive([xviii])([xix]) sub-types of Anaplastic Large Cell Lymphoma([xx])([xxi]) have almost identical histomorphologies, and essentially cannot be differentiated visually([xxii])([xxiii]). This is consequential, since the ALK-positive sub-type has a much better prognosis than the ALK-negative sub-type; the optimum

treatment for each differs: the latter requiring more aggressive treatment; and further ALK-negative sub-type is much more susceptible to relapse.

Deep Learning in Medicine

Deep Learning (machine learning)([xxiv]) ([xxv]), a sub-category of machine learning named for its use of many layered Neural Networks, has made rapid progress in a wide variety of fields since about 2010, and is on the cusp of becoming clinically relevant in multiple medical fields([xxvi])([xxvii]), very much including Pathology([xxviii])and Oncology([xxix])([xxx])([xxxi]).

Medical applications of Deep Learning can be divided into those focusing on research topics and those with direct clinical aims. Deep Learning has to date had much less impact in the clinical domain than it has in research domains (acknowledging that the former is inextricably and causally intertwined with the latter). Notable exceptions include the diagnosis of skin cancer([xxxii]) and eye disease, where the accuracy of Deep Learning is in some cases on a par with human diagnosticians([xxxiii])([xxxiv]).

Deep Learning in Pathology and Oncology

During the past five years in particular, researchers have applied Deep Learning in a variety of novel ways to multiple cancer artifacts including Whole Slide Images([xxxv]) and all kinds of omics data, with diverse objectives and varying degrees of success([xxxvi]). These objectives include, but are not limited to: prediction of cancer driver genes([xxxvii]); identifying cancer associated signalling pathways(ref); predicting gene expression levels from histopathology images([xxxviii]) ([xxxix])([xl]), recapitulating cell morphology from gene expression data([xli]), improve variant calling([xlii]); cancer patient prognosis forecasting([xliii]); quantifying and segmenting Tumor Infiltrating Lymphocytes([xliv])([xlv]); and classifying cancer types and sub-types, of which multiple examples will be discussed in subsequent sections.

Reflecting the immature status of medical applications of machine learning, only a small number of papers specifically address methods and techniques([xlvi])([xlvii])(ref)([xlviii]).

Deep Learning using Whole Slide Images

Many researchers have applied machine learning to histopathology images, inspired perhaps by the success of machine learning in image recognition, classification and segmentation in other domains; and facilitated by the widespread availability of open source implementations of effective machine learning algorithms, including VGG([xlix]), Resnet ([l]) and GoogLeNet/Inception ([li]), and machine learning frameworks, such as Tensorflow([lii]), PyTorch([liii]) and Caffe([liv]).

A typical project of this type uses one of the above mentioned deep learning models to detect cancer and/or classify cancer type or subtypes from sets of very high resolution Whole Slide Images (WSI) scans of Haemotoxylin and Eosin (H&E) stained tissue samples or Core Needle Biopsy (CNB) samples. In most cases, detection and classification tackle focus on one or a small number of cancer types and sub-types, however some projects take the opposite tack and may attempt to classify 20 or more cancer types/sub-types)

Project scopes are frequently defined or constrained by the availability of suitable WSI data, very often making use of the TCGA([lv]), or else take place in the context of a 'Grand Challenge'([lvi]) such as Camelyon17 ([lvii]), or ICIAR/BACH([lviii]) where curated data is made available to participants, and frequently left in place for future researchers to use.

Most but not by no means all projects use publicly available data. Most but not all use slide level truth labels. A small proportion engage pathologists to hand curate slides or tiles (portions of slides). While the latter approach should in principle lead to higher quality outputs, it is also costly and labour intensive, which may be why it is not seen more.

The following paragraphs survey contemporary applications of Deep Learning to Whole Slide Image data:

Campanella et al ([lix]) (2019) classify four types of cancer from H&E stained WSI image samples using slide-level labels (n=17,661) with a ResNet 34 based machine learning system, claiming an accuracy of >98% for all types. Note that this is classification of cancer types, rather than the more challenging problem of classifying sub-types of a particular cancer.

Coudray et al ([lx]) (2018) classify H&E stained WSI samples (n=1634) into either normal, lung adenocarcinoma (LUAD) and squamous cell carcinoma (LUSC) with an InceptionV3 based machine learning system, achieving an AUC of 0.97, which they claim is comparable to the performance of human pathologists.
This experiment encompasses both cancer detection and sub-typing; albeit relatively simple sub-typing.

Brancati et al ([lxi]) (2019) classify three types of invasive ductal carcinomas: Chronic lymphocytic leukemia (CLL); mantle cell lymphoma (MCL) and Follicular Lymphoma (FL); from H&E stained breast tissue samples (n=162) using FusionNet (a type of Residual Convolutional Neural Network); with a claimed accuracy of 98%. This experiment encompasses classifying three distinct B-Cell Lymphoma sub-types.

Le et al ([lxii])(2020) (I) detected and segmented breast cancer tumour regions and (ii) classified and quantified tumour infiltrating lymphocytes in H&E stained breast cancer WSI samples (n=1015) using FusionNet, a type of Residual Convolutional Neural Network.
This experimenter encompasses TIL quantification and visualization in addition to cancer detection (but `does not include classification).

Pantanowitz et al([lxiii]) (2020) detected prostate cancer from H&E stained and core needle biopsy WSI images (n=1627) in the context of a clinical workflow using a Convolutional Neural Network, with a headline claimed AUC of 0.997, and including one case which human pathologists had missed.

Deng at al (2020)(ref) and Komura and Ishikawa (ref) (2018) are rare and valuable examples of review paper on machine learning techniques for histopathological image analysis.

Deep Learning using Omics Data

In presenting a successful method of classifying childhood small-round-blue-cell tumors (SRBCTs) from gene expression signatures [1]using a neural network classifier[2], the highly cited project of Khan et al[lxiv] (2001) must be counted among the very first to have applied a neural network to gene expression data[3].

Applying similar techniques to a slightly different problem, Aliferis et al([lxv]) (2003) used a neural network to detect, classify and subtype (2 subtypes) lung cancer from micro-array gene expression data (12,600 gene expression values). Undoubtedly because of 2003 hardware limitations, they found the neural network classifier to be impractical unless used with small subsets of the dataset. Nonetheless, their approach (three layer fully connected network, gradient descent with momentum, in-training validation testing) is essentially the same as contemporary approaches; the only real difference being that vastly more powerful hardware is now readily available.

After 2003, we see many similar papers. A Google Scholar search reveals 19 further papers with the phrases "cancer classification" and"gene expression" and either "deep learning" or "neural networks" in their title.

Typical of these, Gao et all ([lxvi]) (2019) classified five different PAM50(44) defined breast cancer sub-types using a Fully Connected Network trained with the TCGA RNA-Seq data enriched with the Molecular Signatures Database (MsigDB) ([lxvii]), with a claimed accuracy of 80% (n=456). Lee et all ([lxviii]) (2020) sub-typed five different types of cancer using TCGA RNA-Seq data and the KEGG ([lxix]) pathway database using a Graph Convolutional Network achieving classification accuracy of 91.5% for STAD (four sub-types) and a little less for the other three cancers. Additionally, their innovative use of a pathway database allowed them to add an element of explanation to the sub-typing. Luo et al (*ref*) predict the top n driver colorectal and breast cancer driver genes using a one dimensional Convolutional Neural Network and compare these predictions with those of the literature, with an AUC of 0.984, claimed to be 15% higher than the best competing algorithm. Al Mamun and Al Mamun ([lxx]) (2019) classified eight cancer types using Long Non-coding RNA Data (RNA-Seq) from UCSC Xena ([lxxi]) TCGA sed four kinds of Neural Networks (Fully Connected, Convolutional, Long Short Term Memory and a Deep Autoencoder, achieving accuracy of 94-98%. They noted that lncRNA data appears to be superior in mRNA data for the classification task [a proposition my thesis will be readily able to test].

With multiple kinds of omics data available, omics machine learning research objectives are unsurprisingly more diverse([lxxii]) than is the case for WSI. Levy et all ([lxxiii]) (2020) predicted cancer type/sub-type, age and smoking status (inter alia) from essentially all TCGA methylation data using a comprehensive Deep Learning system ('MethylNet') which is based on Fully Connected Networks and Variational Autoencoders. It achieved 97% accuracy in classifying a mix of 32 cancer types and subtypes, and additionally 95% accuracy (n=1018) for PAM50 classification of breast cancer sub-types.

---

1 cDNA micro-array data: 63 samples x 6,567 gene expression values

2 front-ended by a PCA dimensionality reducer

3 in addition to NNs, they used SVMs and KNNs.

Multi-modal Learning using both Image and Omics Data

'Multi-modal Learning' here refers learning from the example-level integration of than one kind of input data, for example; WSI data+RNA-Seq, or WSI data+methylation data, WSI data+RNA-Seq+methylation data. Specifically, outputs derive from *features* learned from the individual learning modes, rather than directly from input data.

There are currently few examples of Multimodal Learning in Pathology/Oncology, possibly because the research community tends to cleave between microscope/image oriented and genomics oriented. Notwithstanding, this small group includes some ambitious and innovative projects.

The following paragraphs survey contemporary applications of Multimodal Deep Learning to multiple data modalities:

Gundersen et al ([lxxiv]) (2019) use a combination of a Deep Convolutional Generative Network (DCGAN), Fully Connected Network (FCN) and Canonical Correlation Analysis (CCA) to extract mode specific and common mode information from image and RNA-Seq data using NIH's GTEx V6 dataset ([lxxv]).The common mode information is used to recapitulate approximate versions of both the original images and expression data (primary objective); the trained model is also used to classify the 50 types of human tissue (n=2221) represented in the GTEx V6 dataset from common mode signal alone, and achieves p<=0.05 for 13 of the 50 tissue types. Although it has significantly different objectives, the Gundersen system besides the CCA component has many technical similarities with the system required for our experiments. The core learning engine of our system is an adaptation and extension of the Gundersen system.

Carmichael et al ([lxxvi]) (2019) use the combination of a Deep Convolutional Network and a statistical technique analogous to CCA called AJIVE ([lxxvii]) to to extract, analyse and explain biologically features from Multi-modal data drawn from two sources: (i) image and gene expression data from the Carolina Breast Cancer Study (CBSC) (n=1191); and, (ii) four omics data types from the TCGA Breast Cancer data dataset (n=616).

Ash et al ([lxxviii]) (2018) use a combination of a Convolutional Autoencoders and Sparse Canonical Correlation Analysis to discover associations between sets of genes and physical cellular features/attributes; using image and gene expression data from two TCGA datasets and the GTEx dataset.

Cheerla et al ([lxxix]) (2019) predict patients survival rates for 20 cancer types and 11,160 patients using TCGA data by using a customized Multi-modal Neural Network incorporating CNN, and FCNs ,to three data modalities: Whole Slide Images data, RNA-Seq data and clinical data with an overall AUC of 0.78.

Islam ([lxxx]) et al perform PAM50 classification of breast cancer subtypes using two kinds of omics data, viz: RNA-Seq and copy number alteration (n=1925), using a Deep Convolutional Neural Network, achieving an AUC of 0.83.

Algorithmic Classifiers

While there are many good classification algorithms (SVN, Random Forest, K Nearest Neighbours etc), supervised neural networks make particularly multi-class effective classifiers([lxxxi])([lxxxii]); Indeed, supervised Neural Network are currently 'best-in-class' for classification tasks. A further advantage is that they operate essentially identically regardless of whether the input is Image or RNA-Seq or some other kind of data (e.g. Methylation).

Supervised Learning

Supervised learning algorithms have a distinct *training* phase, during which the algorithm learns from a set of examples for which the true class labels are known.

With any kind of supervised neural algorithm:

- experts (e.g. pathologists) supply a 'Training Set' of correctly labelled examples;
- the algorithm iteratively learns from this Training Set, resulting in a model which, hopefully;
- generalizes sufficiently well that it is able to classify new, hitherto unseen examples

Neural Network Classifiers

Figure 2 illustrates a supervised neural network configured as a classifier:

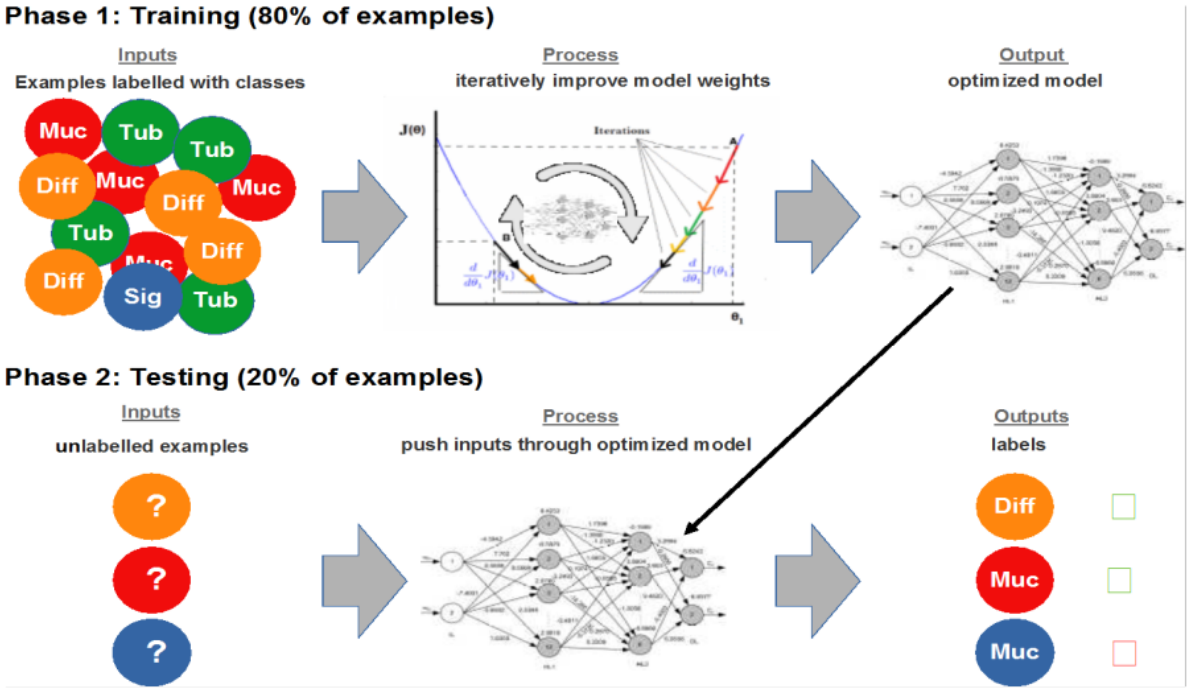


Figure 2: Supervised neural network configured as a classifier

In operation, the network is trained using a set of examples for which the true class is known – here, the 'Tubular', 'Mucinous', 'Diffuse' and 'Signet Ring' subtypes of Stomach Adenocarcinoma - and when trained, challenged to produce the correct classification for examples it has never seen before ('held-out' examples)

During *Training*, shown at top, the network is presented with examples for which the correct subtype is known. After each round of training, the network's weights are updated in a manner which minimizes an error function; and the network is then asked (*Testing)* to predict the subtype of some randomly selected held out examples, as shown in the bottom half of the image.

And this process is repeated over and over

After each training iteration, the network (hopefully) gets better and better at classifying held-out examples, until eventually it will start to get worse. At that point we have our optimised model, or at least as optimised as it's going to get, which is saved and used to classify future examples which the network has never seen before, and for which the true classification is unknown.

With reference now to Figure 3 below, a neural network comprises an arbitrary number of 'layers' with each layer comprising of a number of 'neurons'. Each of the *n* neurons in a layer is 'connected' to *every* neuron in the layer to its left, or to the input *p* in the case of the first layer.

Figure 3: Generalized neural network

Every neuron implements a simple arithmetic task: it multiplies the output of each neuron in the layer to its left by a weight associated with the edge (w_{ij}) connecting it to that neuron; sums these values; passes the sum through a non-linear, so-called *activation function f()* - often a rectified linear unit - producing an output which is used in identical fashion by the neurons of the layer to its right.

Input *p is* a digital representation of a single image: for example, a 64 x 64 colour image is encoded as an integer vector of length 64 x 64 x 3 = 12,288, with each integer taking on a value in the range 0-255[4] corresponding to the intensity of one of the three colors describing a single pixel.

In the case of a *multi-class* classifier - the type we will use for cancer classification and subtyping - the final layer uses a *softmax* function as its activation function. This converts the sum-of-products values (Wp+b) into a set of values in the range (0:1) which add up to 1, and these values are interpreted as probabilities. The predicted class is then the class corresponding to the output neuron which produces the highest probability.

A neural network's ability to 'learn' – i.e iteratively converge on a set of output class labels that closely correspond to the true Training Set class labels - derives from the mechanism used to update the network's weights, such that each successive iteration on average predicts a larger number of correct labels than its predecessor. This mechanism has two parts: an optimiser and a loss function.

The *loss function* quantifies the difference between predicted and true class labels. It is cleverly defined so that no matter how parameterised, it is always convex, that is, has just one minimum. If this were not

4 In practice, these integer values are inevitably scaled and normalised prior to ingestion by the network.

the case, any algorithm using it would inevitably get trapped in one of the many local minima which abound in high dimensional function spaces.

The *optimiser* is always a form of gradient descent, and optimisation works as follows. After a set of inputs have been propagated through the network, the loss (function) is differentiated with respect to the inputs using the chain rule, thereby determining the direction of the minimum value of the loss function for each weight in the network. The weight parameters are then updated by an amount that represents a small step in the direction of the minimum. The size of the step is determined by a user provided *learning rate*, which specifies the proportion of the differential to use when updating each weight.

The overall objective is always to obtain the highest number of correct predictions for *unseen* examples, as represented by held-out examples the Test Set. A neural network can typically achieve near perfect predictions for training data, but such a model will be highly over-fitted, and will normally not generalize well to examples outside the Training Set. Achieving this requires a combination of well chosen hyperparameters and judgment.

It's self evident but nonetheless worth mentioning that if the neural network is to successfully perform its task, the training examples and the class labels must actually be related by *some* underlying correlation – perhaps a complex and highly non-linear function; and further, the training data must be a true representation of the unseen examples that the network is being expected to classify.

In brief summary:

The network is initialised with random weights (wi,i)

- a batch of inputs (p) from the Training Set is propagated through the network
- a loss is calculated for the batch: true 'minus' predicted classes
- the chain rule is applied to the loss, and all weights updated
- try again with new weights

Periodically validate with examples randomly drawn from the Test Set.

Drawbacks of Neural Networks, and Supervised Classifiers Generally

A first drawback of Deep Learning Based classifiers compared to some other forms of classifier is that the models produced are not interpretable. They may well produce extremely good classifications, but one is not able to say *why* a particular classification was achieved. We would very much like to be able to look at the parameters or sub-units of a model and see, for example, that it corresponds to a particular morphological structure (like tubules or pappillae), but this is not currently possible2.

Particularly in the realm of medicine, neural network based tools are best viewed as decision support tools for human experts. They nonetheless have the potential to make the job of a pathologist much easier, by for example providing them with preliminary and potentially highly accurate preliminary assessments of samples for in depth review; and to improve the quality of their assessments: neural networks don't get tired, even after looking at millions of tiles; or have 'bad days'.

A second drawback is that *supervised* classifiers of any type can only, at best, be as good as the 'truth set' used to train them: logically, they cannot achieve a level of classification accuracy higher than that of the humans who provided the labelled training set.

This is a practical problem for this thesis, since it is focused on difficult to subtype cancers, and there is no general remedy for this problem. If we assume, however, that for difficult to subtype cancers, there is a higher degree of variance between highly experienced experts and their more junior colleagues; and between practitioners at research hospitals like PeterMac who are exposed to a higher volume of rare cancers and those at other institutions who are not; then it is still possible for supervised classifiers to make a positive contribution, *provided* they trained on well labelled data from the highly skilled practitioners. In this case they can assist other practitioners who do not have the skill level, or who do not regularly see difficult cancers of the type the network is trained on.

Assessing Classifier Effectiveness

Metrics for measuring and comparing classifier performance are well defined and universally agreed; and they centre on an object called a "Confusion Matrix" and closely associated set of statistics.

A Confusion Matrix is always used in conjunction with one or more of the following statistics: Accuracy, Precision, Recall and F1 Score, which in turn fall straight out of the Confusion Matrix. The most important ones for our purposes are Precision, which indicates how good the model is at not generating false positives; Recall, which indicates how good the model is at not generating false negatives, and F1 score, which is the geometric mean of Precision and Recall, and which puts the overall performance into a single number.

Sources of Experimental Data

Although the objective is to perform experiments on PMCC's own image and genomic data, it is unrealistic to assume that these will be available at an early date in the quantities required. It was therefore necessary to locate sources of public data.

After extensive searching, it became apparent that the only suitable public sources of matched WSI and omics data are (I) the NIH's TCGA and (ii) NIH's GTEx data repositories; and of these two, only NIH TCGA has the volumes and types of data required. Ostensibly there are many other public slide and omics data repositories world wide, but none have matched image and omics data; and further many 'large data sets' turn out to be highly heterogeneous, fragmented collections of small data sets which are often also unstructured or poorly curated.

More positively, the TCGA data repository, an intentional by-product of the decade long Pan Cancer project([lxxxiii]) is *highly* suitable. It contains an enormous volume of uniformly structured and well curated cancer data, and a much smaller but nonetheless sufficient volume of case matched Whole Slide Image data.

Preliminary Experiments (Aim 1)

Purpose

- To validate that the experiment system and classifiers work properly
- To establish a baseline for Aim 2 experiments

Scope

- Use the experiment system developed in Aim 1 to classify TCGA Adenocarcinoma samples into the six TCGA defined subtypes
  - **Whole Slide Image Model**
    - ↘ **228** TCGA STAD **H&E tissue** samples
    - ↘ Each sample → 2500 randomly selected 64x64 pixel tiles
    - ↘ 5,000 randomly selected held-out tiles tested after each of 4 runs = 20,000 total predictions
    - ↘ Neural Network: VGG11 = 11 layer convolutional network; ADAM optimiser; 80% training set + 20% held-out
  - **RNA-Seq Model**
    - ↘ **479** TCGA STAD **mRNA + lnRNA** samples
    - ↘ Each sample → 60,483 Upper Quartile FPKM values
    - ↘ All held-out samples tested after each run yielding 96 x 101 = 9,696 total predictions
    - ↘ Neural Network: 3 layer fully connected network (relu + softmax activation); ADAM optimiser; 80% training + 20% held-out
  - **Dual Mode Model (not yet done – experiment system will support dual mode in Ded 2021)**
    - ↘ **228** case-level matched STAD samples
    - ↘ Each image → 2500 tiles; each tile → embedding using CNN or autoencoder
    - ↘ Each RNA-Seq sample reduced to an embedding using either an FCN or autoencoder
    - ↘ Neural Network: fully connected network (relu + softmax activation); ADAM optimiser; 80% training + 20% held-out

Data

TCGA STAD is used as a reference dataset for the preliminary experiments because:

a) it contains enough subtypes (seven)[5] to make the task of classification a meaningful one

5 One (stomach adenocarcinoma - intestinal adenocarcinoma - papillary type) are represented in very small numbers (8) and will not be used in experiments

b) it contains enough pairs (219) of matched WSI and RNA-Seq data to make multimodal training feasible

c) it contains enough examples of five of the six subtypes to make extraction of a balanced subset feasible[6]

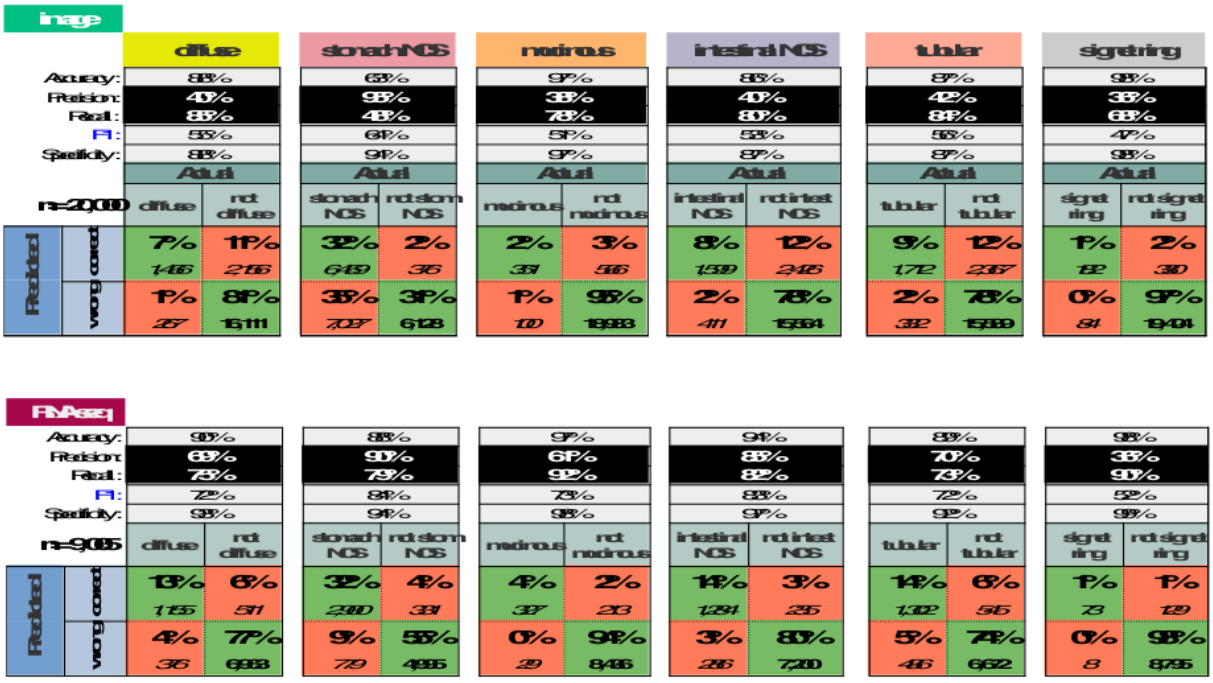
d) background tissue is relatively homogeneous, providing much less of a shortcut clue to cancer subtype compared to say, the TCGA Sarcoma dataset

The six STAD subtypes represented in the samples are:

- stomach adenocarcimoa – diffuse type
- stomach adenocarcinoma - NOS
- intestinal adenocarcinoma – mucinous type
- intestinal adenocarcinoma – tubular type
- intestinal adenocarcinoma – papillary type (*not used – too few examples)*
- intestinal adenocarcinoma - NOS
- signet ring type adenocarcinoma

6 443 cases (228 + 479 samples) made up as follows: 72 x stomach adenocarcinoma diffuse type; 160 x stomach adenocarcinoma NOS; 21 x mucinous type intestinal adenocarcinoma; 82 x intestinal adenocarcinoma NOS; 79 x tubular type intestinal adenocarcinoma ; 13 x signet ring type stomach adenocarcinoma; 8 x papillary type intestinal adenocarcinoma; 3 x degenerate/unknown. The proportions of each class vary for the image and the RNA-Seq subset.

Summary of Results



Confusion Matrices

Observations

- Results appear to be quite good (Accuracy), but this is superficial (Precision, Recall)
  - ↘ Accuracy: 80-90%-ish for both Image and RNA-Seq
  - ↘ Precision: 36-95% for image and 36-90% for RNA-Seq
  - ↘ Recall: 48-85% for image and 73-92% for RNA-Seq
- RNA-Seq model performed better and more consistently than Image model
- Performance varies greatly between subtypes, probably indicating some are easier to classify than others

Conclusions

- The system is validated (Aim 1 Objective)

- Results are typical of what others achieve with single mode classifiers (Aim 1 Objective)
- Constitute a benchmark to used when testing the hypothesis (Aim 1 Objective)

Summary of Progress to Date

Summary

Aim 1 - development and validation of a suitable experiment system - is 95% complete and will be 100% complete by 20 Dec 2020.

- The experiment system is 'all but' complete and is documented elsewhere herein
- Validation testing with *image data* has been completed, as documented elsewhere herein
- Validation testing with *gene expression* data is complete, as documented elsewhere herein
- Base-lining with dual mode WSI + RNA-Seq data has not commenced

Aim 2 and Aim 3 have not yet commenced

Notes:

1. Aim 1 is by far the largest activity in the thesis, in terms of both degree of difficulty and time required
2. Aims 2 and 3 centre on conducting and documenting a series of experiments using the system developed in Aim 1 (testing the hypothesis occurs only in Aims 2 and 3)
3. Aim 1 is on the critical path for Aim 2 and Aim 3. Mitigating this risk, the experiment system has been tested and validated throughout development, such that the evolving classifiers have already been used hundreds of times on many hundreds of stomach cancer, sarcoma and lymphoma examples.

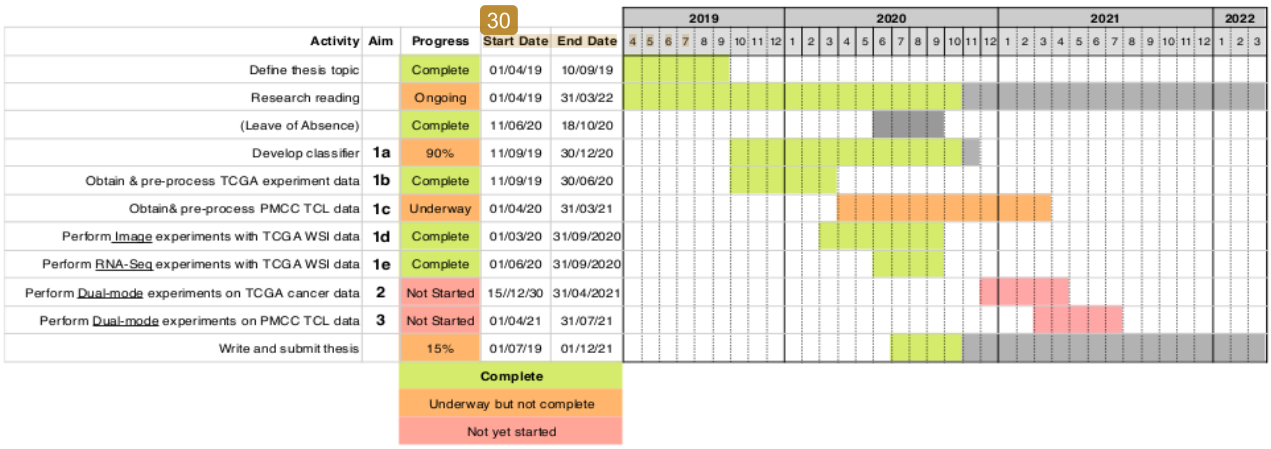
Experiment System

The experiment system I have developed over the past 12 months:

- is a GNU/Linux application suite comprising 48 python files, 13 shell files, and multiple application specific libraries, including the PyTorch machine learning library
- is currently capable of currently of classifying/sub-typing cancers using either whole slide image data or gene expression data
- is an 'end-to-end' system. The classification sub-system is just one component. Data acquisition and pre-processing aspects are largely automated. (These facets are otherwise labour intensive and error prone)

- is able to dispatch & run multiple experiment jobs without user intervention
- produces scalar and graphical results, including spreadsheets and visual charts in real-time
- can run on a standalone compute; in the cloud; or on PeterMac's Machine Learning cluster
- used the following hardware setup for development: 16 x 3.5GHz core AMD Ryzen CPU, 132GB RAM a2 SSD; dual Nvidia Titan GPUs each with 576 tensor cores and 24GB memory

Research Plan



Activity	Aim	Progress	Start Date	End Date
Define thesis topic		Complete	01/04/19	10/09/19
Research reading		Ongoing	01/04/19	31/03/22
(Leave of Absence)		Complete	11/06/20	18/10/20
Develop classifier	1a	90%	11/09/19	30/12/20
Obtain & pre-process TCGA experiment data	1b	Complete	11/09/19	30/06/20
Obtain& pre-process PMCC TCL data	1c	Underway	01/04/20	31/03/21
Perform Image experiments with TCGA WSI data	1d	Complete	01/03/20	31/09/2020
Perform RNA-Seq experiments with TCGA WSI data	1e	Complete	01/06/20	31/09/2020
Perform Dual-mode experiments on TCGA cancer data	2	Not Started	15//12/30	31/04/2021
Perform Dual-mode experiments on PMCC TCL data	3	Not Started	01/04/21	31/07/21
Write and submit thesis		15%	01/07/19	01/12/21

Complete
Underway but not complete
Not yet started

Research Outputs

- Jun 2020 presentation to Bioinformatics Seminar: “Notes on Developing a Multimodal Deep Learning Network to Classify Cancers using Whole Slide Image and Gene Expression Data
- Jan 2020 presentation to Bioinformatics Journal Club on paper: “End-to-end training of deep probabilistic CCA forjoint modeling of paired biomedical observations”
- Jun 2019 presentation to PeterMac machine learning Seminar on paper: “Text-mining clinically relevant cancer biomarkers for curation into the CIViC database”
- Participation in VCCC Molecular Tumour Boards
- Co-organizer of / participation in Bioinformatics Journal Club
- Participation in PMCC Bioinformatics Seminars

Record of attendance

- 2019 Methods in Cancer seminar series
- 2020 Hallmarks of Cancer seminar series
- Post graduate research induction and orientation
- Completion of PMCC Occupational Health And Safety (BP&C) training
- Completion of PMCC Laboratory Safety Expiry training
- Completion of PMCC Emergency Procedures training

Appendix 1: Design of a Suitable Experiment System

Learnings from the Literature Review

These derive from a synthesis of the entire literature review rather than any individual paper; in some cases in combination with the candidate's technical computing know how.

High level Learnings (strategic / system / methods / techniques)

a) Use proven open source learning algorithms supported and maintained by the computer science community. Avoid 'novel' and 'bespoke' algorithms.

b) TCGA is the by far the best (and essentially the only) public source of a suitable volume of labeled multimodal case data.

c) Use Multiple Instance Learning and slide level labels rather than manual tile curation.

d) Data acquisition from TCGA is likely to be repetitive and time consuming, so it should be automated

e) Data pre-processing is likely to be complex and time consuming, so it should be automated

f) The native dimensionality of the TCGA RNA-Seq data is so high the a suitable means of dimensionality reduction may be required (e.g. Autoencoder or)

g) The TCGA RNA-Seq data includes gene expression values for both long non-coding RNA (lnRNA) in addition to protein coding mRNA. The experiment system should allow for the use of either of these sets, and additionally the subset of mRNA corresponding to PMCC gene panel(s) as well as subsets of genes which are already known to be implicated in particular cancers.

h) We hypothesize that the image branch will likely benefit from the ability to accept image tiles with multiple levels of magnification, so the experiment system should cater for this

i) Digital Stain Normalization may be required to compensate for batch variations in staining, so the experiment system should cater for this

j) The experiment system must be able to cater for both H&E stained tissue slides and aspirate smears

k) It is unknown whether data augmentation (e.g. hue variation) will be beneficial, but the experiment system should assume that it will be, and cater for it.

l) We suspect that Transfer Learning (eg. from ImageNet, which some of the projects use) is likely to be of zero or marginal benefit, so it need not be catered for in the first instance

Low Level (software, algorithm and choices of parameters)

a) Use Python/Pytorch rather than Tensorflow/Keras or Caffe as the machine learning engine

b) The ubiquity of Leica/Aperio slide scanning systems means that the system must cater for SVS image format in addition to the TIF image format

c) The image analysis branch should support at least the following network models: VGGnn, INCEPTION and RESNET

d) The rna analysis branch requires only the Fully Connected Network model (/ANN / 'Multi-Layer Perceptron'

e) The optimizer of choice is likely to be ADAM ('Adaptive Moment Estimation'), although others should be tried (eg. ADAGRAD, SGD ...)

f) The classification loss function will certainly be Cross Entropy

experiment system

Testing the hypothesis requires the development of a flexible experimental system capable of being trained multimodally from matched whole slide image and gene expression data. The development of such a system is an key deliverable of the thesis.

A single user defined experiment 'job' defines & runs multiple experiments, each with selectable combination of parameters/hyperparameters. Job level experiment definition has already shown itself to be a critical tool for experiment efficiency, consistency and repeatability.

The experiment system supports, via PyTorch, CUDA(82)/CUDNN(83) compliant GPUs. The system does not require GPUs to operate, but with the size and volumes of data being processed, GPUs are inevitably necessary. If more than one GPU is available, the system will take vantage of all GPUs and perform parallel processing (currently only implemented for Autoencoder mode, but will likely be implemented for image and RNA-Seq mode in due course). Additionally, the Data Analysis subsystem uses the CuPy(82) to take advantage of GPU processors and memory pooling when performing correlation and covariance analysis.

The system builds on and extends open source software made available by other researchers, but the greater proportion of it is the original work of the candidate.

Subsystems and Capabilities:

The experiment system comprises the following following functional subsystems/capabilities:

1 Biodata Acquisition Subsystem

Automatically acquires data from NIH GDC repository (via its RESTful(82)) API in accordance with user provided filters; for example the following command will download not more than 2000 matched STAD(83) image and RNA-Seq files then save them in the structured manner expected by the Pre-processing subsystem.

./gdc_fetch.py --case_filter="filters/TCGA-STAD_case_filter" --file_filter="filters/GLOBAL_file_filter_UQ" --max_cases=2000 --max_files=3 --global_max_downloads=2000 --output_dir=stad

2 Pre-processing Subsystem

- Comprising the following capabilities:

  a) rapid identification and removal of background, degenerate and low contrast tiles using stochastic algorithms

  b) extraction of tiles from each high resolution WSI image sample in accordance with user defined quality parameters

  c) (optional) stain colour normalization(82) ('reinhard(83)' or 'spcn(84)')

  d) (optional) gene set selection using ENSEMBL(82) or PMCC 'cancer genes of interest' or user defined custom gene set

  e) (optional) normalization of images and RNA-Seq data (JUST_SCALE/GAUSSIAN; LOG10PLUS1/ LOG2PLUS1

  f) final generation of Pytorch ready input dataset (including optional generation of matched image+RNA-Seq dataset)

3 Experiment Management Subsystem

- Takes the output of the Pre-processing subsystem and runs experiments on it according to user defined parameters
- Four modes: image, rna, image_rna and autoencoder
- Experiment job definition: arbitrary combinations of the following parameters (as an unrealistic but valid example, in RNA-Seq mode, the underlined parameters would launch a job comprising 2x2x2x2x2x3x2x3x2= 1,152 separate experiments, and the all experiments would then run sequentially without further user supervision)

  a) machine learning models (e.g. "DENSE CONV1D")

  b) optimization algorithms (e.g. "ADAM SGD")

  c) tiles per slide (e.g. "500 1000 2000")

  d) learning rates(s) (e.g. ".007 .001")

  e) number of samples to use (e.g. "100 200")

  f) mini-batch size(s) (e.g. "64 128")

  g) hidden layer neurons for Fully Connected models (e.g. "10000 90000 8000")

  h) topology definition for deep FCN models (one topology per job) (e.g. "3000 2000")

i) optional dropout regularization on layer 1 of DENSE models (e.g. “0.2 0.3”)

j) optional dropout regularization on layer 2 of DENSE models (e.g. “0.2 0.3”)

k) optional stain normalizations (e.g. “NONE REINHARD SPCN”)

l) optional data transformations (rotations, flips, greyscale …)

m) optional data normalizations (e.g. “NONE JUST_SCALE GAUSSIAN”)

n) optional data standardizations (e.g. “NONE LOG10PLUS1”)

o) optional noise and randomization data (for testing purposes: jitter, class randomization)

- Learning models

a) 3 x models & variants for image data ('VGG11/13/16/19', 'RESNET', INCEPTION V3')

b) 2 x models and variants for gene data (DENSE, 1D CNN)

c) 2 x models and variants for autoencoding (FC, DEEP, Variational)

4 Output Instrumentation Subsystem

a) real-time graphical view of results during training of training/test loss loss and accuracy curves

b) real-time console view of experiments as they are running

5 Output Visualization Subsystem

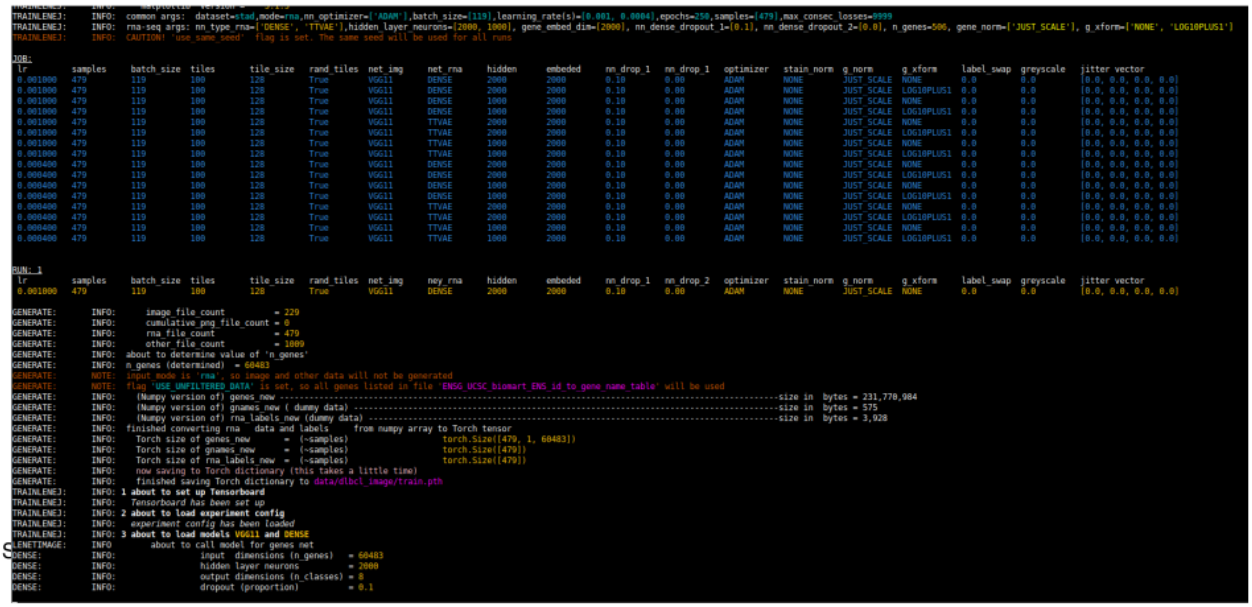
Real-time viewing modes are as follows:

a) test image tiles annotated with predicted labels vs ground truth labels (per batch)

b) test patches annotated with predictions and probabilities for each class

c) heatmaps showing confidence of correct/incorrect predictions

d) expression level correlation and covariance between genes for chosen gene set



6 Data Analysis

Performs correlation and clustering analyses on input gene sets, including at present, correlation, covariance, principle component analysis and K-means clustering and provides graphical visualizations of the outcomes. Used in conjunction with Autoencoder dimensionality reduction.

System Hardware

Most development is carried out on a high end 'MSI GS76 Stealth' laptop with a 6 core 12 thread Intel 9th Gen CPU, NVIDIA GeForce 2080 RTX GPU; 32GB RAM and a Samsung 1TB (soon 3TB) M.2 solid state drive.

Experiments are conducted on a water cooled tower PC equipped with a 16 core / 32 thread AMD Ryzen Threadripper 3950X; 2 x bridged NVIDIA TITAN RTX GPUs; 128GB RAM; a 2TB M.2solid state drive ('SSD') and 4TB of conventional HDD.

Some experiments have also been conducted on Amazon's AWS cloud. In general this proved not to be as convenient for development and experiment efficiency as local hardware.

Should the local experiment system prove to be inadequate for larger scale experiments, they could alternatively be carried out on PMCC's GPU cluster, which is CUDA based. At the moment it's not clear if this will be necessary or not.

Appendix 2: Preparatory Experiments

In reading the charts, most focus should be on curves *'test loss'* and *'percent correct'*

WSI Experiments

a) 479 TCGA STAD RNA-Seq examples

b) Only absolute (rather than differential) gene expression values are used. Subsequent experiments may also use differential gene expression.

c) All 60,483 RNA-Seq values, comprising both protein coding messenger RNA (mRNA) and long non-coding messenger RNA (lnRNA) used[7].

d) 20% of samples exclusively held out for testing

e) 3 layer Fully Connected ('DENSE') network model reducing the dimensionality from 60,483 to the number of classes via a single hidden layer[8]

f) use Log10+1 data normalization

7 Subsequent experiments will use subsets of the 60,483 mRNA + lnRNA, including (i) just protein coding rna (ii) just genes represented in PMCC gene panels (iii) just genes which are known or suspected to be associated with a particular cancer, as defined in the lliterature or by PMCC scientists.

8 Finding the dimensions of the hidden layer is the objective of one the experiments below – it turns out to be around 250. Deeper models were also tested, but at best these achieved results no better than the three layer model.

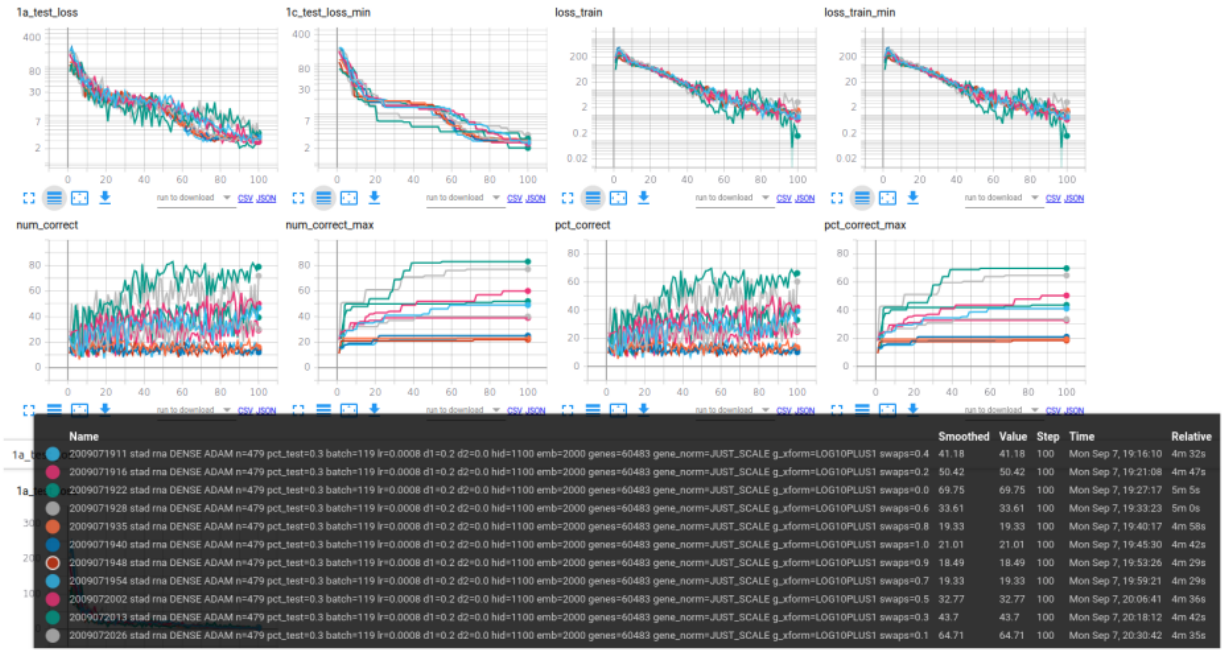
Experiment r1

Objective: ensure the model is genuinely learning

Outcome: the model is genuinely learning

To ensure the network is actually learning from the data and not from hidden internal artifacts (wrong data, bugs), we randomly swap a proportion of the true class labels for class labels chosen at random. With all labels swapped, we'd expect predictions to achieve no better than chance. With *no* labels swapped, we'd expect the network to be genuinely learning to the extent it is able to learn given other parameters. As the proportion of labels swapped reduces, we expect predictions will be intermediate between these two extremes. Each of the 11 runs in this experiment randomize a proportion of the samples' truth labels, ranging from 100% to none of them, in -10% steps.

As can be seen in Figure , the model is genuinely learning. The progression is most easiest to discern from the *num_correct_max* curves.



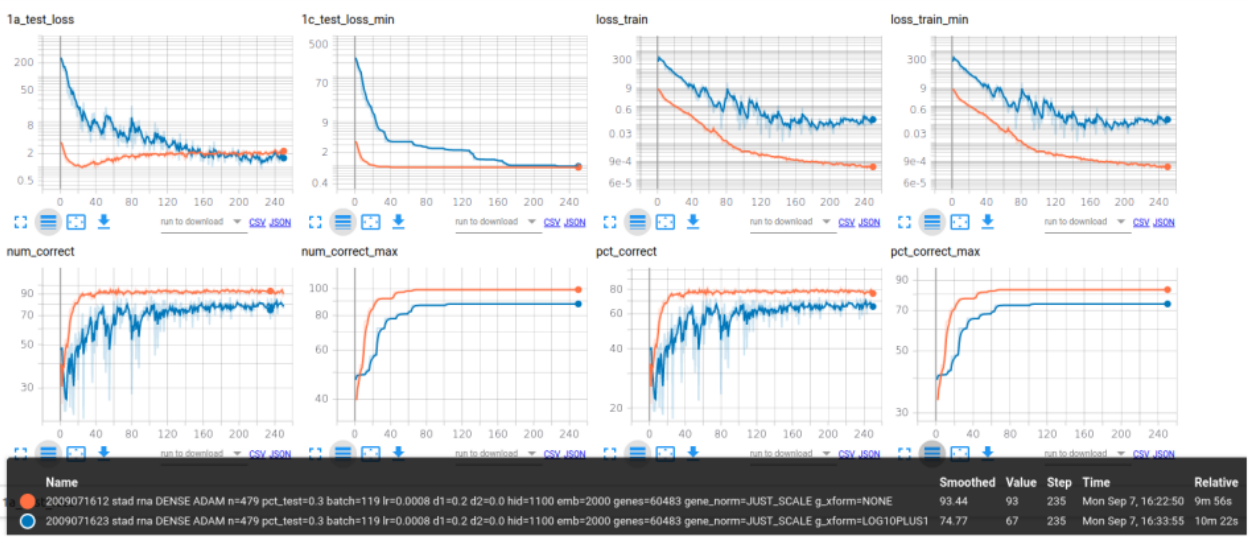
Experiment r2

Objective: Determine if scaling is required and if so what kind scaling should best be used.

Outcome: Log10+1 works the best. Log2+1 (not shown here) less so. Unscaled raw values works poorly.

Figure __ shows the dramatic impact of data scaling on the orders-of-magnitude-in-variance RNA-Seq data. Run 1 has no scaling, run 2 uses log10+1 scaling. All other parameters are identical and relatively well optimised.

With Log10+1 scaling, the lowest test loss occurs around epoch 39 where accuracy is 77%. Without scaling, the model takes much longer to get the test loss down, and achieves about 8% lower accuracy in the 'steady state'.



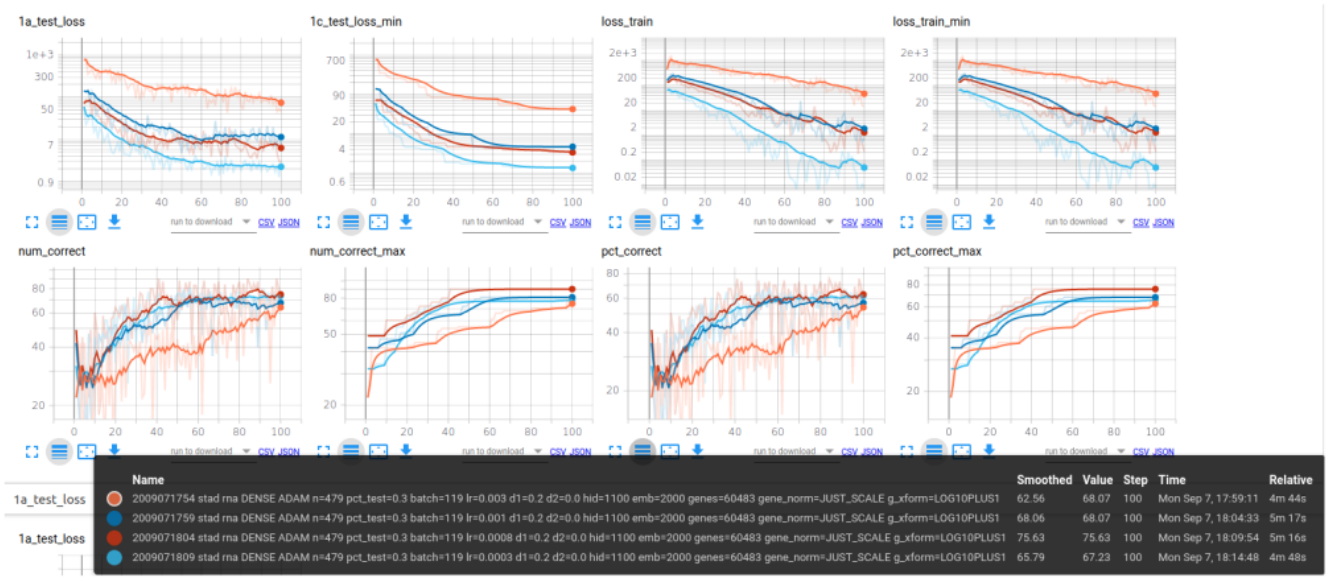
Experiment r3

Objective: Determine the best value for the learning rate hyperparameter

Outcome: The best performing value of those tested was ~0.0008

Learning rate refers to the proportion of the loss function gradient used to update network weights after each batch. The choice of learning rate makes a substantial difference to outcomes.

Running the model with a wide variety of learning rates (not just the ones shown here) reveal that the optimum learning rate is close to 0.0008 (.08%). In experiment 3, four learning rates have been tried, with all other parameters unchanged and reasonably optimised. The the deep red line corresponds to a learning rate of 0.0008.

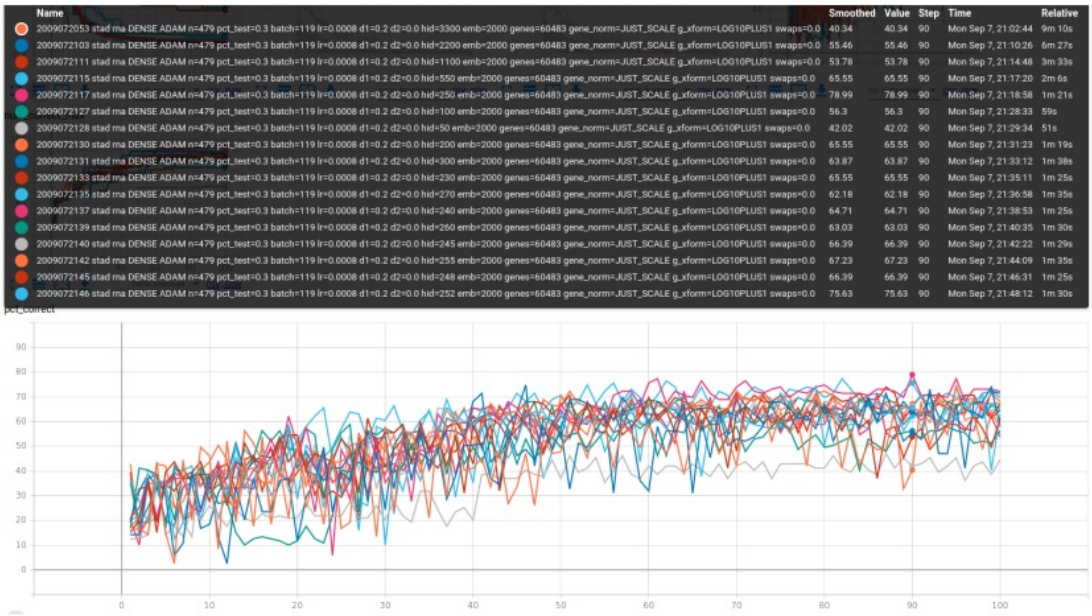


Experiment r4

Objective: Determine the optimum number of neurons to use in the hidden layer of the Fully Connected Network

Outcome: The optimum number neurons to use in the hidden layer is ~250

In this experiment the number neurons in the hidden layer is varied through the



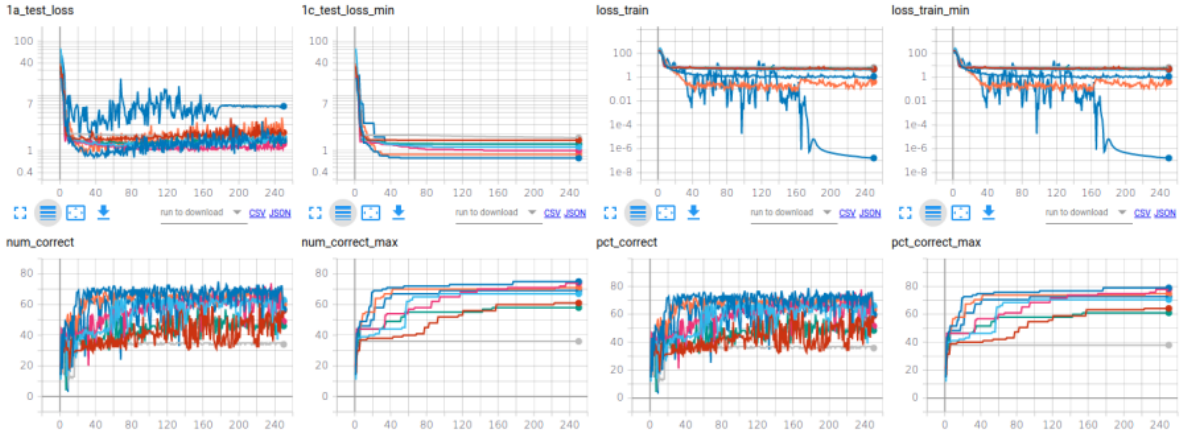
following set: (3300, 2200, 1100, 550, 250, 200, 300, 230, 270, 240, 260, 245, 255, 248, 252). The number of neurons corresponding to the least cost is 250 (pink curve), with lowest cost value occurring at epoch 91, where the test set achieved 79% accuracy – see image __ below.

Experiment r5

Objective: Determine a good value for the drop-out regularization hyperparameter

Outcome: Of those tested , the value of drop-out regularization with least cost was was 0.2

Dropout regularization is a means of regularization whereby a proportion of randomly selected neurons in one layer is dropped after each iteration. It limits the ability of a network to overfit training data and improves generalization. In this experiment, dropout values of 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6 and 0.7 were imposed on the hiddentlayerr all other parameters unchanged and reasonably optimized. The lower blue curve corresponds to 0.2. Lowest cost is at epoch 51 where percent correct = 73%.



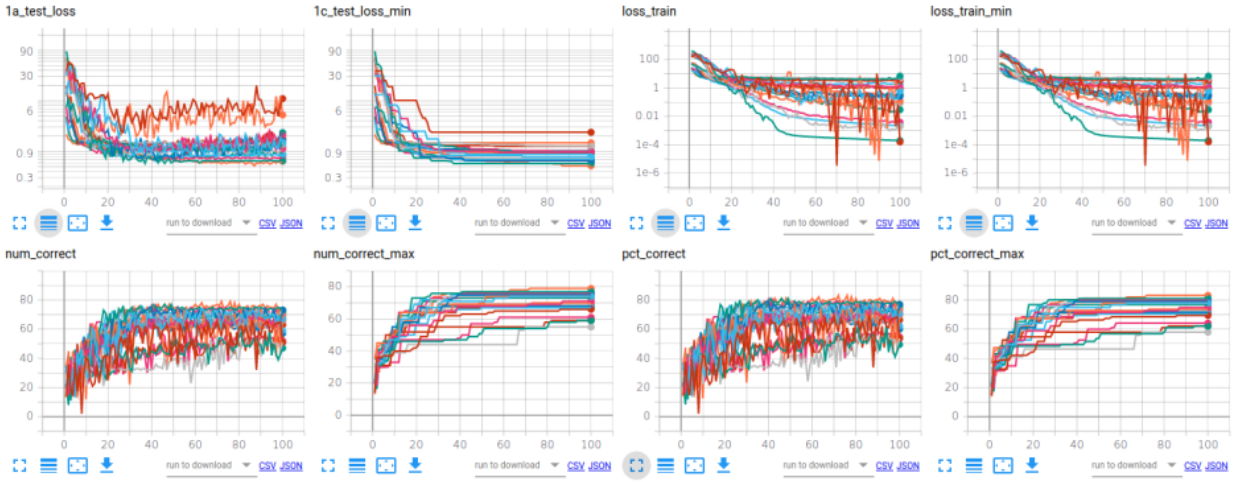
Experiment r6

Objective: Fine tune the hyperparameters deriving from Experiments r1 to r5.

Outcome: an improved result is obtained with drop-out = 0.4 and learning rate = . 0001,. Lowest cost occurs at epoch 87 where accuracy can be seen to have improved to 80%.

Using the nominally optimized hidden layer parameter (250 neurons), bracketing drop-out and learning rates around their nominally optimized values (0.2 and 0.0008 resp) to try and further improve the combination of hyperparameters.

The following set of drop-out values were used: 0.0 0.1 **0.2** 0.3 0.4 and the following set of learning rates (.001 .**0008** .0003 .0001 neurons) yielding an experiment job comprising 5x4=20 runs.



Experiment r7

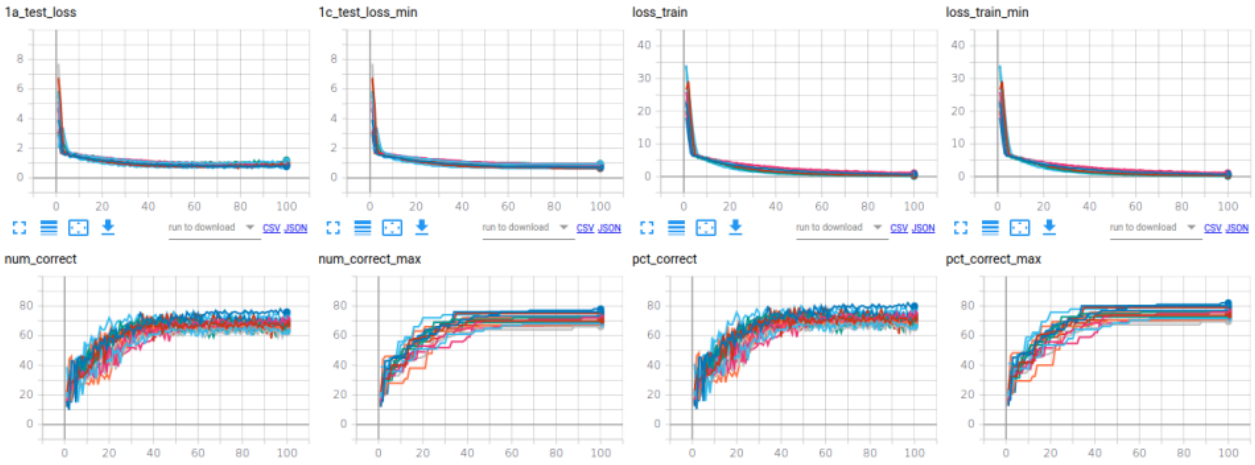
Objective: Verify that the the now notionally optimized model of produces reproducible results

Outcome: The optimized model produces reproducible results

Deep Learning based classification fits multidimensional curves to discrete data, which makes it a fundamentally stochastic procedure. While the held-out test set is sufficiently large (n=96) to ensure that results are unlikely to be influenced much by systematic or random factors, repeating the experiment with the same parameters plus small perturbations thereof would provide further comfort that systematic and random factors are not unduly influencing outcomes.

Therefore, a further set of experiments was conducted with drop-out values: (0.42 0.41 **0.40** 0.39 0.38) and learning rates (.00012 .00011 .**00010** .00009 .00008) yielding an job comprising 5x5=25 runs.

As can be seen from figure __, the cost curves are very similar for all cases. The number/percent correct curves vary somewhat, but this is to be expected since test batches are drawn at random, so every test batch is unique. We also observe a slightly improved accuracy occurring with drop-out = 0.38 and learning rate = 0.00011, however this might not be 'real'



WSI Experiments

The preliminary image experiments use 228 samples; representing 7 subtypes; have 40% of samples held out for testing; and use a 11 layer VGG11 network model. A fixed number (e.g. 4000) of identically dimensioned (e.g. 64 x 64 pixels) tiles are extracted from random positions in each WSI sample, and these tiles are used for training (not the entire image).

Because the number of objects used for training is so large, the image experiments take much longer for images than for RNA-Seq; in the order of hours or small number of days. E.g. with 4,000 tiles per sample every epoch of training must process 228 x 4000 = 916,00 image tiles. This puts a practical limit on the number of experiments that can be conducted.

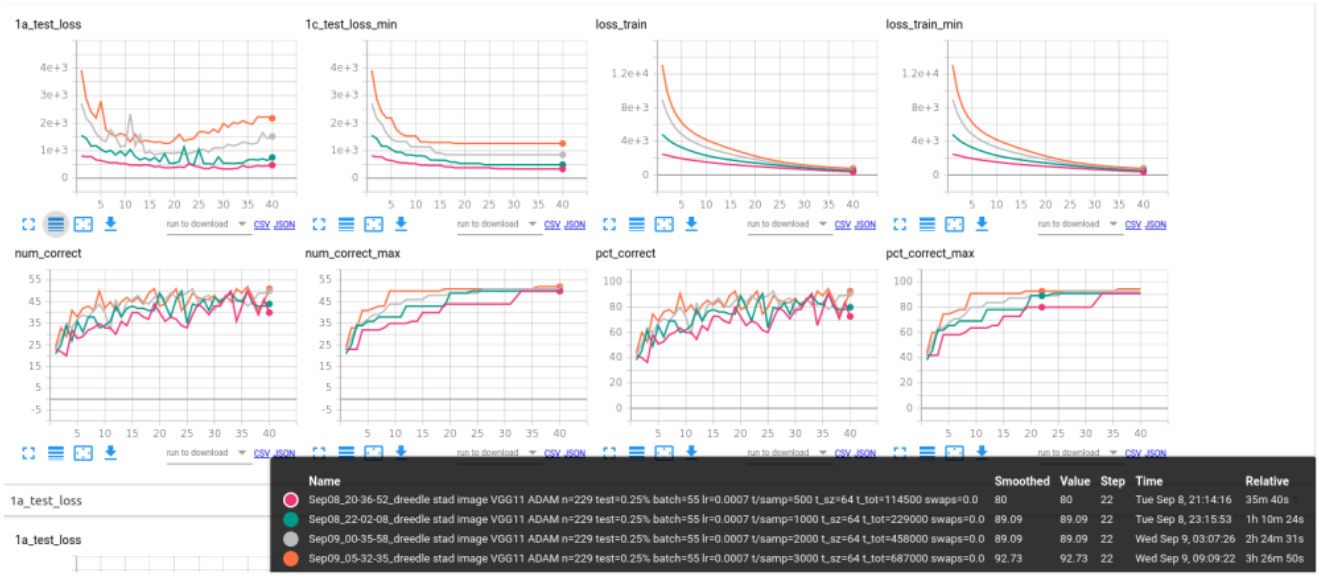
Experiment w1

Tiling is a computationally expensive task, so we wish to discover the minimum number of tiles per sample consistent with a good outcome. Since so many tiles in each sample will contain essentially the same information, it's not necessarily the case that more tiles will equate to higher classification accuracy.

A set of experiments was conducted with these tiles per WSI: (500 1000 2000 3000 4000).

Objective: determine the optimum number of tiles per sample to use

Outcome: As can be seen, the least cost curve (magenta) corresponds to 500 tiles per image, with the lowest cost value occurring at epoch 22, however this corresponds to an accuracy of only 69%. In fact, highest accuracy corresponds to the 3000 tiles per image curve (orange), closely followed by the 2000 columns per image curves (grey), with peak accuracies of 93% and 92% respectively. It's expected that the highest accuracy should occur on the least cost (500 tiles per sample), whereas the best accuracy achieved with the 500 tiles per sample case 89%, at epoch 33; so this requires further investigation. (Note that the size of the held-out test set is sufficiently large that the much higher accuracies obtained in the two cases mentioned is real).



Experiment w2

Predictions are made on a per-tile basis. We need to confirm that model classifications also make sense at a 'macro morphology' level.

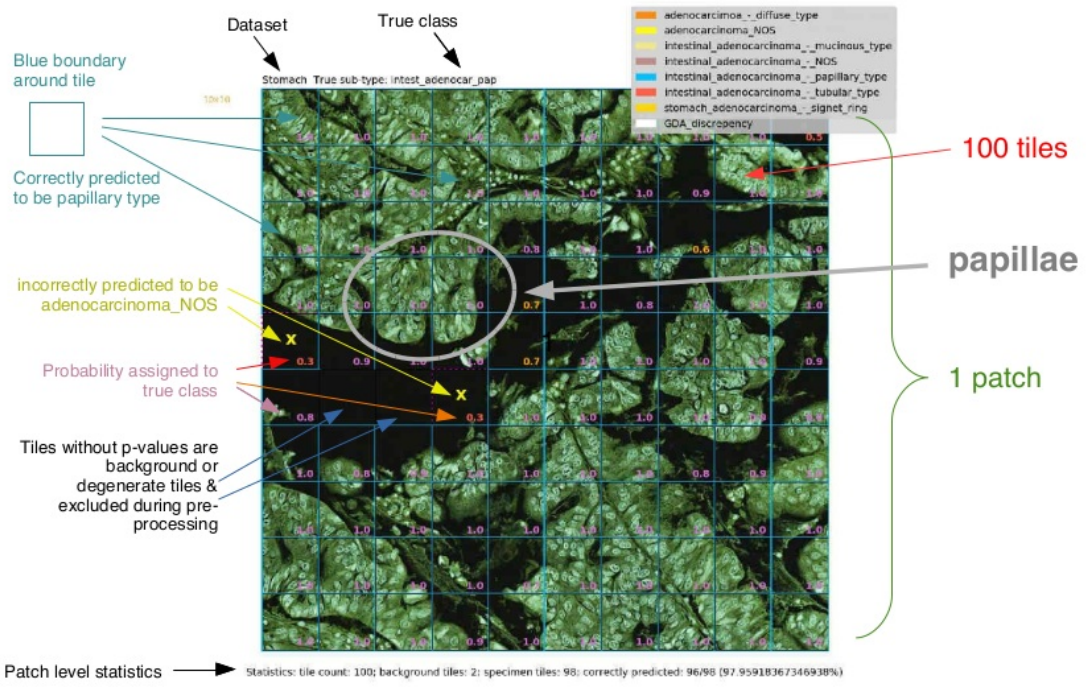
Objective: confirm that classifications produced by the model are 'macro morphologically' sensible.

Method: Extract 2D contiguous sets of tiles (a 'patch') from the samples and process using one of the high accuracy models from Experiment W1; annotating patches with applicable metadata including class predicted, true class and the probability the network assigned to the true class. Ask a PMCC histologist to review these.

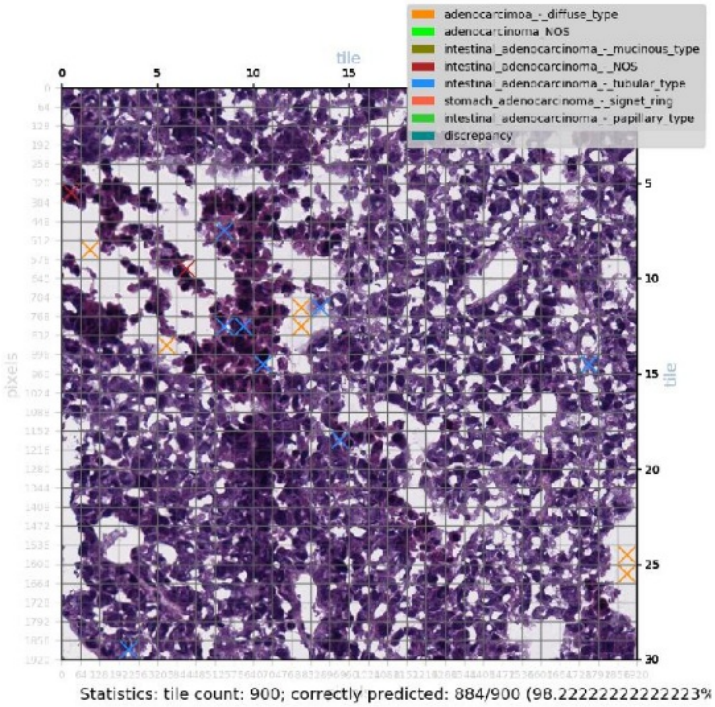
Outcome:

Each result is visualized in a manner similar to figure __.

A subset of 12 of the 225, viz: 3 x confident correct predictions; 3 x unconfident correct predictions and 3 x wrong predictions will be provided to a pathology registrar to assess.

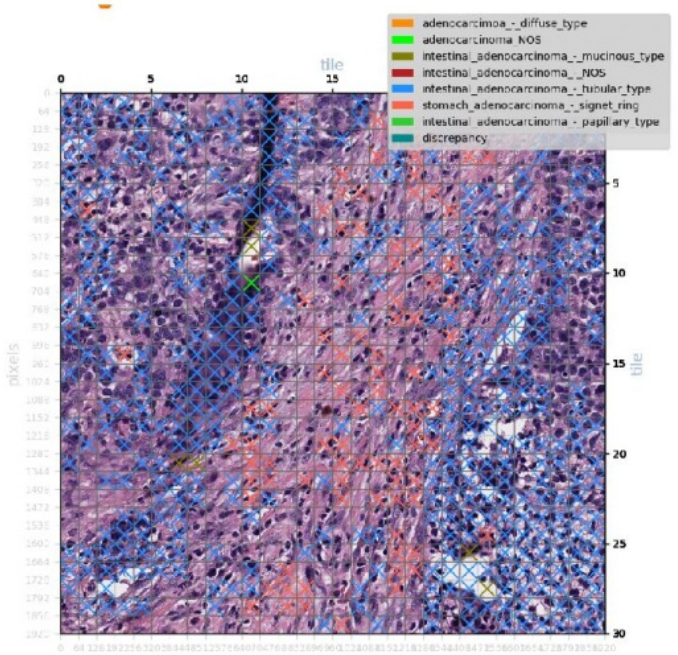


Example of a Correct Prediction Made with High Confidence



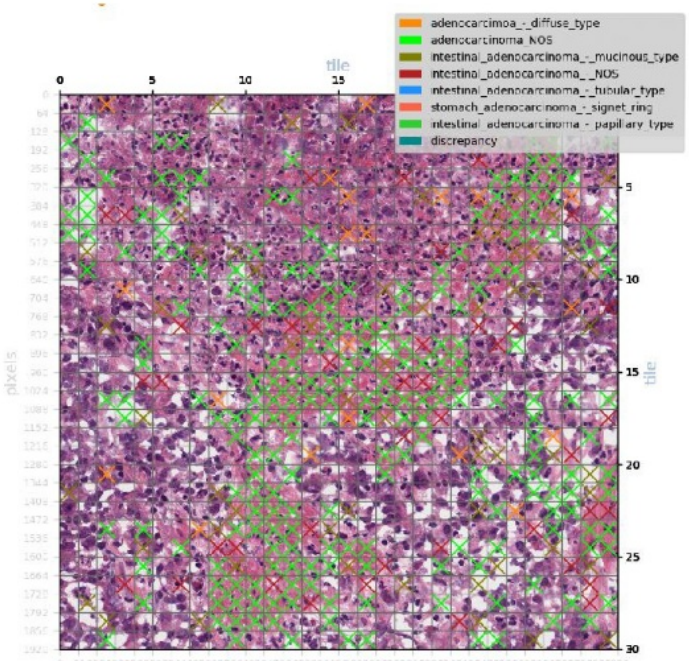
Statistics: tile count: 900; correctly predicted: 884/900 (98.22222222222223%

Example of a Correct but Unconfident Prediction

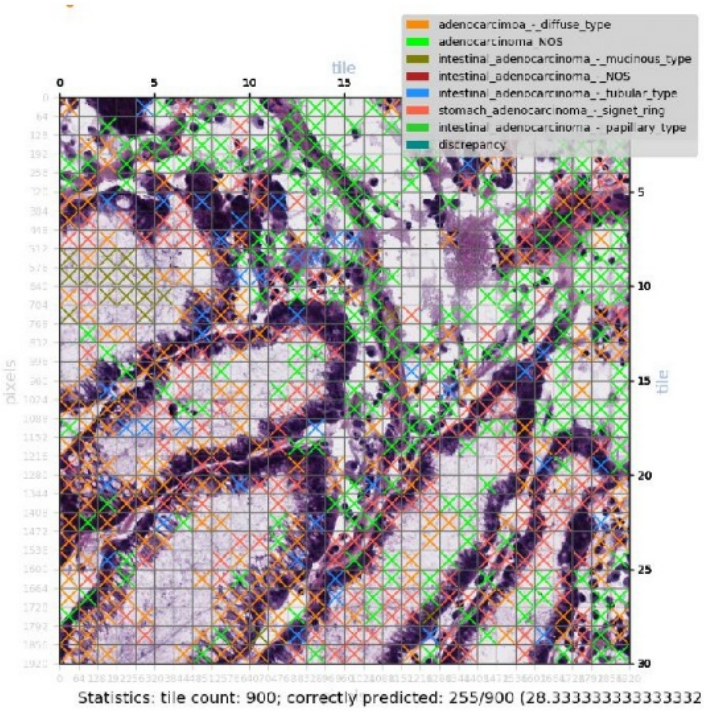


Statistics: tile count: 900; correctly predicted: 408/900 (45.333333333333%

Example of a Wrong Prediction



Statistics: tile count: 900; correctly predicted: 538/900 (59.77777777777777%



Example of a Badly Wrong Predictio

[i] Salto-Tellez and Cree. *Cancer Taxonomy: Pathology Beyond Pathology*. European Journal of Cancer. **115** 57-60 (2019). doi.org/10.1016/j.ejca.2019.03.026

[ii] Bright. *Cancer: Its Classification And Remedies*. S W Butler (1871). Google Books.

[iii] Burney I. *A Historical Tale of Two Lymphomas: Part Ii: Non-Hodgkin Lymphoma*. Sultan Qaboos University Med Journal. **15**(3) 317-321 (2015) Doi: 10.18295/squmj.2015.15.03.003

[iv] Medeiros, Elenitoba-Johnson, *Anaplastic Large Cell Lymphoma.* American Journal of Clinical Pathology, **127** (5) 707–negative722 (2007). DOI: 10.1309/R2Q9CCUVTLRYCF3H

[v] N.L. Harris, E.S. Jaffe, J. Diebold, G. Flandrin, H.K. Muller-Hermelink, J. Vardiman. *Lymphoma Classification – From Controversy to Consensus: the R.E.A.L. and WHO Classification of Lymphoid Neoplasms*. Annals of Oncology, **11**, Supplement 1, S3-S10.(2000) 10.1093/annonc/11.suppl_1.S3

[vi] Golub. *Toward a Functional Taxonomy of Cancer*. Cancer Cell **6**, (2) (2004) 107-108. DOI: 10.1016/j.ccr.2004.08.007

[vii] Faltas. *Lumpers and Splitters: A New Molecular Taxonomy for Cancer*. Science Translational Medicine. **6,** (249) 249-138 (2014). DOI: 10.1126/scitranslmed.3010118

[viii] Diamantidis, Papadopoulos, Kaiafa, G. et al. *Differential Diagnosis and Treatment of Primary, Cutaneous, Anaplastic Large Cell Lymphoma: Not Always an Easy Task.* Int Journal of Hematology **90**, 226–228 (2009). DOI: 10.1007/s12185-009-0365-7

[ix] Erik Peterson, Jason Weed, Kristen Lo Sicco, Jo-Ann Latkowski, *Cutaneous T Cell Lymphoma: A Difficult Diagnosis Demystified*, Dermatologic Clinics **37**, 4 (2019) 455-469, 10.1016/j.det.2019.05.007

[x] Win, Khin Than1; Liau, Jau-Yu2; Chen, Bo-Jung et al. *Primary Cutaneous Extranodal Natural Killer/T-Cell Lymphoma Misdiagnosed as Peripheral T-Cell Lymphoma: The Importance of Consultation/Referral and Inclusion of EBV In Situ*

Hybridization for Diagnosis. Applied Immunohistochemistry & Molecular Morphology. **24**, 2, 105-111(7) (2016) DOI: 10.1097/PAI.0000000000000162

xi Wlodarska, De Wolf-Peeters et al. *The Cryptic inv(2)(p23q35) Defines a New Molecular Genetic Subtype of ALK-Positive Anaplastic Large-Cell Lymphoma*. Blood 1998; **92** (8): 2688–2695 (1998). DOI: 10.1182/blood.V92.8.2688

xii Jaffe, Barr, Smith. *Understanding the New WHO Classification of Lymphoid Malignancies: Why It's Important and How It Will Affect Practice*. American Society of Clinical Oncolology Educational Book. 37 535-546 (2017) doi:10.1200/EDBK_175437

xiii Scott E. Miller. *DNA Barcoding and the Renaissance of Taxonomy*. Proceedings of the National Academy of Sciences. **104** (12) 4775-4776 (2007) . DOI: 10.1073/pnas.0700466104

xiv Diamantidis, Papadopoulos, Kaiafa, G. et al. *Differential Diagnosis and Treatment of Primary, Cutaneous, Anaplastic Large Cell Lymphoma: Not Always an Easy Task.* Int Journal of Hematology **90**, 226–229 (2009). DOI: 10.1007/s12185-009-0365-7

xv Diamantidis, Papadopoulos, Kaiafa, G. et al. Differential Diagnosis and Treatment of Primary, Cutaneous, Anaplastic Large Cell Lymphoma: Not Always an Easy Task. Int Journal of Hematology 90, 226–229 (2009). DOI: 10.1007/s12185-009-0365-7

xvi Win, Khin Than1; Liau, Jau-Yu2; Chen, Bo-Jung et al. *Primary Cutaneous Extranodal Natural Killer/T-Cell Lymphoma Misdiagnosed as Peripheral T-Cell Lymphoma: The Importance of Consultation/Referral and Inclusion of EBV In Situ Hybridization for Diagnosis*. Applied Immunohistochemistry & Molecular Morphology. **24**, 2, 105-111(7) (2016) DOI: 10.1097/PAI.0000000000000162

xvii Savopoulos CG, Tsesmeli NE, Kaiafa GD, et al. *Primary Pancreatic Anaplastic Large Cell Lymphoma, ALK Negative: a Case Report*. World Journal of Gastroenterology. **11** (39) 6221-6224. DOI: 10.3748/wjg.v11.i39.6221

xviii Daniel Benharroch, Zarouhie Meguerian-Bedoyan, Laurence Lamant,et al. *ALK-Positive Lymphoma: A Single Disease With a Broad Spectrum of Morphology*. Blood;**91** (6): 2076–2084. 1998 DOI: 10.1182/blood.V91.6.2076

xix Vassallo, Lamant, Brugieres, et al. *ALK-Positive Anaplastic Large Cell Lymphoma Mimicking Nodular Sclerosis Hodgkin's Lymphoma.* The American Journal of Surgical Pathology **30**, 2, 223-229 (2006) DOI: 10.1097/01.pas.0000179123.66748.c2

xx Yeo-Rye Cho, Jeong-Wan Seo, Sung Yong Oh, Min-Kyoung Pak, Ki-Ho Kim. *The expressions of MUM-1 and Bcl-6 in ALK-negative systemic anaplastic large cell lymphoma with skin involvement and primary cutaneous anaplastic large cell lymphoma*. International Journal of Clinical and Experimental Pathology. 2020; **13**(7): 1682–1687.

xxi Xiaoli Wang , Jingjing Wu, MiAnika Cheerla, Olivier Gevaert. Deep Learning With Multimodal Representation for Pancancer Prognosis Prediction. Bioinformatics, **35**, 14, i446–i454. DOI: 10.1093/bioinformatics/btz342.

xxii Savage, Harris, Vose et al, for the International Peripheral T-Cell Lymphoma Project, *ALK– Anaplastic Large-Cell Lymphoma Is Clinically and Immunophenotypically Different From Both ALK+ ALCLand Peripheral T-Cell Lymphoma, Not Otherwise Specified: Report From the International Peripheral T-Cell Lymphoma Project.* Blood111 (12): 5496–5504 (2008). DOI: 10.1182/blood-2008-01-134270

xxiii Delsol, Brugières, Gaulard et al. Anaplastic Large Cell Lymphoma, ALK-Positive And Anaplastic Large Cell Lymphoma ALK-Negative. Hematology Meeting Reports (formerly Haematologica Reports), **3**(1) (2009) DOI: 10.4081/hmr.v3i1.530

xxiv LeCun, Y., Bengio, Y. & Hinton, G. *Deep Learning. Nature* **521,** 436–444 (2015). DOI: 10.1038/nature14539

xxv Sidike, Alom,Taha, Asari. *The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches.* (2018) https://arxiv.org/abs/1803.01164

xxvi Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones. *Opportunities and Obstacles For Deep Learning In Biology and Medicine.* Journal of the Royal Society Interface. **15**, 20170387 (2018). DOI: 10.1098/rsif.2017.0387

xxvii Geert Litjens, Thijs Kooi, Babak Ehteshami, Bejnordi. *A Survey On Deep Learning In Medical Image Analysis.* Medical Image Analysis **42**, 60-88 (2017). DOI: 10.1016/j.media.2017.07.005

xxviii Serag, Ion-Margineanu, Qureshi, McMillan. *Translational AI and Deep Learning in Diagnostic Pathology.* Frontiers in Medicine (2019). DOI: 10.3389/fmed.2019.00185

[xxix] Cong L, Feng W, Yao Z, Zhou X, Xiao W. *Deep Learning Model as a New Trend in Computer-aided Diagnosis of Tumor Pathology for Lung Cancer.* J ournal iof Cancer **11**(12):3615-3622.(2020) DOI: 10.7150/jca.43268

[xxx] Azuaje, F. *Artificial Intelligence For Precision Oncology: Beyond Patient Stratification*. Nature Precision Oncology **3**, 6 (2019). DOI: 10.1038/s41698-019-0078-1

[xxxi] Trister AD. *The Tipping Point for Deep Learning in Oncology*. *JAMA Oncology* **5,** 10 1429–1430 (2019). DOI: 10.1001/jamaoncol.2019.1799

[xxxii] Zaidan, Zaidan, Albahri et al. *A Review On Smartphone Skin Cancer Diagnosis Apps in Evaluation And Benchmarking: Coherent Taxonomy, Open Issues and Recommendation Pathway Solution*. Health Technologies **8**, 223–238 (2018). DOI: 10.1007/s12553-018-0223-9

[xxxiii] Esteva, A., Kuprel, B., Novoa, R. et al. *Dermatologist-Level Classification Of Skin Cancer With Deep Neural Networks*. *Nature* **542,** 115–118 (2017). DOI: 10.1038/nature21056

[xxxiv] Lo Ying-Chih; Keng-Hung, Lin; Bair, Henry; Sheu Wayne Huey-Herng; Chi-Sen, Chang et al. *Epiretinal Membrane Detection at the Ophthalmologist Level using Deep Learning of Optical Coherence Tomography*. Scientific Reports (Nature Publisher Group); London. **10**, 1 (2020). DOI:10.1038/s41598-020-65405-2

[xxxv] Deng, S., Zhang, X., Yan, W. et al. *Deep Learning in Digital Pathology Image Analysis: A Survey*. Frontiers of Medicine. **14,** 470–487 (2020). DOI: doi.org/10.1007/s11684-020-0782-9

[xxxvi] Jiang, Y, Yang, M, Wang, S, Li, X, Sun, Y. *Emerging role of deep learning-based artificial intelligence in tumor pathology. Cancer Communications.* **40,** 154– 166 (2020). DOI: 10.1002/cac2.12012

[xxxvii] Ping Luo, Yulian Ding, Xiujuan Lei, and Fang-Xiang Wu. *deepDriver: Predicting Cancer Driver Genes Based on Somatic Mutations Using Deep Convolutional Neural Networks*. Frontiers in Genetics. **10**, 13 (2019). DOI: 10.3389/fgene.2019.00013

[xxxviii] Kather, J.N., Heij, L.R., Grabsch, H.I. *et al. Pan-Cancer Image-Based Detection of Clinically Actionable Genetic Alterations. Nat Cancer* **1,** 789–799 (2020). DOI: 10.1038/s43018-020-0087-6

[xxxix] Benoît Schmauch, Alberto Romagnoni, Elodie Pronier et al. *Transcriptomic Learning for Digital Pathology*. BioArxiv (2019). DOI: 10.1101/760173

[xl] Schmauch, B., Romagnoni, A., Pronier, E. *et al*. A Deep Learning Model to Predict RNA-Seq Expression Of Tumours from Whole Slide Images. Nature Communications **11,** 3877 (2020). 10.1038/s41467-020-17678-4

[xli] Gundersen, G, Dumitrascu, B, Engelhardt, B. *End-To-End Training of Deep Probabilistic CCA on Paired Biomedical Observations.* In proceedings of the Conference on Uncertainty in Artificial Intelligence. (2019)

[xlii] Ainscough, B.J., Barnell, E.K., Ronning, P. *et al. A Deep Learning Approach to Automate Refinement Of Somatic Variant Calling from Cancer Sequencing Data*. *Nature Genetics* **50,** 1735–1743 (2018). 10.1038/s41588-018-0257-y

[xliii] Xiaoli Wang , Jingjing Wu, MiAnika Cheerla, Olivier Gevaert. Deep Learning With Multimodal Representation for Pancancer Prognosis Prediction. Bioinformatics, **35**, 14, i446–i454. DOI: 10.1093/bioinformatics/btz342ngzhi Zhang.

[xliv] Le H, Gupta R, Hou L, et al. *Utilizing Automated Breast Cancer Detection to Identify Spatial Distributions of Tumor-Infiltrating Lymphocytes in Invasive Breast Cancer*. *American Journal of Pathology* **190**, 7 1491-1504 (2020). 10.1016/j.ajpath.2020.03.012

[xlv] Aïcha BenTaieb, Ghassan Hamarneh. *Deep Learning Models for Digital Pathology*. Arxiv Computer Vision and Pattern Recognition (2019). https://arxiv.org/abs/1910.12329

[xlvi] Shaver, M.M., Kohanteb, P.A., Chiou, C., Bardis, M.D., Chantaduly, C., Bota, D., Filippi, C.G., Weinberg, B., Grinband, J., Chow, D.S., Chang, P.D. *Optimizing Neuro-Oncology Imaging: A Review of Deep Learning Approaches for Glioma Imaging.* Cancers **11**, 829 (2019) 10.3390/cancers11060829

[xlvii] Aïcha BenTaieb, Ghassan Hamarneh. *Deep Learning Models for Digital Pathology*. Arxiv Computer Vision and Pattern Recognition (2019). https://arxiv.org/abs/1910.12329

xlviii Daisuke Komura, Shumpei Ishikawa. *machine learning Methods for Histopathological Image Analysis.* Computational and Structural Biotechnology Journal, **16**, 34-42 (2018). DOI: doi.org/10.1016/j.csbj.2018.01.001

xlix Karen Simonyan, Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition.* Published as a conference paper at the International Conference on Learning Representations 2015.

l Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. *Deep Residual Learning for Image Recognition*. Published as a conference paper at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). DOI: 10.1109/CVPR.2016.90

li Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich. *Going Deeper With Convolutions.* Published in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). DOI: 10.1109/CVPR.2015.7298594

lii Tensorflow: 'An end-to-end open source machine learning platform' Web site: https://www.tensorflow.org/

liii Pytorch. 'An open source machine learning framework that accelerates the path from research prototyping to production deployment.' PyTorch project web site: https://pytorch.org/

liv 'Caffe is a deep learning framework made with expression, speed, and modularity in mind. It is developed by Berkeley AI Research (BAIR) and by community contributors.' Caffe web site: https://caffe.berkeleyvision.org/

lv The Cancer Genome Atlas (TCGA) Data Portal home page: https://portal.gdc.cancer.gov/

lvi Grand Challenge home page: "A platform for end-to-end development of machine learning solutions in biomedical imaging." https://grand-challenge.org/

lvii CAMELYON17 Grand Challenge home page: https://camelyon17.grand-challenge.org/

lviii Guilherme Aresta, Teresa Araújo, Scotty Kwok, et al. BACH: Grand Challenge on Breast Cancer Histology Images, Medical Image Analysis, **56**, 122-139. (2019) 10.1016/j.media.2019.05.010

lix Campanella, G., Hanna, M.G., Geneslaw, L. et al. *Clinical-Grade Computational Pathology Using Weakly Supervised Whole Deep Learning On Whole Slide Images.* Nature Medicine **25**, 1301–1309 (2019). DOI: 10.1038/s41591-019-0508-1

lx Coudray, N., Ocampo, P.S., Sakellaropoulos, T. et al. *Classification And Mutation Prediction From Non–Small Cell Lung Cancer Histopathology Images Using Deep Learning.* Nature Medicine **24**, 1559–1567 (2018). DOI: 10.1038/s41591-018-0177

lxi N. Brancati, G. De Pietro, M. Frucci and D. Riccio, A Deep Learning Approach for Breast Invasive Ductal Carcinoma Detection and Lymphoma Multi-Classification in Histological Images, in IEEE Access, vol. **7,** pp. 44709-44720, 2019. DOI 10.1109/ACCESS.2019.2908724.

lxii Han Le, Rajarsi Gupta, Le Hou, Shahira Abousamra, Danielle Fassler, Tahsin Kurc, Dimitris Samaras, Rebecca Batiste, Tianhao Zhao, Arvind Rao, Alison L. Van Dyke, Ashish Sharma, Erich Bremer, Jonas S.Almeida, Joel Saltz. *Utilizing Automated Breast Cancer Detection to Identify Spatial Distributions of Tumor Infiltrating Lymphocytes in Invasive Breast Cancer.* American Journal of Pathology. 2020 Jul; **190** (7):1491-1504. DOI: 10.1016/j.ajpath.2020.03.012

lxiii Liron Pantanowitz, Gabriela M Quiroga-Garza, Lilach Bien et al. *An Artificial Intelligence Algorithm For Prostate Cancer Diagnosis in Whole Slide Images of Core Needle Biopsies: A Blinded Clinical Validation and Deployment Study. **2**, 8 E407-E416 (2020). DOI: 10.1016/S2589-7500(20)30159-X*

lxiv Khan J, Wei, J et al. *Classification and Diagnostic Prediction Of Cancers Using Gene Expression Profiling And Artificial Neural Networks*. Nature Medicine **7**, 673–679 (2001). DOI: 10.1038/89044

lxv Aliferis C, Tsamardinos I et al. *machine learning Models for Classification of Lung Cancer and Selection of Genomic Markers Using Array Gene Expression Data*. (2003). Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference (flairs 2003)

lxvi Feng Gao, Wei Wang , Miaomiao Tan et al. *DeepCC: A Novel Deep Learning-Based Framework For Cancer Molecular Subtype Classification*. Oncogenesis **8**, 44 (2019). DOI: 10.1038/s41389-019-0157-8

[lxvii] MsigDB home page. https://www.gsea-msigdb.org/gsea/msigdb/

[lxviii] Sangseon Lee, Sangsoo Lim, Taeheon Lee, Inyoung Sung, Sun Kim. *Cancer Subtype Classification and Modeling By Pathway Attention and Propagation.* Bioinformatics, **36,** Issue 12, 3818–3824 (2020). DOI: 10.1093/bioinformatics/btaa203

[lxix] Minoru Kanehisa, Susumu Goto. *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, **28**, 127–30 (2000). DOI: 10.1093/nar/28.1.27

[lxx] Abdullah Al Mamun, Ananda Mohan Mondal. *Long Non-coding RNA Based Cancer Classification using Deep Neural Networks*. In Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '19) (2019). Association for Computing Machinery. DOI: 10.1145/3307339.3343249

[lxxi] Xena Functional Genomics Explorer home page: https://xenabrowser.net/

[lxxii] Zhiqiang Zhang, Yi Zhao, Xiangke Liao, Wenqiang Shi et al, *deep learning in omics: a survey and guideline*. Briefings in Functional Genomics, **18**, 1, 41–57. (2019) DOI: 10.1093/bfgp/ely030

[lxxiii] Joshua J. Levy , Alexander J. Titus, Curtis L. Petersen, Youdinghuan Chen, Lucas A. Salas, Brock C. Christensen. *MethylNet: an automated and modular deep learning approach for DNA methylation analysis.* BMC Bioinformatics **21**, 108 (2020). DOI: 10.1186/s12859-020-3443-8

[lxxiv] Gundersen, G, Dumitrascu, B, Engelhardt, B. *End-To-End Training of Deep Probabilistic CCA on Paired Biomedical Observations.* In proceedings of the Conference on Uncertainty in Artificial Intelligence. (2019)

[lxxv] Gene Tissue Expression (GTEx) Dataset home page: https://gtexportal.org/home/

[lxxvi] Carmichael, Calhoun, Hoadley. *Joint And Individual Analysis Of Breast Cancer Histologic Images And Genomic Covariates* (2019).

[lxxvii] Qing Feng, Meilei Jiang, Jan Hannig, J.S. Marron. *Angle-Based Joint and Individual Variation Explained*. Journal of Multivariate Analysis, **166**, 241-265 (2018). DOI: 10.1016/j.jmva.2018.03.008

[lxxviii] Ash, Darnell, Munro, Engelhardt. *Joint Analysis of Gene Expression Levels and Histological Images Identifies Genes Associated with Tissue Morphology.* BioArxiv pre-print (2018). DOI: doi.org/10.1101/458711

[lxxix] Anika Cheerla, Olivier Gevaert. Deep Learning With Multimodal Representation for Pancancer Prognosis Prediction. Bioinformatics, **35**, 14, i446–i454. DOI: 10.1093/bioinformatics/btz342

[lxxx] Islam, Huang, Ajwad et al, An Integrative Deep Learning Framework for Classifying Molecular Subtypes of Breast Cancer, Computational & Structural Biotechnology Journal, **18**, 2185-2199, (2020). DOI: 10.1016/j.csbj.2020.08.005

[lxxxi] Martin Thoma, *Comparing Classifiers*. 2016 https://martin-thoma.com/comparing-classifiers/

[lxxxii] C Arun, A Prabhu, M Zeeshan & N Rani. *A Study on Various Classifier Techniques Used in Image Processing*. Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020). DOI: 10.1109/ICICCS48265.2020.9121177

[lxxxiii] Campbell, P.J., Getz, G., Korbel, J.O. et al. *Pan-Cancer Analysis of Whole Genomes.* Nature **578**, 82–93 (2020). 10.1038/s41586-020-1969-6

One Year Confirmation Report

ORIGINALITY REPORT

6%

SIMILARITY INDEX

PRIMARY SOURCES

1	journals.lww.com Internet	62 words — 1%
2	www.nature.com Internet	42 words — < 1%
3	ijcep.com Internet	41 words — < 1%
4	hal.cse.msu.edu Internet	32 words — < 1%
5	insights.ovid.com Internet	30 words — < 1%
6	hprc.tamu.edu Internet	26 words — < 1%
7	www.ijitee.org Internet	24 words — < 1%
8	dokumen.pub Internet	22 words — < 1%
9	img.medscape.com Internet	21 words — < 1%
10	Mehreen Tariq, Sajid Iqbal, Hareem Ayesha, Ishaq Abbas, Khawaja Tehseen Ahmad, Muhammad Farooq Khan Niazi. "Medical image based breast cancer diagnosis: State of the art and future directions", Expert Systems with Applications, 2020	19 words — < 1%

Crossref

11	www.springerprofessional.de Internet	19 words — < 1%
12	mafiadoc.com Internet	19 words — < 1%
13	hal.sorbonne-universite.fr Internet	16 words — < 1%
14	www.freepatentsonline.com Internet	16 words — < 1%
15	S F Shepherd, N D McGuire, B P J de Lacy Costello, R J Ewen, D H Jayasena, K Vaughan, I Ahmed, C S Probert, N M Ratcliffe. "The use of a gas chromatograph coupled to a metal oxide sensor for rapid assessment of stool samples from irritable bowel syndrome and inflammatory bowel disease patients", Journal of Breath Research, 2014 Crossref	12 words — < 1%
16	ir.lib.uwo.ca Internet	12 words — < 1%
17	www.frontiersin.org Internet	11 words — < 1%
18	jamanetwork.com Internet	11 words — < 1%
19	www.tandfonline.com Internet	11 words — < 1%
20	Jianming Yang. "An Easily Implemented, Block-Based Fast Marching Method with Superior Sequential and Parallel Performance", SIAM Journal on Scientific Computing, 2019 Crossref	10 words — < 1%
21	"Computer Vision – ECCV 2020", Springer Science and Business	

	Media LLC, 2020 Crossref	10 words — < 1 %
22	Russell E. Ericksen, Siew Lan Lim, Eoin McDonnell, Wai Ho Shuen et al. "Loss of BCAA Catabolism during Carcinogenesis Enhances mTORC1 Activity and Promotes Tumor Development and Progression", Cell Metabolism, 2019 Crossref	10 words — < 1 %
23	"Trends and Innovations in Information Systems and Technologies", Springer Science and Business Media LLC, 2020 Crossref	10 words — < 1 %
24	www.greengazette.co.za Internet	10 words — < 1 %
25	dblp.uni-trier.de Internet	9 words — < 1 %
26	www.archivesofpathology.org Internet	9 words — < 1 %
27	www.karger.com Internet	8 words — < 1 %
28	Youssef Safi, Abdelaziz Bouroumi. "An evolutionary approach for optimizing three-layer perceptrons architecture", 2012 International Conference on Multimedia Computing and Systems, 2012 Crossref	8 words — < 1 %
29	"Intelligent Computing Theories and Application", Springer Science and Business Media LLC, 2019 Crossref	8 words — < 1 %
30	www.hsrc.ac.za Internet	8 words — < 1 %
31	Jose Dolz, Xiaopan Xu, Jérôme Rony, Jing Yuan et	8 words — < 1 %

al. "Multi-region segmentation of bladder cancer structures in MRI with progressive dilated convolutional networks", Medical Physics, 2018
Crossref

32	arxiv.org Internet	8 words — < 1%
33	doctorpenguin.com Internet	8 words — < 1%
34	Debashis Ghosh. "Penalized Discriminant Methods for the Classification of Tumors from Gene Expression Data", Biometrics, 2003 Crossref	8 words — < 1%
35	export.arxiv.org Internet	8 words — < 1%
36	summit.sfu.ca Internet	8 words — < 1%
37	Mingyi Wang, Wen Luo, Kristine Jones, Xiaopeng Bian et al. "SomaticCombiner: improving the performance of somatic variant calling based on evaluation tests and a consensus approach", Scientific Reports, 2020 Crossref	8 words — < 1%
38	Khin Than Win, Jau-Yu Liau, Bo-Jung Chen, Katsuyoshi Takata, Chi-Cheng Li, Cheng-Hsiang Hsiao, Chiao-Yun Chen, Shih-Sung Chuang. "Primary cutaneous extranodal natural killer/T-cell lymphoma misdiagnosed as peripheral T-cell lymphoma: the importance of EBV in situ hybridization for diagnosis", Pathology, 2014 Crossref	7 words — < 1%
39	Javier Juan Albarracín. "Unsupervised learning for vascular heterogeneity assessment of glioblastoma based on magnetic resonance imaging: The Hemodynamic Tissue Signature", Universitat Politecnica de Valencia, 2020 Crossref Posted Content	7 words — < 1%

40 Geoffrey E. Hinton, Simon Osindero, Yee-Whye Teh. "A Fast Learning Algorithm for Deep Belief Nets", Neural Computation, 2006 — 7 words — < 1%
Crossref

41 Xiaomin Zhou, Chen Li, Md Mamunur Rahaman, Yudong Yao et al. "A Comprehensive Review for Breast Histopathology Image Analysis Using Classical and Deep Neural Networks", IEEE Access, 2020 — 7 words — < 1%
Crossref

42 Ying Wang, Ting Peng, Jiajia Duan, Chuang Zhu, Jun Liu, Jiandong Ye, Mulan Jin. "Pathological Image Classification based on Hard Example Guided CNN", IEEE Access, 2020 — 6 words — < 1%
Crossref

43 Sai Chandra Kosaraju, Jie Hao, Hyun Min Koh, Mingon Kang. "Deep-Hipo: Multi-scale Receptive Field Deep Learning for Histopathological Image Analysis", Methods, 2020 — 6 words — < 1%
Crossref

44 Hilary M. O'Leary. "Update on the World Health Organization classification of peripheral T-cell lymphomas", Current Hematologic Malignancy Reports, 10/2009 — 6 words — < 1%
Crossref

EXCLUDE QUOTES ON
EXCLUDE BIBLIOGRAPHY ON
EXCLUDE MATCHES OFF