A Deep Learning Approach to Tumour Identification in Fresh Frozen Tissues

Hassan Ugail, Maisun Alzorgani and Ali Bukar
Centre for Visual Computing,
Faculty of Engineering and Informatics
University of Bradford,
Bradford BD7 1DP, UK

Humera Hussain, Christopher Burn, Thinzar Min Sein and Samar Betmouni
Therapeutic & Diagnostic Centre, Faculty of Life Science & Digital Health Enterprise Zone Academic, University of Bradford,
Bradford BD7 1DP, UK

Abstract — **The demand for pathology services are significantly increasing whilst the numbers of pathologists are significantly decreasing. In order to overcome these challenges, a growing interest in faster and efficient diagnostic methods such as computer-aided diagnosis (CAD) have been observed. An increase in the use of CAD systems in clinical settings has subsequently led to a growing interest in machine learning. In this paper, we show the use of machine learning algorithms in the prediction of tumour content in Fresh Frozen (FF) histological samples of head and neck. More specifically, we explore a pre-trained convolutional neural network (CNN), namely the AlexNet, to build two common machine learning classifiers. For the first classifier, the pre-trained AlexNet network is used to extract features from the activation layer and then Support Vector Machine (SVM) based classifier is trained by using these extracted features. In the second case, we replace the last three layers of the pre-trained AlexNet network and then fine tune these layers on the FF histological image samples. The results of our experiments are very promising. We have obtained percentage classification rates in the high 90s, and our results show there is little difference between SVM and transfer learning. Thus, the present study show that an AlexNet driven CNN with SVM and fine-tuned classifiers are a suitable choice for accurate discrimination between tumour and non-tumour histological samples from the head and neck.**

Keywords — Tumour Identification, Deep Learning, Convolutional Neural Networks, AlexNet, Histopathological Diagnosis, Fresh Frozen Tissue Analysis.

I. INTRODUCTION

Histopathology is the study of disease in tissue samples. It relies on the clinicopathological correlation between clinical findings - history, clinical examination and special investigations such as radiology - and macroscopic and microscopic appearances of the diseased tissue. In general, histopathologists are primarily concerned with the diagnosis of cancer: samples of tissue, removed at surgery from patients with suspected cancer, are processed to produce thin tissue sections, which are mounted onto glass slides, stained and examined under the microscope by a pathologist. This histopathological diagnosis contributes to the planning of additional molecular investigations and the delivery of appropriate treatment for individual patients. Advances in technology have seen some pathology services adopt digital pathology to improve the efficiency of the diagnostic process. Digital pathology entails the digitisation of histopathology glass slides to produce high resolution digital whole slide images that can be viewed using a web-based platform [1].

In 2017, the US Food and Drug Administration (FDA) approved the marketing of the first digital pathology solution (Philips IntelliSite Pathology Solution) [2], and this has provided an increased impetus for the application of digital pathology in diagnostic practice. The availability of digital images has facilitated the development of image analysis and the application of artificial intelligence/machine learning tools: giving rise to an emerging discipline of computational pathology [3].

Computational pathology has the potential to transform pathology; reducing the volume labour-intensive tasks such as quantification of particular histopathology parameters, e.g. area of tumour, number of mitotic figures [4]. It can also be used where there may be inter-observer variation between pathologists, e.g. requirements to score clinically significant tissue features such as HER2-positivity in breast cancer [5]. It is also possible for machine learning to provide predictive diagnostic information from histological digital image data [6]. The development in automated/predictive diagnostics is important as it attempts to overcome inter- pathologist variation in histopathological scoring and it could also potentially provide predictive information: both important in terms of ensuring that patients receive the most appropriate treatment and follow-up for their condition.

Although it is often not explicitly stated, it would appear that most of the studies in machine learning are carried out in formalin fixed paraffin embedded (FFPE) tissue. This type of tissue is the type that is conventionally used in diagnostic laboratories; tissue fixation ensuring that tissue architecture and morphology are adequately retained to aid diagnosis. Furthermore, institutional pathology archives contain large amounts of FFPE tissue, facilitating access for appropriately consented research.

Advances in molecular pathology have necessitated the use of fresh frozen (FF) tissue as this is considered optimal for gene expression studies. This has been highlighted most recently in the 100, 000 Genome Project, co-ordinated by Genomics England, which demonstrated that the quality of DNA isolated from FFPE tissue was not of sufficiently high quality for whole genome sequencing (WGS) and recommends instead that FF samples are used [7]. This requires pathologists to examine FF tissue sections and confirm that they meet the criteria for inclusion in the WGS study: tumour must account for at least 40% of nucleated cells and that less than 20% of the area being necrotic. The use of FF tissue raises a number

978-1-7281-2741-5/19/$31.00 ©2019 IEEE

of challenges to busy histopathology departments that may not have the capacity to process frozen tissue samples. Furthermore, if FF tissue samples can be processed the next potential point of delay is in confirming that the tissue sample meets the inclusion criteria for WGS.

A. Deep Learning in Medical Image Analysis

In recent years, machine learning methods that use deep Convoluted Neural Networks (CNNs), to learn patterns of the image based on a large training data set, have been used to develop Computer-Aided Diagnosis (CAD) in medical images [8]. CNNs are found to be effective in automatically extracting the necessary set of discriminative features for a given classification task. This has been demonstrated, for example, using the AlexNet model which has a well-established architecture and has shown good performance when used for natural images as well as medical images [9].

The AlexNet model was developed by Krizhevsky and colleagues and is trained on 1.2 million images of the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2010 dataset [10]. The ILSVRC 2010 training set contains 1000 different categories, representing objects such as flowers, vehicles, animals and so on. The structure of the deep AlexNet architecture as shown in Figure 1 and is composed of a stack of eight layers: the first five layers are convolutional and the remaining three are fully-connected. The input layer is configured to take a fixed sized 227x227 RGB image as an input; all the training images are normalised to get the same range of values for each of the input features.

The first convolutional layer filters the input image with 96 kernels (size: 11x11x3) with a stride of 4 pixels. The second convolutional layer filters the output (normalised and pooled) of the first convolutional layer with 256 kernels (size: 5x5x48). The third, fourth, and fifth convolutional layers are connected without any intervening pooling or normalisation layers. The third convolutional layer has 384 kernels (size: 3x3x256), the fourth convolutional layer has 384 kernels (size: 3x3x192), and the fifth convolutional layer has 256 kernels (size: 3x3x192). The fully-connected layers have 4096 neurons each. The output of the last fully connected layer is fed to 1000 categories. A softmax layer produces the probability distribution for the outputs of the last fully connected layer converts them to real values between zero and one.

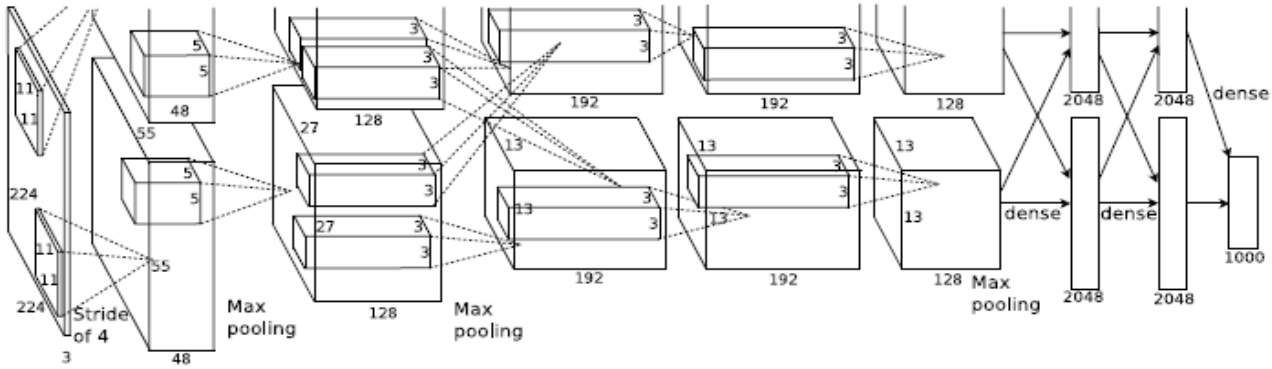

Figure 1. The AlexNet Architecture (as proposed in Krizhevsky et al. 2012).

Wang and colleagues [8] demonstrated that SVM belongs to the kernel-based classifier family, which implicitly maps the input features into a higher dimensional feature space using a kernel function that measures the distance between feature points in the mapped space. Thus an SVM is able to achieve much better classification performance than conventional linear classification methods. Furthermore, Sequential Minimal Optimization (SMO) for SVM with feature selection has been used to overcome on the Quadratic Programming (QP) problem that appears during the training of large dataset [11].

From a mathematical point of view, Fan and colleagues [12], consider a binary classification problem with a dataset $(x_{,i}, y_i), i = 1,2,..n)$, where x_i is an input vector and $y_i = \pm 1$ is a binary label corresponding to it. A soft-margin SVM is trained by solving a quadratic programming problem, which is expressed in the dual form such that,

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} y_i \, y_j \, k(x_i, x_j) \, \alpha_i \, \alpha_{j,} \qquad (1)$$

subject to, $0 \le \alpha_i \le C, for\ i = 1,2, \dots, n, \sum_{i=1}^{n} y_i \, \alpha_i,$

where $K(x_{,i}, x_j),$ is the kernel function, the variables α_i are Lagrange multipliers and C is the penalty parameter of the error term and it controls the cost of the misclassification on the training set. C keeps the allowable values of the Lagrange multipliers α_i in a bounded region.

For solving a large QP optimisation problem [11], SMO breaks this large QP problem into a series of smallest possible problems. These small QP problems are solved analytically, which avoids using the time-consuming numerical QP optimisation as an inner loop. The amount of memory required for SMO is linear in the training set size, which allows SMO to handle very large training sets. Because of the linear equality constraint involving the Lagrange multipliers α_i, the smallest possible problem involves two such multipliers. Then, for any two multipliers α_1 and α_2, the constraints are reduced to [12],

$$0 \le \alpha_1, \alpha_2 \le C, \qquad (2)$$

$$y_1 \alpha_1 + y_2 \alpha_2 = k. \qquad (3)$$

The problem has is considered solved when algorithm tries to modify all the Lagrange multipliers to satisfy the Karush–Kuhn–Tucker (KKT) conditions. Although this algorithm is guaranteed to converge, heuristics are used to choose the pair

of multipliers so as to accelerate the rate of convergence. This is critical for large data sets, since there are $n(n-1)/2$ possible choices for α_i and α_j [12]. Similarly, it has been demonstrated that training set of data should be considerably larger and varied in order to give more accurate results [13,14,15].

Dosovitskiy and colleagues investigated the role of data augmentation in deep learning to get enough different samples which needed to train a CNN from an image dataset [16]. Sokolova and Lapalme extracted these terms from the confusion matrix. A 2×2 confusion matrix was used to represent prediction results of the set of two pathological samples - HNSCC tumour and Normal histology. The standard performance measurements were formulated as depicted in the following equations [17], $accuracy = \frac{tp+tn}{tp+fn+fp+tn}$, to evaluate the overall effectiveness of a classifier, $precision = \frac{tp}{tp+fp}$, to evaluate the class agreement of data labels with the positive labels given by the classifier,

$Sensitivity = \frac{tp}{tp+fn}$, to evaluate the effectiveness of data of a classifier to identify positive labels,

$specificity = \frac{tn}{tn+fp}$, to evaluate the effectiveness of a classifier to identify negative labels and

$AUC = \frac{1}{2}\left(\frac{tp}{tp+fp} + \frac{tn}{tn+fp}\right)$, to evaluate the classifier's ability to avoid false classification,

where tp, tn, fp, fn are true positives, true negatives, false positives and false negatives respectively.

II. Methodology

Our methodology for the binary image classification problem discussed here utilises the AlexNet architecture as described above. We use a sufficient dataset of pre-processed images to train the CNN and then a different sample to test the accuracy of the classification using SVM as well as transfer learning. Figure 2 shows a block diagram of our methodology details of which we describe below.

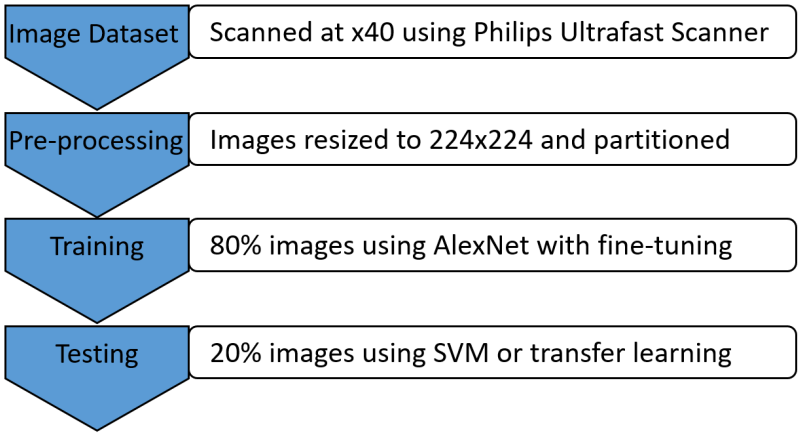

Figure 2. Block diagram showing our proposed methodology for classification.

A. Image Aquistion

All tissue used was from consented, anonymised patients and was provided by the Ethical Tissue University of Bradford (licenced by the Human Tissue Authority). The samples were collected from surgical operations performed between 2010-2012; were snap frozen at the time and stored at -80˚C. Fresh frozen (FF) samples of head and neck tumours (n=8) and normal tissue (n=8) were embedded in Tissue Tek OCT embedding compound and 6 mm sections were cut on a cryostat (Brights) and taken onto 3-Aminopropyl-triesthoxysilane (APES) coated slides.

The sections were air dried and fixed in 10% paraformaldehyde for 10 minutes at room temperature. All sections were stained with haematoxylin and eosin (H&E). The glass slides were scanned using the Philips Ultrafast Scanner at x40 magnification with high resolution image sensor (0.25mm/pixel). Digital histological images were visualised using the Philips Image Management System. In Figure 3, we show the process of acquiring images from the software interface connected to the Philips Ultrafast scanner.

For the experiments we conducted, a total of 529 slides utilised. For the purpose of machine learning, it was essential that the images selected were of similar morphology. Of the slides considered, 47 of them showing cross-sections through the tongue were selected for this project, of which 5 were excluded as they did not meet the criteria of the study, i.e. images that failed to download or were of poor quality.

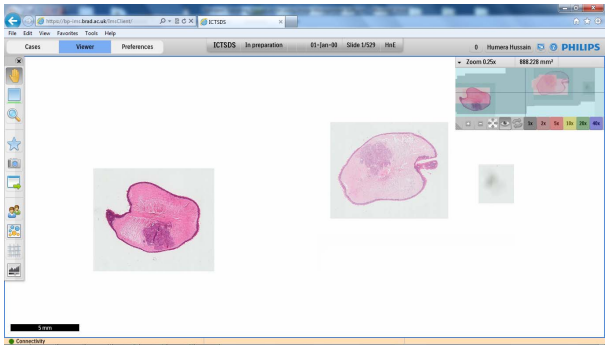

Figure 3. Illustration of the process of acquiring images from the software interface connected to the Philips Ultrafast scanner.

B. Pre-processing

Using the above mentioned process, a total of 1,424 histological images, derived from sections of FF tissue, were used to train and test the AlexNet CNN model. In order to deal with the difference in the number of images among two type classes (tumour and non-tumour), data augmentation was

introduced. The images in the dataset were rotated (90, 180 and 270), flipped left-to-right horizontally and then vertically to create a larger sample size and to make the approach recognise tumour in different orientations, as shown in Figure 4. The study was performed in a MATLAB environment using methods described in [13,14]. For the machine learning, 80% of data were randomly chosen for training, and the remaining 20% were used for testing.

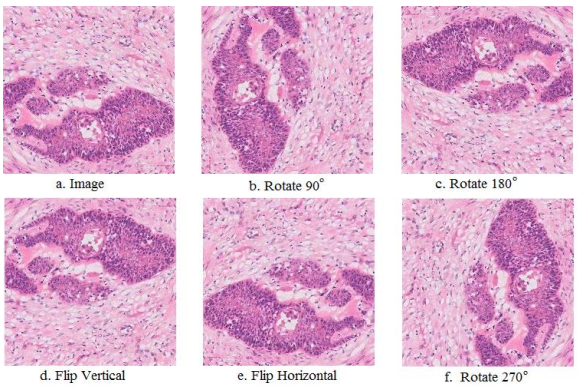

Figure 4. Illustration of data augmentation of an original image in preparation for training the AlexNet CNN.

The AlexNet model requires input images of size 227x227. The training and test images were therefore resized to height 227 and width 227, before they are inputted to the pre-training AlexNet network. The SVM classifier and transfer learning network were evaluated in terms of standard performance such as tp, tn, fp, fn as described earlier.

III. Experiments and Results

In this section, we discuss some of the experiments and the corresponding results for the classification problem at hand whereby the above discussed setup is applied.

A. Using the Binary SVM Classifier

The SVM classifier uses a pre-trained AlexNet CNN as a feature extractor by using the layer activations as features. In this study, we used 10-fold cross-validation to evaluate the classifier.

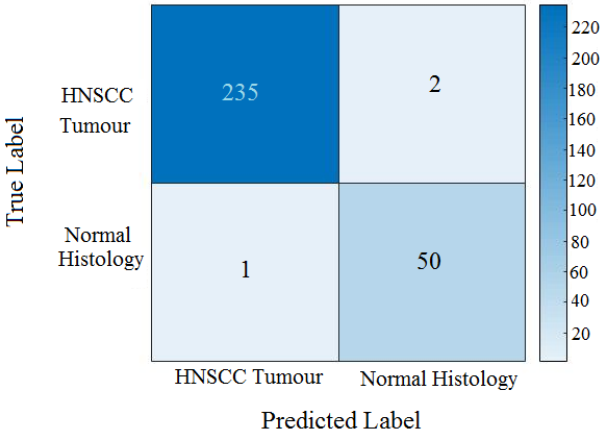

Figure 5. Description of the confusion matrix for the SVM Classifier.

For each of the cross-validation, the performance values were calculated for each feature set based on the nine folds of training samples, via grid search in the parameter space. Therefore, each cross-validation might have slightly different values, and the average optimal value was reported. SVM is used a linear function as the kernel function whereby Layer 'fc7' is used to extract features from the images and the SMO method is used to find the separating hyper plane, with an average kernel size was 2.0.

The SVM classifier was tested and was found to be performing at an accuracy rate of 0.9896 within approximately 46 seconds on a standard laptop computer. The corresponding 2x2 confusion matrix is shown in Figure 5. For HNSCC tumour, the tp is 235, fp is 2, tn is 50 and fn is 1. In Table 1, we report the results of this experiment.

Table 1. The standard performance indicators for the SVM classifier.

	Acc(%)	P	S	Spe	AUC
Results	**98.96**	**.9860**	**.9787**	**.9787**	**.9823**

Using the Transfer Learning Network

The transfer learning strategy takes layers from a pre-trained CNN on a large data to fine tune on a new data set. The last three layers of the AlexNet model are configured for 1000 classes. In order to fine tune these three layers for the new binary classification problem, we extract all layers, except the last three layers (Fully Connected 'fc8', Softmax 'prob' and Classification Output 'output'). These three layers were replaced by new layers.

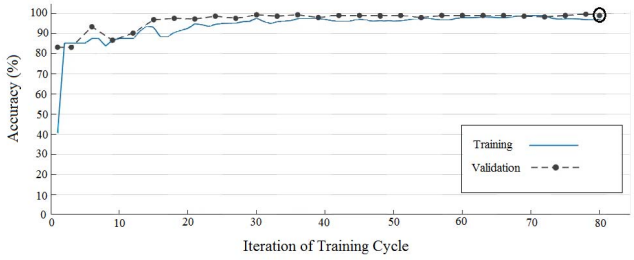

Figure 6. Illustration of the training process for the transfer learning network.

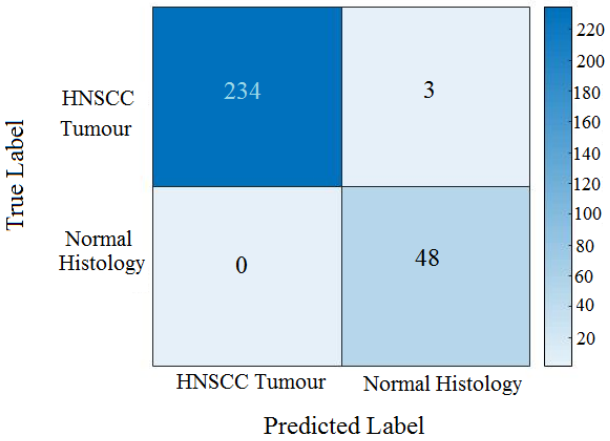

Figure 7. Description of the confusion matrix for the transfer learning network.

Table 2. The standard performance indicators for the transfer learning network.

	Acc(%)	P	S	Spec	AUC
Results	**98.95**	**.9937**	**.9706**	**.9706**	**.9706**

To retrain the AlexNet model, the new layers can be employed with FF histological images and the final fully connected layer can be set to have two classes. Therefore, during retraining, the weights of layers can be preserved

while the last fully connected layer can be updated continuously. The initial learning rate for training is set at 0.0001 as a starting point, the minimum batch size is 128, Verbose Frequency is 50 and the optimisation technique used is Stochastic Gradient Descent (SGD) with a momentum of 0.9. The algorithm validates the network every three iterations during training. For each training cycle, 10 epochs and 8 iterations per epoch were chosen. As a result of the test made, the accuracy rate for the classification was obtained as 0.9895 after 17 minutes and 40 seconds with 10 epochs of training.

Figure 6 shows the training process of the transfer learning model. It can be observed, from the 4th epoch, the accuracy rate of 98% was achieved, and a mini batch loss was obtained near to zero. The reason for ending Epoch 10 is that the error falls slowly from the 4th epoch. In Figure 7, we show a 2x2 confusion matrix for the HNSCC tumour where we can see that tp is 234, fp is 0, tn is 48 and fn is 0. The performance values are reported in Table 2.

IV. CONCLUSIONS

This study was carried out to investigate the use of machine learning algorithms in the prediction of tumour content in Fresh Frozen (FF) histological samples of head and neck tumours. These images of normal and cancerous tissues display a very complex distinctive geometric structure. The selected algorithm for identifying tumour should be able to extract discriminative features. In this paper, a pre-trained convolutional neural network (CNN) was adapted for performing tumour classification on FF image datasets.

Deep AlexNet CNN was utilised to build two common machine learning classifiers. For the first classifier, the pre-trained AlexNet network was used to extract features from the activation layer and then the SVM classifier was trained by using these extracted features. The second classifier replaces the last three layers of pre-trained AlexNet network which is configured for 1000 classes by new layers for binary classes (normal and tumour) and then fine-tune these layers on FF histological images.

In this study, a total of 1,424 histopathological images of FF tissue were used to detect the tumour cells in samples. The performance of the classifiers was evaluated according to standard performance criteria. Although the number of images was minimal, the results show very good accuracy among both classifiers and appears to be performing in a consistent manner.

Thus, from the results we have obtained, the use of AlexNet CNN with SVM or transfer learning for classification of tumours in images from FF appears to be promising. Despite being a small scale project, several steps were taken to ensure that the best results were achieved and potential bias removed. This included using high quality images, increasing the number of images; ensuring that images were augmented, using the holdout method to split the training and testing images ensuring that the test dataset was significantly smaller than the training dataset.

The results indicate that, there was little difference between SVM and transfer learning for classification. Thus, both appear to have the high classification accuracy of 98.96% and 98.95% for SVM and transfer learning, respectively. Thus, the present study shows that the AlexNet CNN with both the chosen classifiers are the suitable for classifying tumour content in FF histological samples of head and neck tumours.

Recently, studies appear to be emerging which observes the effect of CAD technology on disease diagnosis, providing promising results. However, even if they are being deployed, they are currently being used alongside conventional methods. The next step in studies of this nature would be to develop fully automated systems that can be incorporated into the pathology workload. With technological advancements and a growing interest in automated diagnostics, the use of fully automated systems, for example based on what we have proposed here, appears quite tenable in the faceable future.

ACKNOWLEDGMENT

This work was supported in part by the European Union's Horizon 2020 Programme H2020-MSCA-RISE-2017, under the project PDE-GIR with grant number 778035.

REFERENCES

[1] C. A. Bertram and R. Klopfleisch, The Pathologist 2.0: An Update on Digital Pathology in Veterinary Medicine, *Vet. Pathol,*, 54(5), 756-766, (2017).

[2] FDA, FDA allows marketing of first whole slide imaging system for digital pathology, (2018). https://www.fda.gov/newsevents/newsroom/pressannouncements/ucm552742.htm

[3] T.J. Fuchs and J.M. Buhmann, Computational pathology: challenges and promises for tissue analysis, *Comput, Med. Imaging. Graph.*, 35(7-8), 515-530, (2011).

[4] D. Komura, and S. Ishikawa, Machine Learning Methods for Histopathological Image Analysis, *Computational and Structural Biotechnology*, 16, 34-42, (2018).

[5] M. E. Vandenberghe, M. L. J. Scott, P. W. Scorer, M. Söderberg, D. Balcerzak and C. Barker, Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer, *Scientific Reports*, 7, 45938, (2017)

[6] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. Velázquez Vega, D. J. Brat, and L.A.D. Cooper, Predicting cancer outcomes from histology and genomics using convolutional networks, *Proceedings of the National Academy of Sciences*, (2018).

[7] 1,00,000 Genome Project Update, (2019). https://www.genomicsengland.co.uk/100000-genomes-project-update/

[8] H. Wang, Z. Zhou, Y. Li, Z. Chen, P. Lu, W. Wang, W. Liu and L. Yu, Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images, *EJNMMI Research*, 7(1), 11, (2017).

[9] T. Araujo, G. Aresta, E. Castro, J. Rouco, P. Aguiar, C. Eloy, A. Polonia, and A. Campilho, Classification of breast cancer histology images using Convolutional Neural Networks, PLoS One, 12(6), e0177544, (2017).

[10] A. Krizhevsky, I. Sutskever and G.E. Hinton, ImageNet classification with deep convolutional neural networks, NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, 1, 1097-1105, (2012).

[11] A. Urmaliya and J. Singhai, Sequential minimal optimization for support vector machine with feature selection in breast cancer diagnosis, *2013 IEEE Second International Conference on Image Information*

Processing (ICIIP-2013), Shimla, 481-486, (2013).
[12] R.-E. Fan, P.-H. Chen, and C.-J. Lin, Working Set Selection Using Second Order Information for Training Support Vector Machines, *J. Mach. Learn. Res.,* 6, 1889-1918, (2005)..
[13] A. Elmahmudi and H. Ugail, Deep face recognition using imperfect facial data, *Future Generation Computer Systems*, 99, 213-225, (2019).
[14] A. Elmahmudi and H. Ugail, Experiments on Deep Face Recognition Using Partial Faces, *2018 International Conference on Cyberworlds (CW)*, 357-362, (2018).
[15] A. Bukar, H. Ugail and D. Connah, Automatic age and gender classification using supervised appearance model, *Journal of Electronic Imaging*, 25(6), 061605, (2016).
[16] A. Dosovitskiy, J.T. Springenberg and T. Brox, Unsupervised feature learning by augmenting single images, (2018). https://arxiv.org/abs/1312.5242
[17] M. Sokolova and G. Lapalme, A systematic analysis of performance measrues for classification tasks, *Information Processing and Management*, 45, 427-437, (2009).