Dual structured convolutional neural network with feature augmentation for quantitative characterization of tissue histology

Mira Valkonen⋆, Kimmo Kartasalo, Kaisa Liimatainen, Matti Nykter, Leena Latonen, Pekka Ruusuvuori
Faculty of Medicine and Life Sciences,
BioMediTech, University of Tampere, Tampere, Finland
Tampere University of Technology, Finland
⋆mira.valkonen@staff.uta.fi

Abstract

We present a dual convolutional neural network (dCNN) architecture for extracting multi-scale features from histological tissue images for the purpose of automated characterization of tissue in digital pathology. The dual structure consists of two identical convolutional neural networks applied to input images with different scales, that are merged together and stacked with two fully connected layers. It has been acknowledged that deep networks can be used to extract higher-order features, and therefore, the network output at final fully connected layer was used as a deep dCNN feature vector. Further, engineered features, shown in previous studies to capture important characteristics of tissue structure and morphology, were integrated to the feature extractor module. The acquired quantitative feature representation can be further utilized to train a discriminative model for classifying tissue types. Machine learning based methods for detection of regions of interest, or tissue type classification will advance the transition to decision support systems and computer aided diagnosis in digital pathology. Here we apply the proposed feature-augmented dCNN method with supervised learning in detecting cancerous tissue from whole slide images. The extracted quantitative representation of tissue histology was used to train a logistic regression model with elastic net regularization. The model was able to accurately discriminate cancerous tissue from normal tissue, resulting in blockwise AUC=0.97, where the total number of analyzed tissue blocks was approximately 8.3 million that constitute the test set of 75 whole slide images.

1. Introduction

In digital pathology, analysis of histopathological images is mainly time-consuming manual labor and prone to subjectivity, and sometimes a challenging task even for an expert [30]. Automated image analysis provides methods to analyze these images in a quantitative, objective, and efficient way [5, 22]. The development of accurate image analysis tools can advance the transition to decision support systems and computer aided diagnosis in digital pathology [15].

Two common approaches to supervised learning based image analysis include, traditional feature engineering methods combined with machine learning, and deep learning based methods, such as deep convolutional networks.

Traditional feature engineering methods rely on expertise and knowledge on the important discriminative properties and methods to manually engineer these features for a certain application. Quantitative feature representation combined with machine learning methods provide powerful image analysis tools. One advantage of these type of models is that they are relatively easy to interpret by relating the features of the classifier model to the biological information. These traditional methods have been used for years in digital pathology and in many different applications, such as, cell detection and classification [1], grading prostate cancer both in human [7], and mouse model [21] samples. Additionally, these traditional approaches have been successfully used for quantitatively describing characteristics of histology [26, 6]. However, previous studies have left room for improved feature engineering and classification performance.

Neural network based methods, on the other hand, do not need any specific feature engineering, these methods learn both the feature representation as well as the classification model. In recent years, these methods have become the state-of-art methods in many image classification and detection tasks [24, 31, 2, 9]. While the the deep learning approach has been shown to outperform traditional approaches in a variety of tasks, the model interpretation and link to biological information remains largely unresolved.

Convolutional neural networks and deep learning can be

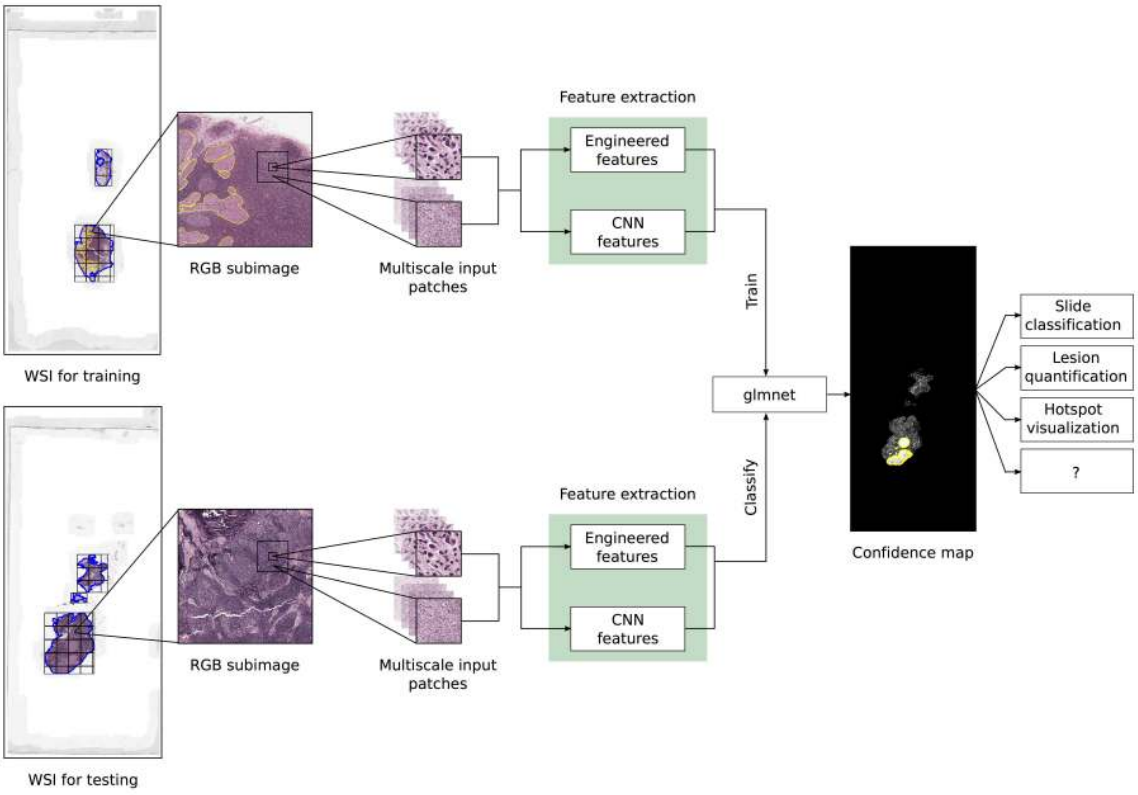

Figure 1. Image analysis system workflow. The upper half presents the training of the model and the lower half presents the steps for classifying an unseen image with the trained model. The feature extraction module consists of two parts, extraction of deep dCNN features and extraction of engineered features. As an output, the model provides confidence map presenting the probability of certain tissue region to contain cancerous tissue.

also used for extracting features. This provides the opportunity to combine these two common approaches. The accuracy of deep learning combined with engineered features and machine learning could provide classification models for image analysis with high level of performance and increased model interpretability. With network architecture visualizations, these deep CNN feature representations can be better understood and linked to spatial information. There are studies that have established the benefits of combining these two methods, and achieving even better performance than either of the methods alone [32, 16, 23].

To mimic the way that a pathologist view and analyze histological sample, first from distance and then with a closer look, we implemented a method to characterize tissue histology in a multi-scale manner. A dual convolutional neural network (dCNN) architecture was developed for extracting multi-scale features from histological tissue images. The dual structure consists of two identical convolutional neural networks working with different image scales that are merged together and stacked with two fully connected layers. In addition, manually engineered features were integrated to the model by concatenating vector of manually engineered features with the output vector of last fully connected layer of the dCNN. We evaluated the proposed feature-augmented dCNN method with supervised learning in detecting cancerous tissue from whole slide images.

2. System overview

In this work, an image analysis system was implemented that utilize deep dual CNN with feature augmentation to automatically extract quantitative characteristics from histological images. Supervised machine learning was then applied to train a discriminative model based on these quantitative histological characteristics. The system was implemented using python programming language and extends the feature engineering based method presented by Valkonen et al. [27, 28]. Figure 1 presents an overview of the implemented feature-augmented dCNN method.

The data used in this study consists of two independent hematoxylin and eosin stained whole slide image (WSI)

datasets, first collected at the Radboud University Medical Center (Nijmegen, the Netherlands) and the second set at the University Medical Center Utrecht (Utrecht, the Netherlands). A total of 140 WSIs present normal lymph node sections and 130 present pathologists annotated WSIs of breast cancer micro- and macro-metastases. The WSIs and the corresponding annotation masks were provided as multi-resolution pyramids in Phillips BigTIFF format. The pixel size of the images at the full resolution level was 243 nm. Both datasets were provided for Camelyon16 challenge [8]. This Challenge was organized in conjunction with the 2016 IEEE International Symposium on Biomedical Imaging (ISBI-2016).

The image data was divided into training set and test set. The training set consisted of approximately 80 percent of the images, 195 WSIs in total. The test set included the remaining 75 WSIs.

In order to simplify the classification task and to reduce the amount of data, first a coarse segmentation step was performed for each image to detect the lymph node tissue while excluding the background and most of the adipose tissue.

The visual appearance of the two independent image sets is quite notable due to different scanner and staining protocols. Therefore, histogram matching was applied to correct this color variation across the WSIs. Histogram matching was applied separately to each color channel. A training image Tumor 015.tif was selected as the reference based on visual examination.

For convenient handling of the image data during model training and classification, the images were divided into smaller subimages and stored in JPEG2000 format, along with information of their location within the WSI. Each resulting subimage had approximately the dimensions of 8192×8192 pixels. The tissue segmentation and ground truth masks were processed similarly and saved in TIF format.

Color deconvolution [20] was applied before extracting engineered features. In this study, scikit-image [29] implementation of the color deconvolution algorithm was used. The deep dCNN features were extracted from RGB images.

Each subimage was processed blockwise in a way that each 128×128 block is presented with vector of 1344 feature values. These features include 256 deep dCNN features and 1088 engineered features extracted from tissue image with a multi-scale perspective. Thus, the quantitative representation of a 128×128 pixel block includes also the information about its neighboring area.

The quantitative feature representations was then used to train a logistic regression model for detecting cancerous tissue. The trained model provides a confidence map that presents the probability of each tissue block to belong to the positive (tumor) class.

3. Dual structured convolutional neural network

3.1. Network architecture

The architecture of the dCNN is presented in figure 2. Keras module [3] was used to build two parallel sequential models merged into one model. The network takes two RGB images as input: patch of 128×128 image blocks and patch of 640x640 image blocks. Both of the input images were downsampled into size of 32×32×3 pixels. The smaller block covers approximately a $31\mu m$ physical area of the tissue and the wide neighborhood covers a $155\mu m$ area of the tissue.

Both of the parallel networks include four two-dimensional convolutional layers followed by rectified linear unit (ReLU) activation and max pooling layer. The kernel size was 3×3 for each convolutional layer and 2×2 for max pooling. These two parallel networks are merged together with two fully connected layers of length 256 followed by ReLU activation. The output layer consists of one fully connected output node with sigmoid activation. Total number of trainable parameters in the network is 3 758 337.

Binary cross-entropy was used as loss function and the optimization procedure was performed using stochastic optimization algorithm Adam [14]. For Adan algorithm, step size was set to 0.001, exponential decay rates for the moment estimates (β_1 and β_2) were set to 0.9 and 0.999, and the used fuzz factor was $1 * 10^{-8}$. Learning rate decay over each update was set to 0.

3.2. Training of the dCNN

The dCNN was trained with sample blocks from the 195 training WSIs. The training set was divided into 12 sets to prevent the dCNN model from overfitting for one training set. Each training set included same proportion of images from both of the independent datasets. The whole dataset included more normal tissue area than cancerous tissue. Therefore, to balance the training data, random sampling of the normal samples was done to get even number of tumor samples and normal tissue samples. Each of the 12 sets was used to train the CNN model for 5 epochs with a batch size of 100 samples. The model was trained for 60 epochs, and in total, with approximately 6.4 million tissue sample blocks. Only samples with 50% coverage of the tissue mask or tumor mask within the smaller scale block were accepted for training. The training accuracy and loss are presented in figure 3.

The final dense layer before output layer in the deep dCNN architecture was used as deep dCNN feature vector. This is visualized in figure 2. Thus, one 128×128 pixel block was represented with 256 deep dCNN features.

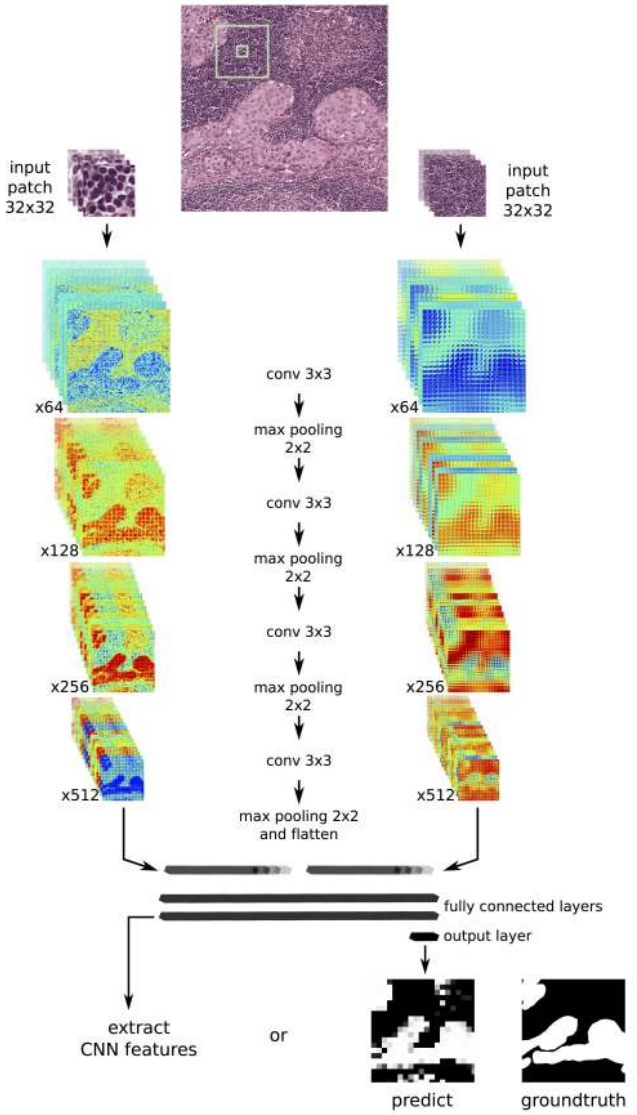

Figure 2. The implemented deep dCNN architecture is composed of two identical and parallel convolutional neural networks that are merged into one model and stacked with two fully connected layers.

4. Feature augmentation

Engineered features have shown to capture important characteristics of tissue structure and morphology [12, 17, 27]. Therefore, the dCNN model was augmented with module for extracting set of manually engineered features, 272 different features in total. This same feature extraction module was applied four times, for tissue blocks of both hematoxylin and eosin channels with two different scales. In total, one 128×128 pixel block can be presented with 1088 manually engineered features. Scikit-image implementations for extracting manually engineered features were used in this study [29].

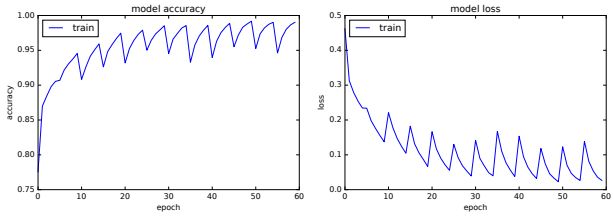

Figure 3. The training accuracy and loss obtained after training the dCNN model with 12 training sets, 5 epochs each.

The engineered features included first and second order statistical texture features derived from image histogram and gray-level co-occurrence matrix (GLCM [13]). These features included for example, mean intensity, variance, skewness, angular second moment, etc. These co-occurring gray-level value properties were calculated at offset distance of one pixel and with respect to four angles: $[0°, 90°, 180°, 270°]$.

Engineered features were extracted also using local binary patterns (LBP [18]). The basic idea of the LBP operator is to transform a local circular neighborhood into a binary pattern by thresholding the neighborhood with the gray value of the center pixel. This circularly symmetric neighborhood is determined by assigning parameters that control the quantization of the angular space and radius of the neighborhood. For the implemented method, the LBP responses were calculated for two different circularly symmetric neighborhood: radius of 3 pixels and angular space of 24 points, and radius of 5 pixels and angular space of 40 points. The response image to these both LBP operators over a sample block present uniform LBP patterns in the sample block. Properties of a discrete occurrence histogram of these response images were used as features.

Histogram of Oriented Gradients (HOG [4]) were also used as engineered features. This method counts occurrences of gradient orientation in local portions of an image. First, gradient image is computed and then histograms of gradient orientation angles are computed within smaller cells of the image. Finally, the orientation histograms of all cells are flattened into feature vector.

Finally, we included also the so called Daisy features [25]. Daisy is an efficient dense descriptor for extracting local features from an image. Daisy descriptor is a combination of Scale-invariant feature transform (SIFT) and Gradient Location and Orientation Histogram (GLOH) descriptor. The Daisy descriptor is mainly based on computing Gaussian convolutions. Number of circles was set to two. From each circle, altogether six histograms were calculated with eight orientation bins.

5. Generalized linear model

The probability to belong to the group of cancerous tissue was predicted with trained generalized linear model for all tissue blocks of each WSI in the test set. The model applied for prediction was a logistic regression model with elastic net regularization.

Python package for fitting generalized linear model (glmnet [11]) was used in this study. This model is fitted by minimizing the negative binomial log-likelihood. This loss function for the model can be written as

$$\min_{\beta_0,\beta} -\left[\frac{1}{N}\sum_{i=1}^{N} y_i * (\beta_0 + x_i^T\beta) - \log(1 + e^{(\beta_0 + x_i^T\beta)})\right] + \lambda P_\alpha(\beta) \quad (1)$$

where y_i presents the log-likelihood, N is the number of observations, $P_\alpha(\beta)$ the elastic net regularization term, and λ controls the overall strength of the regularization. The elastic net regularization can be written as

$$P_\alpha(\beta) = (1-\alpha)\left\|\beta\right\|_2^2/2 + \alpha\left\|\beta\right\|_1 \quad (2)$$

A quadratic approximation of the log-likelihood is used, which results in penalized weighted least-squares problem. Finally, coordinate descent is used as an optimization procedure for the model weight update by minimizing the weighted least-squares problem of loss function.

The α was set to 0.5, therefore, the elastic net regularization uses both lasso and ridge. The glmnet algorithm fits series of models to determine the λ value that controls the overall strength of the regularization. In total, 100 values for λ was computed. After computing the path of λ values, the performance of the model is analyzed using 3-fold cross validation. The value of λ which achieved the best performance in cross validation was used in prediction. The best value for λ was 0.0002.

The feature data was normalized to the range $[0\ 1]$. The number of training samples for the logistic regression was approximately 100 000 positive samples and 100 000 negative samples. The samples for training were randomly sampled from all the available training data. However, this was done in a way that the train set included samples from each WSI in the training set. In total, the train set included 773 subimages containing cancerous tissue and 10 404 subimages of normal tissue. Consequently, 130 sample blocks were randomly sampled from each tumor subimage and 10 sample blocks from each normal subimage.

After the confidence map for each subimage was obtained, the confidence subimages are stitched back into whole slide images for the purpose of visual inspection.

6. Results

The implemented feature-augmented dCNN method was evaluated with test set of 75 WSIs of lymph node sections. In addition to classifying the test WSIs using feature-augmented dCNN model with logistic regression, the test set was evaluated using the trained dCNN as a classifier, and also using logistic regression model trained with only the engineered features. Each test WSI was scored with confidence levels using three different models trained with samples from 195 WSIs.

Examples of class predictions performed by dCNN classifier, logistic regression with engineered features, and feature-augmented dCNN model with logistic regression are presented in figure 4. The example images include five 8192×8192 subimages of the tissue histology, the corresponding tumor annotation, and confidence map generated with each model. Blockwise receiver operating characteristics (ROC [10]) curve was calculated for each example image and the corresponding area under the curve (AUC) measures are also presented in the figure 4.

In order to evaluate the performance of the methods numerically, all of the confidence values from the test set were collected and blockwise ROC curve was calculated for each model. These ROC curves are presented in the first column of figure 5. Background blocks were excluded and only tissue blocks within the tissue mask area were included in the ROC analysis. The obtained AUC measure for logistic regression with engineered features was 0.969 (95% CI [0.9683, 0.9690]), for dCNN classifier the AUC measure was 0.886 (95% CI [0.8827, 0.8883]), and 0.968 (95% CI [0.9678, 0.9684]) for the feature-augmented dCNN model with logistic regression. The 95% confidence intervals were calculated using logit based method [19]. Further, the differences of confidence values of metastatic tissue blocks and normal tissue blocks were visualized with boxplots. These boxplots are presented in the second column of figure 5. On each box, the central mark presents the median value, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The dashed line extend to the last data points that are not considered as outliers. The outliers are plotted individually using the plus symbol.

Model	Glmnet: ef	dCNN prediction	Glmnet: dCNNf+ef
Classification accuracy (%)	87.9	92.8	91.3
Sensitivity (%)	87.6	94.0	91.2
Specificity (%)	93.4	70.9	91.7
F-score	0.93	0.96	0.95

Table 1. Classification performances.

In total, the test set included approximately 8.3 million

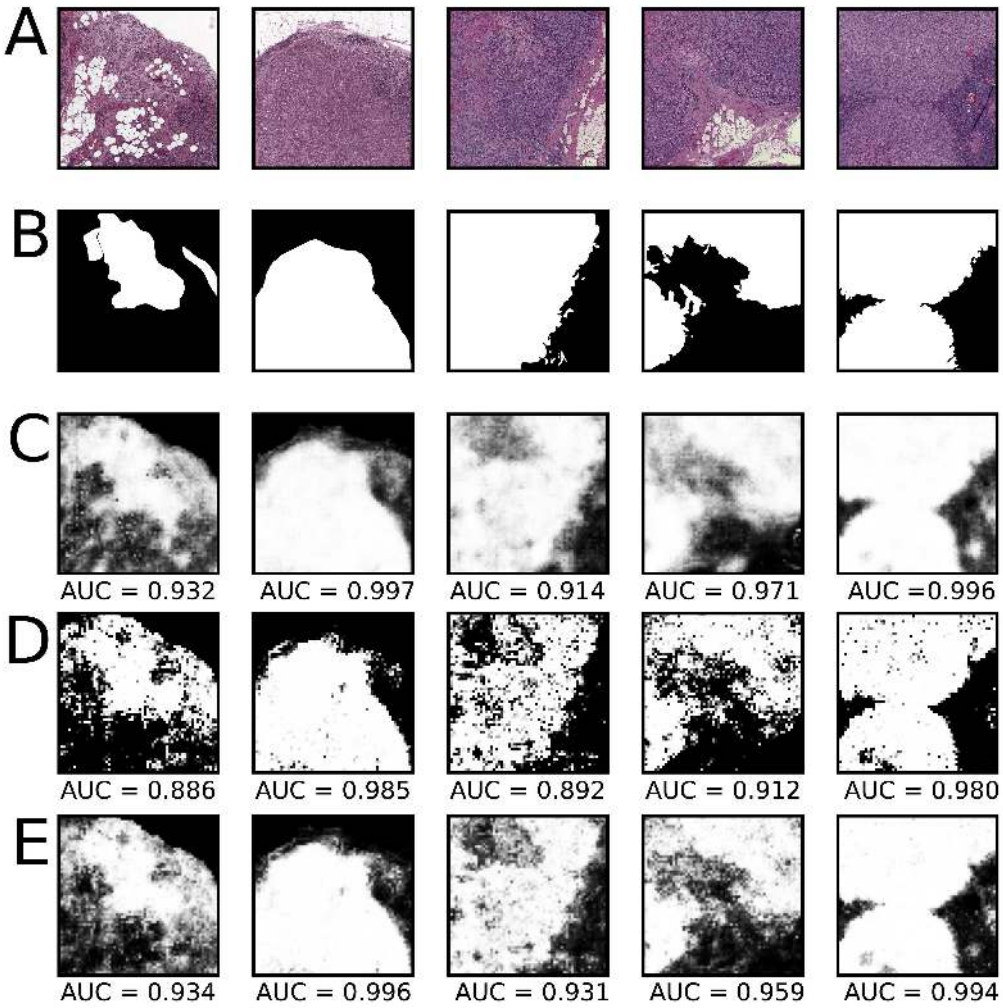

Figure 4. Confidence maps predicted for five different tissue subimage. The tissue histology for each color corrected example is presented in row **A**. Row **B** presents the groundtruth annotations, where white mask area stands for metastatic tissue. The class predictions generated using glmnet model with engineered features are presented in row **C**, predictions using the dCNN as classifier are presented in row **D**, and feature-augmented dCNN model predictions are presented in row **E**. The confidence denote the probability of a certain tissue block to belong to the group of cancerous tissue. To support the visual examination of these examples, ROC curve was calculated for each example confidence map and the corresponding AUC value is presented below each confidence map.

sample blocks (420 000 tumor sample blocks and 7 900 000 normal tissue sample blocks). Percentage of correctly classified samples, classification sensitivity and specificity, and F-score were calculated for logistic regression with engineered features, dCNN classifier, and feature-augmented dCNN model. These results are presented in table 1. The percentage of correctly classified samples was $87,9\%$ for logistic regression with engineered features, $92,8\%$ for dCNN classifier, and $91,3\%$ for feature-augmented dCNN model with logistic regression.

7. Conclusions and discussion

In this work, feature-augmented dCNN method was presented for extracting multi-scale features for quantitative characterization of tissue histology. Supervised learning was applied to train a logistic regression model with elastic net regularization to detect cancerous tissue from whole slide images. The method was able to accurately discriminate cancerous tissue from normal tissue (AUC=0.97).

To evaluate the performance and input of different features of the proposed method, classification performance was compared with two additional models. The perfor-

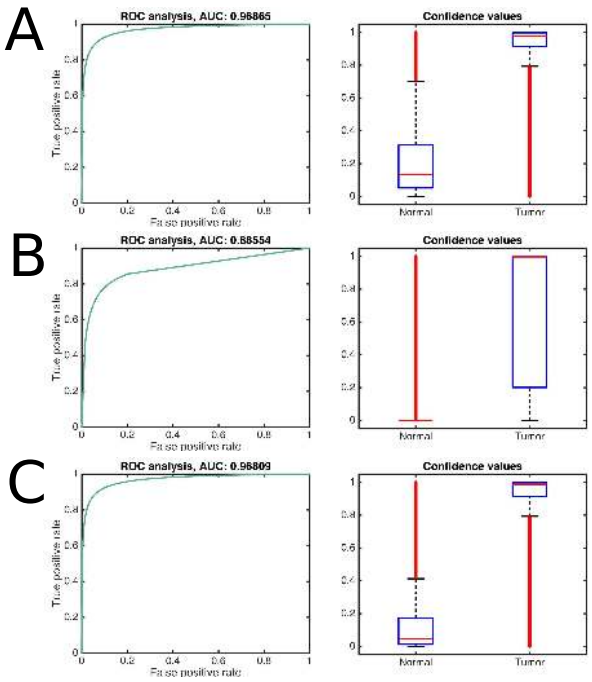

Figure 5. All of the confidence values predicted with each model were collected and blockwise ROC curve and AUC measure were calculated. Further, the differences of confidence values gathered from metastatic tissue blocks and normal tissue blocks are visualized in the boxplots. On each box, the central mark presents the median value, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The outliers are plotted individually using the plus symbol. Row **A** presents the ROC curve and boxplot for glmnet model using engineered features. Row **B** presents the ROC curve and boxplot for dCNN predictions, and row **C** presents the same plots for feature-augmented dCNN model.

mance of the proposed logistic regression model trained with engineered features and dCNN features was compared with logistic regression model trained with engineered features alone, and with the dCNN model. Each method gives as an output a confidence map that presents the probability of a certain tissue region to contain cancerous tissue. Example outputs of these three different models are presented in figure 6, where the cancerous area is annotated with magenta line. This type of output map could be beneficial for example in a usage scenario where a pathologist uses the confidence values as a guidance to potential hot-spot regions within the tissue sample.

The numerical evaluation of the these three different models are presented in table 1. The table lists the classification performance of the logistic regression model with engineered features, dCNN model, and feature-augmented dCNN model. The percentage of correctly classified samples is quite high for each case. When considering only the percentage of correctly classified samples, the dCNN model reaches the best accuracy. The second best classification performance is obtained with the feature-augmented dCNN model. The small drop in sensitivity, correctly classified samples, and F-score with feature augmentation can be explained by the weaker performance of engineered features alone considering these metrics. The most significant features provided by the glmnet model included both dCNN features and engineered features. Since glmnet gives more weight to certain significant features, the resulting model is a compromise of most discriminative features selected from both dCNN features and engineered features. Although there was small drop in some metrics, the specificity of feature-augmented dCNN model increases significantly compared to plain dCNN model, and therefore, the overall performance improves considering all accuracy metrics. The improved performance can be seen also from the ROC analysis and confidence value boxplots. These analyses confirm that the feature-augmented dCNN method provides most distinct confidence values for normal and cancerous tissue.

In addition to classification performance evaluation, several misclassification cases were considered. Most false positive signals were detected where normal lymph node medulla was misinterpreted as cancerous tissue, this can be caused by the fact that the normal lymph node stroma has similar color tones and size of nuclei as certain breast cancer cell phenotypes. Many misclassifications were also detected in the border of healthy tissue and cancerous tissue. This is caused by the fact that only samples with 50% coverage of the tissue mask or tumor mask within the smaller scale block were accepted for training. This could have been improved for example by using positive training samples with 100% coverage of the tumor mask and negative training samples with 50% coverage of the tissue mask. Most false negative signals were detected in small metastases, where single or only a few cancer cells are surrounded by lymphocytic cells.

The evaluation of the results is not an easy and straightforward task [27]. From blockwise classification point of view, certain measures can be calculated to evaluate the performance of the implemented method. However, from the pathologists viewpoint, these measures dont necessarily give real accuracy of the method. This type of blockwise evaluation gives more weight on larger tumor regions in the final scoring, since they consist of a larger number of pixels than smaller lesions. For the pathologist, this is a problem, since large macrometastases can often be spotted more easily than the small ones.

If a realistic usage of the implemented method is considered, it would be used as a decision support system, not as a tool for diagnosis. In this case, the high sensitivity of the

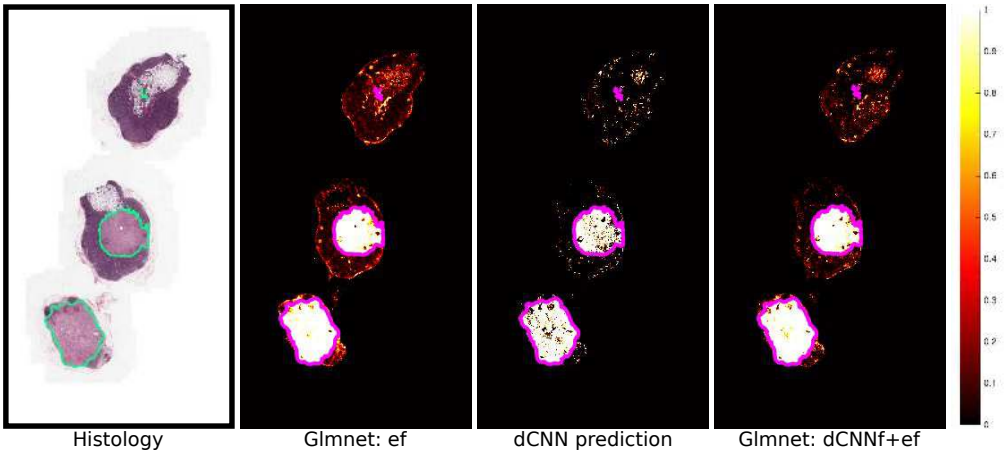

Figure 6. The confidence maps for an example WSI predicted using three different models: the logistic regression model with only engineered features, the dCNN model, and the feature-augmented dCNN model.

method, list of most discriminative features, and the confidence map output show a potential that the method could be used in the future as a software for computer aided diagnosis. High sensitivity is useful for ruling out normal tissue, since 100% sensitivity would mean that the method recognises all of the areas with metastatic tissue. Therefore, workload of a pathologist could be reduced if an automated image analysis system with close to 100% sensitivity could be used to pre-screen histological slides and to exclude even part of the slides.

Considering the results of this study from a classification point of view and the evaluation problem regarding a realistic application and its requirements for digital pathology, selecting a best method is rather inconclusive. However, the study provides proof-of-principle for using feature-augmented neural network approach in analysis of histopathological images. The main advantage of the proposed feature-augmented method, compared to accurate state-of-art deep learning methods, is the model interpretability. A list of significant features considering the classification problem is provided by the model and these features provide insight into the quantitative characteristics of tissue histology that separate metastatic tissue from normal tissue. Most of the engineered features are intuitive as such, and application specific features can be designed based on prior knowledge, such as, histogram properties that can be directly linked to stains highlighting particular biological features. Therefore, even if the feature-augmentation might bias the discovery and classification, the tradeoff with the provided information is beneficial. Certainly, the important higher-order dCNN features can be also connected to the spatial location of the tissue with network visualization. However, linking these features to any relevant biological information may be challenging. Therefore, with the combination of these two methods, advantages of both methods can be obtained; high level of classification performance and increased model interpretability.

In addition, the implemented method is modular and can be easily extended with new features or more than two support areas for the multi-scale perspective. Therefore, we anticipate that similar approach could be applied in a wide variety of biomedical image analysis problems in addition to the breast cancer metastasis detection task presented here, provided that a large set of annotated image data is available for training. Transferring the models from prediction task to another with substantially smaller amount of training data will be one of our future research goals.

References

[1] F. S. Abas, H. N. Gokozan, B. Goksel, J. J. Otero, and M. N. Gurcan. Intraoperative neuropathology of glioma recurrence: Cell detection and classification. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 9791, 2016.

[2] R. Chen, Y. Jing, and H. Jackson. Identifying metastases in sentinel lymph nodes with deep convolutional neural networks. *arXiv preprint arXiv: 1608.01658*, 2016.

[3] F. Chollet. Keras. Accessed: 2016-09-01.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 886–893. IEEE, 2005.

[5] G. Danuser. Computer vision in cell biology. *Cell*, 147(5):973–978, 2011.

[6] J. Diamond, N. H. Anderson, P. H. Bartels, R. Montironi, and P. W. Hamilton. The use of morphological characteristics and texture analysis in the identification of tissue compo-

sition in prostatic neoplasia. *Human Pathology*, 35(9):1121–1131, 2004.

[7] S. Doyle, M. Hwang, K. Shah, A. Madabhushi, M. Feldman, and J. Tomaszeweski. Automated grading of prostate cancer using architectural and textural image features. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1284–1287. IEEE, 2007.

[8] B. Ehteshami Bejnordi and J. van der Laak. Camelyon16: Grand challenge on cancer metastasis detection in lymph nodes. Accessed: 2017-04-05.

[9] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[10] T. Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[11] J. Friedman, T. Hastie, and R. Tibshirani. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1, 2010.

[12] S. B. Ginsburg, G. Lee, S. Ali, and A. Madabhushi. Feature importance in nonlinear embeddings (fine): applications in digital pathology. *IEEE transactions on medical imaging*, 35(1):76–88, 2016.

[13] R. M. Haralick, K. Shanmugam, et al. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, (6):610–621, 1973.

[14] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv: 1412.6980*, 2014.

[15] S. Kothari, J. H. Phan, T. H. Stokes, and M. D. Wang. Pathology imaging informatics for quantitative analysis of whole-slide images. *Journal of the American Medical Informatics Association*, 20(6):1099–1108, 2013.

[16] G. Li and Y. Yu. Visual saliency based on multiscale deep features. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5455–5463, 2015.

[17] A. Madabhushi and G. Lee. Image analysis and machine learning in digital pathology: challenges and opportunities. 2016.

[18] T. Ojala, M. Pietikäinen, and T. Mäenpää. Gray scale and rotation invariant texture classification with local binary patterns. In *European Conference on Computer Vision*, pages 404–420. Springer, 2000.

[19] G. Qin and L. Hotilovac. Comparison of non-parametric confidence intervals for the area under the roc curve of a continuous-scale diagnostic test. *Statistical Methods in Medical Research*, 17(2):207–221, 2008.

[20] A. C. Ruifrok, D. A. Johnston, et al. Quantification of histochemical staining by color deconvolution. *Analytical and Quantitative Cytology and Histology*, 23(4):291–299, 2001.

[21] P. Ruusuvuori, M. Valkonen, M. Nykter, T. Visakorpi, and L. Latonen. Feature-based analysis of mouse prostatic intraepithelial neoplasia in histological tissue sections. *Journal of pathology informatics*, 7, 2016.

[22] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature Methods*, 9(7):671, 2012.

[23] A. A. A. Setio, A. Traverso, T. de Bel, M. S. Berens, C. v. d. Bogaard, P. Cerello, H. Chen, Q. Dou, M. E. Fantacci, B. Geurts, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *arXiv preprint arXiv: 1612.08012*, 2016.

[24] K. Sirinukunwattana, J. P. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, et al. Gland segmentation in colon histology images: The glas challenge contest. *Medical Image Analysis*, 35:489–502, 2017.

[25] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5):815–830, 2010.

[26] V. J. Tuominen, S. Ruotoistenmäki, A. Viitanen, M. Jumppanen, and J. Isola. Immunoratio: A publicly available web application for quantitative image analysis of estrogen receptor (er), progesterone receptor (pr), and ki-67. *Breast Cancer Research*, 12(4):R56, 2010.

[27] M. Valkonen, K. Kartasalo, K. Liimatainen, M. Nykter, L. Latonen, and P. Ruusuvuori. Metastasis detection from whole slide images using local features and random forests. *Cytometry Part A*, 2017.

[28] M. Valkonen, P. Ruusuvuori, K. Kartasalo, M. Nykter, T. Visakorpi, and L. Latonen. Analysis of spatial heterogeneity in normal epithelium and preneoplastic alterations in mouse prostate tumor models. *Scientific Reports*, 7, 2017.

[29] S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, and the scikit-image contributors. Scikit-image: Image processing in Python. *PeerJ*, 2:e453, 6 2014.

[30] M. Veta, J. P. Pluim, P. J. Van Diest, and M. A. Viergever. Breast cancer histopathology image analysis: A review. *IEEE Transactions on Biomedical Engineering*, 61(5):1400–1411, 2014.

[31] D. Wang, A. Khosla, R. Gargeya, H. Irshad, and A. H. Beck. Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv: 1606.05718*, 2016.

[32] H. Wang, A. Cruz-Roa, A. Basavanhally, H. Gilmore, N. Shih, M. Feldman, J. Tomaszewski, F. Gonzalez, and A. Madabhushi. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1(3):034003–034003, 2014.