

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Pölönen, Ilkka; Rahkonen, Samuli; Annala, Leevi; Neittaanmäki, Noora

Title: Convolutional neural networks in skin cancer detection using spatial and spectral domain

Year: 2019

Version: Accepted version (Final draft)

Copyright: © Society of Photo-Optical Instrumentation Engineers (SPIE), 2019.

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Pölönen, I., Rahkonen, S., Annala, L., & Neittaanmäki, N. (2019). Convolutional neural networks in skin cancer detection using spatial and spectral domain. In B. Choi, & H. Zeng (Eds.), Proceedings of SPIE Volume 10851 : Photonics in Dermatology and Plastic Surgery 2019 (Article 108510B). SPIE, The International Society for Optical Engineering. SPIE conference proceedings, 10851. <https://doi.org/10.1117/12.2509871>

Convolutional neural networks in skin cancer detection using spatial and spectral domain

Ilkka Pölönen^a, Samuli Rahkonen^a, Leevi Annala^a, and Noora Neittaanmäki^b

^aFaculty of Information Technology, University of Jyväskylä, Mattilanniemi 2, Jyväskylä, Finland

^bDepartments of Pathology and Dermatology, Institutes of Biomedicine and Clinical Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

ABSTRACT

Skin cancers are world wide deathly health problem, where significant life and cost savings could be achieved if detection of cancer can be done in early phase. Hyperspectral imaging is prominent tool for non-invasive screening. In this study we compare how use of both spectral and spatial domain increase classification performance of convolutional neural networks. We compare five different neural network architectures for real patient data. Our models gain same or slightly better positive predictive value as clinicians. Towards more general and reliable model more data is needed and collection of training data should be systematic.

Keywords: Hyperspectral imaging, convolutional neural network, skin cancer, melanoma

1. INTRODUCTION

Skin cancers are constantly increasing problem world wide. Traditionally this has been concern of people whose skin is relatively lightly coloured and annual portion of sunlight is high. Because of increased traveling and ageing of the population, melanoma is increasing problem also in the Nordic countries. For example in Sweden,¹ 50 % of all the annual skin cancer related costs are caused by melanomas.

There is a need for tools, which are able to detect early stage skin cancers and delineate them properly from healthy tissue. With proper detection it is possible to reduce amount of re-surgeries, when part of the malignant tissue has been left to the patient in original tumor removal. This is highlighted by the fact that overall positive predictive value of clinical melanoma diagnosis is 33 %.² In non-specialised clinics this is even lower. For every melanoma removal there will be 9 to 30 non-melanoma lesions removed depending on how specialised clinic is.³ Thus, early detection will lower the treatment costs and will ensure higher survival rate.

Hyperspectral imaging is method where hundreds narrow wavebands of light are imaged simultaneously. This method will provide almost continuous spectrum for each pixel of the image as figure 1 is showing. Hyperspectral imaging is non-invasive imaging modality, because it is using only visible and near infra-red illumination to capture images. Previously we have used it in delineation of tumor border and distinguish in-situ melanoma from malignant melanoma.^{4,5}

If you look at closely two spectra in the figure 1 , it is quite easy to see that in clear cases melanoma and healthy skin have characteristic spectra. Unfortunately this is not so in all the cases. In figure 2 we have spectral distributions of malignant melanoma, lentigo-maligna, dysplastic nevus and benign nevus. We can see that these distributions are overlapping. This means that if the melanoma is hard to recognise in clinical study, it will be hard distinguish using just spectral information. Thus, it seems natural that we also utilize spatial domain in the classification task.

Further author information: (Send correspondence to Ilkka Pölönen)

Ilkka Pölönen: E-mail: ilkka.polonen@jyu.fi, Telephone: +358 400 248 140

Samuli Rahkonen: E-mail: samuli.rahkonen@jyu.fi

Leevi Annala: E-mail: leevi.a.annala@jyu.fi

Noora Neittaanmäki: E-mail: noora.neittaanmaki@fimnet.fi

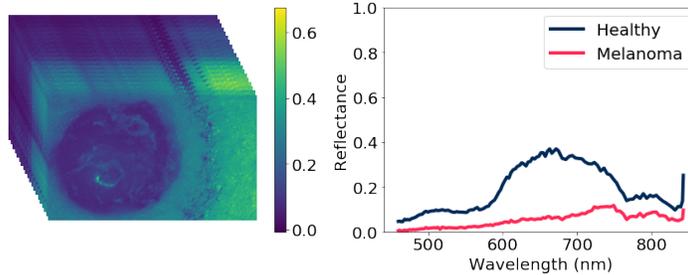


Figure 1. Hyperspectral image consist of intensity images, which are narrow wavebands from visible and near-infra red region of the electromagnetic spectrum. Each pixel in the image is spectrum through spectral domain.

Convolutional neural networks have shown great success in different kind of pattern recognition tasks.⁶⁻⁸ They have also been recently used in classifying melanomas and other skin cancers from dermatoscope and regular color images.⁹ In these cases, results are given for the whole images. Because of such binary classification we don't actually have an opportunity to determinate lesion's borders from analysis or what kind of other irregularities there are in the tumor.

There are multiple strategies to utilize convolutional neural networks. We are introducing efficient strategy, which contains utilization of both spectral and spatial domain. With hyperspectral data containing wavebands from visible to infra-red region, we are able to gather more information from each pixel than using regular imaging systems.^{4,5} Using sliding window method over captured spectral image, we will have spectral and spatial domain for further analysis. In this study we describe some incremental steps, which are taking us closer to automatic skin cancer detection, identification and delineation.

2. MATERIAL AND METHODS

In this study we have small data set ($n=61$) of hyperspectral images covering narrow wavebands from 450 to 850 nm. Data set consist of several lesions, which were imaged and diagnosed by histopathology. Lesions consist of malignant melanomas, melanoma in-situs, dysplastic nevi and bening nevi. All patients have volunteered to participate in the study. The study protocol has followed the Declaration of Helsinki and it was approved by the local Ethics committee. Patient were recruited and imaged by the Department of Dermatology and Allergology of Helsinki University Hospital, Helsinki, Finland and by the Päijät-Häme Central Hospital, Lahti, Finland, between June 2016 and October 2017.

All hyperspectral images were collected with two identical hyperspectral imagers (Revenio Prototype 2016). Spectral separation of the imager is based on Fabry-Pérot interferometer (FPI). Use of FPI enables fast scanning in the spectral domain. The imager works on wavebands from 450 nm to 850 nm. The imager captures 120 wavebands within few seconds. Full width of each waveband's half maximum (FWHM) vary from 5 to 15 nm. Variation in FWHM comes as a function of wavelength. Another source of the variation comes from which multiple of FPI's is used. Imaging system contains a broadband halogen light source, which produces diffuse illumination to the imaged region of the interest (ROI). At the imager there is covering tube, which blocks illumination from other sources. Image acquisition is done with color cmos machine vision camera, which is integrated to the imager. The used machine vision camera is capable to take images in 1920×1200 pixel resolution. This corresponds approximately to $15 \mu\text{m}/\text{pixel}$ spatial resolution.

The spectral imager produces a raw data cube, which is calibrated to the radiance by following method of Saari et. al (2013).¹⁰ There was some indeterminated fluctuation at the end of recorded spectra. Thus, twenty last wavebands were left outside from further analysis. For each data cube there was captured white reference target. This was used to convert imaged radiance to reflectance $R = I/I_0$, where I is imaged region of interest and I_0 is data cube from white reference. To improve quality of the data in spectral domain and reduce memory consumption in further processing, the data is downsampled. This was done by averaging nearest pixels of every fifth pixel. Also, only every second waveband was used in further analysis. By these operations data cubes size reduced to $384 \times 240 \times 50$ pixels.

Training of the classifier needs labelled data. For each image there were annotated areas, which indicated either healthy skin, lesion or used marker. From each image’s annotated areas 1000 data sets (or less if annotated area contained less than 1000 pixels) were selected for training purposes. These data sets contained annotated pixel and its 10×10 neighborhood. Figure 2 shows distribution of spectra of melanoma, lentigo maligna, dysplastic nevus and benign nevus. As we can see that there are overlapping in the distributions and some of the spectra has deviation. To reduce these effects and some problems from the vignetting and lightning irregularities, each imaged spectrum was subtracted by its average in spectral domain.

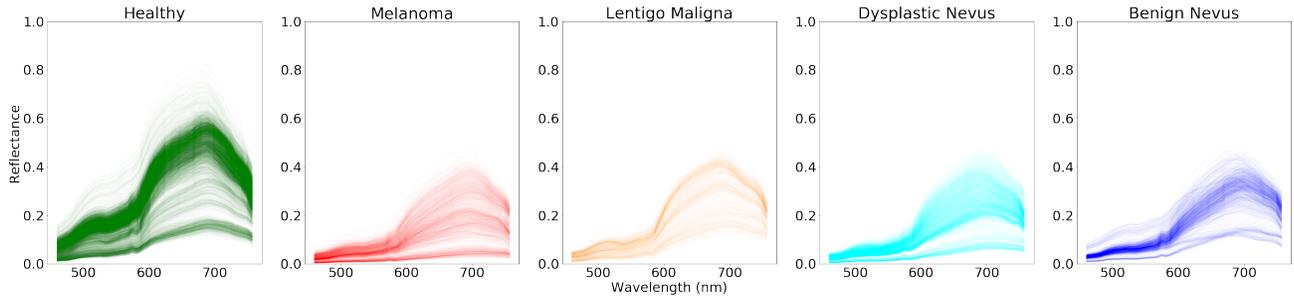


Figure 2. The distributions of spectra of melanoma, lentigo maligna, dysplastic nevus and benign nevus.

During recent years, deep neural networks have made new records in pattern recognition.⁶ Our aim was to use both spatial and spectral domain simultaneously. Convolutional neural networks have been used to classify melanomas and other skin cancers from dermatoscope and regular color images.⁹ Spectral data cube has three-dimensional nature, thus, standard 2D convolutional neural network might not be enough to utilize spectral data.

In deep learning and especially with convolutional neural networks classification task has two parts. In the feature learning we are calculating features using convolution operations with different weights. By tuning weights during back-propagation we will eventually achieve optimized feature space for our classification task. The actual classification model is just a deep regular multi-layer perceptron network. This structure is illustrated in the figure 3.

In this study we tested three different kind of feature learning structures - 1D, 2D and 3D convolutions. We will also have basically two different types of inputs. Single spectra and small window surrounding this spectra. As figure 3 shows, a 1D convolution input takes a single spectra. For 2D and 3D convolution input will be a subset’s of spectral cube. Difference between 2D and 3D convolution is that 2D case doesn’t operate over spectral domain while 3D does.

Used deep neural networks consist from two parts. First part executes feature learning by using the convolutional operator by the Conv layers. The Maxpooling layers reduces data’s dimensionality. Machine learning part and actual classification is done with deep feed-forward neural network, which consist of six Dense layers. Convolutional and dense layers use rectified linear unit activation (ReLU) function. The single Dropout layer is added to avoid overfitting of the model. Last dense layer does final classification using Softmax activation function. Parameters of each layer are shown in the figure 3.

By varying the described architecture we tested five different kind of networks - 1D, 2D and 3D convolutional neural networks and two combinations where feature learning was executed by using 3D+1D convolutions and 3D+2D+1D convolutions.

The actual training data was sampled randomly from annotated points, so that there were 10000 data points from each class. An annotation was based on its histopathological results. Whole lesions were marked the same way. Annotation was done by a non-expert.

For annotated points, data augmentations was utilized so, that each training cube was mirrored and flipped horizontally and vertically. These operations fourfold the number of the inputs in the training phase. Training

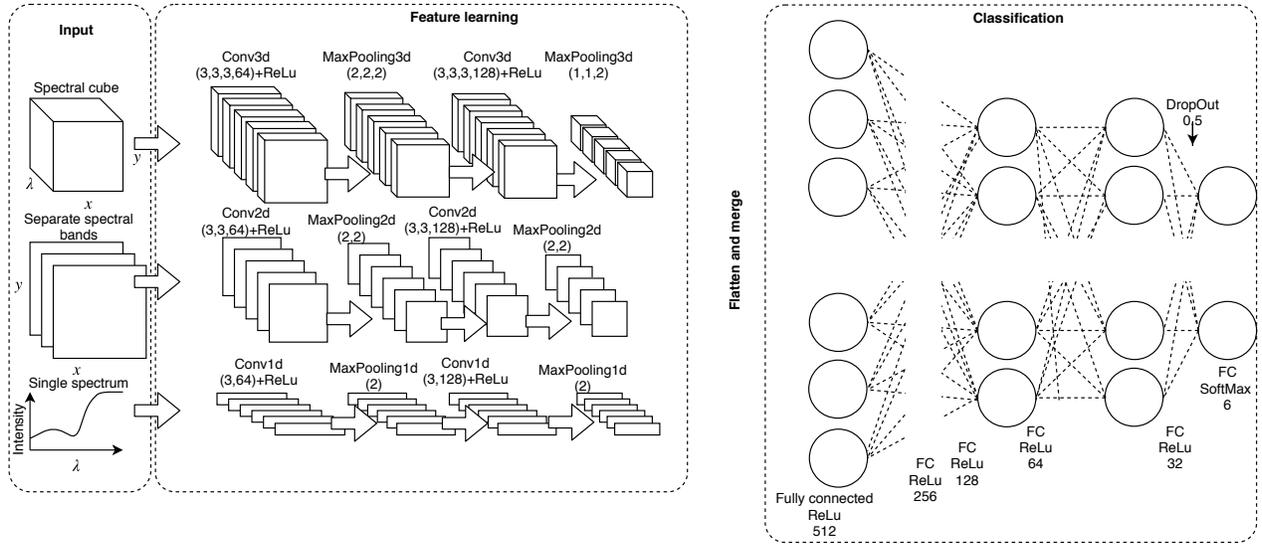


Figure 3. Schematic structure of used convolutional neural networks. Best results were gained using all inputs and all three different convolutional feature learning parts simultaneously.

set consisted of approximately 240 000 data points. For the optimization we used Adam, which is a first-order gradient-based optimization method of stochastic objective functions. The used hyperparameters for the optimization was the learning rate of 0.001, $\beta_1 = 0.9$ and $\beta_2 = 0.999$, while the learning rate decay over each update stayed at 0. The used cost function was categorical cross-entropy.

Our implementation used Keras with Tensorflow backend and Python 3.6 . All calculations were executed using IBM PowerAI platform, which includes two Nvidia Tesla V100-SXM2 16 GB GPU units.

There were only 61 imaged lesions (15 malignant melanoma, 6 lentigo maligna, 26 dysplastic nevus and 14 benign nevus). Thus, leave-one-out cross-validation was used. In this procedure classifier is trained 61 times for each image separately. This will guarantee that training set does not include data points from the image which is currently under classification.

3. RESULTS

Our ground truth consist of the results of histopathology. This meant that whole lesion was labelled based on most dangerous diagnosis. Because our approach gives us pixel wise information we ended situations where one lesion had several differently classified pixels. This actually might be quite realistic situation. Malignant lesions can have non-malignant parts. Thus, final classification for each lesion was made based on most dangerous pixel, which was found from lesion. If there was even a single pixel, which was classified as melanoma, whole lesion was classified to melanoma. In melanoma detection with this approach we will gain relatively high sensitivity, but low specificity which is seen in table 1 and in figure 8. And as opposite for benign nevus will have high specificity and low sensitivity.

Table 1. Sensitivity, specificity and positive predictive value of different classifiers for the melanoma classification

	CNN 1D	CNN 2D	CNN 3D	CNN 3D+1D	CNN 3D+2D+1D
Sensitivity	1	1	0.93	0.93	0.93
Specificity	0.15	0.12	0.14	0.14	0.21
Positive predictive value	0.34	0.35	0.32	0.32	0.34

When we are looking at sensitivities of different classifiers, we can see, that all 15 melanoma cases actually were classified correctly using only 1D and 2D convolution networks. 14 of 15 melanoma cases were classified correctly when 3D convolution was utilised. These metrics are actually misleading. If we look at actual classification results as shown in figures 4, 5, 6 and 7, we can see that actually classification results based on single spectra are often noisy. Figure 4 was confirmed to be dysplastic nevus in histopathology. All single source convolutional neural networks fail to classify it correctly. On lesion boundaries there is quite typical error, where trained model for some reason mis-classify lesion to melanoma. Here best result is achieved using multiple inputs and three different kind on CNN's.

In general level it seems that spatial features give more reliable looking results. When we combine those with spectral domain, results get better, because specificity increases. If we take closer look one false positive case in the figure 7 we can see that majority of the pixels in the lesion is actually classified correctly. This is promising result because with more training cases we might have chance to train better models.

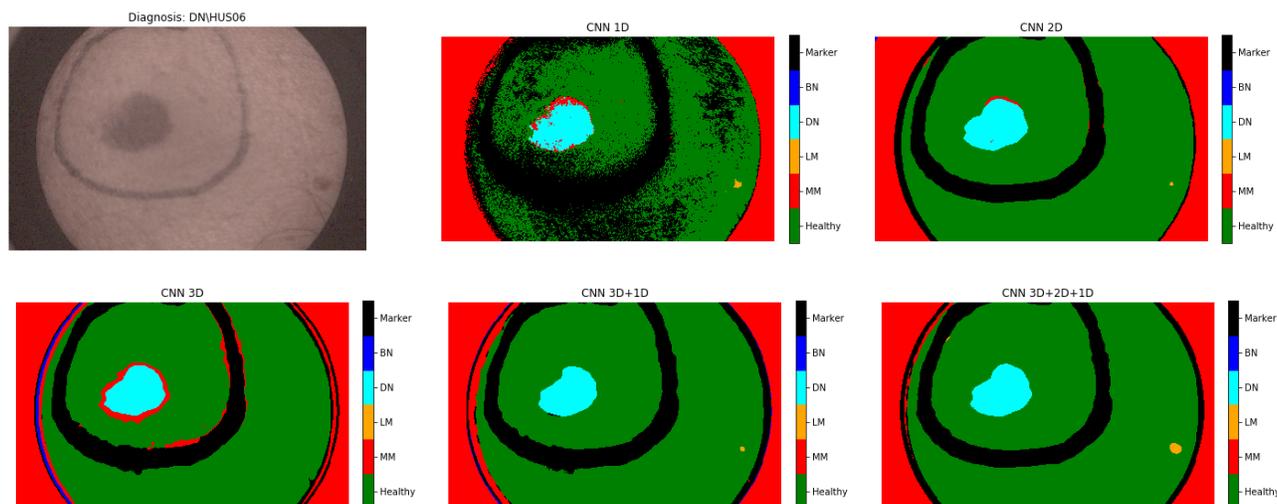


Figure 4. Classification results of five different classifiers for dysplastic nevus.

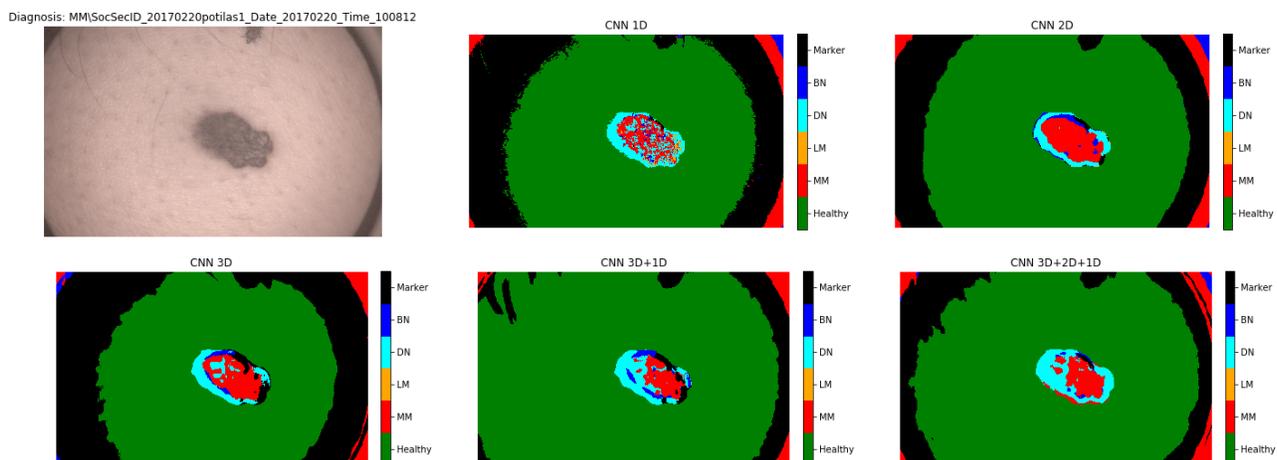


Figure 5. Classification results of five different classifiers for malignant melanoma.

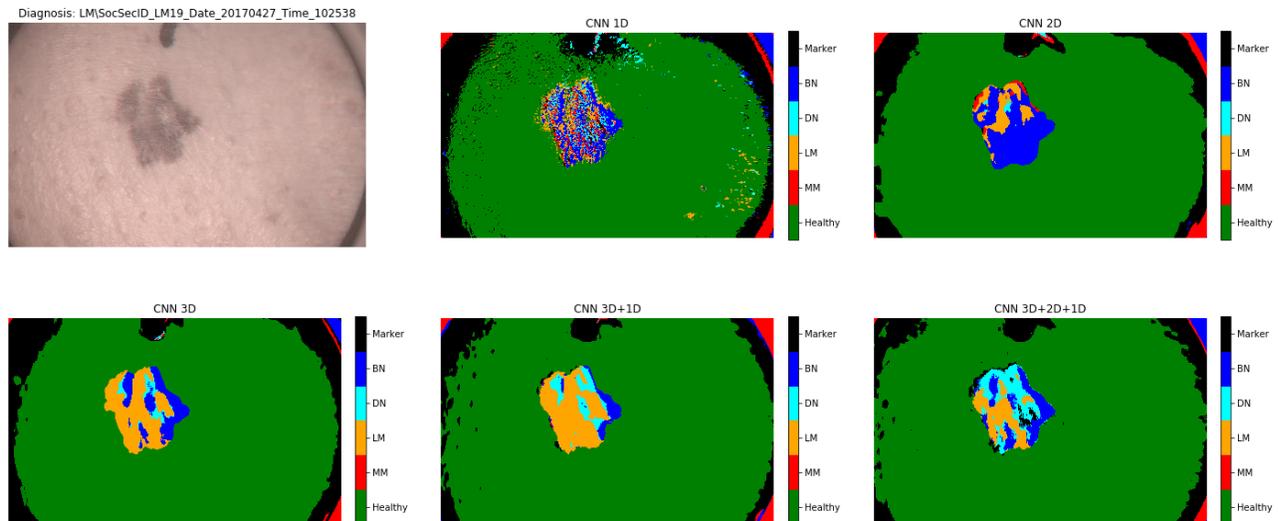


Figure 6. Classification results of five different classifiers for lentigo maligna.

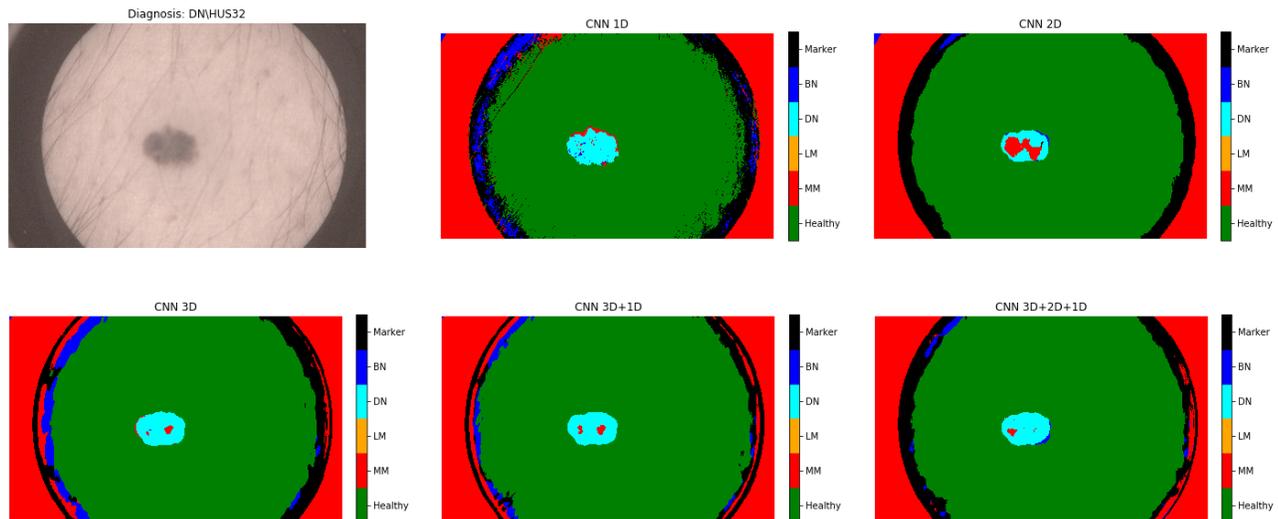


Figure 7. Classification results of five different classifiers for dysplastic nevus. Here all classifiers give false positive as a result. Even though majority of pixels are classified correctly, the end result will be false positive for whole lesion.

4. DISCUSSION

Shown results are promising. With all classifiers we achieved same positive prediction value (PPV) as clinicians. It is shown that utilisation of the spectral and spatial domain increases classification performance.

There is work to be done to gain higher specificity and PPV. We could play around with detection probabilities provided by the softmax layer and take some threshold probabilities, which would be concerned during classification (for example only classification results over 90% confident would be recognised). Or we could calculate which class has majority of pixels on lesion area. Unfortunately both of these approaches would actually decrease the sensitivity and the number of false negatives would rise.

Our study's first limitation comes from the small data set. Even though we had over hundred million pixels at our disposal, we eventually had only 61 different lesions. This is a quite limited data set and more data is

needed to develop and calculate a more robust and accurate neural network model. This would mean that we will need multi center studies, where patient data is gathered in several countries simultaneously. For example Finnish population is too small to produce enough patients to train enough general models.

Another limitation is that the ground truth labeling is based on histopathological diagnosis of whole lesion. There is a great possibility that a lesion can include several classes. Thus, our ground truth contains bias and this bias is also transferred to our training data. What we actually should do is that we should have several biopsied training points from each lesion so that we could use those spots in our training data. This would decrease bias in the training data, but it would also lead to reduced training data size.

Process of validating results and gathering training data should be similarly iterative as training of the neural network itself. When a hyperspectral imager and a classification model is used in a clinical study, we should take biopsies based on results. The spatial locations of these biopsies should be saved and the model should be updated using histopathological results of these studies afterwards.

The approach to use spectral and spatial domains seems feasible. Our next ideas are to add more features to the data. By modifying the illumination source we can take photogrammetric stereo images. From these images it is possible to calculate surface normals, a digital elevation model and skin's albedo as a function of wavelength. Each of these can be used as new features in cancer classification and delineation.

5. CONCLUSION

We have shown that use of spectral and spatial domain will increase classification performance of convolutional neural network. Our results show that with a relative small data set we are able to get same or slightly better positive prediction values as clinicians. This information was achieved by using a novel hyperspectral imager prototype in a clinical setup and train five different neural network models based on histopathological diagnoses. Because of the climate change proportion direct of sun radiation seems to grow, thus non-invasive automatic skin cancer detection and delineation systems will be needed even more in the future. These results are incremental steps towards this goal.

REFERENCES

- [1] Eriksson, T. and Tinghög, G., "Societal cost of skin cancer in Sweden in 2011," *Acta dermatovenereologica* **95**(3), 347–348 (2015).
- [2] Heal, C. F., Raasch, B. A., Buettner, P., and Weedon, D., "Accuracy of clinical diagnosis of skin lesions," *British Journal of Dermatology* **159**(3), 661–668 (2008).
- [3] Argenziano, G., Cerroni, L., Zalaudek, I., Staibano, S., Hofmann-Wellenhof, R., Arpaia, N., Bakos, R. M., Balme, B., Bandic, J., Bandelloni, R., et al., "Accuracy in melanoma detection: a 10-year multicenter survey," *Journal of the American Academy of Dermatology* **67**(1), 54–59 (2012).
- [4] Neittaanmäki-Perttu, N., Grönroos, M., Jeskanen, L., Pölönen, I., Ranki, A., Saksela, O., and Snellman, E., "Delineating margins of lentigo maligna using a hyperspectral imaging system," *Acta dermatovenereologica* **95**(5), 549–552 (2015).
- [5] Neittaanmäki, N., Salmivuori, M., Pölönen, I., Jeskanen, L., Ranki, A., Saksela, O., Snellman, E., and Grönroos, M., "Hyperspectral imaging in detecting dermal invasion in lentigo maligna melanoma," *Br J Dermatol* (2016).
- [6] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in *[Advances in neural information processing systems]*, 1097–1105 (2012).
- [7] Lawrence, S., Giles, C. L., Tsoi, A. C., and Back, A. D., "Face recognition: A convolutional neural-network approach," *IEEE transactions on neural networks* **8**(1), 98–113 (1997).
- [8] Kim, Y., "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882* (2014).
- [9] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S., "Dermatologist-level classification of skin cancer with deep neural networks," *Nature* **542**(7639), 115 (2017).
- [10] Saari, H., Pölönen, I., Salo, H., Honkavaara, E., Hakala, T., Holmlund, C., Mäkynen, J., Mannila, R., Antila, T., and Akujärvi, A., "Miniaturized hyperspectral imager calibration and UAV flight campaigns," in *[Sensors, Systems, and Next-Generation Satellites XVII]*, **8889**, 88891O, International Society for Optics and Photonics (2013).

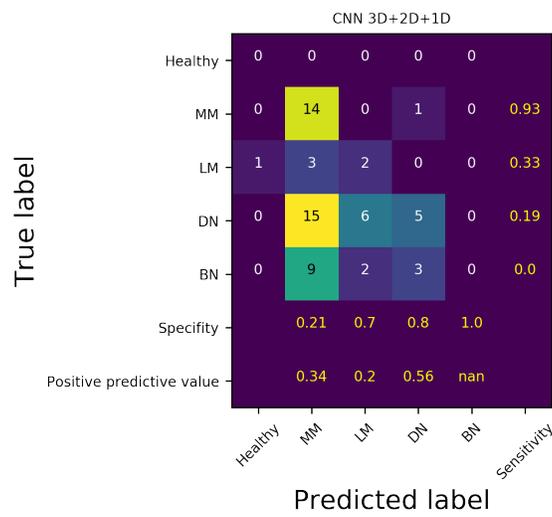
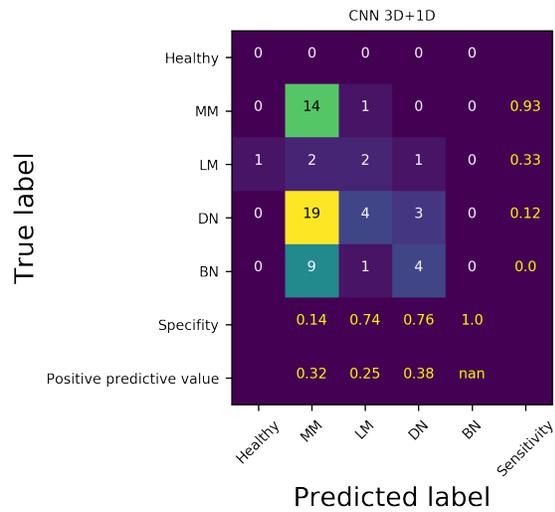
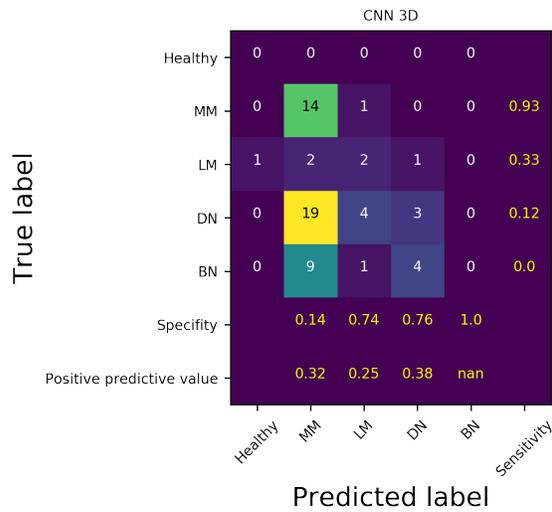
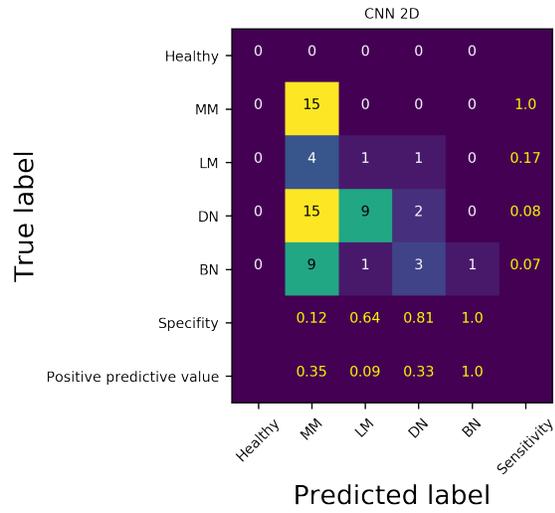
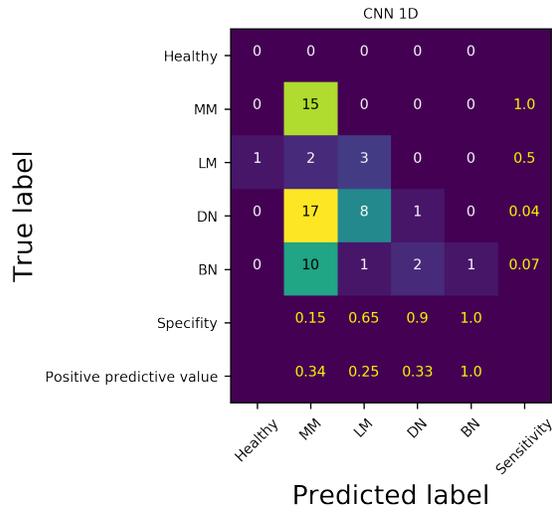


Figure 8. Confusion matrices for the different convolutional neural network models.