CAPSULE NEURAL NETWORKS IN CLASSIFICATION OF SKIN LESIONS

Evgen Goceri
Biomedical Engineering Department, Engineering Faculty, Akdeniz University
Dumlupinar Boulevard, 07058, Antalya, Turkey

ABSTRACT
Classification of skin lesions is a difficult issue even for highly experienced dermatologists and pathologists because of several reasons such as low contrast between lesions and surrounding skin tissue, high noise, fuddled lesion boundaries, visual similarities of skin lesions. On the other hand, early and accurate classification of lesions is crucial for timely and accurate treatment of skin diseases. Therefore, automated methods have been developed to perform objective, quantitative and reproducible results. Recent methods in pattern recognition and image classification is based on deep networks, particularly convolutional neural networks. However, pooling layers providing down-sampling in these networks lead to data loss and cause low performance in generalization. Also, convolutional neural networks cannot transfer spatial information and instantation parameters (e.g., pose of low-level features to each other, deformation and texture information). To overcome these problems with dynamic routing, capsule neural networks have been proposed. Capsules can transfer pose parameters and part-whole relationship using likelihood and spatial information between low-level features. In this work, capsule networks applied for skin lesion classification have been proposed. It has been observed that although capsule networks can overcome deficiencies of convolutional neural networks, there are only four techniques based on capsule networks in the literature to achieve automated classifications of skin lesions. Motivated by the lack of articles on this topic, a comprehensive assessment of the capsule networks proposed for skin lesion classification is presented in this paper. Also, strengths and weakness of those four techniques have been presented to help the researchers interested in this area.

KEYWORDS
Capsule Network, CapsNet, Skin Disease, Lesion Classification, Deep Neural Networks, Pattern Recognition

1. INTRODUCTION
Skin diseases can occur due to various reasons (e.g., environmental factors, aging, depression, trauma, etc.). They are serious widespread disorders affecting about 1.9 million people worldwide (Liu et al. 2020) and have an important effect on the quality of life. Existence of more than three thousand kinds of them identified with different severity and symptoms is known (Fritsch et al. 2006; Tizek et al., 2019). For example, malignant melanoma, which may cause death, is a kind of skin cancer grown from pigments containing cells (melanocytes). Those affected cells are generally located in the skin. They can also be found in the eye, mouth, on women’s legs or men’s backs. Also, they may arise from a mole with several changes (e.g., color change, un-even edges, irritation, an escalation in size, and skin break-down). Another example is psoriasis that is a chronic and also relapsing disorder (Marks et al., 2012). It is very challenging to distinguish psoriasis and erythema-squamous disorders, such as eczema lichen planus, seborrheic dermatitis, tinea corporis, leprosy, pityriasis, etc. (Pal et al., 2016, Pal et al., 2018, Roy et al., 2017). Psoriasis occurs when the immune-system causes over-production of novel skin cells by sending faulty signals mistakenly. Then, nucleated cells and neutrophils can infiltrate in stratum corneum. The diagnostic marker of psoriasis in clinical pathology is Munro's microbes (Marks et al., 2012). Challenge in detecting Munro's microbes lies in the fact that because of variations in staining, neutrophils in stratum corneum are generally mis-classified due to their dark staining as nucleated-keratinocytes. In addition, various artefacts caused by imaging process increase the difficulties. Therefore, accurate diagnosis of lesions by naked-eye according to visual cues is a challenging issue, even for vastly experienced clinicians and pathologists.
Deep Neural Networks (DNNs), particularly Convolutional Neural Networks (CNNs) as special types of DNNs, are effective architectures for image classification and accelerate decision making in clinical environments (Litjens et al., 2017). However, they can classify only when a certain feature does exist in the image from the testing dataset by ignoring spatial relations. This causes false negatives. Also, the same lesions viewed from different angles can be assigned to different classes. This causes false positives. Capsule Networks (CapsNets) have been designed to overcome these problems.

A CapsNet structure is constructed using one capsule layer as the final layer and many convolution layers. It uses capsules (which are vectors representing an element with size and direction) other than pooling. A CapsNet includes dynamic routing and layer-based squashing. By this way, CapsNets can help in avoiding false negatives and false positives.

A basic CapsNet including 2 convolutional layers and a capsule layer is shown in Figure 1 (Sabour et al., 2017). In this architecture, the 1st convolution layer is used to obtain low-level features from the image with size 28x28x1. There are 256 channels and each of them uses a 9x9 convolution kernel. The 2nd layer is constructed as a convolutional capsule layer including a total of 6x6x32 capsules which are called as primary capsules. Every capsule produces an eight-dimensional vector. Every group of those capsules in the 2nd layer is convolved by eight convolution units. The 3rd layer, which is called as digital capsule layer, is a fully connected layer including 10 capsules. Each capsule in this layer is 16-dimensional vector and performs the classification process. The output of these digital capsules is sent to three fully connected layers (i.e., decoder) to minimize the summation of the squared difference of the corresponding pixels between the input and the reconstructed image. The length for each capsule (i.e., the classification possibility) is calculated in the final layer and a loss function is used at the end of the network. In the training stage, if correct digital capsule cannot be predicted by a vector, then zero value is assigned to the vector. Input image is constructed only by the digital capsules providing correct predictions.

![Figure 1. A basic CapsNet architecture](Image)

Although several advantages of CapsNets such as fast convergence speed without reduction in accuracy, there are only four techniques based on CapsNets in the literature to achieve automated classifications of skin lesions. According to the best of our knowledge, in the literature, there is no any work focused on assessment of capsule networks proposed for skin lesion classification. Therefore, we believe that this work will fill this gap in the literature and be helpful for the researchers who are working or interested in this area and want to improve those approaches by aware of their strengths, weaknesses and potential challenges to solve the classification problem.

2. RELATED METHODS AND MATERIALS

In the literature, only four techniques based on CapsNets have been proposed to solve lesion classification problem and their authors have handled different types of skin diseases. They have also used different types of images. These techniques and the image data sets used by them have been examined in detailed in this section.
2.1 Classification of benign nevus and malignant melanoma (Tiwari et al., 2021)

The data sets used for classification of benign nevus and malignant melanoma have been constructed with non-dermoscopic images which have been provided from the Groningen Medical Center University. The number of images including nevus (100) and melanoma (70) samples is 170. Example images are shown in Figure 2. To increase the size of image data sets to provide efficiency in the training stage, original images have been rotated and 2040 images have been obtained.

![Example images for nevus (a) and melanoma (b) (Tiwari et al., 2021)](image)

Three classification models, which are Multi-Layer Perceptron (MLP), CNN and CapsNet, have been applied. Default functions (such as, maximum pooling, adaptive moment estimation (adam), rectified linear unit (ReLU) and cross-entropy loss function) have been used in these models. (More details can be read from Tiwari et al., 2021).

100 epochs have been enough to train these three models. 70% of the data has been used for training whereas 30% of the data has been used for validation. According to the reported results, the graphs of validation accuracies (Figure 3) indicate that the CapsNet outperforms than MLP and CNN.

![Validation accuracies obtained from MLP, CNN and CapsNet (Tiwari et al., 2021)](image)

The results in terms of average accuracy, average precision, F1 scores and recall values have been presented in Table 1 (Tiwari et al., 2021). These quantitative values indicate better performance of CapsNet than MLP and CNN architectures.
Table 1. Accuracies, precisions, F1 scores and recall values

<table>
<thead>
<tr>
<th>Models</th>
<th>Average accuracy</th>
<th>Average precision</th>
<th>F1 scores</th>
<th>Recall values</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>0.76</td>
<td>0.78</td>
<td>0.78</td>
<td>0.76</td>
</tr>
<tr>
<td>CNN</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>CapsNet</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
</tbody>
</table>

2.2 Classification of Stratum Corneum to Diagnose Psoriasis (Pal et al., 2018)

Stratum corneum classification has been performed using skin biopsy images (Figure 4), which have been taken from 120 patients. If stratum corneum layer has been lost during tissue processing, then the image has been ignored. 2 experts have labelled these images.

![Figure 4](image.png)

Figure 4. Example image patches showing neutrophils, which are dark blue colored and circular shaped cells, and nucleus which are light blue colored and oval shaped cells (a,b); Image shows only nucleus (c); Image shows neither neutrophils nor nucleus (d) (Pal et al., 2018)

Affected tissues have been obtained in 10% formalin by experienced dermatologists after clinical confirmation of psoriasis. Thin sections (5μM) have been selected to prepare slides and stained by using biomarkers (i.e., hematoxylin and eosin). These slides have been saved under a microscope with the highest magnification (i.e., 10X magnification) to fit whole-biopsy samples efficiently and the images (1936x2584 pixels) have been obtained by the microscope.

The CapsNet architecture designed in this work includes 2 sections called as primary and secondary capsule. The architecture has been constructed so that the receptive fields can avoid crowding. Here, the term crowding corresponds to existence of multiple-samples of the same entity in the receptive fields in a capsule.

In a region of interest of the image patch $I$, the probability ($K$) of neutrophils is denoted by the length of the output vector that is obtained from secondary capsule. Several neutrophils can be found in the image patch. Therefore, the probability of neutrophils has been obtained by taking average of the $K$ probabilities.

If the probability of neutrophil in a patch $I$ is represented by $p_I$ and the default loss function (binary cross-entropy) $L$ is used in the training stage, the function to be minimized is written by $y_I$, which has zero value when the patch does not include a neutrophil and one otherwise, as:

$$L = -y_I \log(p_I) - (1 - y_I) \log(1 - p_I)$$

The results of patch classifiers with ResNet-50 and the proposed method have been presented with several metrics. They are given in Table 2 for comparative evaluations.

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracy</th>
<th>Precision</th>
<th>F1 score</th>
<th>Recall</th>
<th>Parameters (Million)</th>
<th>Load Time (in second)</th>
<th>Prediction Time (second/image)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet-50</td>
<td>0.94 ±0.02</td>
<td>0.95 ±0.015</td>
<td>0.91 ±0.036</td>
<td>0.87 ±0.078</td>
<td>23.5</td>
<td>2.15</td>
<td>0.021</td>
</tr>
<tr>
<td>The network in (Pal et al., 2018)</td>
<td>0.92 ±0.013</td>
<td>0.94 ±0.027</td>
<td>0.88 ±0.027</td>
<td>0.84 ±0.064</td>
<td>0.1</td>
<td>0.26</td>
<td>0.087</td>
</tr>
</tbody>
</table>
2.3 Classification of Skin Cancer Lesions (Cruz et al., 2020)

Automated classification of lesions to diagnose skin cancer using a CapsNet architecture has been performed with publicly available image sets (Figure 5). These data sets have been provided from BioGPS (Web_1, 2021) and the HAM-10000 dataset (Tschandl et al., 2018), which contains 10015 dermatoscopic images of 7 pigmented lesions, (benign keratosis-like lesions, melanocytic nevi, basal cell carcinoma, dermatofibroma melanoma, vascular skin lesions, and actinic keratoses)

Figure 5. Example cancer images showing skin lesions (Cruz et al., 2020)

Transfer learning has been used to train the network architecture to identify cancerous cells from images. The network has been trained and tested by using 7010 and 3005 images respectively. In the CapsNet architecture (Figure 1), ReLU, softmax adam and cross-entropy functions have been used. The number of epochs, batch sizes and learning rate have been determined as 50, 10 and 0.001 respectively.

Epochs and the accuracies obtained from validation and training steps are shown in Figure 6a. The validation and training loss are presented with red and blue color respectively in Figure 6b.

Figure 6. Testing and training accuracy are presented with red and blue respectively (a); Validation and training loss are presented with red and blue color respectively (b) (Cruz et al., 2020)
2.4 Classification of Melanoma Type of Skin Cancer (Boaro et al., 2020)

Skin cancers can be grouped as non-melanoma and melanoma. For this classification, image data sets including 4018 images (1084 melanoma cases and 2934 non-melanoma cases) have been provided from ISIC database (Web_2, 2021). Mirroring and rotations have been applied to increase the number of images. After these operations, 5509 images have been obtained totally and they have been re-sized to 600x408.

Binary masks obtained from ISIC have been used to achieve high classification performance by highlighting only the lesion areas and excluding irrelevant regions. Figure 7 shows an example image (Figure 7a), a binary mask (Figure 7b) and the lesion region (Figure 7c) generated by using the mask image.

![Example image showing a lesion (a), binary mask (b), the lesion region (c)](image)

Figure 7. Example image showing a lesion (a), binary mask (b), the lesion region (c) (Boaro et al., 2020)

The network architecture consists of VGG16 network (which has been pre-trained using commonly known ImageNet (Russakovsky et al., 2015)), and the original CapsNet (Figure 1). In the combined architecture, image features have been obtained by the VGG16 and relationships in these feature sets have been obtained by the CapsNet.

The results obtained from the original CapsNet and the combined architecture are given in Table 3 as average accuracies, recalls, specificities and also areas under the Receiving Operating Characteristic (ROC) curves (Boaro et al., 2020).

<table>
<thead>
<tr>
<th>Architecture</th>
<th>Accuracies</th>
<th>Recalls</th>
<th>Specificities</th>
<th>Area under the ROC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CapsNet</td>
<td>0.86</td>
<td>0.85</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>VGG16 and CapsNet</td>
<td>0.92</td>
<td>0.91</td>
<td>0.93</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Validation loss and validation accuracy graphics have also been shown in Figure 8 to indicate performance of the combined network structure.

![Validation loss (a) and accuracy (b) of CapsNet (blue) and the combined model (yellow)](image)

Figure 8. Validation loss (a) and accuracy (b) of CapsNet (blue) and the combined model (yellow) (Boaro et al., 2020)
3. CONCLUSION AND DISCUSSION

In this work, the role of CapsNets in classification of skin lesions has been examined by addressing those related works proposed in the literature.

Skin lesion classifications with CapsNets provide quite promising results when they are compared to the results obtained from related works.

Experimental works to determine presence or absence of neutrophils in an image patch to diagnose psoriasis showed that the CapsNet based model, although it has with less parameters, achieves comparable performance to ResNet-50.

Experiments for melanoma detection indicated that the performance of classification with CapsNet can be improved when high-level features are extracted using a pre-trained VGG16. Also, studies on the application of multi-layer perceptron, CNN and CapsNet to classify benign nevus and malignant melanoma showed that the CapsNet is the most appropriate classification architecture for these images.

Therefore, in the literature, although different skin diseases and image data sets having different characteristics have been handled, it has been observed that CapsNet based architectures provide efficiency in terms of accuracy and parameters required for lesion classification from skin images. However, further researches with increased number of images are still needed in this area.

Also, datasets are usually unbalanced which can compromise the learning step of the network. Therefore, balanced number of images should be used in the data sets to avoid biasing. To increase the number of images, operations such as rotation and mirroring should not be used since they only lead to repeat the same image in the data set. More efficient image augmentation algorithms should be applied to generate new synthetic images.

REFERENCES


