



Severity Assessment of Psoriatic Plaques Using Deep CNN Based Ordinal Classification

Anabik Pal¹(✉), Akshay Chaturvedi¹, Utpal Garain¹, Aditi Chandra²,
Raghunath Chatterjee², and Swapan Senapati³

¹ CVPR Unit, Indian Statistical Unit, Kolkata 700108, West Bengal, India
anabik_r@isical.ac.in

² Human Genetics Unit, Indian Statistical Unit, Kolkata 700108, West Bengal, India

³ Consultant Dermatologist, Uttarpara, Hooghly 712258, West Bengal, India

Abstract. Development of computer-aided diagnosis (CAD) tool for severity assessment of psoriatic plaques is important to assist the dermatologists to overcome the human limitation. In this paper, a pioneering attempt is made to build a Convolutional Neural Network (CNN) model to classify a skin image with respect to its severity class. However, the commonly used loss functions like categorical cross entropy and mean square error ignores the underlying ordinal class relationships (distance between predicted and actual class) which are important for the present problem. In this paper, the Earth Mover's Distance based loss function is proposed for training CNN since it takes into account the corresponding ordinal class relationships. Separate CNNs are trained for severity scoring corresponding to three plaque characteristics- erythema (redness), scaling (silveryness) and induration (elevation). Mean accuracy (MA), mean absolute error (MAE) and Kendall's τ_b are used for performance evaluation. The experimental result shows that the proposed ordinal classification technique outperforms the traditional approaches.

Keywords: Psoriasis image · Psoriatic plaque · Erythema · Scaling
Induration · Convolutional Neural Network · Ordinal loss · Dataset

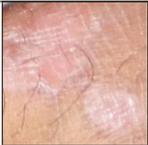
1 Introduction

Psoriasis is a chronic, immune-mediated, relapsing, inflammatory skin disease and usually associated with itch. The prevalence of psoriasis varies 1%–12% among different populations worldwide [1]. This disease develops when the immune system mistakes a normal skin cell for a pathogen and sends out faulty signals that cause overproduction of new skin cells. This disease can be diagnosed by visual and haptic inspection. The visual changes of outer skin surface due to this disease include development of elevated red scaly dry patches with well-demarcated borders on the skin surface. However, the shape, size, color and

distribution of these patches vary. In dermatology, these patches are termed as psoriatic plaque [13].

No drug is available yet to cure psoriasis completely but the severity can be controlled by suitable drug doses. As the drug response varies among different patients thus development of reliable severity assessment procedure is required to decide the type and dose of the drugs as well as measure disease progress and drug's efficacy. Dermatologists use Psoriasis Area Severity Index (PASI) [5] for estimating severity. PASI considers two major aspects of the disease: ratio of body surface area affected by this disease and the severity of the plaques formed on the skin surface. The severity of the plaque is determined by the visual disorder formed on the affected skin regions. Three different aspects are considered for severity of the plaques: degree of redness or erythema, thickness or induration and scaling or desquamation. All aspects are scored with a value between 0–4. Table 1 contains a sample image for every severity class.

Table 1. Visualization of psoriasis plaques with different severity scores.

Severity Factor	Absent (0)	Mild (1)	Moderate (2)	Severe (3)	Very Severe (4)
Erythema (Redness)					
Scaling (Silvery-ness)					
Induration (Elevation)					

The severity factors are determined by the dermatologists in an eye estimation technique. The severity assessment procedure suffers from both inter- and intra-observer variability. Hence, development of an automated and robust system for severity assessment of psoriatic plaque is necessary for clinical studies. Some approaches have already been proposed for automatic scoring of scaling [2] and erythema [4, 6, 7, 12]. In [14], an image based system is also proposed to compute the aggregated severity score according to plaque characteristics. In [11], an attempt has been made to assess the erythema, scaling and induration scores from psoriatic plaque images. However, all of these approaches consider the present severity grading task as an image classification problem but fail

to capture the underlying ordinal relationship among the severity labels. This motivates us to develop CNN based ordinal classifiers for severity assessment of psoriatic plaques.

To summarize, the key contributions of this paper are: (i) a pioneering attempt towards developing a deep convolutional neural network based ordinal classifier for predicting severity score of psoriatic plaque, (ii) a new loss function is used for training a CNN which can capture the ordinal relationship among the class labels, (iii) two pre-trained CNN models (namely, ResNet-50 and Mobile Net) trained on imagenet dataset are fine-tuned to develop the severity assessment classifiers, and finally, (iv) the performance of the proposed CNN is compared with several baselines.

2 Methodology

2.1 Convolutional Neural Network

Nowadays, Convolutional Neural Network (CNN) is widely used for image classification tasks as it relieves the researchers from designing hand-engineered feature descriptors and automatically develops powerful mathematical models directly from the training images. These models are made up of multiple processing units and each processing unit consists of trainable weights and biases. In the training phase, the network parameters are updated by comparing the distribution of predicted class labels with the actual class labels of the training images. A brief description of the traditional categorical cross entropy (CCE) loss and the mean square error (MSE) loss functions are given below.

Suppose, for a C -class ($C > 2$) single-label image classification problem, the ground truth of a particular image is given by a binary vector G of length C such that $G_i = 1$ whenever $i = k$ and 0 otherwise. The output of the CNN is a probability distribution P of length C such that its i^{th} entry (P_i) represents the predicted probability of the i^{th} class. Now the definition of CCE loss and the MSE loss are given in Eqs. 1 and 2.

$$\mathcal{L}_{CCE} = - \sum_{i=1}^C G_i \ln(P_i) \quad (1)$$

$$\mathcal{L}_{MSE} = \sum_{i=1}^C (P_i - G_i)^2 \quad (2)$$

2.2 Ordinal Classification and Limitation of CCE and MSE Loss

In the present severity assessment task, there exists an ordinal relationship among the severity grades. Suppose, the actual and predicted severity score of a misclassified image is K and K_1 respectively. Then, we would prefer the classifier to have the least possible absolute difference $|K - K_1|$. But it can be seen from Eqs. 1 and 2, CCE and MSE loss ignores this relationship since CCE only considers the probability of the correct class and MSE is invariant to permutation of probabilities of incorrect classes.

2.3 Proposed Loss Function

Motivated from [9], for the present classification task, we used the Earth Mover’s Distance (EMD) based loss function. Let X_i^{CDF} denote the i^{th} element of the cumulative distribution of X then the loss function is as follows:

$$\begin{aligned}
 \mathcal{L}_{EMD} &= \sum_{i=1}^C (P_i^{CDF} - G_i^{CDF})^2 \\
 &= \sum_{i=1}^C \left(\sum_{j=1}^i P_j - \sum_{j=1}^i G_j \right)^2 \\
 &= \underbrace{\sum_{i=1}^{k-1} \left(\sum_{j=1}^i P_j \right)^2}_{\mathcal{A}} + \underbrace{\sum_{i=k}^C \left(\sum_{j=1}^i P_j - 1 \right)^2}_{\mathcal{B}} \tag{3}
 \end{aligned}$$

where k is the correct class. According to Eq. 3, when $i < k$, increasing the value of P_i increases the value of \mathcal{A} whereas when $i \geq k$, increasing the value of P_i decreases the value of \mathcal{B} . Since, in \mathcal{A} , P_i occurs $(k - i)$ times hence, the value of \mathcal{L}_{EMD} increases as $|i - k|$ increases. Similarly, in \mathcal{B} , for $i \geq k$, P_i occurs $(C - i)$ times hence, the value of \mathcal{L}_{EMD} increases as $|i - k|$ increases. Thus the proposed loss function trains the network in such a way that the class label farthest from actual class gets less probability.

3 Experimental Setup

Dataset: In this research, an image dataset of seven hundred seven (707) psoriatic plaque images having expert annotated severity scores for erythema, scaling and induration is used. This dataset is built by cropping sub-images from a dataset of psoriasis images collected from 80 patients. The original images are collected in an uncontrolled environment by layman photographers with different view angle, distance, lighting condition and varying background. Apart from photographic limitation and skin color tone variation, the presence of several artefacts like hair, wrinkle etc. make the severity assessment task challenging.

Network: As the data volume is small, the training of a Convolutional Neural Network (CNN) from scratch does not produce satisfactory performance. Fine-tuning of pre-trained network is opted for the present classification task. Two pre-trained networks ResNet-50 [8] and Mobile Net [10] trained on imagenet dataset are considered for fine-tuning. ResNet-50 is chosen due to its impressive performance on imagenet classification. The mobile net is chosen as it contains comparatively fewer parameters but produces good performance on imagenet classification.

Training: In this paper, the performance of the developed system is reported on the basis of 7-fold cross validation. The model is trained with stochastic gradient

descent optimizer using a batch size of 4 images, momentum of 0.9, weight decay of 10^{-6} and with the learning rate of 0.001. For every fold, the network is trained 10 times and the trained model which ends with minimal loss is chosen for prediction of test images. Horizontal and vertical flipping augmentation is used for improving the generalization ability of the classifiers.

Baselines: In this paper, the performance of the CNN trained with proposed ordinal loss minimization is compared with four baselines. First two CNNs are trained with traditional categorical cross entropy (CNN_{CCE}) and mean-square error (CNN_{MSE}) loss minimization. In the third approach (CNN_{Regr}), the severity scores are projected into C equal partitions in $[0, 1]$ and the CNN is trained in such a way that the i^{th} class ($i = 1, 2, \dots, C$) image outputs a value in $[\frac{i-1}{C}, \frac{i}{C}]$. The last approach is the decomposition ($\text{CNN}_{\text{Decomp}}$) of the C class classification problem into $C - 1$ binary classification problems where the i^{th} classifier predicts whether an image has classification label more than i or not. Then these trained classifiers are used to predict class labels of the test images. It is worth mentioning that the binary CNNs are trained with binary cross-entropy loss minimization. Among all considered baselines, only the last two classifiers can capture the ordinal relationship among the labels.

Performance Evaluation Metrics: The performance of the trained CNN is measured with three different evaluation metrics- (i) Mean Accuracy (MA), (ii) Mean Absolute Error (MAE) and (iii) Kendall's τ_b . The value of MA lies in $[0, 1]$ and a higher value represents better performance. A lower value of MAE represents better performance. On the other hand, Kendall's τ_b measures the association or rank correlation between two measured quantities. The τ_b value lies in $[-1, +1]$, where, $+1$ is the maximum agreement between the prediction and the ground truth class labelling, 0 represents no correlation between them and -1 represents maximum disagreement. MAE and Kendall's τ_b are used since MA ignores the ordinal relationship between predicted and actual class for a misclassified image.

Suppose, there are N test images having a discrete class label in $[1, C]$. Let Y_i^p, Y_i^g represent the predicted and the ground-truth class label of the i^{th} test image respectively. Then the mathematical expressions of these metrics are shown in Eqs. 4, 5, 6.

$$MA = \frac{1}{N} \sum_{i=1}^N \delta(Y_i^g, Y_i^p); \quad \delta(x, y) = \begin{cases} 1, & \text{if } x = y \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i^g - Y_i^p| \quad (5)$$

$$\tau_b = \frac{\sum_{i,j=1}^N \hat{C}_{i,j} C_{i,j}}{\sqrt{\sum_{i,j=1}^N \hat{C}_{i,j}^2 \sum_{i,j=1}^N C_{i,j}^2}}, \text{ where}$$

$$\hat{C}_{ij} = \begin{cases} 1, & \text{if } Y_i^p > Y_j^p \\ -1, & \text{if } Y_i^p < Y_j^p \\ 0, & \text{otherwise.} \end{cases} \quad C_{ij} = \begin{cases} 1, & \text{if } Y_i^g > Y_j^g \\ -1, & \text{if } Y_i^g < Y_j^g \\ 0, & \text{otherwise.} \end{cases} \quad (6)$$

4 Results and Discussion

The average performance (metrics described in Sect. 3) of Mobile Net and ResNet-50 for erythema, scaling and induration scoring using considered approaches are listed in Table 2. According to Table 2, the performance of the chosen networks trained with proposed loss function outperforms the same network trained with CCE or MSE loss minimization. However, the networks trained with CCE and MSE loss minimization produce comparable performance. We receive poor performance when the CNN is trained for regression (Regr) output. This justifies the fact that the sensitivity of this method towards presence of noise in test images affects the performance badly. So, this approach is unsuitable for the present task. On the other hand, binary decomposition approach outperforms the CNN models trained with CCE and MSE loss minimization. However, in most cases, this approach is beaten by the proposed method. Obviously, the success of the binary decomposition approach depends on the robustness of all decomposed classifiers and a weak classifier may affect the whole classification scheme adversely. According to Table 2, among all considered approaches, the best performance is achieved when ResNet-50 is fine-tuned with EMD loss minimization. Some images in our dataset along with their actual and predicted severity scores with respect to erythema, scaling and induration predicted by the best models are given in Fig. 1.

Table 2. Experimental Result

	Method	Erythema			Scaling			Induration		
		MA (%)	MAE	τ_b	MA (%)	MAE	τ_b	MA (%)	MAE	τ_b
MobileNet	<i>CNN_{MSE}</i>	57.14	0.521	0.763	57.28	0.523	0.775	59.97	0.443	0.757
	<i>CNN_{CCE}</i>	57.57	0.501	0.779	57.14	0.502	0.790	58.84	0.457	0.748
	<i>CNN_{Regr}</i>	47.52	0.605	0.737	45.54	0.634	0.734	55.45	0.504	0.721
	<i>CNN_{Decomp}</i>	56.01	0.508	0.777	58.84	0.471	0.804	60.54	0.431	0.766
	Proposed	59.69	0.488	0.781	59.26	0.478	0.800	60.11	0.446	0.758
ResNet-50	<i>CNN_{MSE}</i>	58.27	0.478	0.791	58.56	0.487	0.790	58.84	0.444	0.764
	<i>CNN_{CCE}</i>	59.26	0.467	0.796	59.83	0.457	0.809	61.39	0.410	0.772
	<i>CNN_{Regr}</i>	45.97	0.632	0.728	43.71	0.655	0.729	48.23	0.581	0.686
	<i>CNN_{Decomp}</i>	58.56	0.474	0.792	59.83	0.465	0.803	62.52	0.409	0.775
	Proposed	61.10	0.440	0.812	62.66	0.430	0.820	63.51	0.390	0.782

The psoriasis image dataset developed for [11] is reused in our research. In [11], the best models for erythema and induration were obtained from the AlexNet based MTL network and for scaling it was from the AlexNet based



Fig. 1. Psoriasis images and their ground-truthed (GT) and Predicted (Pred) severity scores achieved from the best classifiers. The scores are given in (Erythema, Scaling, Induration). The errors are highlighted in yellow. (Color figure online)

STL network. The performance was evaluated with average correct classification accuracy, without and with ± 1 tolerance¹, combined average classification accuracy² without and with ± 1 tolerance (see footnote 1). In Table 3, the first row contains the previous best result and the second row contains the result produced by our best model. According to Table 3, our model redefines the current state of the art.

Table 3. Comparison with the state of the art. WoT refers **without tolerance** and WT refers **with tolerance**.

Method	Individual Mean Accuracy						Overall Mean Accuracy	
	Erythema		Scaling		Induration		WoT	WT(± 1)
	WoT	WT(± 1)	WoT	WT(± 1)	WoT	WT(± 1)		
[11]	60.68	93.64	58.98	94.77	61.10	93.78	27.58	86.28
Our	61.10	95.47	62.66	95.19	63.51	97.45	28.43	89.11

5 Conclusion

A novel loss function is designed to make CNN suitable for ordinal classification and used for automatic severity assessment of psoriatic plaques. The use of such

¹ Maximum deviation of predicted scores for an image lies in $[-1, 1]$.

² Erythema, scaling and induration all scores are correct.

loss function is a pioneering attempt. The proposed learning scheme successfully improves the classification performance. Specifically, improvement of MAE and τ_b in comparison to the considered baselines justifies the advantage of training a CNN with ordinal loss minimization. The proposed loss minimization in CNN training can be employed for other image based severity prediction from medical images, age group estimation from face images [3] etc.

References

1. Chandran, V., Raychaudhuri, S.P.: Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. *J. Autoimmun.* **34**(3), J314–J321 (2010)
2. Delgado, D., Ersbøll, B., Carstensen, J.M.: An image based system to automatically and objectively score the degree of redness and scaling in psoriasis lesions. In: Proceedings fra den 13. Danske Konference i, p. 130 (2004)
3. Escalera, S., et al.: Chalearn looking at people 2015: apparent age and cultural event recognition datasets and results. In: 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), pp. 243–251, December 2015
4. Fadzil, M.H.A., Ihtatho, D.: Modeling psoriasis lesion colour for PASI erythema scoring. In: International Symposium on Information Technology, ITSIm 2008, vol. 2, pp. 1–6. IEEE (2008)
5. Fredriksson, T., Pettersson, U.: Severe psoriasis: oral therapy with a new retinoid. *Dermatologica* **157**, 238–244 (1978)
6. Gupta, M.D., Srinivasa, S., Madhukara, J., Antony, M.: Random forest based erythema grading for psoriasis. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), pp. 819–823, April 2015
7. Hani, A.F.M., Prakasa, E., Nugroho, H., Asirvadam, V.S.: Implementation of fuzzy c-means clustering for psoriasis assessment on lesion erythema. In: 2012 IEEE Symposium on Industrial Electronics and Applications (ISIEA), pp. 331–335. IEEE (2012)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016
9. Hou, L., Yu, C.-P., Samaras, D.: Squared earth mover’s distance-based loss for training deep neural networks. arXiv preprint [arXiv:1611.05916](https://arxiv.org/abs/1611.05916) (2016)
10. Howard, A.G., et al.: Mobilenets: efficient convolutional neural networks for mobile vision applications. CoRR, abs/1704.04861 (2017)
11. Pal, A., Chaturvedi, A., Garain, U., Chandra, A., Chatterjee, R.: Severity grading of psoriatic plaques using deep CNN based multi-task learning. In: 23rd International Conference on Pattern Recognition (ICPR 2016), December 2016
12. Raina, A., Hennessy, R., Rains, M., Allred, J., Diven, D., Markey, M.K.: Objective measurement of erythema in psoriasis using digital color photography with color calibration. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 3333–3336, August 2014
13. Roenigk, H.H.: Psoriasis Basic and Clinical Dermatology. Informa Healthcare (1998)
14. Shrivastava, V.K., Londhe, N.D., Sonawane, R.S., Suri, J.S.: A novel approach to multiclass psoriasis disease risk stratification: machine learning paradigm. *Biomed. Signal Process. Control.* **28**, 27–40 (2016)