Clustering and Classification of Dermatologic Data with Self Organization Map (SOM) Method
Dermatolojik Verilerin Self Organizing Map (SOM) Yöntemi ile Sınıflandırılması

Ugur FIDAN[1], Nese OZKAN[1], Ismail CALIKUSU[2]
[1]Biomedical Engineering, Afyon Kocatepe University, Afyonkarahisar, Turkey
{ufidan,nozkan}@aku.edu.tr
[2]Biomedical Device Technology, Nevşehir Hacı Bektaş Veli University, Nevşehir, Turkey
{ismailcalikusu}@nevsehir.edu.tr

***Abstract*—Nowadays, skin diseases have increased due to various external effects such as chemical, radiological and so on. In dermatology, the distinguished diagnosis of Erythemato-squamos diseases is a situation that doctors often confront. When many skin diseases are examined, it is seen that many of them are quite similar in shape and appearance although their reasons of emergence are different. Doctors try to distinguish diseases from each other and diagnose by evaluating the clinical findings with pathological parameters. It is observed that many researchers have conducted studies on the Erythemato-Squamous diseases to develop decision support systems using different classification algorithms for detection and diagnosis. Unlike the cited studies in literature, the aim of the present study is to extract Self Organization Maps (SOM) of clinical and pathological findings and investigate cluster of condition various diseases from reduced data. SOM is a size reduction process which aim to simplify the problem. Basically, SOM provides less size reduction output using multidimensional input. In this study, the clinical and pathological classification was realized separately and together. As the result, classification of six types of Erythamato-Squamos skin disease was performed with SOM artificial intelligence application. In addition, clinical and pathological effects of SOM application was seen clearly by showing as a graphically display instead of a matrix. As a result, in the diagnosis of Erythemao-Squamos diseases, it was determined that a dermatologist diagnose mostly depending on the clinical findings although pathological findings contain quantative data.**

***Keywords*—SOM, Artificial Intelligience, Clustering , Dermatology, Erythemato-Squamos**

***Özetçe* —İnsan vücudunun en büyük organı olan deri, kimyasal, radyoaktif vb. gibi birçok dış kaynaklı etkiye maruz kalmasından dolayı günümüzde deri hastalıklarının arttığı görülmektedir. Dermatolojide, Erythemato-squamos hastalıklarına ayırt edici tanı koyulması doktorların sıkça karşılaştığı bir durumdur. Birçok deri hastalığı incelendiğinde birçoğu ortaya çıkış sebepleri farklı olmasına karşın şekil ve görünüş açısından benzerlik taşımaktadır. Doktorlar klinik bulgular ile patolojik parametreleri birlikte değerlendirerek hastalıkları birbirinden ayırt etmeye ve teşhis koymaya çalışmaktadır. Birçok araştırmacının Erythemato-Squamous deri hastalıklarının farklı sınıflandırma algoritmaları kullanarak teşhis ve tanı için karar destek sistemleri geliştirmek üzere çalışmalar yaptığı görülmektedir. Bu çalışmada, literatürde belirtilen çalışmalardan farklı olarak klinik ve patolojik bulguların öz düzenleyici haritalarının (SOM) çıkarılması ve indirgenmiş verilerden farklı hastalıkları kümeleme durumu araştırılması hedeflenmektedir. SOM, problemin basitleştirilmesini amaçlayan bir boyut azaltma işlemidir. Temel olarak SOM çok boyutlu girdilerin daha az boyuttaki çıktılara indirgenmesini sağlar. Sonuç olarak 6 tür Erythamato-squamos deri hastalığının klinik ve patolojik ayrı ayrı ve birlikte sınıflandırması SOM yapay zekâ uygulaması ile gerçekleştirilmiştir. SOM uygulanması matris gösterim yerine grafik olarak gösterilerek sınıflandırmada klinik ve patolojik etkileri daha net görülebilmiştir. Erythemao-Squamos hastalıklarına tanı konulmasında her ne kadar patolojik bulgular nicel veriler taşımış olsa da daha çok dermatologların klinik bulgularına bağlı olduğu belirlenmiştir.**

***Anahtar Kelimeler*—SOM, Yapay Zekâ, Kümeleme, Dermatoloji, Erythemato-Squamos**

I. Introduction

Skin is the largest organ of the human body which is exposed to various external effects such as chemical, radioactive, etc. , and therefore skin diseases have increased. The distinguished diagnosis of Erythemato-squamos diseases is a situation often faced by doctors. When many skin diseases are examined, it is seen that although the emergences are different, they are quite similar in shape and appearance. Therefore, doctors try to distinguish and diagnose a disease by evaluating clinical findings and pathological parameters. Detecting the distinguishing factors between these diseases by classification using various techniques and algorithms are important to assist in the diagnosis of a disease in terms of dermatologists. Analysing studies on the subject, first study shows that by the West. In this study, the SOM network structure model and complex expert systems were applied on dermatologic data, and the results were compared with each other. Seborrheic dermatitis and pityriasis rosea were tried to be distinguished with SOM structure through 6 illness using dermatologic data, and the error rate was found to be 0.01159 in the SOM building [4].

Radwan E. Abdel-Aal and his friends worked on the development of the classification of multiple diseases with the analysis of problems. They classified dermatologic complex

978-1-5090-2386-8/16/$31.00

data sets in the divide and win logic dividing data separated into subgroups. Group Method of Data Handling algorithm (GDMH) and inductive inference mechanism were implemented on the data. It can be said that GDMH algorithm is an adequate model analysing complex structure and high nonlinear structure. GDMH algorithm provides the benefit to classification that protect the system against both of the errors which occur by human and less load on automated analysis system. They divide dermatologic data as training (258) and test (100).

As a result of classification, while the highest achievement was 99%, classification success was obtained as 91% in this study [1]. Kenneth Revett et al. performed a rough clustering approach using dermatologic data set in feature space in their study. When clustering, data is examined as clinical, histopathological and all data. In their rough clustering method, the accuracy of data value was found as a result of the study by removing the criterion of age. Accuracy in the cluster was demonstrated as 0.79 for in clinical data, 0.85 in histopathological data and 0.95 in all data [3]. Haryanto and friends used the Self Organization Map method to determine the Erythemato-Squamous type dermatological diseases by working on the same data set. In the study, 30 of which was for training and 201 was for testing 231 data were separated. SOM structure was used for six dermatologic diseases separately by doing nine experiments with the SOM method, and it demonstrated accuracy percentages of experience for each disease. As a result of studies, each disease has reached a successful result of which is that experience using the best accuracy of SOM structure [7].

It is seen that many researchers have studied to develop decision support systems for detection and diagnosis of Erythemato-Squamous skin diseases using different classification algorithms. In this study, unlike the cited literature studies, the process of obtaining the SOM of clinical and pathological findings and the status of different disease clusters from status of reduced data were investigated. SOM is a size reduction process aiming at simplifying the problem. Basically, SOM provides a reduction of output lesser than the multi-dimensional input. In the result of study, concrete data were put forward about diseases of clinical, and pathological findings were mixed regarding the classification or what the diseases were.

II. Material and Methods

A. Erythemato-Squamous Skin Diseases

The data of this study used Erythematos- Squamos skin diseases were taken from the UCI Machine Learning Database. These data were taken from 375 people aged between 7 and 75 and with some qualifications relating to dermatologic diseases. Erythamato-Squamos dermatological diseases from the database are shown in Figure 1 [8].

B. Clinical and Pathological Findings

All clinical and pathological findings in the database (Table 1) are classified examined by specialist dermatologists. In the diagnosis of the diseases, a total of 34 features are used as evaluation criteria, including 12 clinical and 22 pathological

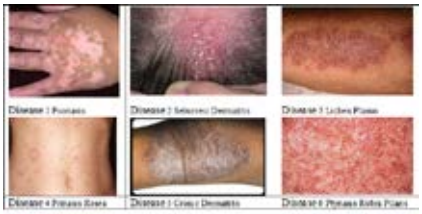

Figure 1: Types of Erythemato-Squamos skin diseases.

parameters. Histopathological findings are quantitative characteristics formed under the microscope while the clinical findings are qualitative parameters used by doctors to diagnose examination.

Clinical Findings		Histopathological Findings			
1	Erythema	1	Melanin Incontinence	13	Munro Mibroabcess
2	Scalling	2	Eosinophils in the Infiltrate	14	Focal Hipergranulosis
3	Definite Border	3	PNL Infiltrate	15	Disappearance of the Granular Layer
4	Itchy	4	Fibrosis of the papillary dermis	16	Vacuolisation and Damage of Basal Layer
5	Koebner Phenomenon	5	Exocytocis	17	Spongiosis
6	Polygonal Papules	6	Aconthosis	18	Saw-tooth Appearance of Retes
7	Follicular Papules	7	Hyperkeratosis	19	Follicular Horn Plaque
8	Oral Mucosal Involvement	8	Parakeratosis	20	Perifollicular Parakeratosis
9	Knee and Elbow Involvement	9	Clubbing of the ridges	21	Inflammatory Mononuclear Infiltrate
10	Scalp Involvement	10	Elongation of the Rete Ridges	22	Bant-like Infiltrate
11	Family History	11	Thinning of the Suprapaillary Epidermis		
12	Age	12	Spongiformpustule		

Table I: Clinical and Pathological Findings

C. Self- Organization Map (SOM)

SOM is a structure of artificial neural network whose input data are placed in the geometric structure with unsupervised learning capable of expressing with a finite number of units. Neighbourly relations of geometric structure arranged unit considered that during updating of the weight vectors which define the unit, the units at the close neighbourhood represent similar data. Thus, the topological properties of the space formed by the input data are transferred to one or two dimensional geometry. Because of this topological feature learning rule that leads to the protection, SOM varies other supervised learning structures used in clustering problem [5]; [2].

SOM algorithm (Figure 2) is composed from appointment of the first values and the number of iterations, determining neighbourhood, presenting data, determining the winner neuron, updating the winner neuron and its neighbourhood and termination steps.

Assignment of the first values: Weight vectors $\omega j \in Rm, j = 1, \ldots, l$ are assigned randomly. Thus, the first value is determined in l neurons against the central block of the data.

$$d_{ij}^{distance} = ||d_i - d_j||, j = 1, \ldots, l, i = 1, .., l \qquad (1)$$

Determiningg the neighborhood: Neighbourhood scheme is determined by placing l neurons a desired in one or two dimensional geometry and calculate the physical distance $d_{ij}^{distance}$ between each neuron.

Presenting Data: Data $xk \in Rm, k = 1, \ldots, p$ is selected from the training set.

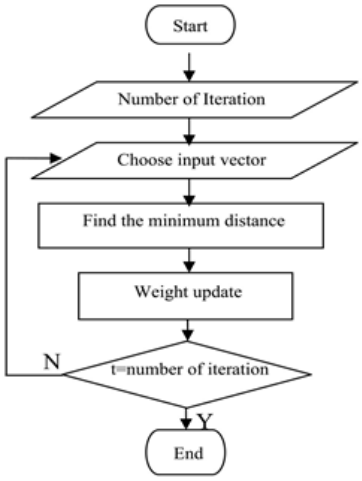

Figure 2: Flow diagram of SOM algorithm.

Determining the winner neuron: $d_{ij}^{similarity}$ is calculated with equation 2 to determine the most similar k.data to in l neuron selected from the training set.

$$d_{ki}^{similarity} = (\sum_{j=1}^{m}(x_{kj} - w_{ij})^2)^{\frac{1}{2}}, i = 1, \dots l \quad (2)$$

The winner neuron is determined with k.data.

$$indis^{winner}(k) = min_i\{d_{ki}^{similarity}\}, i = 1, \dots .l \quad (3)$$

Updating the weights of the winning neuron and its neighbours: Winner weight to a maximum of neurons, the weight of the winning neuron neighbourhood is updated with Equation 4 according to determined neighbourhood function $Q(d_{ij}^{distance}, n)$.

$$w(n+1) = w(n) + Q(d_{ij}^{distance}, n)\mu(n)(x(n) - w(n)) \quad (4)$$

Where, $\mu(n)$ corresponded to the learning speed expressed as Equation 5 and it decreased exponentially during sequential updates.

$$\mu(n) = \mu_0 exp(\frac{-n}{\lambda}) \quad (5)$$

In another consecutive term depends update is neighbourhood function and contracted for sequential updates.

Termination: For the same data in the training set, bunching is formed as started to winner same neurons, but after this stage education process continues a time that the number of steps until the stage up to about four times for neurons corresponding to heap centres to enable the features that represent data in a heap [6]. Weights are changed with the presentation of data in the training set to the network again and again by adapting to the weight of neurons. It is transformed into a same neuron space of discrete neurons by protecting characteristics of the input where the data are located.

D. Clustering of Erythemato-Squamos Skin Diseases with SOM

Diagnostic parameters of Erythemato-Squamos skin diseases taken from Machine Learning Repository (UCI) were separated into two groups according to clinical and pathological data. Only clinical, only pathological and just for all input parameters 20x20 SOM clustering were implemented using "Neural Network Clustering Tool-nctool" development application tool of MATLAB 2013 software program. MATLAB functions are created for SOM determined weight between training is completed and neurons.

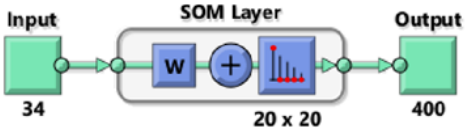

Figure 3: The Structure of SOM Network

III. Results and Discussions

Thirty-four findings including clinical and pathological data of Erythemato-Squamos disease groups were used as input parameters . The width of the SOM map was set to 20X20 and map of the network was optained. Topologically clustering of all input parameters in Table 1 in the network are obtained as in Figure 4.

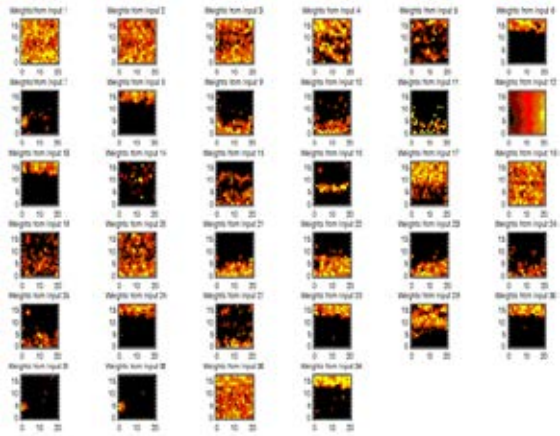

Figure 4: The clustering of all input parameters in SOM

In the second stage, SOM clustering is done using 22 pathological findings shown in Table 1. The width of the SOM map is set to 20X20 and the map of the network was optained.

SOM clustering is made using 12 clinical findings seen Table 1. The width of the SOM map was set to 20X20 and topological map of the network was optained. Likewise, clustering structure in the network (Figure 5) is obtained using the clinical parameter as input.

Figure 4 and Figure 5 show the input parameters based on the properties of the cluster. However, it is necessary to

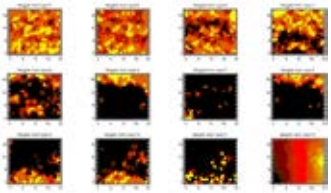

Figure 5: Clustering of clinical input parameters in SOM

identify the Erythemato-Squamos patients removed from the map of attributes for the classification of diseases.

Entry data classified by SOM set one cell as "1" and set other cells as "0" in a 20X20 matrix. SOM was applied to clinical, pathological and the functions created for all entries by classifying diseases in the database. For six types of Erythemato-Squamos disease in 20x20 matrix, which index number of active cell determined by written script program. Figure 6 shows the SOM output index belonging to six different groups of diseases which uses all input parameters.

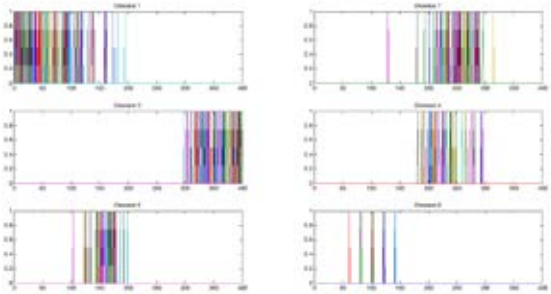

Figure 6: Classification of Erythemato-Squamos diseases for all input parameters

When examined in Figure 6, Lichen Planus can be seen except among themselves the other diseases. Especially, chronic dermatitis with Psoriasis and Seborreic Dermatitis with Pitriasis Rosea are mixed with each other. These results indicate the cause of the low performance of studies in the literature.

In Figure 7, different disease groups of SOM output index are seen for clinical and pathological findings. According to the pathological findings made in clustering, it is seen that all diseases are mixed among all themselves. However, Seborrheic Dermatitis and Chronic Dermatitis diseases are possible by mixing the Dermatitis. As shown in Figure 6, although pathological findings have moved quantitative data, it is seen to be dependent on more dermatologists of clinical symptoms.

IV. Summary and Future Expectations

When many skin diseases are examined, it can be seen that many of them are quite similar in shape and appearance although their reasons of emergence are different. It is observed that many researchers have

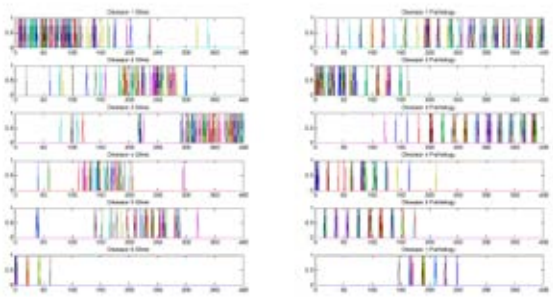

Figure 7: Classification of Clinic and Pathological Parameters for Erythemato-Squamos Diseases

done studies on the Erythemato-Squamous diseases to develop decision support systems using different classification of algorithms for detection and diagnosis. In this study, unlike the cited studies in literature, it was conducted extracting Self Organization Maps (SOM) of clinical and pathological findings and investigating cluster of condition various diseases from reduced data. As a result, it was implemented clinical and pathological classification of 6 types of Erythemato-Squamos with SOM skin diseases separately and together.The clinical and pathological effects could be seen more clearly in classification by showing SOM application graphically instead of matrix display. In diagnosing Erythemato-Squamos diseases, although pathological findings include quantitative data it was determined due to more clinical findings by dermatologists. In the future studies, it will try to improve the SOM classification performance using networks and other artificial intelligence techniques together in the classification of the point where the SOM network failed.

Reference

[1] Abdel-Aal, R. E., Abdel-Halim, M. R.,& Abdel-Aal, S. (2006). Improving the classification of multiple disorders with problem decomposition. Journal of biomedical informatics, 39(6), 612-625.

[2] Haykin, S. S., Haykin, S. S., Haykin, S. S., & Haykin, S. S. (2009). Neural networks and learning machines (Vol. 3). Upper Saddle River: Pearson Education.

[3] Revett, K., Gorunescu, F., Salem, A. B., & El-Dahshan, E. S. (2009). Evaluation of the feature space of an erythematosquamous dataset using rough sets. Annals of the University of Craiova-Mathematics and Computer Science Series, 36(2), 123-130.

[4] West, D., & West, V. (2000). Improving diagnostic accuracy using a hierarchical neural network to model decision subtasks. International journal of medical informatics, 57(1), 41-55.

[5] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological cybernetics, 43(1), 59-69.

[6] Nouri, F., & Şengör, N. S. (2012).Öz-düzenlemeli Ağ Yapısı ile Farklı Yaklaşımların Sınanması Testing Different Approaches by Self Organizing Map.

[7] Haryanto, H., Ulum, M., Rahmawati, D. R., Joni, K., Ubaidillah, A., Alfita, R., ... & Khotimah, B. K. (2015). The Erythemato-Squamous Dermatology Diseases Severity Determination using Self-Organizing Map. IPTEK Journal of Proceedings Series, (1).

[8] Fidan,U.,Ozkan,N.(2016).The Effect of Clinical And Histopathological Parameters In The Classification Of Dermatological Disesases With Artificial Intelligence.IMUCO'16.103-111.