A Hybrid Rough-Neuro model For Diagnosing Erythemato-Squamous Diseases

Shahenda Sarhan* Enas Elharir Magdi Zakaria*

* Faculty of Computers and Information, Mansoura University, Egypt.

Abstract

In this paper, a Rough-Neuro hybrid methodology of the diagnostic process is proposed as a means to achieve accurate diagnosing of Erythemato-Squamous diseases. The methodology incorporates a two-stage hybrid mechanism. Rough sets Johnson Reducer for reduction of relevant attributes and artificial neural network Levenberg-Marquardt algorithm for the classification of the diseases. The model achieved really good results in the diagnosing process that approached 98.8% diagnosing accuracy.

Keywords: *Dermatology, Erythemato-Squamous diseases, Rough set, Reducts, Artificial Neural Networks.*

1. Introduction

Skin is the largest organ of the body and is the most visible. Although many skin diseases are isolated, some are manifestations of an internal disease [17]. Skin diseases are usually caused by infection with viral or bacterial organisms, it can also occur due the imbalance in the physiological condition of the skin.

Erythemato-Squamous is a class of dermatological disorders which presented by redness of the skin (Erythema) due to the loss of skin cells (Squamous). Erythemato-Squamous usually occurred due to genetic or environmental cause, and is common in late childhood and early adolescence [6].

The similarity of the clinical features of Erythemato-Squamous diseases with Erythema and scaling make differential diagnosis of Erythemato-Squamous diseases real difficult [5]. Among Erythemato-Squamous diseases, there are six groups [4] [5], Psoriasis, Seboreic Dermatitis, Lichen Planus, Pityriasis Rosea, Chronic Dermatitis and Pityriasis Rubra Pilaris. These groups of diseases are frequently observed in the outpatient dermatology departments.

These cases usually present similar clinical features with those of Erythema and scaling with slight variations. When manifested carefully [5], some patients have the typical clinical features of the disease at the predilection sites while another group has typical localizations.

Patients were subjected to evaluation using 33 features divided into 12 clinical features as the degree of Erythema and scaling, whether the borders of lesions are definite or not and The formation of papules, whether the oral mucosa, elbows, knees and the scalp are involved or not [2][3]. The rest 21 are histopathological features for example the stage of the disease. While some diseases show their characteristic signs only in late stage the others show interfering signs with other diseases in early stages.

From all of the above we had the motive to try to solve the Erythemato-Squamous diagnoses problem through proposing a rough-neuro model. The proposed Erythemato-Squamous Rough-Neuro model combines and optimizes first a rough set attribute reduction algorithm, second a neural network classifier.

1.1 Rough Set

Rough set (RS) theory [7][14] is a fairly new intelligent technique for managing uncertainty that is used for the discovery of data dependencies, to evaluate the importance of attributes, to discover patterns in data, to reduces redundancies, and to recognize and classify objects. Moreover, it is being used for the extraction of rules from databases where one advantage is the creation of readable if-then rules. Such rules have a potential to reveal previously undiscovered patterns in the data; furthermore, it also collectively functions as a classifier for unseen samples.

Unlike other computational intelligence techniques, rough set analysis requires no external parameters and uses only the information presented in the given data [15][16]. One of the nice features of rough set theory is that it can tell whether the data is complete or not based on the data itself. If the data is incomplete, it will suggest that more information about the objects is required.

On the other hand, if the data is complete, rough sets are able to determine whether there are any redundancies in the data and find the minimum data needed for classification. This property of rough sets is very important for applications where domain knowledge is very limited or data collection is expensive/laborious because it makes sure the data collected is just sufficient to build a good classification model without sacrificing the accuracy or wasting time and effort to gather extra information about the objects [14][15][16].

1.2 Neural Networks

Artificial neural networks (ANN) are computational models of nervous systems [10][12], a neural network is a system composed of many simple processing elements called neurons operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes.

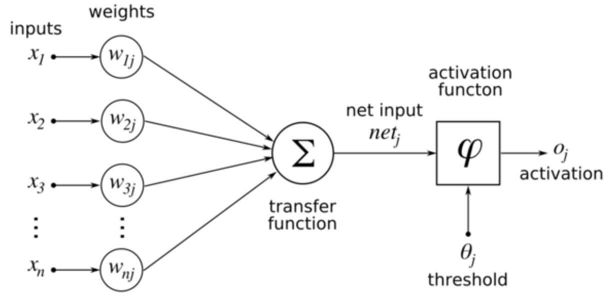

Fig 1: Neural Network Neuron [18]

Neurons are grouped into layers or slabs. The neurons in each layer are the same type. Each neuron is connected to other layers by means of interconnections or links with an associated weight [3][9]. The behavior of an ANN depends on both the interconnections and the input-output function (transfer function) that is specified for the units.

Neural networks is trained rather than programmed. They have excellent capability of learning the relationship between input-output mapping from a given dataset without any knowledge or assumptions about the statistical distribution of data. This capability of learning from data without any prior knowledge makes neural networks particularly suitable for classification and regression tasks in practical situations and in most financial and manufacturing applications.

Neural networks are also inherently nonlinear [10] which makes them more practical and accurate in modeling complex data patterns as opposed to many traditional methods which are linear.

The paper is organized as follows. In Section (2) we will present a brief overview of the previous Erythemato-Squamous diagnosing models. In Section (3) a detailed description of our proposed Rough-Neuro model. Model results will be in Section (4) and finally in Section (5), conclusions are drawn.

2. Background

Many algorithmic and architectural approaches have been proposed to solve Erythemato-Squamous diagnosis uncertainty problem. Some of them are overly complicated or costly to implement. It is still an open and challenging problem to find practical solutions.

Parthiban (2009) introduced an approach based on Coactive Neuro-Fuzzy Inference System (CANFIS) for the detection of the Erythemato-Squamous. The proposed CANFIS model combined the neural network adaptive capabilities and the fuzzy logic qualitative approach which are then integrated with genetic algorithm. [8]

Revett et. al., (2009) presented a differential diagnosis of Erythemato-Squamous using rough sets. The results indicated that the histopathological feature space provided a much more significant classification rate relative to clinical features. In addition, the results of this study indicated that only a small subset of the histopathological feature space is required for maximal classification accuracy [6].

While, Ubeyli and Doğdu (2010) introduced a new approach based on the implementation of k-means clustering presented for automated detection of Erythemato-Squamous diseases. The algorithm was used to detect the five Erythemato-Squamous diseases and to determine an optimum classification scheme for this problem [13].

In (2012) Aruna and others presented a diagnostic model based on Naive Bayes developed to diagnose Erytemato-Squamous diseases. The hybrid feature selection method, named Information Gain and Sequential Backward Floating Search, combined the advantages of filters and wrappers to select the optimal feature subset from the original feature set. [11]

Finally Xie and others (2013) molded two-stage hybrid feature selection algorithms for diagnosing Erythemato-Squamous diseases. The algorithms adopt Support Vector Machines (SVM) as a classification tool, the extended Sequential Forward Search (SFS), Sequential Forward Floating Search (SFFS), and Sequential Backward Floating Search (SBFS), respectively, as search strategies, and the generalized F-score (GF) to evaluate the importance of each feature [4].

All the challenges that face the Erythemato-Squamous diagnoses in the previous work were concerning the diagnosing features how to measure them accurately to get a correct diagnosing of your disease as measuring 33 features is not an easy job besides its money and time consuming. In the next section we will introduce our proposed model called Rough-Neuro Erythemato-Squamous diagnosing Model (RNESD) that tried to solve the huge amount of features values using a hybrid of the rough sets and Neural Networks. Next we will discuss the proposed model clarifying its stages in details

3. The Rough-Neuro Erythemato-Squamous Diagnosing Model

Before starting to go through the proposed model we need first to clarify that the dataset were taken from UCI

Repository containing 33 features values for 366 patients' records (200 records for training and the rest for testing). Patients were first evaluated clinically with 12 features and histopathologically with 21 features.
In the dataset, the family history feature has the value 1 if any of these diseases has been observed in the family and 0 otherwise. Every other feature (clinical and histopathological) was given a degree in the range of 0 to 3. Here, 0 indicates that the feature was not present, 3 indicates the largest amount possible, and 1, 2 indicate the relative intermediate values. We will now start to introduce the proposed model two stages as following:

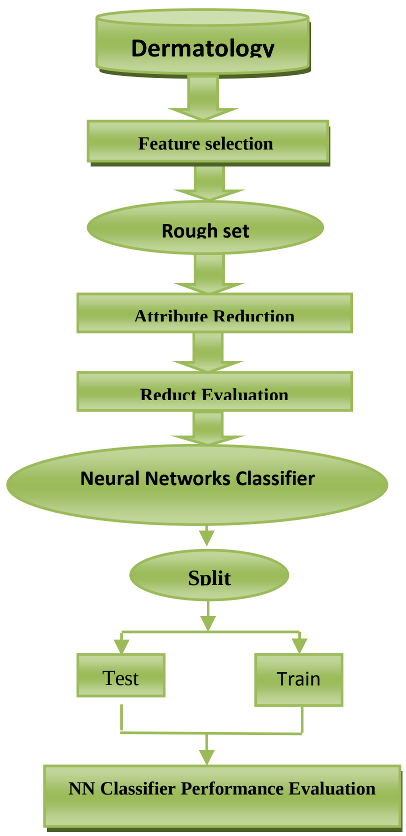

Fig. 2: Erythemato-Squamous Rough-Neuro Model Steps

3.1 Rough Set Phase

Using RSs attributes reduction (RSAR) algorithms programmed in Aleksander Øhrn ROSETTA [1], we tried to reduce the number of measured attributes through applying them on each of the five reduction algorithms, generating each algorithm reducts and rules.

The data is available as Microsoft Access database, loaded in ROSETTA as an ODBC [1] and applied in each reduction algorithm. Table 1 displays the evaluation measurements of reducts and rules produced by each algorithm.

From table (1) we found that Holte1RReducer has the lowest number of rules but it also has the lowest support which eliminates it from our list of choice.
After Holte1RReducer elimination we found that Johnson Reducer have the lowest number of rules, the lowest cardinality and has the highest support. This tells us that Johnson Reducer (Figure 4) is the best choice for our system.

Table 1: The Evaluation Measurements of Reducts and Rules.

Reduction algorithm	No. of reducts	No. of rules	cardinalities of reduct	Support of reduct
SAVGeneticReducer	212	73601	11,13,14,15,16,17 18,19,20,21	100
JohnsonReducer	1	323	8	100
Holte1RReducer	33	129	1	1
ManualReducer	1	363	26	0
Genetic algorithms	207	73328	11,12,13,14,15,16, 17, 18,19,20,21	100

From table (1) we can also see that Johnson Reducer provides only one reduct with eight attributes that gives the same decision as the whole attributes but with 76% reduction in complexity (Figure 3)

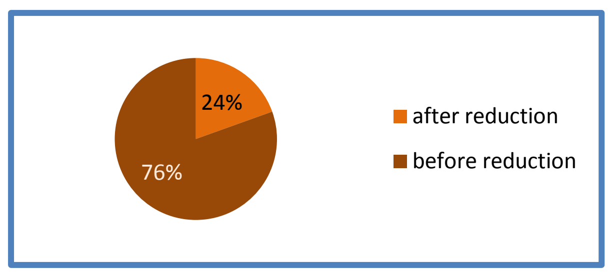

Fig. 3: Displays the Attribute Before and After Reduction

Let A be a "universal " set of n elements, $\$=\{S_1,S_2,\ldots.S_k\}$ a collection of subset U formaing a cover for it, and $c:\$\rightarrow Q^{+}$ a cost function . Johnson's approximation algorithm finds a sub-collection of \$ covring all the elements of U at minimal cost
Input : C =0, T=0
Output : T

1. Step1 let C=0 , T =0
2. Step2 while C≠U do
   - Find S € \$ such that c(S)\|S\C| is minimum
   - ¥ x € s, define cost(x)= C(S)\ |S\C|.
   - C←C U S , T ←C U {S}
3. Step3 Result = T

Fig. 4 : Johnson Reducer Algorithm

3.2 Neural Network Classifier Phase

The Neural Network Classifier in our model consists of three layers

Table 2: Confusion Matrix Of The Classifier Using The Full Dataset

Desired output / Output results	psoriasis	seboreic dermatitis	lichen planus	pityriasis rosea	cronic dermatitis	pityriasis rubra pilaris
Psoriasis	49	0	0	0	0	0
seboreic dermatitis	0	28	0	0	0	0
lichen planus	0	0	32	0	0	0
pityriasis rosea	1	0	0	22	0	0
cronic dermatitis	0	0	0	0	24	0
pityriasis rubra pilaris	0	0	0	1	0	9

- Input layer: have 8 input neurons.
- Hidden layer: the number of neurons in hidden layer varied from the ½ of the number of inputs to the double number of inputs + 1. So we had to test our data in all different cases of hidden layer number of neurons from 4 hidden neurons to 18 hidden neurons. The logsig activation function was used in this layer.
- Output layer: the output layer had six neurons depending on having six diseases as output. The linear activation function was used in this layer.

The training and testing of the network was performed with the Neural Networks Toolbox in MATLAB environment by the Levenberg-Marquardt algorithm in two cases at learning rate LR=0.5 and LR = 1.
The training was performed at 1500, 2000 and 2500 epochs where different random initial weights are used in each trial, the network is trained until the mean square error reaches zero. Figures 5 and 6 display the MSE test results for FFNN classifiers with different number of hidden neurons.

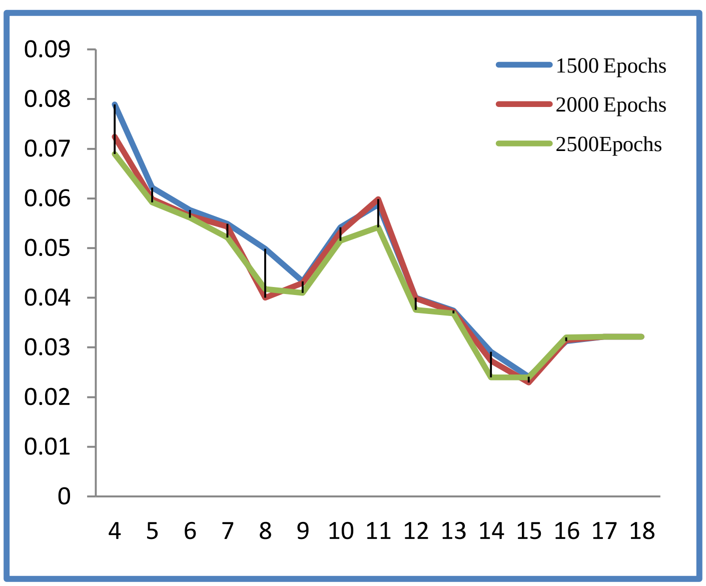

Fig. 5: Displays After Reduction MSE Results for LR=0.5

From figures 5 and 6 we can observe that the least MSE value 0.0229145 appeared at a classifier with 15 hidden neurons trained at 2000 epochs with LR =0.5 while the largest one was 0.24456 at a classifier with 5 hidden neurons trained at 1500 epochs with LR =1.

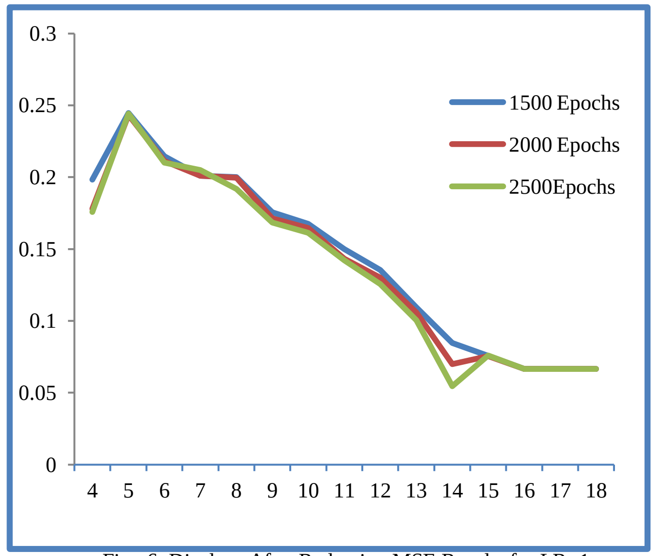

Fig. 6: Displays After Reduction MSE Results for LR=1

From all of the above we can finally conclude that the proposed model classifier is a feed forward neural network (8 Inputs – 15 Hidden Neurons – 6 Outputs) trained at 2000 epochs with the Levenberg-Marquardt algorithm.

As a final step we tried to evaluate the proposed model through constructing a confusion matrix to measure the accuracy of the system diagnosing ability as demonstrated in table (2) that the system accuracy approached 98.8% which is considered a great result.

4. Conclusions

Through this study we used a combination of rough sets and artificial neural networks to provide a useful and simple methodology for solving the similarity of the features of Erythemato-Squamous diseases which led to difficult diagnosing of the Erythemato-Squamous diseases. The proposed solution is straightforward and simple due to the two-stage approach. Rough sets allow us to easily have an effective reduct of our data and its belonging rules using Johnson's reducer algorithm programmed in Aleksander Øhrn application ROSETTA.

Using ROSETTA we reduced 33 attributes to only eight (76% reduction in complexity) . This reduction solves the neural network problem with the increasing size of training data. As the 8 inputs neural network classifier produces the same accuracy as the 33 inputs one does besides saving memory and time. The rough sets-based reduction of the attribute space improves the efficiency of the neural classifier showing that the classification error is at least comparable, if not smaller, when the set of inputs reduced.

REFERENCES

[1]. A. Øhrn, (2001). "ROSETTA Technical Reference Manual". Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway. http://www.idi.ntnu.no/~aleks/rosetta/>

[2]. E. Derya and I. Guler, (2005)."Automatic Detection of Erythemato-Squamous Diseases Using Adaptive Neuro-Fuzzy Inference Systems".cxComputers in Biology and Medicine, Vol. (35), pp. 421–433.

[3]. H. xiao, X. zhang & Y. Du, (2008). "A Comparison of Neural Network, Rough Sets and Support Vector Machine on Remote Sensing Image Classification". 7th WSEAS Int. conf. on applied computer & applied computational science (ACACOS '08), Hangzhou, Vol. (7), pp. 552-558.

[4]. J. Xie, J. Lei, W. Xie, Y. Shi and X. Liu, (2013). "Two-Stage Hybrid Feature Selection Algorithms for Diagnosing Erythemato-Squamous Diseases". Health Information Science and Systems, Vol. (**1**), Issue (10).

[5]. J. Xie, W. Xie, Ch. Wang, (2010). "A Novel Hybrid Feature Selection Method Based on IFSFFS and SVM for the Diagnosis of Erythemato- Squamous Diseases". Workshop on Applications of Pattern Analysis. pp. 142-151.

[6]. K. Revett, F. Gorunescu, A. Salem and E. El-Dahshan, (2009). "Evaluation of the Feature Space of an Erythematosquamous Dataset Using Rough Sets". Annals of University of Craiova, Math. Comp. Sci. Ser. Vol. (36), Issue (2), pp. 123-130. ISSN: 1223-6934.

[7]. L. Polkowski, (2003). "Rough Sets: Mathematical Foundations". Physica- Verlag.

[8]. L. Parthiban, R. Subramanian, (2009). "An Intelligent Agent for Detection of Erythemato-Squamous Diseases Using Co-Active Neuro-Fuzzy Inference System and Genetic Algorithm". 978-1-4244-4711-©2009 IEEE.

[9]. R. Sassi, L. da Silva & E. Hernandez, (2006)." Neural Networks and Rough Sets: A Comparative Study On Data Classification". Department of Electronic Systems Engineering University of Sao Paulo, Brazil.

[10]. R. Stathacopoulou, et. al., (2004)." Neuro-Fuzzy Knowledge Processing in Intelligent Learning Environments for Improved Student Diagnosis". Information Sciences 170 (2-4), pp. 273-307. ISSN 0020-0255.

[11]. S. Aruna, L. Nandakishore and S. P. Rajagopalan, (2012). "A Hybrid Feature Selection Method based on IGSBFS and Naïve Bayes for the Diagnosis of Erythemato - Squamous Diseases". International Journal of Computer Applications (0975 – 8887). Vol. (41), No.(7).

[12]. T. Smolinski, D. Chenoweth, and J. Zurada, (2003). "Rough Set and Artificial Neural Networks-based Hybrid Approach to Financial Time Series Forecasting". The IASTED International Conference on Neural Networks and Computational Intelligence, pp. 108–111.

[13]. E. Ubeyli , E. Doğdu , (2010). "Automatic detection of erythemato-squamous diseases using k-means clustering'. Department of Electrical and Electronics Engineering, Faculty of Engineering". Springer US, Journal of Medical Systems . Vol.(34), No. (2), pp. 179-184, ISSN 0148-5598.

[14]. Z. Pawlak, (1991). "Rough Sets- Theoretical aspect of Reasoning about Data". Kluwer Academic Publishers.

[15]. Z. Pawlak, (1982). "Rough sets". International J. Comp. Inform. Science, Vol. (11), pp.341-356.

[16]. Z. Pawlak, J. Grzymala-Busse, R. Slowinski and W. Ziarko, (1995). "Rough Sets". Communications of the ACM, Vol.(38), No.(11), pp.88-95.

[17]. Overview of Dermatology. http://www .Patientsmedical.com/healthaz/dermatology/default.aspx.

[18].WWW. Wikipedia.org

Shahenda Sarhan is a lecturer at computer science department –Mansoura University – Mansoura – Egypt. She had her M.Sc in E-learning and Ph.D in Computer Games from Mansoura University. Currently she is a coordinator of two research groups on NLP and computer games in Mansoura University.

Enas El-Harir is a master student in the Department of Computer Science - Mansoura University – Egypt. She received her Diploma and B.Sc. in Computer Science.

Magdy Zakaria is a professor in Artificial Intelligence in Computer Science Department in Mansoura University. He has supervised many PhDs and masters mostly specialized in Artificial Intelligence and its applications related to real life.