

Identification of Erythemato-Squamous Skin Diseases using Support Vector Machines and Extreme Learning Machines: A Comparative Study towards Effective Diagnosis

S. O. Olatunji¹ and Hossain Arif²

¹*Computer Science Department, College of Computer Science and Information Technology,
University of Dammam, Saudi Arabia*

²*School of Engineering and Computer Science, BRAC University, Dhaka–Bangladesh*

¹olulolunji.aadam@gmail.com; ¹osunday@uod.edu.sa

²hossain.arif@gmail.com; ²hossain.arif@bracu.ac.bd

ABSTRACT

Extreme Learning Machine (ELM) is a recently introduced learning algorithm for single hidden-layer feedforward neural network. Compared with classical learning algorithms in neural network, e.g. Back Propagation, ELM can achieve better performance with much shorter learning time. In the existing literature its better performance and comparison with Support Vector Machine (SVM), over regression and general classification problems, have caught the attention of many researchers. In this work, a comparison between ELM and SVM on identification of Erythemato-Squamous skin diseases is investigated. Detailed comparative studies were carried out through adequate experimentation. Experimental results indicated that ELM outperformed SVM. The effect of varying the size of training and testing sets on the performance of classifiers was also investigated in this study. The two techniques compared proved to be viable tools in this germane field of medical diagnosis.

Keywords: Extreme Learning Machine; Support Vector Machine; Erythemato-squamous skin diseases.

1 Introduction

The differential diagnosis of erythemato-squamous diseases is a real and difficult problem in dermatology. There are six groups in erythemato-squamous diseases. They are psoriasis, seboric dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris. These diseases are frequently seen in the outpatient dermatology departments. Since all erythemato-squamous diseases share the clinical features of erythema and scaling with very little differences, it is very challenging to identify which particular variant of the diseases is present in a particular patient.

With the recent advancement in Artificial Intelligence (AI) based techniques and their successful application in real life engineering and medical diagnosis, an interesting research problem is to look at how these diseases can be identified easily based on such AI techniques. Although the diagnosis of erythemato-squamous diseases have recently caught the attention of AI researchers leading to the use of some AI techniques [1-4], no one has yet implemented extreme learning machines in the identification of these diseases and compared its performance with support vector machines.

SVM is acclaimed by many researchers for its leading performance [3, 4]. Therefore, it has been widely used for different classification purposes. Most recently, a new learning algorithm, extreme learning machine (ELM), is available for the training of single layer feedforward neural network. The inventors of ELM have done a set of comprehensive experiments in regression and general classification to compare its performance with SVM [4-7]. The experimental results show that compared with classical learning algorithms in neural network, e.g. Back Propagation, ELM can achieve better performance with much shorter learning time [8]. Compared with SVM, ELM is sometimes better than SVM in terms of accuracy, though not always. But since the number of neurons available for each ELM machine is the only parameter to be determined, ELM is much simpler for parameter tuning compared with SVMs whose kernel functions are nonlinear, e.g. RBF functions, thus saving tremendous time in searching for optimal parameters.

Therefore, this work is set to present a comparative study of the performance of ELM and SVM on the identification of erythemato-squamous skin diseases. The effect of varying size of training and testing sets on the performance of classifiers was also investigated in this study. To do this, we partition the available dataset into training and testing sets based on four different ratios, which include 80:20, 70:30, 60:40 and 50:50 percentages of training and testing respectively. Each of these ratio partitioning was used to carry out experiments involving training and testing the proposed classifiers to be compared.

Results from experiments carried out indicated that ELM outperformed SVM in both training and testing results. The testing set is the most crucial, as it is where the ability to identify unseen sample is tested.

The remaining part of this paper is organized as follows: section two contains the literature review; section three describes the proposed models, while empirical works and discussion of results are presented in section four. Conclusion and recommendations are presented in section five.

2 Literature Review

Juanying Xie et al. proposed hybrid feature selection algorithms to build efficient diagnostic models based on a new accuracy criterion, generalized F-score (GF), and SVM [1]. The hybrid algorithms adopt Sequential Forward Search (SFS), and Sequential Forward Floating Search (SFFS), and Sequential Backward Floating Search (SBFS), respectively, with SVM to accomplish hybrid feature selection with the new accuracy criterion to guide the procedure. These hybrid methods are called modified GFSFS, GFSFFS and GFSBFS, respectively. These combine the advantages of filters and wrappers to select the optimal feature subset from the original feature set to build efficient classifiers. Xie et al. conducted 10-fold cross validation experiments on training subset as well as on the whole erythemato-squamous diseases datasets to obtain the best and statistically meaningful classifiers. Experimental results showed that their proposed hybrid methods construct efficient diagnosis classifiers with high average accuracy when compared with traditional algorithms. The results of 10-fold cross validation experiments on erythemato-squamous diseases dataset show that the proposed hybrid feature selection algorithms, modified GFSFS, GFSFFS, and GFSBFS, have obtained the average classification accuracies of 99.17%, 98.33%, and 95.28% with the average size of selected feature subsets of 22, 13.2, and 17.1 respectively. While the classification accuracy of GFSFS, GFSFFS, and GFSBFS are 98.89%, 98.06%, and 95.81%, and the size of selected feature subset of them are 22.2, 13.3, and 19.3 respectively.

In [2], Xie and Wang developed a diagnosis model based on support vector machines (SVM) with a novel hybrid feature selection method to diagnose erythemato-squamous diseases. They proposed an

improved hybrid feature selection method, named improved F-score and Sequential Forward Search (IFSFS), which is a combination of filter and wrapper methods to select the optimal feature subset from the original feature set. The new IFSFS method improved the original F-score from measuring the discrimination of two sets of real numbers to measuring the discrimination between more than two sets of real numbers. The improved F-score and Sequential Forward Search (SFS) are combined to find the optimal feature subset in the process of feature selection. The best parameters of kernel function of SVM are found out by grid search technique. Xie and Wang then conducted experiments on different training-test partitions of the erythemato-squamous diseases dataset taken from UCI (University of California Irvine) machine learning database. Their experimental results show that the proposed SVM-based model with IFSFS achieves 98.61% classification accuracy and contains 21 features. The authors conclude that their method is very promising compared to the previously reported results.

Kecman and Kikec present the results of using Support Vector Machines (SVMs) and Radial Basis Function Neural Networks (RBF NNs) for diagnosing erythemato-squamous diseases [3]. The paper shows an application of RBF NN and SVM for diagnosing skin diseases by transforming a K-class problem into K two class problems (one-vs-all approach for multiclass problems). The data set contains 358 data pairs of 34 dimensional input records of patients with six known diagnosis (outputs). Thus, the data set is sparse and fairly unbalanced. The paper also discusses the strategies for training SVMs. Both networks design six different one-against-other classifier models which show extremely good performance on previously unseen test data. The training and the test sets are obtained by randomly splitting the dataset into two groups ensuring that each group contains at least one patient for each disease. 100 random split trials (equivalent to performing 10-fold-crossvalidation 10 times independently) were carried out for estimating the tests error rates. SVM models perform better than RBF NN ones, and the SVM models using both and polynomial kernels can perfectly classify, during the training unseen, test data. For a given data set SVMs using the polynomials of second order were particularly efficient and accurate. They use only between 5% and 10% of training data as the support vectors achieving perfect, error-less, diagnosis. Kecman and Kikec claim that the results shown are the best known to date for diagnosing erythemato-squamous diseases which represent difficult dermatological problems.

In their research paper, Ubeyli and Guler presented a new approach for detection of erythemato-squamous diseases based on adaptive neuro-fuzzy inference system (ANFIS) [4]. Neuro-fuzzy systems are fuzzy systems which use artificial neural networks (ANNs) theory in order to determine their properties (fuzzy sets and fuzzy rules) by processing data samples. Neuro-fuzzy systems harness the power of two paradigms: fuzzy logic and ANNs, by utilizing the mathematical properties of ANNs in tuning rule-based fuzzy systems that approximate the way man processes information. The ANFIS learns features in the data set and adjusts the system parameters according to a given error criterion. The six ANFIS classifiers were used to detect the six erythemato-squamous diseases when 34 features defining six disease indications were used as inputs. Each of the ANFIS classifiers was trained so that they are likely to be more accurate for one type of erythemato-squamous disease than the other diseases. The predictions of the six ANFIS classifiers were combined by the seventh ANFIS classifier. To improve diagnostic accuracy, the seventh ANFIS classifier (combining ANFIS) was trained using the outputs of the six ANFIS classifiers as input data. The performances of the ANFIS model were evaluated in terms of training performances and classification accuracies, and the results confirmed that the proposed ANFIS

model has good potential in detecting erythemato-squamous diseases. The total classification accuracy of the ANFIS model was 95.5%. The ANFIS model achieved accuracy rates which were higher than that of the stand-alone neural network model.

3 Proposed Modeling Techniques

In order to facilitate comparing the results of ELM experiments with those of SVM, SVM and ELM techniques will be briefly described.

3.1 Support Vector Machines

SVMs are modern learning systems that deliver state-of-the-art performance in real world Pattern Recognition and data mining applications such as Text Categorization, Hand-Written Character Recognition, Image Classification, Material Identification and Bioinformatics, to mention but a few. Support Vector Machines have been recently proposed as a new intelligence framework for both prediction and classification based on both structure risk minimization criterion and soft margin hyperplane. This new framework deals with kernel neuron functions instead of sigmoid-like ones, which allows projection to higher planes and solves more complex nonlinear problems. It has featured in a wide range of medical and business journals, often with promising results.

Generally, in prediction and classification problems, the purpose is to determine the relationship among the set of input and output variables of a given dataset $D = \{Y, X\}$ where $X \in R^p$ represents the n-by-p matrix of p input variables. It may be noted that $Y \in R$ for forecasting problems and $Y \subseteq R$ for classification problems. Suppose $D = \{y_i, x_{i1}, \dots, x_{ip}\}$ is a training set for all $i = 1, \dots, n$ of input variables, X_j where $[X_j = (x_{1j}, \dots, x_{nj})^T]$ for $j = 1, \dots, p$, and the output variables, $Y = (y_1 \dots y_n)^T$. The lower case letters $x_{i1}, x_{i2}, \dots, x_{ip}$ for all $i = 1, \dots, n$ refer to the values of each observation of the input variables, and $y = k$ to the response variable Y to refer to class A_k for all $k = 1, 2, \dots, c$, where $c \geq 2$.

Here we briefly describe the basic ideas behind SVM for pattern recognition, especially for the two-class classification problem, and refer readers to [13, 14] for a full description of the technique.

The goal is to construct a binary classifier or derive a decision function from the available samples which has a small probability of misclassifying a future sample. SVM implements the following idea: it maps the input vectors $\vec{x} \in \mathbb{R}^d$ into a high dimensional feature space $\Phi(\vec{x}) \in \mathbf{H}$ and constructs an Optimal Separating Hyperplane (OSH), which maximizes the margin, the distance between the hyper plane and the nearest data points of each class in the space \mathbf{H} . Different mappings construct different SVMs. The mapping $\Phi(\cdot)$ is performed by a kernel function $K(\vec{x}_i, \vec{x}_j)$ which defines an inner product in the space \mathbf{H} . The decision function implemented by SVM can be written as:

$$f(\vec{x}) = \text{sgn} \left(\sum_{i=1}^N y_i \alpha_i \cdot K(\vec{x}, \vec{x}_i) + b \right) \quad (1)$$

Where the coefficients α_i are obtained by solving the following convex Quadratic Programming (QP) problem:

$$\text{Maximize} \quad \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j \cdot y_i y_j \cdot K(\vec{x}_i, \vec{x}_j)$$

Subject to

$$0 \leq \alpha_i \leq C \quad (2)$$

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad i = 1, 2, \dots, N.$$

In the equation (2), C is a regularization parameter which controls the tradeoff between margin and misclassification error. These \vec{x}_j are called Support Vectors only if the corresponding $\alpha_i > 0$.

Several typical kernel functions are:

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^d, \quad (3)$$

$$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\gamma \|\vec{x}_i - \vec{x}_j\|^d\right) \quad (4)$$

Equation (3) is the polynomial kernel function of degree d which will revert to the linear function when d = 1. Equation (4) is the Radial Basis Function (RBF) kernel with one parameter γ .

Other kernel functions are:

$$\text{Linear: } K(x_i, x_j) = x_i^T x_j$$

And

$$\text{Sigmoid: } K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r).$$

Here $\gamma, r, \text{ and } d$ are kernel parameters.

3.2 Extreme Learning Machines

Extreme learning machine was introduced not-long-ago as a learning algorithm for single-hidden layer feed-forward neural networks (SLFNs), which randomly chooses hidden nodes and analytically determines the output weights of SLFNs. In general, the learning rate of feed-forward neural networks (FFNN) is time-consuming than required and this has become bottleneck in their applications. According to [5], there are two main reasons behind this behavior; one is slow gradient based learning algorithms used to train neural network (NN) and the other is the iterative tuning of the parameters of the networks by these learning algorithms. To overcome these problems, [5, 14-15] proposed a learning algorithm called extreme learning machine (ELM) for single hidden layer feed-forward neural networks (SLFNs). It is stated that "In theory, this algorithm tends to provide the best generalization performance at extremely fast learning speed since it is a simple tuning-free algorithm" [15]. Therefore, it is an

interesting option to be considered in predictive modeling applications, particularly in reservoir characterization applications, where better generalization ability is often sought for.

The introduction of ELM has been considered as a good and welcome development because, in the past, it seems that there exists an unbreakable virtual speed barrier which classic learning algorithms cannot break through and therefore feed-forward network implementation then take a very long time to train itself, irrespective of the application type whether simple or complex. Also ELM tends to reach the minimum training error as well as it considers magnitude of weights which is opposite to the classic gradient-based learning algorithms which only intend to reach minimum training error but do not consider the magnitude of weights[5]. Also unlike the classical gradient-based learning algorithms which only work for differentiable activation functions, ELM learning algorithm can be used to train SLFNs with non-differentiable activation functions [14]. According to [5], "Unlike the traditional classic gradient-based learning algorithms, like back-propagation method, facing several issues like local minimum, improper learning rate and over-fitting, etc, the ELM is a simple tuning-free three-step algorithm that tends to reach the solutions straightforward without such trivial issues".

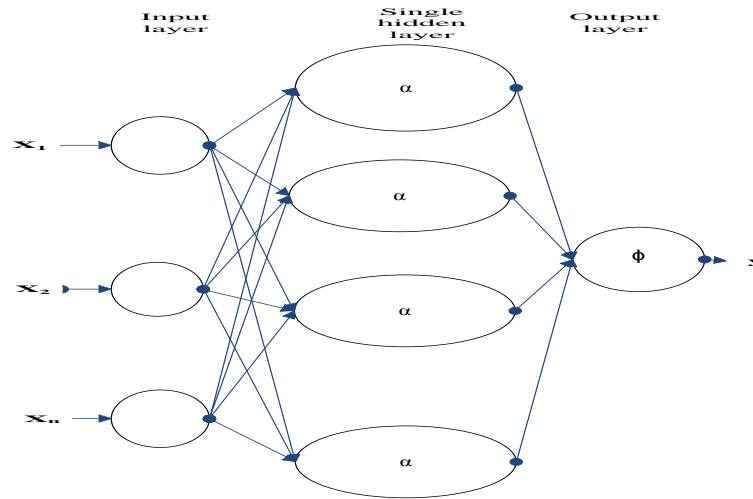


Figure 1: Schematic Representation of ELM ($x_1 \dots x_n$ are the input values, α is the activation function, and y are the output values).

3.2.1 The learning process for the proposed model of Erythematous-Squamous skin diseases based on ELM Framework

Let us first define the standard SLFN (single-hidden layer feed-forward neural networks). If we have N samples (x_i, t_i) , where $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$, then the standard SLFN with \tilde{N} hidden neurons and activation function $g(x)$ is defined as:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = o_j, j = 1, \dots, N \quad (5)$$

Where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]^T$ is the weight vector that connects the i^{th} hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$ is the weight vector that connects the i^{th} neuron and the output neurons, and b_i is the threshold of the i^{th} hidden neuron. The " \cdot " in $w_i \cdot x_j$ means the inner product of w_i and x_j .

SLFN aims to minimize the difference between o_j and t_j . This can be expressed mathematically as:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i \cdot x_j + b_i) = t_j, j = 1, \dots, N \quad (6)$$

Or, more compactly, as:

$$H \beta = T \quad (7)$$

where

$$H(w_1, \dots, w_{\tilde{N}}, b_1, \dots, b_{\tilde{N}}, x_1, \dots, x_N) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \dots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad T = \begin{bmatrix} T_1^T \\ \vdots \\ T_{\tilde{N}}^T \end{bmatrix}_{N \times m}$$

As proposed by Huang and Babri [16], H is called the neural network output matrix.

With the above SLFN specification background, thus the training procedures for the proposed ELM based model for erythemato-squamous disease identification can be summarized in the following algorithmic steps. See [5, 15] for further details on the workings of ELM algorithm.

Input - The inputs to the system, include clinical attribute obtained from patients, which represent the inputs parameters (input $x_i \in \mathbf{R}^n$ and target $t_i \in \mathbf{R}^m$), activation function, and the number of hidden neuron, \tilde{N} .

Output - The outputs of the ELM system is the target classes identified as the skin diseases type presented and then the weights of the layer.

Mathematically, given a training set

$$N = \{(x_i, t_i) | x_i \in \mathbf{R}^n, t_i \in \mathbf{R}^m, i = 1, \dots, N\},$$

activation function $g(x)$, and the number of hidden neuron = \tilde{N} , then, do the following:

Step 0: **Initialization**. Assign random values to the input weight w_j and the bias b_j , $j = 1, \dots, \tilde{N}$

Step 1: Find the hidden layer output matrix H.

Step 2: Find the output weight β as follows:

$$\beta = H^+ T$$

Where β , H and T are defined in the same way they were defined in the SLFN specification above (equations 5, 6, and 7).

4 Empirical Studies, Results and Discussions

4.1 Dataset Description

In this work, the University of California Irvine (UCI) machine learning database containing erythemato-squamous diseases dataset was used and analyzed. There are 366 samples in this database and each sample has 34 attributes. Each sample contains 12 clinical features and 22 histopathological features. Patients were first evaluated clinically with 12 features. Afterwards, skin samples were taken for the evaluation of 22 histopathological features. The values of the histopathological features are determined by an analysis of the samples under a microscope. These attributes are listed in Table 1. The family history feature has the value '1' if any of these diseases has been observed in the family and '0' otherwise. The age feature represents the age of the patient. It has been missed in some samples, so it is removed in the experiments. Every other feature (clinical and histopathological) was given a degree in the range of '0' to '3' so that '0' indicates that the feature was not present, '3' indicates the largest amount possible, and '1', '2' indicate the relative intermediate values.

The dataset contains 34 attributes, 33 of which are linear valued and one of them is nominal. The differential diagnosis of erythemato-squamous diseases is a real problem in dermatology. They all share the clinical features of erythema and scaling, with very little differences. The six diseases in this group are psoriasis, seboric dermatitis, lichen planus, pityriasis rosea, chronic dermatitis, and pityriasis rubra pilaris. The class distribution of the diseases is shown in Table 2.

Table 1: Features in the UCI erythemato-squamous diseases dataset

Clinical Features (take values 0, 1, 2, 3, unless otherwise indicated)	Histopathological Features (take values 0, 1, 2, 3)
Feature 1: Erythema	Feature 12: Melanin incontinence
Feature 2: Scaling	Feature 13: Eosinophils in the infiltrate
Feature 3: Definite borders	Feature 14: PNL infiltrate
Feature 4: Itching	Feature 15: Fibrosis of the papillary dermis
Feature 5: Koebner phenomenon	Feature 16: Exocytosis
Feature 6: Polygonal papules	Feature 17: Acanthosis
Feature 7: Follicular papules	Feature 18: Hyperkeratosis
Feature 8: Oral mucosal involvement	Feature 19: Parakeratosis
Feature 9: Knee and elbow involvement	Feature 20: Clubbing of the rete ridges
Feature 10: Scalp involvement	Feature 21: Elongation of the rete ridges
Feature 11: Family history (0 or 1)	Feature 22: Thinning of the suprapapillary epidermis
Feature 34: Age (linear)	Feature 23: Spongiform pustule
	Feature 24: Munro microabscess
	Feature 25: Focal hypergranulosis
	Feature 26: Disappearance of the granular layer
	Feature 27: Vacuolisation and damage of basal layer
	Feature 28: Spongiosis
	Feature 29: Saw-tooth appearance of retes
	Feature 30: Follicular horn plug
	Feature 31: Perifollicularparakeratosis
	Feature 32: Inflammatory mononuclearinfiltrate
	Feature 33: Band-like infiltrate

Table 2: Class distribution of erythematous-squamous diseases.

Class code	Class	Number of instances
1	Psoriasis	112
2	Seboric dermatitis	61
3	Lichen planus	72
4	Pityriasis rosea	49
5	Chronic dermatitis	52
6	Pityriasis rubra pilaris	20

4.2 Experimental setup and implementation process

As for the implementation of the two methods, we made use of MATLAB environment for the coding, while also calling some functions made available online, relating to ELM and SVM.

To evaluate performance of the two classifiers based on SVM and ELM modeling schemes, the acquired dataset described earlier is divided, using the stratified sampling approach, into 80% training set and 20% testing set for estimating how the investigated model performed on new unseen data. For testing and evaluation of both SVM and ELM framework and to carry out effective comparisons, the classification accuracy was calculated using Percent Correct measure, which is a measure of the percentage of correctly classified target classes.

As for the SVM, parameter settings include kernel function set to be “Gaussian”, C set to be 450 and labda set to be 1e-7, based on parameters search outcomes. As for the ELM, hardlim activation function was chosen based on parameter search while the hidden neuron was set to 500.

To further investigate the effect of varying the size of training and testing sets on the performance of the two classifiers, we partition the available dataset into training and testing sets based on four different ratios, which include 80:20, 70:30, 60:40 and 50:50 percentages of training and testing respectively. Each of these ratio partitioning was used to carry out experiments involving training and testing the two compared techniques.

4.3 Results and Discussions

Experimental results are presented in tables 3 through 6 representing the performance measure for both SVM and ELM classifiers. The results for the four different cases investigated, based on four different percentages of training and testing sets, are presented. These include training and testing sets in ratios 80:20, 70:30, 60:40 and 50:50 percentages of training and testing respectively.

Table 3: Experimental Results with Training-Testing partition ratio 80:20

	Training results	Testing Results
SVM	100	97.26
ELM	100	98.36

Table 4: Experimental Results with Training-Testing partition ratio 70:30

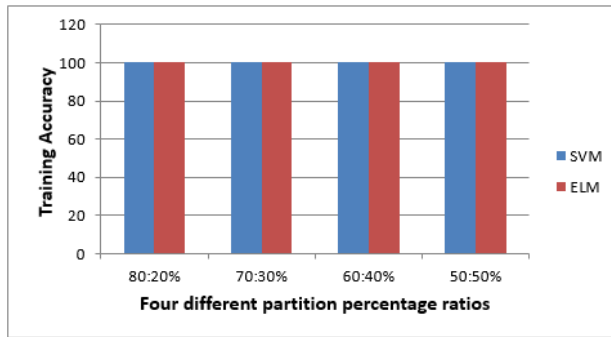
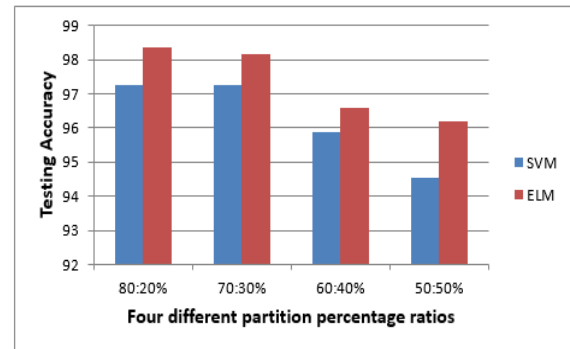
	Training results	Testing Results
SVM	100	97.25
ELM	100	98.17

Table 5: Experimental Results with Training-Testing partition ratio 60:40

	Training results	Testing Results
SVM	100	95.89
ELM	100	96.58

Table 6: Experimental Results with Training-Testing partition ratio 50:50

	Training results	Testing Results
SVM	100	94.54
ELM	100	96.18

**Figure 2: Summary of Training Results for all the Training-Testing partition ratios****Figure 3: Summary of Testing Results for all the Training-Testing partition ratios**

Figures 2 and 3 summarize the training and testing results, respectively.

It can be easily noticed from the tables and figures that ELM outperformed SVM in all the cases considered, for the testing set. For the training set, both ELM and SVM achieved the same high performance. As for the testing set, ELM performed better than SVM for all the four cases by a small margin.

Another important trend that is worth noticing and discussing is the pattern of difference in the performances displayed by each of the two methods for different data partition ratios used. The performance of both ELM and SVM for the testing set slightly decreases as the training set size decreases while testing set size increases. This is only applicable to the testing set. As for the training set, the performance of both ELM and SVM remain constant irrespective of the partition ratio used.

5 Conclusion and Recommendations

The popular SVM and ELM techniques have been proposed and compared as two new identification models to better identify erythemato-squamous skin diseases. The two models based on ELM and SVM have been compared looking into the effect of varying the ratio of training set to testing set. Experimental results indicated that ELM outperformed SVM in all fronts for the testing set case. Both the proposed classifiers have proven to be viable tools in this germane field as demonstrated by high accuracy and consistency of results even in the face of varying training-testing dataset percentages. Thus, SVM and ELM have been shown to demonstrate their unique ability to achieve excellent results in the field of biomedical diagnosis.

Another unique discovery made in this work is the ability of both SVM and ELM to maintain consistent results in the face of varying percentages of training and testing set partition ratio for the training set.

Although there was minimal decrease in the testing set performance as the training set decreases while the testing set increases in size, the reduction in performance is very minimal and acceptable as excellent results were still achieved. In addition, these two models based on each of the SVM and ELM models can be applied to other relevant biomedical diagnosis problems in the future.

REFERENCES

- [1] Xie, J., et al., *Novel Hybrid Feature Selection Algorithms for Diagnosing Erythematous-Squamous Diseases*, in *Health Information Science*, J. He, et al., Editors. 2012, Springer Berlin Heidelberg. p. 173-185.
- [2] Xie, J. and C. Wang, *Using support vector machines with a novel hybrid feature selection method for diagnosis of erythematous-squamous diseases*. Expert Systems with Applications, 2011. **38**(5): p. 5809-5815.
- [3] Vojislav Kecman and Mirna Kikec, *Erythematous-squamous diseases diagnosis by support vector machines and RBF NN*, *Proceedings of the 10th International Conference on Artificial intelligence and Soft Computing: Part I*, pp. 613-620, 2010, in Springer -Verlag Berlin Heidelberg 2010, e.a. L. Tutkowski, Editor. 2010. p. 613-620.
- [4] Übeyli, E.D. and İ. Güler, *Automatic detection of erythematous-squamous diseases using adaptive neuro-fuzzy inference systems*. Computers in Biology and Medicine, 2005. **35**(5): p. 421-433.
- [5] Huang, G.B., Q.Y. Zhu, and C.K. Siew, *Extreme learning machine: a new learning scheme of feedforward neural networks*, in *International Joint Conference on Neural Networks (IJCNN2004)*. 2004: Budapest, Hungary. p. 985 - 990.
- [6] Liu, H., J. Li, and L. Wong, *Use of Extreme Patient Samples for Outcome Prediction from Gene Expression Data*. Bioinformatics, 2005. **21**(16): p. 3377-3384.
- [7] Mahmoud, S.A. and S.O. Olatunji, *Automatic Recognition of Off-line Handwritten Arabic (Indian) Numerals Using Support Vector and Extreme Learning Machines*. International Journal of Imaging, 2009. **2**(A09): p. 34-53.
- [8] Olatunji, S.O., *Comparison of Extreme Learning Machines and Support Vector Machines on Premium and Regular Gasoline Classification for Arson and Oil Spill Investigation*. Asian Journal Of Engineering, Sciences & Technology, 2011. **1**(1): p. 1-7.
- [9] Olatunji, S.O., I.A. Adeleke, and A. Akingbesote, *Data Mining Based on Extreme Learning Machines for the Classification of Premium and Regular Gasoline in Arson and Fuel Spill Investigation*. Journal Of Computing, 2011. **3**(3): p. 130-136.

- [10] Olatunji, S.O., S. Ali, and A. Abdul Azeez. *Modeling Permeability Prediction Using Extreme Learning Machines*. in *Fourth Asia International Conference on Mathematical/Analytical Modelling and Computer Simulation, AMS2010*. 2010. Kota- Kinabalu, Malaysia: IEEE.
- [11] Olatunji, S.O., et al., *Extreme Learning Machine as Maintainability Prediction model for Object-Oriented Software Systems*. *Journal of Computing*, Volume 2, Issue 8, August 2010, 2010. **2**(8): p. 42-56.
- [12] Teddy Mantoro, et al., *Extreme learning machine for user location prediction in mobile environment*. *International Journal of Pervasive Computing and Communications*, 2011. **7**(2): p. 162 - 180.
- [13] Duda, R.O., P.E. Hart, and D.G. Stock, *Pattern Classification*. 2001, New York: John Wiley and Sons. 654.
- [14] Huang, G.B., et al., *Can threshold networks be trained directly?* *IEEE Trans. Circuits Syst. II*, 2006. **53**(3): p. 187-191.
- [15] Huang, G.B., Q.Y. Zhu, and C.K. Siew, *Extreme learning machine: Theory and applications*. *Neurocomputing*, Elsevier, 2006. **70**(1-3): p. 489-501.
- [16] Huang, G.B. and H.A. Babri, *Feedforward neural networks with arbitrary bounded nonlinear activation functions*. *9(1):224–229. IEEE Trans Neural Network*, 1998. **9**(1): p. 224-229.