Research Journal of Applied Sciences, Engineering and Technology 7(18): 3883-3887, 2014
DOI:10.19026/rjaset.7.746
ISSN: 2040-7459; e-ISSN: 2040-7467
© 2014 Maxwell Scientific Publication Corp.
Submitted: November 22, 2014 Accepted: December 17, 2013 Published: May 10, 2014

Research Article

A Survey on Utilization of the Machine Learning Algorithms for the Prediction of Erythemato Squamous Diseases

N. Badrinath and G. Gopinath
Department of Computer Engineering and Applications, Bharathidasan University,
Tiruchirapalli, Tamil Nadu, India

Abstract: The aim of this study list the contributions of various machine learning algorithms for the prediction of Erythemato Squamous Diseases (ESDs) and it is very useful for the budding researchers to do research in this field. In the advent of ozone depletion the ultra violet radiation is the major cause of many skin diseases, which are leading to skin cancer. Early detection of skin cancer is more important to avoid human loses and especially the white skinned people are more affected. The Asian and African race people are less affected as they have melanin in their skin. The American's are directly and more widely affected by the ozone depletion, due to this ESD, which is predominant among the skin diseases. Due to technology advancements a large amount of data are deposited. In these data the information is hidden as raw data and with latest methodologies and technologies like Data Mining, neural networks, fuzzy systems, Genetic and Evolutionary computing a pattern can be evolved to study them. Guvenir *et al*. (1998) studied about ESDs and contributed 366 patients data with 34 features consisting of clinical and histopathological data in the dermatology dataset (The data taken from School of Medicine in Gazi University and the department of Computer Science in Bilkent University, Turkey; and it is available in the URL (http://archive.ics.uci.edu/ml/datasets/Dermatology) in the year 1998. This survey study gives a brief description about the contribution of what in the field of ESDs in Chronological order from the year 1998 till 2013. In this study we intend to contribute various machine learning algorithms dealing with ESDs.

Keywords: AdaBoost, Adaptive Neuro-Fuzzy Interference Systems (ANFIS), ESD, Extreme Learning Machine (ELM), fuzzy based ELM, fuzzy logic, preprocessing techniques, Support Vector Machine (SVM)

INTRODUCTION

The detection of ESDs is very difficult and it is a Herculean task as these diseases share common features like clinical and histopathological features with very minor differences. Due to ozone depletion, the sun's ultra violet radiation cause many skin diseases and the predominant skin disease is ESD. Many factors influence skin diseases like increasing bacteria involvement, climatic conditions like dampness or humidity, dryness, exposure of more sunlight's ultra violet radiation, fungal involvement, food habits, allergic to gases and chemicals, external infections, dead skin, dust, unwanted secretions, oral involvement and so on.

The dataset of ESDs consists of 366 patients data with 34 features consisting of clinical and histopathological data in the http://archive.ics.uci.edu/ml/datasets/Dermatology. The clinical features can be easily identified visually but for histopathological features a biopsy of the patient's sample is needed and the dataset consists of values from 0 to 3 except for age feature. The value '0' represents there is no occurrence of the feature and '3' represents the high occurrence. The intermediate values represent the severity of the features. The clinical features are of 12 and that of histopathological are 22. The value of family history is either 0 or 1 and the feature is not considered much. For the missing parameter either median values is considered or omitted.

The objective of this study is to collect all the work related to the diagnosis of ESDs which can be particularly used in the machine learning algorithms. The collection is starting from 1997 to till date and definitely this survey will help the researchers who start their work in this field.

PREDICTION OF ESDs USING MACHINE LEARNING ALGORITHMS

Some of the important machine learning algorithms and its analysis is given below.

Guvenir *et al*. (1998) developed a new classification algorithm, namely VF15 and implemented to differential diagnosis of ESDs. It has short training and classification times and the algorithm

Corresponding Author: N. Badrinath, Department of Computer Engineering and Applications, Bharathidasan University, Tiruchirapalli, Tamil Nadu, India

Table 1: The comparison of the percentage of the incidences for some common skin diseases in Riyadh and other regions of Saudi Arabia

Skin disorder	Riy	Ab[a]	Jou[b]	Ha[b]	Jed[b]	Naj[b]	E.PR[b]	Mak[a]	Asr[b]
Dermatitis	21.29	25.68	34.14	16.30	18.64	37.00	19.60	23.80	25.70
Acne	11.88	5.45	9.57	12.40	9.48	12.80	13.80	4.80	5.40
Viral warts	8.64	2.49	2.85	8.40	6.78	6.00	11.90	2.50	2.50
Bacterial infections	2.97	13.19	10.87	2.80	7.65	5.00	4.80	11.20	13.20
Vitiligo	2.69	3.03	3.35	3.90	3.12	7.00	5.00	0.70	3.00
Psoriasis	2.47	2.10	5.33	3.60	3.01	1.50	3.40	1.80	2.10
Lichen planus	1.14	1.32	1.21	1.20	0.64	1.10	1.70	0.70	1.30

a: Francesco (2011), b: Guvenir and Emeksiz (2000), Riy: Riyadh; Ab: Abha; Jou: Al-Jouf; Ha: Hail; Jed: Jeddah; Naj: Najran; E.PR: Eastern Province; Mak: Makkah; Asr: Asir

proved the robustness in noisy training instances and missing feature values. The missing feature values are ignored during training and test instances. Guvenir and Emeksiz (2000) proposed a Graphical User Interface (GUI) tool for diagnosing ESDs which is based on three classification algorithms, namely Nearest Neighbor Classifier (NNC), Naive Bayesian Classifier using Normal Distribution (NBCND) and VF15. Again, the team of Guvenir and Emeksiz (2000) has also developed an expert system based GUI for diagnosing ESDs.

Zhu *et al*. (2002) aimed at studying and identifying the pattern distribution and intrinsic correlations in large data sets by partitioning the data points into similarity clusters and developed a new algorithm called CoFD and it is based on a distance based clustering algorithm for high dimensional spaces.

Castellano *et al*. (2003) proposed a multistep learning strategy called KERNEL (Knowledge Extraction and Refinement by Neural Learning). Data sets are split into 10 subsets and out of which nine are used for training and one for testing purposes. The age feature is removed as it contains missing values and also the results vary with the inclusion of age.

Ubeyli and Inan (2005) proposed a new approach for estimating the ESDs based on ANFIS. For improving the higher accuracy, they have used seven classifiers instead of six classifiers. The seventh classifier is taking all the output of the six classifiers respectively for six ESDs as inputs and it classified the exact disease. The proposed approach is based on fuzzy input values with neural network capabilities. The proposed system achieved more accuracy rates than that of simple neural network model. The results of six classifiers are combined and given to the seventh classifier. The classification accuracy of this model is raised to 95.5%.

Fabien and Alfred (2005) proposed an automated “flood-fill segmentation” method and it is a new clustering method of the U*-matrix of a Self Organizing Map (SOM) after training. They found that U*F method is performing better for the wide set of critical dataset types and also proved that U*F the computation time of the SOM phase is negligible and it does not require apriori knowledge on the number of clusters, making it a real “cluster-mining” algorithm. The “cluster-mining” algorithm was not able to distinguish two of the actual categories. Again, they proved that U*F clustering method is an alternative method to other clustering algorithms like single-linkage, K-means, ward and so on.

Abdel-Aal *et al*. (2006) proposed that divide and conquer principle can be used effectively for differential diagnosis of dermatology through decomposing into simpler sub-problems and each is solved separately. In the same year, Loris (2006) proposed an ensemble of SVM based on Random Subspace (RS) and feature selection is developed and applied to ESDs. The results showed that the average predictive accuracy obtained by a “stand-alone” SVM or by a RS ensemble of SVM is less when compared to the proposed method. The classification accuracy is raised from 97.22 to 98.3%.

Al-Zoman *et al*. (2008) had done a retrospective study on major skin diseases in the period 2001-2005 in the central region of Saudi Arabia. The patients’ details were collected from Riyadh Military Hospital and out of 58450 cases women (58.38%) were most affected than man (41.62%) (Agarwal, 1997; Raddadi *et al*., 1999). Most of the diagnoses were done by clinical method and the large volume of patients affected in the age group of 41-50 years. ESDs were only 4.61% and out of six diseases Psoriasis is 2.47% predominant. The most common in the Eczema dermatitis group is seborrheic dermatitis (29.76%), patient’s origin is from Arab (98.01%) and non Arabs is (1.99%). Psoriasis was 53.4%, Lichen Planus was (24.7%). The comparison of the percentage of the incidences for some common skin diseases in Riyadh and other regions of Saudi Arabia and it is given in Table 1. White skin persons are widely affected by all the forms of Skin Tumors which is given in Table 2. The skin diseases do not affect people based upon the age as in Table 3.

Abu Naser and Akkila (2008) proposed an expert system using Artificial Intelligence (AI), to help dermatologists in diagnosing some of the skin diseases with the help of an interface engine and a knowledge base. This proposed expert system is not for a specialized disease but it can be used to diagnose nine skin diseases. Übeyli and Erdoğan (2010) proposed a new approach for detection of ESDs based upon K-means clustering of data mining methods. The proposed methodology is used only for classification and in this study only 33 features are considered and only five diseases are considered for which Pityriasis rubra pilaris is omitted. The classification result shows 94.22% total classification accuracy out of 346 patients.

Kenneth *et al*. (2009) proposed the Rough Sets of data-mining technique and the classification accuracy is

Table 2: Distribution of patients according to race (Fabien and Alfred, 2005)

Race	Total	Male	Female	(%)
Africans	78	31	47	0.13
Americans	25	9	16	0.04
Arabs	57286	24058	33228	98.01
Asians	801	159	642	1.37
Australians	24	4	20	0.04
Europeans	208	53	155	0.36
Others	28	13	15	0.05
Total	58450	24327	34123	100

Table 3: Distribution of age based on sex (Fabien and Alfred, 2005)

Age group	Total	(%)	Male	Female
1-10	5763	9.86	3122	2641
11-20	9079	15.53	3723	5356
21-30	10706	18.32	3933	6773
31-40	9359	16.01	3711	5648
41-50	19156	32.77	7924	11232
51-60	2429	4.16	927	1502
61-70	1280	2.20	596	684
71-80	486	0.83	275	211
81-90	154	0.26	90	64
91-100	30	0.05	22	8
101-110	8	0.01	4	4
Total	58450	100	24327	34123

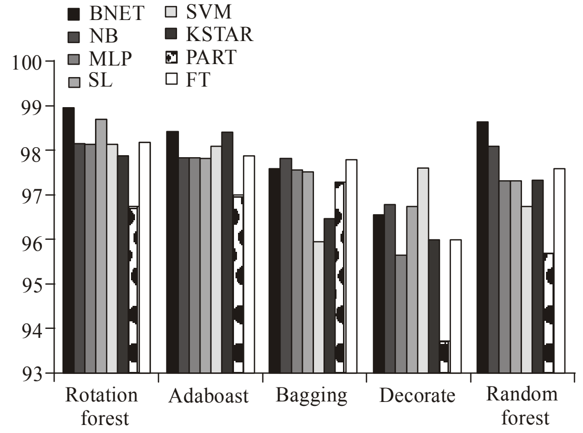

Fig. 1: Comparison of ensemble feature selection algorithms depending on accuracies of classifiers (Nani, 2006)

high with values more than 98% in some instances. The classification results shows that the clinical feature shows only 50-60% accuracy when age feature is added and histopathological features produced only 85% approximately. Übeyli (2009) proposed the use of Combined Neural Networks (CNNs) model for diagnosis of Erythemato Squamous and also Multilayer Perceptron Neural Networks (MLPNNs) is also tested. The network is trained using Levenberg-Marquardt algorithm. The CNN (97.77%) showed more accuracy than MLPNN (85.47%).

Juanying and Chunxia (2011) and Juanying *et al.* (2010) developed a model based upon SVM with IFSFFS (Improved F-score and Sequential Forward Floating Search) which selects optimal feature subset by considering the advantages of wrappers and filters. Finally, the classification accuracy is evaluated with the help of 14 features, which includes both histopathological and clinical features.

Akin and Arif (2012) proposed a new multi-class feature selection method based on rotation forest meta-learner algorithm. This proposed system eliminates the redundant attributes and the accuracy of this model is varying between 98 and 99%. They concluded that the classification accuracy between different feature selections is given as Fig. 1.

Davar *et al.* (2011) introduced a Catfish binary particle swarm optimization (Catfish BPSO), Kernelized SVM (KSVM) and Association Rules (AR) for feature selection method. The proposed model achieves 99.09% classification accuracy using 24 features as their inputs. Francesco (2011) suggested that the optimized k-Nearest Neighbor Classifier (k-NNC) and PEL-C. Finally, he found that the proposed methodologies provide good accuracy. For decision support systems the instance based optimized k-nearest classifier and Prototype Exemplar Learning Classifier is the best for knowledge extraction.

Aruna *et al.* (2012) proposed a hybrid feature selection method namely IGSBFS (Information Gain and Sequential Backward Floating Search), which combines the advantages of filters and wrappers to select the optimal feature subset from the original feature set based on a diagnostic model of Naïve Bayes. The classification accuracy of the proposed method is 98.9% with only 10 features.

Akın and Arif (2013) suggested a Genetic Algorithm (GA) based FS algorithm combined in parallel with a BN classifier based on GA wrapped Bayesian Network (BN) Feature Selection (FS). The accuracy of BN algorithm is increased due to GA based heuristic search of 10-fold cross-validation and the classification accuracy is 99.20%. The accuracy of other classifiers are SVM-98.36%, Multi-Layer Perceptron (MLP) -97.00%, Simple Logistics (SL) -98.36% and Functional decision Tree (FT) -97.81%.

Abdi and Giveki (2013) developed a model which is based on Particle Swarm Optimization (PSO), SVMs and ARs. The proposed system achieves 98.91% classification accuracy using 24 features as their inputs.

Juanying *et al.* (2013) proposed two-stages for the detection of ESDs. The stage one is called data preprocessing and they used following feature selection algorithms: Extended Sequential Forward Search (SFS), Sequential Forward Floating Search (SFFS) and Sequential Backward Floating Search (SBFS). The stage two is called classifier and here they used SVM.

Ravichandran *et al.* (2013) proposed a new approach based upon Fuzzy Extreme Learning Machine (FELM). The proposed system showed more efficiency both in time and accuracy. The classification accuracy of FELM based approach is reached to 94% before data preprocessing and the same is increased to 99.02% after data preprocessing. But, the total computational time is less than 1 sec, where as the average computational time for other machine learning algorithms is 124 sec. Again, Badrinath *et al.* (2013) developed AdaBoost and Hybrid classifier methodologies along with Apriori and ARs data preprocessing. In this case, the classification

Table 4: Performance analysis of various machines learning algorithm

Author	Method	Percentage of the accuracy
Guvenir *et al*. (1998)	VF15	96.20
Guvenir and Emeksiz (2000)	NNC, NBC and VF15	99.20
Lopes *et al*. (2001)	Genetic programming	96.64
	C4.5	89.12
Ubeyli and Inan (2005)	ANFIS	95.50
Nani (2006)	LSVM	97.22
	RS	97.22
	B1-5	97.50
	B1-10	98.10
	B1-15	97.22
	B2-5	97.50
	B2-10	97.80
	B2-15	98.30
Polat and Gunes (2009)	C4.5 and one-against-all	96.71
Ubeyli (2009)	CNN	97.77
Übeyli and Erdoğan (2010)	k-means clustering	94.22
Lekkas and Mikhailov (2010)	Evolving fuzzy classification	97.55
Xie and Wang (2011)	IFSFS and SVM	98.61
Davar *et al*. (2011)	KSVM and AR	99.06
Aruna *et al*. (2012)	IGSBFS hybrid method	98.90
Akın and Arif (2013)	Genetic Algorithm (GA) based FS algorithm	99.20
Badrinath *et al*. (2013)	AdaBoost and its hybrid algorithms	99.26

accuracy is increased to 99.26% and the computational is very high than FELM.

RESULT ANALYSIS

Some of the important machine learning algorithm listed above and its performance measures are given in Table 4.

Table 4 shows that the performance accuracy is varying from 89% to little bit above of 99%. Clearly, the performance of accuracy is reached to 93% before data preprocessing and the performance of accuracy is reached to above of 99% if data preprocessing is used.

CONCLUSION

In this study, machines learning algorithms used in determining the ESDs have been discussed. The data taken from medical and skin related information involves numerous and complex datasets. The data preprocessing helped us to reduce dimensionality of the given datasets considerably and hence it indirectly helped us to reduce the time complexity. In this study, we have consolidated some of the important works which are related to machine learning algorithms based on the prediction of ESDs. Since the machine learning algorithms have exponentially used in the last one and half decades in many of the medical and skin related problems and hence in this study, we have summarized some of the important works related to machine learning algorithms based on the prediction of ESDs. The performance analysis in terms of diagnosis of the accuracy of the ESDs is presented in Table 4.

REFERENCES

Abdel-Aal, R.E., M.R. Abdel-Halim and S. Abdel-Aal, 2006. Improving the classification of multiple disorders with problem decomposition. J. Biomed. Inform., 39(6): 612-625.

Abdi, M.J. and D. Giveki, 2013. Automatic detection of erythemato-squamous diseases using PSO-SVM based on association rules. Eng. Appl. Artif. Intel., 26: 603-608.

Abu Naser, S.S. and A.N. Akkila, 2008. A proposed expert system for skin diseases diagnosis. J. Appl. Sci. Res., 4(12):1682-1693.

Agarwal, P.K., 1997. Pattern of skin diseases in the Al Jouf region. Ann. Saudi Med., 17: 112-114.

Akin, O. and G. Arif, 2012. A robust multi-class feature selection strategy based on rotation forest ensemble algorithm for diagnosis of erythemato-squamous diseases. J. Med. Syst., 36: 941-949.

Akin, Ö. and G. Arif, 2013. Genetic algorithm wrapped Bayesian network feature selection applied to differential diagnosis of erythemato-squamous diseases. Digit. Signal Process., 23: 230-237.

Al-Zoman, A.Y., M.D. Facharizt and A.K. Al-Asmair, 2008. Pattern of skin diseases at Riyadh military hospital. Egypt. Dematol. Online J., 4(2-4): 01-10.

Aruna, S., L.V. Nandakishore and S.P. Rajagopalan, 2012. A hybrid feature selection method based on IGSBFS and Naïve Bayes for the diagnosis of erythemato-squamous disease. Int. J. Comput. Appl., 41(7): 114-123.

Badrinath, N., G. Ganapathy and K.S. Ravichandran, 2013. Estimation of automatic detection of erythemato-squamous diseases through Adaboost and its hybrid classifiers. J. Artif. Intell. Rev., (under review).

Castellano, G., C. Castiello, A.M. Fanelli and C. Leone, 2003. Diagnosis of dermatological diseases by a neuro-fuzzy system. Proceedings of the 3rd Conference of the European Society for Fuzzy Logic and Technology. Zittau, Germany, September 10-12, 2003.

Davar, G., S. Hamid, A.B. Amir and K. Younes, 2011. Detection of erythemato-squamous diseases using AR-Catfish BPSO-KSVM. Signal Image Process. Int. J., 2(4): 57-72.

Fabien, M. and U. Alfred, 2005. U*F Clustering: A New Performant Cluster-Mining method based on Segmentation of Self-Organizing Maps. Proceeding of the 5th Workshop on Self-Organizing Maps (WSOM'2005), Paris: France, pp: 25-31.

Francesco, G., 2011. Instance-based classifiers applied to medical databases: Diagnosis and knowledge extraction. Artifi. Intell. Med., 52: 123-139.

Guvenir, H.A. and N. Emeksiz, 2000. An expert system for the differential diagnosis of erythemato-squamous diseases. Expert Syst. Appl., 18: 43-49.

Guvenir, H.A., G. Demiroz and N. Ilter, 1998. Learning differential diagnosis of erythemato-squamous diseases using voting feature intervals. Artif. Intell. Med., 13: 147-165.

Juanying, X. and W. Chunxia, 2011. Using support vector machines with a novel hybrid feature selection method for diagnosis of Erythemato-squamous diseases. Expert Syst. Appl., 38: 5809-5815.

Juanying, X., X. Weixin, W. Chunxia and G. Xinbo, 2010. A novel hybrid feature selection method based on IFSFFS and SVM for the diagnosis of erythemato-squamous diseases. Proceeding of JMLR: Workshop and Conference, pp: 142-151.

Juanying, X., L. Jinhu, X. Weixin, S. Yong and L. Xiaohui, 2013. Two-stage hybrid feature selection algorithms for diagnosing erythemato-squamous diseases. Health Inform. Sci. Syst., pp: 1-10.

Kenneth, R., G. Florin, S. Abdel-Badeeh and E.S. El-Dahshan, 2009. Evaluation of the feature space of an erythemato squamous dataset using rough sets. Math. Comp. Sci. Ser., 36(2): 123-130.

Lekkas, S. and L. Mikhailov, 2010. Evolving fuzzy medical diagnosis of Pima Indians diabetes and of dermatologica diseases. Artif. Intell. Med., 50: 117-126.

Lopes, C.C., H.S. Freitas and A.A. Bojarczuk, 2001. Data mining with constrained-syntax genetic programming: Applications in medical data set. Proceeding of Data Analysis in Medicine and Pharmacology (IDAMAP-2001), a Workshop at Medinfo-2001, London, UK, 2001.

Loris, N., 2006. An ensemble of classifiers for the diagnosis of erythemato-squamous diseases. Neuro Comput., 69: 842-845.

Nani, L., 2006. An ensemble of classifiers for the diagnosis of erythemato-squamous diseases. Neurocomputing, 69: 842-845.

Polat, K. and S. Gunes, 2009. A novel hybrid intelligent method based on C4. 5 decision tree classifiers and one-against-all approach for multi-class classification problems. Expert Syst. Appl., 36(2): 1587-1592.

Raddadi, A.A., S.A. Abdullah and Z.B. Damanhouri, 1999. Pattern of skin diseases at King Khalid National Guard hospital: A 12-month prospective study. Ann. Saudi Med., 19(5): 453-454.

Ravichandran, K.S., B. Narayanamurthy, G. Ganapathy, S. Ravalli and J. Sindhura, 2013. An efficient approach to an automatic detection of erythemato-squamous diseases. Neural Comput. Appl., DOI: 10.1007/s00521-013-1452-5.

Übeyli, E.D., 2009. Combined neural networks for diagnosis of erythemato-squamous diseases. Expert Syst. Appl., 36: 5107-5112.

Ubeyli, E.D. and G. Inan, 2005. Automatic detection of erythemato-squamous diseases using adaptive neuro-fuzzy inference systems. Comput. Bio. Med., 35: 421-433.

Übeyli, E.D. and D. Erdoğan, 2010. Automatic detection of erythemato-squamous diseases using k-means clustering. J. Med. Syst., 34:179-184.

Xie, J. and C. Wang, 2011. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. Expert Syst. Appl., 38(5): 5809-5815.

Zhu, S., T. Li and M. Ogihara, 2002. An algorithm for non-distance based clustering in high dimensional spaces. Lect. Notes Comput. Sci., 2454: 52-62.