Use of Data Mining Techniques to Predict Short Term Adverse Events Occurrence in NB-UVB Phototherapy Treatments

S. Mohamed, B. Q. Huang, M-T. Kechadi
Insight Centre for Data Analytics
University College Dublin
Dublin, Ireland
sharifa.mohamed@ucdconnect.ie

Abstract—The prediction of short term adverse events occurrence in phototherapy treatment is important for the dermatologists who administrate phototherapy to adjust the treatment and standardize the clinical outcomes. Recently, a modeling technique which can detect the potential short term adverse events occurrence in phototherapy treatments is required for clinicians. Based on data mining, this study tends to explore the significant features and the class distribution of training data for the short term adverse events occurrence prediction in NB-UVB phototherapy treatments. The experimental results highlight that acceptable prediction accuracy can be achieved by using the significant features and the performance of the classifiers can be significantly improved by sampling 40% of negative class samples in training data, hyper parameter tuning of classifiers and use of stacked classifiers in creating prediction models.

***Index Terms*—Adverse Events, Classification, Datamining, Dermatology, Phototherapy, Prediction**

I. INTRODUCTION

Phototherapy involves repeated exposure of the skin to ultraviolet (UV) light to treat various inflammatory skin conditions such as psoriasis, eczema, and vitiligo. This therapy is one of the oldest treatment modalities in dermatology, dating back to the ancient Egyptians, who used natural light in combination with herbal extracts to treat skin diseases [37], [41]. Phototherapy continues to be a highly preferred treatment by dermatologists [27], [41].

There are three main types of phototherapy used for the treatment of psoriasis: broadband ultraviolet B (BB-UVB), narrowband ultraviolet B (NB-UVB), and psoralen plus ultra-violet A (PUVA) [41]. UVB is the most commonly preferred first-line treatment for moderate-to-severe psoriasis in healthy adults [27]. NB-UVB is the most commonly used phototherapy modality today as it has a wider application across various dermatologic conditions, it's easier to use, and has fewer adverse events when compared with BB-UVB or PUVA [8], [41]. In addition, these three main types of phototherapy cause some acute short term adverse events, including erythema and burning, lesional blisters, pruritus etc. [9], [11], [19].

Concerns regarding skin cancer development and adverse events occurrence in phototherapy have become a common source of legal claims in dermatology and have emphasized the importance of fastidious monitoring of its delivery [24]. It is identified as a yellow flag action recommended service standard by the British Association of Dermatologists to compare and standardize number of episodes/patient treatments/year for each grade of symptomatic erythema compared to published standards. Published rates vary but include: less 0.8% of all treatments result in an acute adverse event (0.6% for NB-UVB [0.5% excluding Grade I Erythema], 1.3% for systemic PUVA, and 0.8% for local PUVA); for severe adverse events: 0.05% for NB-UVB and 0.3% for systemic PUVA [43].

If clinicians know the beforehand prospective of treatment, they can adjust the treatment and standardize the clinical outcomes. Therefore, a model which can predict whether a treatment may cause acute adverse events is useful for dermatologists who administrate phototherapy. Recently, data mining techniques have been applied to health-care areas [42], [38], [36] and particularly in dermatology [40].

Artificial neural networks (ANN) have been successfully used for the diagnosis of skin diseases. Yoon, Brobst, Bergstresser and Peterson [2] used artificial neural networks with back propagation algorithm for the diagnosis of papulosquamous skin diseases. For the diagnosis of erythematous squamous diseases, Übeyli [18] used combined neural networks (CNN) and achieved 97.77% of accuracy. Chang and Chen [17] received highest accuracy of 80% by ANN model, outperforming models made of decision tree, combination of decision tree and artificial neural network, decision tree with sensitivity analysis and artificial neural network with sensitivity analysis, Karlik and Harman [30] used supervised back propagation with 95% accuracy. Olatunji and Arif (2013) used Artificial Neural Network and Extreme learning machine, Sarhan, Elharir and Zakaria [35] used artificial neural network Levenberg-Marquardt algorithm with Rough sets Johnson Reducer attribute reduction technique with 98.8% accuracy.

Support Vector Machines (SVM) has been used since 2006 to create models for diagnosis of skin diseases. Nanni [14] proposed an ensemble of Linear SVM based on Random Subspace and feature selection that improved the average predictive accuracy gained by a standalone SVM or by a RS ensemble of SVMs. Übeyli [15] used a hybrid technique which combined multiclass SVM with error

correcting output code which achieved an accuracy of 98.32%. Xie and Wang [23] achieved an accuracy of 98.61% for the model they implemented using SVM and IFSFFS (Improved F-score and Sequential Forward) feature selection method. Giveki, Salimi, Bitaraf and Khademian [22] proposed a model based on Catfish Binary Particle Swarm Optimization (CatfishBPSO), Kernelized Support Vector Machines (KSVM), and Association Rule feature selection method which gained a 99.09% accuracy. Abdi and Giveki[33] also proposed a hybrid method of particle swarm optimization, Support vector machine, and association rules which achieved 98.91% accuracy.

Mroczek, Paja, Piatek and Wrzesie [16] used ID3 decision tree as one of the models for diagnosis and classification of melanocytic skin lesions. Polat and Güneş [20] used C4.5 decision tree classifier with one-against-all approach with 84.48% accuracy and Tran (2008, April) used Gini index based decision tree for erythematosquamous diseases diagnosis.

Manjusha, Sankaranarayanan and Seena [34] used a Naïve Bayesian classifier to predict eight different dermatological conditions while Aruna, Nandakishore and Rajagopalan [29] used a hybrid feature selection method with Naïve Bayes classifier which achieved 98.9% accuracy and Danjuma and Osofisan [39] obtained highest accuracy of 97.4% from Naïve Bayes classifier outperforming Multilayer Perception and J48 decision tree for the Diagnosis of Erythemato - Squamous skin Diseases.

Cataloluk and Kesler [26] created a diagnostic software tool for skin diseases with basic and weighted K-NN and gained 96.36% accuracy when Manhattan Distance was used for weighted K-NN.

Ensembles made by combining different classification techniques also have been used by many researchers to create skin disease diagnostic models. Elsayad [21] has used an ensemble model created by combining multilayer neural network, decision tree and linear discriminant analysis (LDA) techniques and got 98.23% accuracy. Sharma and Hota [31] proposed a hybrid ensemble model by combining Support Vector Machine and Artificial Neural network and obtained 98.99% test accuracy.

Although there has been a high level of interest in implementing skin disease diagnostic models, there was very rare literature found which used data mining techniques for phototherapy data analysis, dermatological treatment outcome prediction or dermatological adverse events occurrence prediction.

This paper introduces a prediction model to detect the acute adverse events of a treatment using data mining techniques. Based on the NB-UVB phototherapy data-set, the proposed prediction model first selects a number of attributes, then pre-pares data, finally applies classification algorithms to predict adverse events occurrence.

II. METHODOLOGY

A. Data-set Collection

The data set used in this paper was obtained from The Adelaide and Meath Hospital, Dublin, known as Tallaght hospital's PuvaMate UVB phototherapy database. The data set was professionally anonymized by OpenApp Computer Support and Services after obtaining ethical approval from Tallaght hospital research ethics committee.

The UVB phototherapy database consisted of 29836 treatment records of 897 patients treated since end of September 2003. The information of each patient includes patient personal details (e.g. gender, year of birth, skin type) and treatment details. There were 464 females and 434 males among the patients. These patients were treated for psoriasis, eczema, granuloma annulara, acne, nodular prurigo, mycosis fungoides, ple, urticaria pigmentosa, morphea, lichen spinulosa, pityrasis lichenoideschronicu and vitiligo. Among all these records, psoriasis, eczema and nodular prurigo treatment records were studied separately in order to predict the adverse event occurrence of treatments as these were the top 3 diseases treated. Table I shows the percentage of acute adverse event occurrences for the above said diseases. If any of the short term acute adverse events including erythema, burning, lesional blisters or pruritus was noted following the treatment, it was recorded as positive occurrence of adverse events and otherwise marked negative. These were used as the prediction classes. As the negative class records were much higher in number than the positive cases, negative class was the majority class in these cases.

B. Methods

RStudio with R version 3.3.1 was used in a 64-bit Windows operating system to conduct experiments with the help of mlr machine learning package.

The prediction model to detect the acute adverse events of treatments consists of the preprocessing and classification processes. The preprocessing process first selects significant attributes from the psoriasis, eczema and nodular prurigo data sets, then filters the noise data and normalizes the data. The classification process applies modelling algorithms to predict whether each treatment causes adverse events. These two processes are described in the following.

1) Preprocessing: The goal of this phase is to provide cleaned data for the classification step. In the preprocessing phase, we first derived new attributes from the data-set and applied the information gain technique [5] to select the significant attributes or features. Then imputed the missing values based on domain knowledge or mode of the attribute in which the missing value is replaced by the value that makes the most sense or in other cases by the value that is most common. Next, we used Local Outlier Factor technique (LOF) [7] to deal with the local outliers. Finally normalized the data-set [13]. In this step, new binary attributes were created for categorical attributes and numerical attribute values were scaled to fall between 0 and 1.

The features selected by information gain technique for the classification process are summarized below. Dosage, dose difference between previous and current treatment, ratio of dosage to med, previous treatment dosage, course cumulative dose, dose difference percentage between previous and current treatment, date difference between previous and current treatment, if the treatment dosage was

increased, reduced or repeated compared to previous treatment, total cumulative dose, skin type and gender were among the attributes which gave non-zero information gain for psoriasis records.

TABLE I. ADVERSE EVENTS OCCURRENCE IN PSORIASIS, ECZEMA AND NODULAR PRURIGO DATASETS.

Dataset	Tot.samples	N.adverse samples	N.normal samples
Psoriasis	23819	3170 (13.3%)	20649 (87%)
Eczema	4238	377(8.9%)	3861 (91%)
Nod. prurigo	791	84(10.6%)	707 (89%)

All of the above-mentioned attributes except dose difference percentage between previous and current treatment, date difference between previous and current treatment attributes gave non-zero information gain for eczema records and except dose difference percentage between previous and current treatment, date difference between previous and current treatment and total cumulative dose attributes gave non-zero information gain for nodular prurigo records.

2) Classification: Patients treated for psoriasis, eczema and nodular prurigo were analyzed separately and 3 experiments were conducted on each of data-set.

a) Experiment 1: Classification algorithms [28], [32] were used with default parameters to predict the occurrence of adverse events.

For each data-set, eXtreme Gradient Boosting (XGB), Linear Discriminant Analysis (LDA) Adaboost, C50 (C5.0 decision tree), Generalized Boosted Regression Modeling (GBM), K Nearest neighbors classifier(IBk) J48 (C4.5 decision tree), JRip (propositional rule learner based on association rules with reduced error pruning), naïve bayes (NB), Neural Network (NNet), OneR (generates one rule for each predictor in the data, then selects the rule with the smallest total error as its "one rule") PART (uses partial decision trees), Random Forest (RF: an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees) and SVM were applied to predict whether each treatment causes adverse events. The 3-fold cross-validation [4] was used to evaluate each classification model.

The class distribution was imbalanced with only 13.3% of positive cases (occurrence of short term adverse event) in psoriasis data-set, 8.9% of positive cases in eczema data-set and 10.6% of positive cases in nodular prurigo data-set. However, these learning techniques were designed and attempt to find an accurate performance over a full range of samples, based on the balanced classes of training data-set. If learning from the data-set with the highly imbalanced class distribution, these learning techniques tend to be overwhelmed by the majority class and ignore the minority class, and consequently might provide poor classification results [25].

To solve this problem, we adjusted the class distribution of training data by under-sampling the majority class samples of the training data. The method of under-sampling data used here is the "farthest distance" technique [10]. When we applied the 3-fold cross-validation for a data-set, we under-sampled the majority class samples to 60%, 50% and 40% in the training data of psoriasis, eczema and nodular prurigo data sets.

b) Experiment 2: We chose the overall best performing undersampled data sets of psoriasis, eczema and nodular prurigo patients and applied parameter tuning of classifier algorithms to improve the accuracy of the classifiers. Performance was evaluated using 3-fold cross validation.

c) Experiment 3: We used the parameter tuned classifiers from experiment 3 and created stacked classifiers of size 2 to check if the accuracy can be further improved. L1-Regularized Logistic Regression classifier was used as the super learner when creating stacked classifiers. Again, the performance was evaluated using 3-fold cross validation.

III. RESULTS AND DISCUSSION

This paper uses the area under curve, overall accuracy and f1-score to evaluate the classification models. The area under curve (AUC) is an abbreviation for the area under the ROC (receiver operating characteristic) curve, based on the minor class. It is the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one [1]. The accuracy rate is the proportion of correctly classified samples of all samples. The f1-score is the harmonic mean of precision and recall. The detail of these terms can be found in [12].

The classification results of experiment 1 is presented in the 3 tables (Tables II, III and IV). In these tables, AUC, acc and f1 represents the average area under the curve, average accuracy and average f1-score of 3-fold cross validation obtained by the classifier algorithms namely XGB, LDA, Adaboost, C5.0, GBM, IBk, J48, JRip, NB, NNet, OneR, PART, RF and SVM when the majority class instances were under-sampled to 60%, 50% and 40% of its original size in psoriasis data-set (Table II), eczema data-set (Table III) and nodular prurigo data-set (Table IV).

As shown in Table II, when the majority class was under-sampled to 60% of its original size in psoriasis data-set, Adaboost classification algorithm gave the highest AUC of 0.86 and highest accuracy of 0.90 was recorded by Adaboost, C5.0, RF and SVM algorithms and highest f1-score was recorded by SVM algorithm with a 70%. The lowest AUC was recorded as 0.53 by OneR algorithm. Lowest accuracy by NB algorithm (0.45) and lowest f1-score (0) by GBM algorithm. When the majority class instances were under-sampled to 50% and 40% of its original size, the overall performance of the classification algorithms improved. Highest AUC of 0.91 was recorded by Adaboost algorithm, highest accuracy (0.92) was recorded by Adaboost, XGB, C50, J48, JRip, NNet, PART, RF and SVM.

TABLE II. PERFORMANCE OF CLASSIFICATION ALGORITHMS WHEN MAJORITY CLASS INSTANCES UNDER SAMPLED TO 60%,50% AND 40% OF ITS ORIGINAL SIZE IN PSORIASIS RECORDS.

	Majority Class*60%			Majority Class*50%			Majority Class*40%		
alg.	***AUC***	***acc***	***f1***	***AUC***	***acc***	***f1***	***AUC***	***acc***	***f1***
XGB	0.84	0.89	0.66	0.86	0.89	0.73	0.90	0.92	0.84
LDA	0.77	0.81	0.39	0.84	0.86	0.69	0.88	0.89	0.80
Adaboost	0.86	0.90	0.68	0.88	0.90	0.74	0.91	0.92	0.84
C50	0.82	0.90	0.68	0.84	0.90	0.74	0.87	0.92	0.83
GBM	0.74	0.79	0.00	0.78	0.76	0.00	0.85	0.71	0.00
IBk	0.76	0.85	0.63	0.79	0.85	0.68	0.84	0.87	0.77
J48	0.82	0.89	0.68	0.83	0.90	0.74	0.87	0.92	0.84
JRip	0.75	0.89	0.66	0.80	0.90	0.74	0.86	0.92	0.84
NB	0.80	0.45	0.40	0.84	0.50	0.46	0.90	0.56	0.55
NNet	0.83	0.89	0.68	0.86	0.90	0.75	0.90	0.92	0.84
OneR	0.53	0.77	0.17	0.71	0.80	0.56	0.85	0.89	0.80
PART	0.83	0.88	0.66	0.86	0.89	0.73	0.90	0.92	0.84
RF	0.85	0.90	0.69	0.87	0.90	0.75	0.90	0.92	0.84
SVM	0.82	0.90	0.70	0.84	0.91	0.76	0.88	0.92	0.85

The best f1-score 0.85 was recorded by SVM when majority class was under-sampled to 40% of its original size. When we consider the overall performance, Adaboost, RF and SVM were the top 3 algorithms worked best for the psoriasis data-set and GBM, NB and OneR performed poorly.

For eczema data-set as in Table III, Adaboost recorded 0.82 as the highest AUC, SVM, RF, Adaboost, PART and C5.0 recorded 0.92 as best accuracy and SVM recorded 0.64 as highest f1-score when majority class instances were under-sampled to 60% of its original size. Likewise, in psoriasis data-set, classifier performance on eczema data-set improved when majority class instances were under-sampled to 50% and 40%. RF recorded the highest AUC of 0.87, Adaboost, RF, SVM, PART, C50 and J48recorded the highest accuracy of 0.92 and highest f1-score of 0.77 was by RF and SVM when majority class instances were under-sampled to 40% of its original size. Overall best performers of eczema data-set were Adaboost, RF and SVM which are the top 3 performers of psoriasis data-set as well. Likewise, GBM, OneR, NB performed poorly.

When we consider the results for nodular prurigo data-set shown in Table IV, we can note that when the majority class instances were under-sampled to 60% of its original size, SVM recorded highest AUC of 0.92, and highest accuracy of 0.93. Best f1-score was by IBK recorded as a 0.73. Similar to both psoriasis and eczema data-sets, performance of classifiers improved when majority class instances were under-sampled to 50% and 40%. RF recorded 0.95 as the best AUC, SVM and PART scoring 0.93 as the best accuracy, SVM and IBK scored 0.84 as the best f1-score

TABLE III. PERFORMANCE OF CLASSIFICATION ALGORITHMS WHEN MAJORITY CLASS INSTANCES UNDER SAMPLED TO 60%,50% AND 40% OF ITS ORIGINAL SIZE IN ECZEMA RECORDS.

	Majority Class*60%			Majority Class*50%			Majority Class*40%		
alg.	***AUC***	***acc***	***f1***	***AUC***	***acc***	***f1***	***AUC***	***acc***	***f1***
XGB	0.77	0.90	0.55	0.83	0.90	0.65	0.85	0.91	0.72
LDA	0.74	0.85	0.01	0.77	0.87	0.44	0.82	0.85	0.64
Adaboost	0.82	0.92	0.63	0.85	0.92	0.70	0.86	0.92	0.76
C50	0.79	0.92	0.63	0.80	0.91	0.67	0.82	0.92	0.75
GBM	0.71	0.86	0.00	0.73	0.83	0.00	0.78	0.80	0.00
IBk	0.77	0.88	0.59	0.79	0.88	0.63	0.82	0.88	0.70
J48	0.78	0.91	0.62	0.80	0.92	0.69	0.81	0.91	0.75
JRip	0.70	0.91	0.57	0.75	0.91	0.64	0.79	0.91	0.73
NB	0.77	0.52	0.34	0.82	0.64	0.42	0.85	0.61	0.49
NNet	0.80	0.91	0.60	0.78	0.90	0.58	0.83	0.90	0.66
OneR	0.51	0.85	0.04	0.60	0.83	0.34	0.73	0.85	0.58
PART	0.78	0.92	0.62	0.78	0.92	0.69	0.83	0.92	0.75
RF	0.81	0.92	0.62	0.85	0.92	0.70	0.87	0.92	0.77
SVM	0.80	0.92	0.64	0.83	0.92	0.70	0.85	0.92	0.77

TABLE IV. PERFORMANCE OF CLASSIFICATION ALGORITHMS WHEN MAJORITY CLASS INSTANCES UNDER SAMPLED TO 60%,50% AND 40% OF ITS ORIGINAL SIZE IN NODULAR PRURIGO RECORDS.

	Majority Class*60%			Majority Class*50%			Majority Class*40%		
alg.	***AUC***	***acc***	***f1***	***AUC***	***acc***	***f1***	***AUC***	***acc***	***f1***
XGB	0.79	0.88	0.57	0.79	0.85	0.57	0.82	0.88	0.72
LDA	0.88	0.88	0.55	0.89	0.89	0.70	0.91	0.92	0.80
Adaboost	0.87	0.90	0.61	0.93	0.91	0.72	0.93	0.91	0.77
C50	0.76	0.90	0.60	0.86	0.91	0.74	0.88	0.92	0.80
GBM	0.80	0.83	0.00	0.79	0.81	0.00	0.81	0.77	0.00
IBk	0.84	0.91	0.73	0.89	0.92	0.81	0.91	0.92	0.84
J48	0.78	0.90	0.61	0.87	0.93	0.78	0.89	0.92	0.80
JRip	0.71	0.89	0.57	0.73	0.89	0.61	0.80	0.89	0.72
NB	0.85	0.60	0.46	0.89	0.58	0.47	0.89	0.60	0.51
NNet	0.87	0.89	0.68	0.87	0.92	0.77	0.91	0.92	0.83
OneR	0.64	0.82	0.40	0.66	0.82	0.45	0.79	0.89	0.72
PART	0.78	0.90	0.61	0.85	0.92	0.76	0.89	0.93	0.82
RF	0.87	0.91	0.64	0.92	0.91	0.71	0.95	0.92	0.81
SVM	0.92	0.93	0.71	0.92	0.94	0.81	0.92	0.93	0.84

when majority class instances were under-sampled to 40% of its original size. Un-like for psoriasis and eczema data-sets, overall best performers of nodular prurigo data-set were SVM, IBk and NNet. But GBM, NB and OneR were the worst performers for Nodular prurigo data-set as same as for psoriasis and eczema data-set.

As psoriasis, eczema and nodular prurigo records gave best performance when the majority class instances were undersampled to 40% of its original size, these data sets were used to check if we can further improve the accuracy in experiment 2 and 3.

The results of experiment 2 is presented in Tables VI, VII and VIII. Hyperparameters of the 14 classifiers namely XGB, LDA, Adaboost, C50, GBM, IBk, J48, JRip, NB, NNet, OneR, PART, RF and SVM were tuned to obtain the highest accuracy using the parameter settings shown in Table V.

When we compare the accuracy of psoriasis records where majority class instances were undersampled to 40% under default classifier settings as shown in Table II and

when hyperparameters tuned in classifiers as shown in Table VI, we can see an improvement of accuracy in two classifiers namely GBM and IBk. Accuracy of GBM classifier improved from 0.71 to 0.77 when hyper parameters were tuned. Likewise, accuracy of IBk improved from 0.87 to 0.92. None of the other classifiers showed an improvement in accuracy when hyper parameters were tuned for psoriasis records where majority class instances were undersampled to 40% of its original size.

By comparing Table III and VII we can learn that IBk, J48, JRip, NB and NNet algorithms improved accuracy when hyper parameters were tuned classifiers were applied on majority class undersampled to 40% of its original size eczema data set.

LDA, GBM and RF algorithms did not improve performance while PART classifier showed a drop in accuracy from 0.93 to 0.92 when hyper parameters were tuned classifiers were applied on majority class undersampled to 40% of its original size nodular prurigo data set. (See Table IV and VIII).

Table IX, X and XI illustrates the results of experiment 3 in which hyper parameter tuned stacked classifiers were used on psoriasis, eczema and nodular prurigo records where majority class instances were under sampled to 40% of its original size. L1-Regularized Logistic Regression classifier was used as the super learner.

As shown in Table IX, we can see that all the stacked classifier combinations except the stacked classifiers made of OneR and GBM base learners and OneR and NB base learners recorded accuracy of 0.92 for psoriasis records. OneR and GBM combination recorded 0.89 accuracy. OneR and NB combination recorded 0.91 accuracy which is an improvement when compared to 0.89 and 0.56 recorded for experiment 2 in Table VI.

Most of the stacked classifiers scored higher than 0.9 accuracy for Eczema records as shown in Table X. GBM and LDA, NB and LDA, OneR and LDA, NB and GBM, NB and OneR classifier combinations recorded 0.86, 0.88, 0.89, 0.89 and 0.87 accuracy which shows an improvement than using them as single classifiers as in Table III. Stacked classifier made up of OneR and GBM recorded 0.85 accuracy and did not improve compared to Table III.

Also for nodular prurigo records we can see that most the stacked classifiers scored 0.9 or higher accuracy as shown in Table XI. OneR and GBM combination showed a drop in performance from 0.9 to 0.89 when used in stacked classifier compares to results in Table IV. But NB and GBM combination could improve their accuracy to 0.87.

IV. Conclusion and Future Work

This paper introduced the prediction of short term adverse events in NB-UVB phototherapy treatments using data mining techniques. We identified the significant feature sets for psoriasis, eczema and nodular prurigo data-sets and used 14 learning algorithms to classify the occurrence of short term adverse events in the data sets in experiment 1. Then we tried to improve the accuracy of these classifiers by tuning hyper parameters in experiment 2. Experiment 3 made use of these hyperparameter tuned classifiers to create stacked classifiers of size 2. The findings of this paper are:

1) The most effective features that models the occurrence of adverse events.

2) when only 40% of negative classes with the farthest distance to the positive classes were used to train models, we could significantly improve the performance of the classifiers.

3) Adaboost, RF and SVM performed best for psoriasis and eczema data sets while SVM, IBk and NNet performed well for nodular prurigo. GBM, OneR and NB algorithms were the least performers for all 3 data-sets.

However unlike in the PuvaMate data-set, if all the necessary features that are required to represent a phototherapy record have been captured, in the future, we might be capable of building a more generalized model that would enable better prediction of adverse event occurrence. When the important attributes that are currently missing from the data sets such as Psoriasis Area and Severity Index (PASI) [6], Dermatology Life Quality Index (DLQI) [3], etc. have been collected for a considerable amount of records, the experiments need to be carried out in the future to check if there's an impact in the performance of prediction. In order to explore the relationship among patients, social network analysis techniques with clustering algorithms need to be used for these applications.

References

[1] J. A Hanley, and B.J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve". Radiology, vol. 143, no. 1, 1982, pp. 29-36.

[2] Y. Yoon, R. W. Brobst, P. R. Bergstresser, and L. L. Peterson, "Automatic generation of a knowledge-base for a dermatology expert system". In Computer-Based Medical Systems, Proceedings of Third Annual IEEE Symposium on, 1990, June, pp. 306-312.

[3] A. Y. Finlay, and G. Khan, "Dermatology Life Quality Index (DLQI) - a simple practical measure for routine clinical use". Clinical and experimental dermatology, vol. 19 no.3, 1994, pp. 210-216.

[4] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection". In Ijcai, vol.14, no. 2, 1995, August, pp. 1137-1145.

[5] Mitchell and M. Tom, "Machine Learning", The Mc-Graw-Hill Companies, Inc, 1997

[6] P. C. M. Van de Kerkhof, "The Psoriasis Area and Severity Index and alternative approaches for the assessment of severity: persisting areas of confusion", 1997

[7] M.M. Breunig, H.P. Kriegel, R.T. Ng, J. Sander, "LOF: identifying density-based local outliers", ACM sigmod record, 2000

[8] H. Honigsmann, "Phototherapy for psoriasis". Clin Exp Dermatol., vol. 26, no.4, 2001, pp.343-50.

[9] S. Laube, S. A. George, "Adverse effects with PUVA and UVB phototherapy", Journal of Dermatological Treatment, vol. 12, no.2, 2001, pp. 101-105.

[10] Y. M. Chyi, "Classification analysis techniques for skewed class distribution problems", Department of Information Management, National Sun Yat-Sen University, 2003

[11] Y. Matsumura, H.N. Ananthaswamy, "Toxic effects of ultraviolet radiation on skin" Toxicol. Appl. Pharmacol., vol. 195, 2004, pp. 298-308.

[12] C. Goutte, and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation". ECIR ,vol. 5, March 2005, pp. 345-359

[13] L Al Shalabi, Z Shaaban and B Kasasbeh ,"Data mining: A preprocessing engine" Journal of Computer Science, 2006

[14] L. Nanni, "An ensemble of classifiers for the diagnosis of erythemato-squamous diseases". Neurocomputing, vol. 69 no. 7, 2006, pp. 842-845.

[15] E.D. Übeyli, "Multiclass support vector machines for diagnosis of erythemato-squamous diseases". Expert Systems with Applications, vol. 35, no. 4, 2008, pp. 1733-1740.

[16] T. Mroczek, W. Paja, L. Piatek and M. Wrzesie, "Classification and synthesis of medical images in the domain of melanocytic skin lesions". Human System Interactions, 2008 May, pp. 705-709.

[17] C. L. Chang, and C. H. Chen, "Applying decision tree and neural network to increase quality of dermatologic diagnosis", Expert Systems with Applications, vol. 36 no. 2, 2009, pp.4035-4041.

[18] E. D. Übeyli, "Combined neural networks for diagnosis of erythemato-squamous diseases", Expert Systems with Applications, vol. 36, no. 3, 2009, pp. 5107-5112.

[19] Health Quality Ontario, "Ultraviolet Phototherapy Management of Moderate-to-Severe Plaque psoriasis: An Evidence-Based Analysis". Ontario Health Technology Assessment Series., vol.9, no.27, 2009, pp.1-66.

[20] K. Polat and S. Güneş, "A novel hybrid intelligent method based on C4. 5 decision tree classifier and one-against-all approach for multi-class classification problems". Expert Systems with Applications, vol. 36, no. 2, 2009, pp. 1587-1592.

[21] A.M. Elsayad, "Diagnosis of erythemato-squamous diseases using ensemble of data mining methods". ICGST-BIME Journal, vol. 10, no. 1, 2010, pp. 13-23.

[22] D. Giveki, H. Salimi, A.A. Bitaraf and Y. Khademian, "Detection of erythemato-squamous diseases using AR-CatfishBPSO-KSVM". Signal and Image Processing, vol. 2, no. 4, 2011, p. 57.

[23] J. Xie and C. Wang, "Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases". Expert Systems with Applications, vol. 38, no. 5, 2011, pp. 5809-5815.

[24] S. Ong, I. Coulson, "Legal claims in English dermatological practice". British Journal of Dermatology, vol. 164, 2011, pp. 217-219

[25] B. Huang, M.T. Kechadi, and B. Buckley. "Customer churn prediction in telecommunications." Expert Systems with Applications, vol. 39, no.1, pp. 1414-1425, 2012

[26] H. Cataloluk and M. Kesler, "A diagnostic software tool for skin diseases with basic and weighted K-NN". Innovations in Intelligent Systems and Applications (INISTA), 2012, July, pp. 1-4.

[27] J. Wan, K. Abuabara, A.B. Troxel, et al., "Dermatologist preferences for first-line therapy of moderate to severe psoriasis in healthy adult patients" J Am Acad Dermatol., vol. 66, no.3, 2012, pp. 376-86.

[28] R. Kumar, and R. Verma, "Classification algorithms for data mining: A survey". International Journal of Innovations in Engineering and Technology (IJIET), vol. 1, no.2, 2012, pp.7-14.

[29] S. Aruna, L.V., Nandakishore and S.P. Rajagopalan, "A hybrid feature selection method based on IGSBFS and naïve bayes for the diagnosis of erythemato-squamous diseases". International Journal of Computer Applications, vol. 41, no. 7, 2012.

[30] B. Karlik, and G. Harman, "Computer-aided software for early diagnosis of eerythemato-squamous diseases". In Electronics and Nanotechnology (ELNANO), IEEE XXXIII International Scientific Conference, 2013, April, pp. 276-279.

[31] D. K. Sharma and H.S. Hota, "Data Mining techniques for prediction of different categories of dermatology diseases". Journal of Management Information and Decision Sciences, vol. 16, no. 2, 2013, p. 103.

[32] Y. Huang, and T. Kechadi. "An effective hybrid learning system for telecommunication churn prediction." Expert Systems with Applications, vol. 40, no.14, pp. 5635-5647, 2013

[33] M.J. Abdi, and D. Giveki, "Automatic detection of erythemato-squamous diseases using PSO–SVM based on association rules". Engineering Applications of Artificial Intelligence, vol. 26, no. 1, 2013, pp. 603-608.

[34] K. K. Manjusha, K. Sankaranarayanan and P. Seena, "Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification". International Journal of Advanced Research in Computer Science and Software Engineering, vol. 4, no. 1, 2014.

[35] S. Sarhan, E. Elharir and M. Zakaria, "A Hybrid Rough-Neuro model For Diagnosing Erythemato-Squamous Diseases". IJCSI Int. J. Comput. Sci. Issues, vol. 11 no. 1, 2014.

[36] Ahmad, Parvez, S. Qamar, and S.Q.A. Rizvi, "Techniques of data mining in healthcare: a review." International Journal of Computer Applications, vol. 120, no.15, 2015.

[37] E. Racz and E.P. Prens, "Phototherapy and photochemotherapy for psoriasis" Dermatol Clin., vol. 33, no.1, 2015, pp. 79-89.

[38] Jothi, Neesha, and Wahidah Husain. "Data mining in healthcare- A review." Procedia Computer Science, vol. 72, 2015, pp. 306-313.

[39] K. Danjuma and A.O. Osofisan, "Evaluation of Predictive Data Mining Algorithms in Erythemato-Squamous Disease Diagnosis". arXiv preprint arXiv:1501.00607, 2015

[40] Rambhajani, Madhura, W. Deepanker, and N. Pathak. "A survey on implementation of machine learning techniques for dermatology diseases classification." International Journal of Advances in Engineering and Technology, vol. 8, no.2, 2015, pp.194

[41] M. Nakamura, B. Farahnik and T. Bhutani, "Recent advances in phototherapy for psoriasis" F1000Research, vol. 5, 2016.

[42] Patel, Sheenal, and H. Patel. "Survey of data mining techniques used in healthcare domain". International Journal of Information vol. 6, 2016.

[43] Phototherapy Service Guidance (2016, October) Retrieved from http://www.bad.org.uk/shared/get-file.ashx?itemtype=documentandid=4151.

TABLE V. HYPER PARAMETER TUNINING SETTINGS

Classifier	Param name	Param type	Param descrption	Tuned values
XGB	booster	discrete	booster type	gbtree,gblinear
	eta	numeric	step size of each boosting step	0.1-0.7
	nthread	integer	number of thread used in training	1-20
LDA	tol	numeric	A tolerance to decide if a matrix is singular	0.0001-0.001
Adaboost	loss	discrete	loss type	exponential,logistic
	type	discrete	type of boosting algorithm to perform	discrete,real,gentle
	iter	integer	number of boosting iterations to perform	10-100
C50	winnow	logical	should predictor winnowing (i.e. feature selection) be used?	True, False
	noGlobalPruning	logical	should global pruning step used?.	True, False
GBM	distribution	discrete	the distribution type	bernoulli,Adaboost,huberized
	n.trees	integer	the total number of trees to fit.	10-100
IBk	K	integer	number of nearest neighbors to be used	1-250
J48	C	numeric	confidence threshold for pruning.	0.1-0.5
	M	integer	minimum number of instances per leaf	1-20
JRip	N	numeric	minimal weights of instances	2-10
	O	integer	number of runs of optimizations	2-20
NB	laplace	numeric	provides a smoothing effect	0-10
NNet	maxit	integer	maximum number of iterations	10-100
	size	integer	number of units in the hidden layer	2-50
OneR	B	integer	minimum number of objects in a bucket	2-10
PART	C	numeric	confidence threshold for pruning.	0.1-0.5
	M	integer	minimum number of instances per leaf	1-10
RF	ntree	integer	number of trees to grow	2- 500
	mtry	integer	no. of variables used as candidates at each split	2-9
SVM	kernel	discrete	kernel function used in training and predicting	rbfdot, polydot, tanhdot, laplacedot, besseldot, anovadot
	scale	numeric	used with "tanhdot" and "polydot" kernels	1-10
	offset	numeric	used with "tanhdot" and "polydot" kernels	1-10
	sigma	numeric	used with "besseldot", "anovadot", "rbfdot" and "laplacedot" kernels	1-10
	degree	integer	used with "besseldot", "anovadot" and "Polydot" kernels	1-6
	order	integer	used with "besseldot" kernel	1-10

TABLE VI. BEST HYPER PARAMETER VALUES AND ACCURACY OF CLASSIFIER ALGORITHMS WHEN PARAMETERS TUNED ON MAJORITY CLASS UNDERSAMPLED TO 40% PSORIASIS RECORDS

Algorithm	Best hyperparameter values	Accuracy
XGB	booster=gbtree, eta= 0.6333333, nthread=20	0.92
LDA	tol=8e-04	0.89
Adaboost	loss=logistic, type=gentle,iter=60	0.92
C50	winnow=true, noGlobalPruning=true	0.92
GBM	distribution=huberized,n.trees=100	0.77
IBk	K=56	0.92
J48	C=0.1444444, M=1	0.92
JRip	N= 7.333333, O=6	0.92
NB	laplace=4.444444	0.56
NNet	maxit=40, size=50	0.92
OneR	B=7	0.89
PART	C=0.4111111, M=10	0.92
RF	ntree=500, mtry=3	0.92
SVM	kernel=laplacedot, sigma=0.1	0.92

TABLE VII. BEST HYPER PARAMETER VALUES AND ACCURACY OF CLASSIFIER ALGORITHMS WHEN PARAMETERS TUNED ON MAJORITY CLASS UNDERSAMPLED TO 40% ECZEMA RECORDS

Algorithm	Best hyperparameter values	Accuracy
XGB	booster=gbtree, eta= 0.2333333, ,nthread=3	0.91
LDA	tol=0.001	0.85
Adaboost	loss=exponential, type=gentle,iter=90	0.92
C50	winnow=false, noGlobalPruning=false	0.92
GBM	distribution=bernoulli,n.trees=100	0.80
IBk	K=29	0.91
J48	C=0.3222222, M=5	0.92
JRip	N=2, O=10	0.92
NB	laplace=1.111111	0.67
NNet	maxit=50, size=29	0.93
OneR	B=6	0.85
PART	C=0.4111111, M=1	0.92
RF	ntree=223, mtry=2	0.92
SVM	kernel=laplacedot, sigma=1.2	0.92

TABLE VIII. BEST HYPER PARAMETER VALUES AND ACCURACY OF CLASSIFIER ALGORITHMS WHEN PARAMETERS TUNED ON MAJORITY CLASS UNDERSAMPLED TO 40% NODULAR PRURIGO RECORDS

Algorithm	Best hyperparameter values	Accuracy
XGB	booster=gbtree, eta= 0.6333333, nthread=1	0.90
LDA	tol=0.0006	0.92
Adaboost	loss=exponential, type=discrete, iter=60	0.92
C50	winnow=false, noGlobalPruning=true	0.93
GBM	distribution=bernoulli, n.trees=40	0.77
IBk	K=1	0.93
J48	C=0.5, M=1	0.93
JRip	N=2,O=10	0.93
NB	laplace=5.555556	0.64
NNet	maxit=80, size=2	0.93
OneR	B=7	0.90
PART	C=0.5, M=1	0.92
RF	ntree=334, mtry=3	0.92
SVM	kernel=laplacedot, sigma=0.1	0.95

TABLE IX. ACCURACY OF HYPER PARAMETER TUNED STACKED CLASSIFIERS ON MAJORITY CLASS UNDERSAMPLED TO 40% PSORIASIS RECORDS

	XGB	LDA	Adaboost	C50	GBM	IBk	J48	JRip	NB	NNet	OneR	PART	RF	SVM
XGB	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LDA	0.92	-	-	-	-	-	-	-	-	-	-	-	-	-
Adaboost	0.92	0.92	-	-	-	-	-	-	-	-	-	-	-	-
C50	0.92	0.92	0.92	-	-	-	-	-	-	-	-	-	-	-
GBM	0.92	0.92	0.92	0.92	-	-	-	-	-	-	-	-	-	-
IBk	0.92	0.92	0.92	0.92	0.92	-	-	-	-	-	-	-	-	-
J48	0.92	0.92	0.92	0.92	0.92	0.92	-	-	-	-	-	-	-	-
JRip	0.92	0.92	0.92	0.92	0.92	0.92	0.92	-	-	-	-	-	-	-
NB	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	-	-	-	-	-	-
NNet	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	-	-	-	-	-
OneR	0.92	0.92	0.92	0.92	0.89	0.92	0.92	0.92	0.91	0.92	-	-	-	-
PART	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	-	-	-
RF	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	-	-
SVM	0.92	0.92	0.92	0.92	0.92	0.92	0.93	0.92	0.92	0.92	0.92	0.92	0.92	-

TABLE X. ACCURACY OF HYPER PARAMETER TUNED STACKED CLASSIFIERS ON MAJORITY CLASS UNDERSAMPLED TO 40% ECZEMA RECORDS

	XGB	LDA	Adaboost	C50	GBM	IBk	J48	JRip	NB	NNet	OneR	PART	RF	SVM
XGB	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LDA	0.91	-	-	-	-	-	-	-	-	-	-	-	-	-
Adaboost	0.92	0.92	-	-	-	-	-	-	-	-	-	-	-	-
C50	0.92	0.92	0.92	-	-	-	-	-	-	-	-	-	-	-
GBM	0.91	0.86	0.92	0.91	-	-	-	-	-	-	-	-	-	-
IBk	0.92	0.92	0.93	0.92	0.92	-	-	-	-	-	-	-	-	-
J48	0.92	0.92	0.92	0.92	0.92	0.92	-	-	-	-	-	-	-	-
JRip	0.92	0.92	0.93	0.92	0.92	0.93	0.92	-	-	-	-	-	-	-
NB	0.92	0.88	0.92	0.92	0.89	0.92	0.92	0.92	-	-	-	-	-	-
NNet	0.92	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	-	-	-	-	-
OneR	0.91	0.89	0.92	0.92	0.85	0.91	0.91	0.92	0.87	0.92	-	-	-	-
PART	0.92	0.92	0.92	0.93	0.92	0.92	0.92	0.92	0.92	0.92	0.92	-	-	-
RF	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.93	0.92	0.92	-	-
SVM	0.93	0.93	0.93	0.92	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	-

TABLE XI. ACCURACY OF HYPER PARAMETER TUNED STACKED CLASSIFIERS ON MAJORITY CLASS UNDERSAMPLED TO 40% NODULAR PRURIGO RECORDS

	XGB	LDA	Adaboost	C50	GBM	IBk	J48	JRip	NB	NNet	OneR	PART	RF	SVM
XGB	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LDA	0.93	-	-	-	-	-	-	-	-	-	-	-	-	-
Adaboost	0.92	0.93	-	-	-	-	-	-	-	-	-	-	-	-
C50	0.91	0.93	0.91	-	-	-	-	-	-	-	-	-	-	-
GBM	0.91	0.92	0.93	0.90	-	-	-	-	-	-	-	-	-	-
IBk	0.93	0.94	0.95	0.94	0.93	-	-	-	-	-	-	-	-	-
J48	0.92	0.92	0.91	0.94	0.94	0.94	-	-	-	-	-	-	-	-
JRip	0.92	0.93	0.92	0.92	0.90	0.94	0.92	-	-	-	-	-	-	-
NB	0.90	0.92	0.92	0.92	0.87	0.95	0.93	0.91	-	-	-	-	-	-
NNet	0.92	0.93	0.91	0.93	0.92	0.93	0.93	0.92	0.93	-	-	-	-	-
OneR	0.90	0.93	0.91	0.91	0.89	0.93	0.91	0.92	0.90	0.93	-	-	-	-
PART	0.93	0.92	0.92	0.91	0.92	0.94	0.92	0.91	0.93	0.92	0.91	-	-	-
RF	0.93	0.94	0.92	0.93	0.92	0.94	0.93	0.93	0.93	0.93	0.92	0.92	-	-
SVM	0.94	0.95	0.94	0.95	0.94	0.95	0.95	0.95	0.94	0.94	0.95	0.95	0.94	-

S. Mohamed is currently a research masters student at University College Dublin (UCD). She obtained her BSc in Computer Science international degree offered by UCD at National School of Business Management (NSBM) Sri Lanka. She was awarded with a Higher Diploma in Computer Based Information Systems and a Diploma in Computer System Design by National Institute of Business Management (NIBM) Sri Lanka.

She has worked as a System Analyst at NIBM Sri Lanka in 2014 and as a Researcher for the Irish Skin Foundation for the last two years. Her current research interests includes application of data mining and machine learning techniques on healthcare data.

Dr. B. Huang is a Senior Researcher at The Insight Centre for Data Analytics. He obtained Ph.D degree in 2007.

He has worked in University College Dublin and Chinese Academy of Sciences as a Research Specialist since 2007. He experiences professional with data mining, machine learning, and statistical data analysis, is interested in applying the knowledge and advanced analytical skills to solve problems in Healthcare, financial services, Telecom, software development, and other business services.

M-Tahar Kechadi is professor of Computer Science, University College Dublin (UCD), Ireland. He was awarded PhD and a Masters degree in Computer Science from University of Lille 1, France.

He is currently Principal Investigator in the INSIGHT Centre for Data Analytics. His research interests span the areas of big data analytics, distributed data analytics, heterogeneous distributed systems, Cloud computing, and forensic computing and cybercrime investigations. The core and central focus of his research for the last decade is how to manage and analyse data quickly and efficiently at larger scale.

He is in the editorial board of the Journal of Future Generation of Computer Systems and of IST Transactions of Applied Mathematics-Modelling and Simulation. He is a member of the International Knowledge Cloud Consortium (IKCC). He is full member at CERN. He is a member of ACM and IEEE.