



PERFORMANCE EVALUATION OF FEATURE SELECTION ALGORITHMS ON SKIN DISEASE PREDICTION



O. C. Abikoye^{1*}, R. G. Komolafe¹ and T. O. Aro²

¹Department of Computer Science, University of Ilorin, Kwara State, Nigeria

²Department of Mathematical and Computing Sciences, KolaDiasi University, Ibadan, Oyo State, Nigeria

*Corresponding author: kemi_adeoye@yahoo.com

Received: November 30, 2018 Accepted: April 11, 2019

Abstract: This paper investigates the influence of feature selection approach in the prediction of skin diseases using data mining techniques. Principal Component Analysis (PCA), Information Gain (GA) and Chi-square were the feature selection algorithms used to reduce features of skin diseases dataset. The classification was done using Random Forest, C4.5 Decision Trees and Functional Tree (FT). Experimental results of the developed predictive model on skin diseases have revealed that the feature selection algorithms did not necessarily improve the accuracy and sensitivity of these algorithms and in situation where they brought an improvement; it was just a little about 1 percent.

Keywords: Feature selection, principal component analysis, information gain, decision tree

Introduction

An approach of Data mining in medical area is swiftly increasing due to the enhancement effectiveness of these approaches for classification and prediction (Banaee *et al.*, 2013). Report given by patients in biomedical diagnosis may contain irrelevant, similar symptoms and signs especially when a patient suffers from more than one type of disease of the same category, then physicians are not getting patient perfectly diagnosed (Toombs, 2013). Computing with data mining intelligent algorithms can produce prediction in clinical datasets with many inputs (Bindushree, 2016). Data mining techniques have a good effect in changing enormous data into definite and more important information for knowledge penetration and prediction purpose (Chandna, 2014). The approach of data mining in healthcare serves as a tool of knowledge finding in database that is used for taking out data associated with several diseases from dataset in order to assist easier prognosis of diseases (Adeyemo & Adeyeye, 2015).

Data mining approaches in medical area is increasing swiftly due to the effective development of these approaches for classification and prediction (Lebbe *et al.*, 2014). Moreover, many researches are using data mining to the diagnosis and classification of diseases to make the process quicker, less and easier (Aneeshkumar & Venkateswaran, 2012; Jin *et al.*, 2014; Kaur and Khamparia, 2015; Patil *et al.*, 2017). Skin is the main organ in a human body that protects the muscles and all parts of the body. Skin in human body has a great value because; any problem on the skin can affect the whole body. Skin is opened to outer environment, thus make it prone to disease and infection, this makes it important to take care of the skin from any disease (Sumithra *et al.*, 2015).

Skin disease is a disease that can have an effect on people of all age groups, examples of such are ring worm, eczema, psoriasis, skin cancer, fungal infection along with others, if skin disease is not treated early, it generally leads to difficulties and unnecessary situations as it extend on the skin, and transmitted to others and harmfully affect the confidence and social status of the carrier of such disease (Ginghi *et al.*, 2017). Skin diseases are not easy to diagnose as they are based upon a number of parameters which are not easy to identify. Different skin diseases often have related symptoms and can lead to an erroneous diagnosis if not properly examined (Parikh *et al.*, 2015).

This study is aimed at investigating the role of feature selection and extraction in the prediction of skin diseases using data mining algorithms that is based on the model of tree.

The remaining part of this paper is arranged as follows: Section 2 focuses on related work in skin disease diagnosis. Section 3 highlights the method used to achieve the result.

Section 4 presents the results and discussion to show the significant impact of feature selection algorithms on the diagnosis of the skin disease. Finally, Section 5 concludes the paper.

Feature selection

According to Aruna *et al.* (2012), it was stated that in dimension reduction for classification, feature selection holds a vital role and the feature selection algorithm task is to give a computing clarification to the issue stimulated by a description of importance. It can achieve classification performance by lowering the computing utilisation with the reduction in dimension. In this study, three different feature selection algorithms were used, they are; Information gain, Principal component analysis and Chi-square.

- (a) **Information gain:** Information gain (IG) measures the amount of information in bits about the class prediction, if the only information available is the presence of a feature and the corresponding class distribution. Concretely, it measures the expected reduction in entropy (uncertainty associated with a random feature) (Mitchell, 1997).
- (b) **Principal component analysis (PCA):** Principal component analysis (PCA) is broadly used in the field of computer science and it is known as a popular method of transformation, where the result that transformed is not openly correlated to an only feature element of the original sample. So, PCA is used in size reduction of the feature space after holding much information, and it is popularly accepted as a transformation method, that the result is not openly to only feature element of the original sample. Therefore, PCA has ability to play the role of feature selection and also able to decide on a number of significant individuals that are from the feature elements (Jolliffe, 2011).
- (c) **Chi-square:** The Chi-Square test of independence is a statistical test to determine if there is a significant relationship between 2 categorical variables. In simple words, the Chi-Square statistic will test whether there is a significant difference in the observed and the expected frequencies of both variables. The chi-square test can also be used with numerical variables by converting them into nominal or categorical types (Chen *et al.*, 2018).

Classification

It is the means of passing on a data item to a yet to defined category and it entails the innovation of an analytical learning task that makes data item to form many predefined classes. Classification involves investigating newly presented object features and assigning it with a predefined class, also

prediction is seen as the structure and assesses the class of an unlabeled sample with the use of a model, or to evaluate the value of an attribute that a specified example is likely to get. Examples are support vector machine, quadratic classifier, K-nearest neighbour, neural networks, learning vector quantization, but in this study, three that were used are C4.5 decision tree, functional tree and random forest.

- (a) **C4.5 decision tree:** Decision tree is one of the techniques in data mining for classification and it is engaged for machine learning process and decision support system (Seema *et al.*, 2012). Generally, the decision tree structure allows the relevant to realize the organization of trained knowledge models and it is the one that a test is being represented by a non-terminal node, where a selection of a particular branch depends on the conclusion of test. Therefore, for a particular data item to be classified, it starts at root node and trail the assertions down till it reaches a terminal node, where a decision is made when a terminal node is approached. Decision trees can be referred to as a form of rule set that is regarded by the hierarchical organization of rules (Karim *et al.*, 2015).
- (b) **Functional tree:** Functional trees are a framework for constructing multivariate trees for problem classification and regression (Gama, 2004). They are used for classification which can have logistic regression functions at the leaves. FT can handle both binary and multi-class variables.
- (c) **Random forest:** According to Breiman (2001) random forest is an enhancement on bagging that builds an enormous gathering the trees that are not correlated and put them on averages. Random forest is actually an ensemble of trees, and can be used for both classification and regression (Hastie *et al.*, 2008). Random forest tree is grown without pruning using classification and regression trees methodology to its maximum size, then the subspace randomization scheme is blended with bagging so as to resample with replacement the training data every time a new tree is grown (Biau, 2012). As the name implies, the feature in Random forest are randomly selected in each decision split while the tree is being built (Ali *et al.*, 2012).

Priya and Sathya (2019) developed classification and prediction of dermatitis dataset using naïve bayes and value weighted naïve bayes algorithms. The study passed through different dermatological diseases with similar symptoms, data were gathered from UCI repository contains the 230 instances with 22 attributes. The experimental results showed that performance of Naïve Bayes and weighted Naïve Bayes methods In order to compare the accuracy of both methods hence weighted Naïve Bayes gave the better results for the prediction process.

Handge *et al.* (2019) proposed system image of the skin disease that will be captured using smart-phone camera. After capturing image pre-processing and segmentation will be performed on the captured image. Feature extraction method is used to capture basic features such as colour, texture and shape or Domain specific features from image for indexing and retrieval. After feature extraction, classification of features can be done. In Feature Extraction, the system will compare the captured image with the training data set using optimization image processing techniques and decides whether a skin suffers from diseases or not.

Pathan *et al.* (2018) developed a computerized image analysis method for dermoscopy; it provides significant information about the lesion, which can be of pertinence for the clinicians and a stand-alone warning implement. In the study, they stated

that Computer-based diagnostic systems require dedicated image processing algorithms to supply mathematical descriptions of the supposed regions; such systems can be useful in oncology. It performed a review of the techniques used in computer-aided diagnostic systems, by giving the domain aspects of melanoma followed by the prominent techniques used in each of the steps. The steps include dermoscopic image pre-processing, segmentation, extraction and selection of peculiar features, and relegation of skin lesions. It also presents cognizance to evaluate the integrity of every methodology used, in addition to the corresponding results obtained in this perspective.

Sreena and Lijiya (2018) in their investigation, proposed an approach for automatic segmentation and classification of skin lesions. Originally, skin images are filtered to remove unwanted hairs and noise, then the segmentation process is done to extract lesion areas. For segmentation, a region growing method is added by automatic initialization of seed points. The performance of the segmentation is calculated with different known measures and the results are significant, afterward the extracted lesion areas are represented by color and texture features. Support Vector Machine (SVM) and k-NN classifiers were used with their fusion for the classification and the extracted features. The performance of the system is tested on the private dataset of 726 instances from 141 images that consist of 5 different classes of diseases. The results are promising with 46.71 percent and 34 percent of F-measure using SVM and k-NN classifiers, respectively and with 61 percent of F-measure for fusion of SVM and k-NN. The limitation is that private data was used.

Yang *et al.* (2018) introduced a level dataset for clinical skin disease. The dataset is the largest for visual recognition of skin diseases. It contains 6,584 images from 198 classes, which vary according to scale, color, shape and structure. This level dataset encourages additional research on visual skin disease classification. In addition, the successes of much computer vision correlated tasks are due to the taking on convolution Neural Networks (CNNs); they also performed broad analyses on this dataset using known methods including CNNs.

Ginghi *et al.* (2017) experimented on a system for identifying three types of skin diseases from skin images observed, the images were pre-processed and segmented using histogram analysis then features were extracted which are then used to classify the type of skin disease in the image. Results are sometimes related and not perfectly accurate.

Karthik *et al.* (2017) presented a study of predicting skin diseases using Naïve Bayesian classifier and concluded that it was a good classifier in predicting skin diseases. The orange data mining tool was used and the dataset was obtained from the UCI repository. Parikh *et al.* (2015) worked on diagnosing common skin disease by applying support vector machines and artificial neural network, they used a private dataset obtained from a hospital in India, It was established that ANN with 2 hidden layers has an enhanced accuracy to ANN with 1 hidden layer, SVM with RBF kernel and polynomial kernel of degree 3. The research also established the viability of using data mining algorithms in diagnosing the type of skin disease. The result cannot be easily compared because of private data used.

Sudha *et al.* (2017) developed a mathematical model for skin disease prediction by using response surface methodology. It focuses majorly on clinical and histopathological attributes influences on psoriasis disease of human body. Disease diagnosis is one of the applications of data mining. Prediction used to calculate the relationship by using regression equation. It starts the relationship among input and response attributes for improving disease diagnosis in medical area. Surface methodology was investigated.

Danjuma and Osofisan (2015) in their research evaluated the prediction of data mining algorithms in erythemato-squamous

disease diagnosis. It identified the most performing predictive data mining algorithms applied in the diagnosis of Erythematous-squamous diseases. The study used Naive Bayes, Multilayer Perceptron and J48 decision tree induction to build predictive data mining models on 366 instances of Erythematous-squamous diseases datasets. The result shows that the Naive Bayes performed best with accuracy of 97.4%, Multilayer Perceptron was second with accuracy of 96.6%, and J48 was the worst with accuracy of 93.5%. The evaluation of these classifiers on clinical datasets gave an indication into the predictive ability of different data mining algorithms applicable in clinical diagnosis especially in the diagnosis of Erythematous-squamous diseases.

is based on the model of tree, the skin disease data was obtained from the University of California, Irvine dataset repository and passed through feature selection algorithms namely; Information Gain, Chi-square and Principal Component Analysis. C4.5 decision tree, Functional Tree and Random Forest, classification algorithms were applied on the reduced and unreduced data. Feature selection algorithms were used and their effect studied using the unreduced data classification results as the base result. Implementation was done using the WEKA data mining tool. Evaluation was done based on Accuracy, Sensitivity and Root Mean Squared. The model architecture is shown in Fig. 1.

Materials and Methods

This study investigated the role of feature selection in the prediction of skin diseases using data mining algorithms that

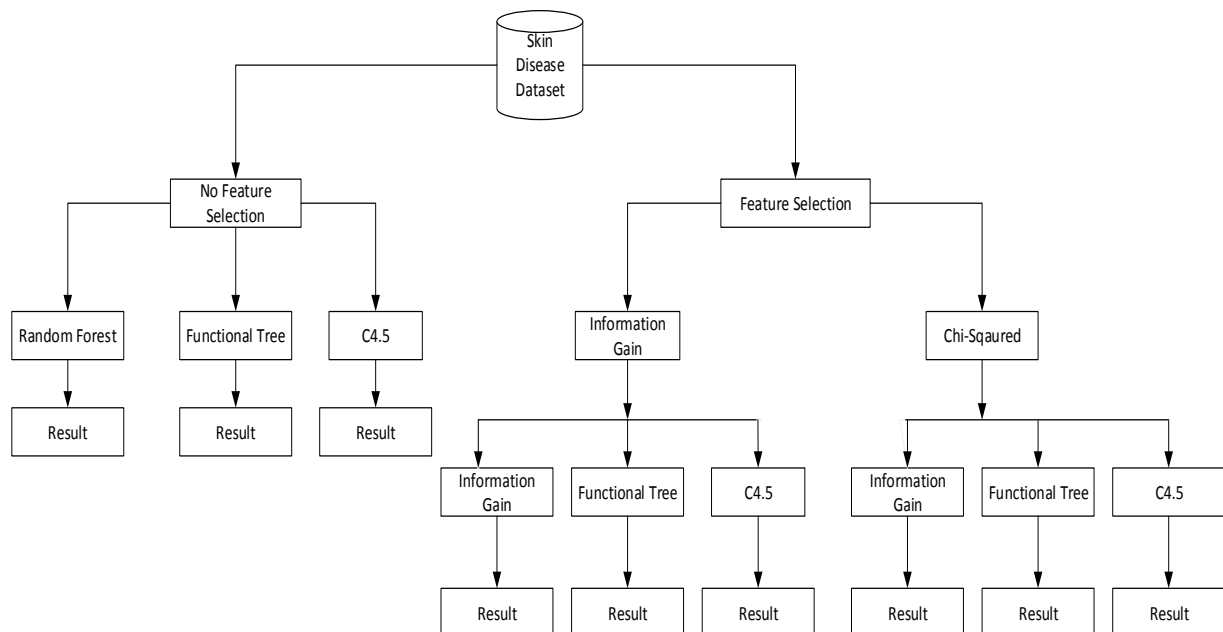


Fig. 1: Model architecture

Table 1: Attributes of the dataset

Classes of Diagnosis	Clinical (features f)	Histopathological
C1: Psoriasis	F1: Erythema	F12: Melanin incontinence
C2: Seboric dermatitis	F2: Scaling	F13: Eosinophils in the infiltrate
C3: Lichen planus	F3: Definite borders	F14: PNL infiltrate
C4: Pityriasis rosea	F4: Itching	F15: Fibrosis of the papillary dermis
C5: Chronic dermatitis	F5: Koebner phenomenon	F16: Exocytosis
C6: Pityriasis rubra pilaris	F6: Polygonal papules	F17: Acanthosis
	F7: Follicular papule	F18: Hyperkeratosis
	F8: Oral mucosal involvement	F19: Parakeratosis
	F9: Knee and elbow involvement	F20: Clubbing of the rete ridges
	F10: Scalp involvement	F21: Elongation of the rete ridges
	F11: Family history (0=no, 1=yes)	F22: Thinning of the suprapapillary epidermis
	F34: Age (linear)	F23: Spongiform pustule
		F24: Munro microabcess
		F25: Focal hypergranulosis
		F26: Disappearance of the granular layer
		F27: Vacuolisation and damage of basal layer
		F28: Spongiosis
		F29: Saw-tooth appearance of retes
		F30: Follicular horn plug
		F31: Perifollicular parakeratosis
		F32: Inflammatory mononuclear infiltrate
		F33: Band-like infiltrate

Description of datasets

The dataset used in this study was obtained from the University of California, Irvine dataset repository and it is originally from the Dermatology Database Gazi University, School of Medicine. 366 skin disease data were obtained which was used for prediction of six skin diseases that shows similar symptoms. The disease data is analysed as follows; 112 instances were psoriasis, 61 instances were seboreic dermatitis, 72 instances were lichen planus, 49 instances were pityriasis rosea, 52 instances were chronic dermatitis and 20 instances were pityriasis rubra pilaris. It was also discovered that 34 attributes are in the database, one is nominal and 33 are linear valued, which contain attribute of the kind of skin disease. They all share the clinical features of erythema and scaling, with next to no distinctions. Typically a biopsy is important for the conclusion yet tragically these ailments share numerous histopathological features. In the dataset, the family history features has the esteem 1 if any of these illnesses has been seen in the family and 0 generally. The age highlight basically speaks to the age of the patient. Each and every other component (clinical and histopathological) was given a degree in the scope of 0 to 3. Here, 0 shows that the element was absent, 3 demonstrate the biggest sum conceivable, and 1, 2 demonstrate the relative middle of the road esteems. The names and id quantities of the patients were removed from the database. Table 1 shows the attributes used in the dataset.

Performance evaluation parameters

(i) **Accuracy:** The correctly and incorrectly classified instances show the percentage of test instances that are correctly and incorrectly classified while the unclassified instances show the percentage of test instances classified incorrectly. The percentage of instances that are correctly classified is called accuracy as shown in Equation 1 while the percentage of instances that are not correctly classified is obtained by subtracting the correctly classified instances from 100.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (1)$$

Where: TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative

(ii) **Sensitivity:** This is the proportion of people who have the disease and was rightly classified as having the disease. It is also known as recall or true positive rate. Sensitivity is an important measure to use when predicting diseases so as to know the capability of the algorithm in predicting right as many diseases as possible.

$$Sensitivity = \frac{TP}{TP+FP} \quad (2)$$

(iii) **Root mean squared error:** The differences between values predicted by a model and the values truly observed are measured by root mean squared error and the square root of the average measures observed. Smaller values indicate a better performance.

Results and Discussion

The results of the skin diseases predictive model were evaluated using accuracy, sensitivity and Root Mean Square Error.

Accuracy of skin diseases predictive model

(a) **Accuracy of random forest:** The accuracy of Random Forest algorithm without any feature selection algorithm is 95.9%. When information gain was used on the data before classifying with random forest, the accuracy is 96.72%. Also, when PCA was used with Random forest, the accuracy was 96.72%. Furthermore, when Chi-square was used with Random forest, the accuracy was 96.18%. For Random forest, all the feature selection algorithms

improved the accuracy of the algorithm slightly but none of the improvement was up to a 1% margin. This indicates that feature selection on skin diseases generally does not have improved algorithms accuracy. Fig. 2 shows the accuracy of the random forest with and without feature algorithms.

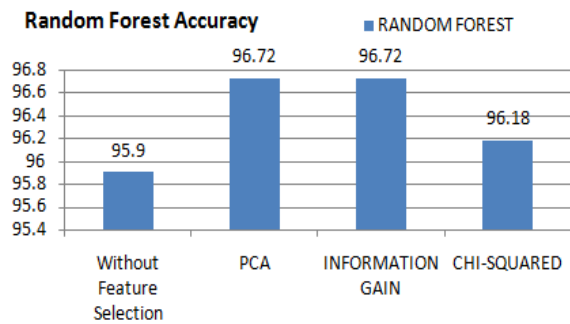


Fig. 2: Accuracy of random forest

(b) **Accuracy of functional tree:** Also, without any feature selection algorithm, the accuracy of Functional Tree algorithm is 95.9%, when PCA was used with FT, the accuracy was 96.99%. Also, 95.9% is the result when information gain was used with Functional Tree. Moreover, when Chi-square was used with Functional Tree, the result is 95.9% as shown in Fig. 3.

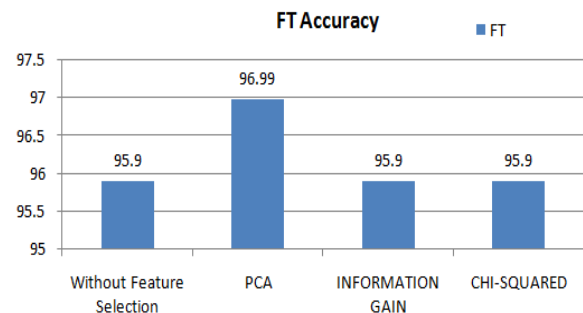


Fig. 3: Accuracy of functional tree

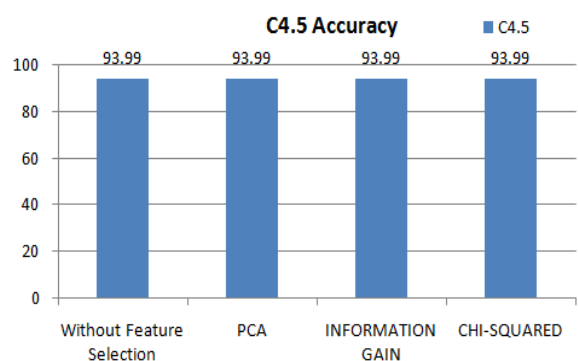


Fig. 4: Accuracy of C4.5 algorithm

(c) **Accuracy of C4.5 decision tree:** Accuracy of C4.5 algorithm without passing through any feature selection algorithm is 93.9%, and the application of Principal Component Analysis (PCA) with C4.5 gave an accuracy of 93.9%. Also, the result was 93.9% when information gain was used with C4.5. Furthermore, the result was the same when Chi-squared was used with C4.5, as shown in Fig. 4.

Sensitivity of the skin diseases predictive model

(a) **Sensitivity of C4.5 decision tree:** The model follows a similar pattern of accuracy in their sensitivity; when the datasets were not passing through any feature selection and when it passed through the feature selection, the same result of 94% was obtained; this shows that feature selection does not improve the sensitivity of C4.5 and it is shown in Fig. 5.

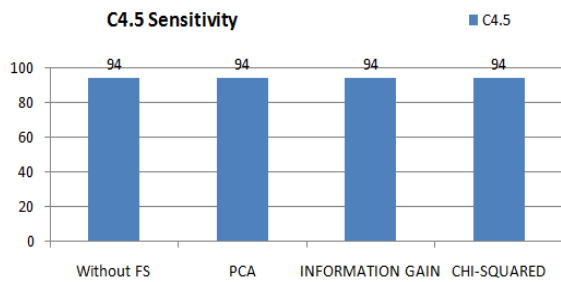


Fig. 5: Sensitivity of C4.5 algorithm

(b) **Sensitivity of functional tree:** Functional Tree sensitivity is improved with 97% when the datasets were used with PCA, while the result is the same for other feature selection algorithms as shown in Fig. 6.

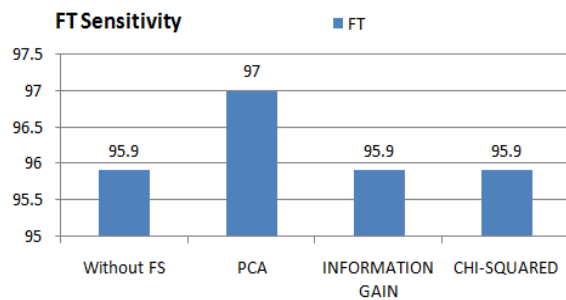


Fig. 6: Sensitivity of FT algorithm

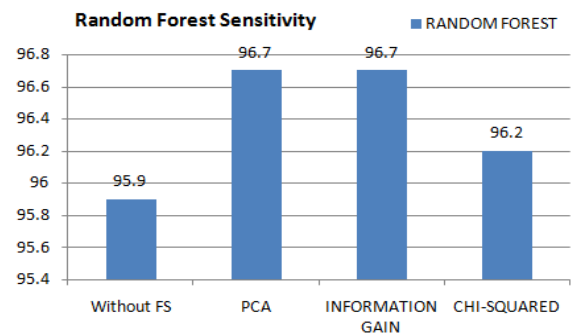


Fig. 7: Sensitivity of random forest algorithm

(c) **Sensitivity of random forest:** Random forest had an increase of about 97% when PCA and Information Gain were applied. This is also an indication that feature selection algorithms do not improve the sensitivity of predicting skin diseases as shown in Fig. 7.

Root mean squared error of skin diseases predictive model

Root mean squared error is the measure of the differences between values predicted by a model or an estimator and the values actually observed. It is the square root of the average measures observed. Smaller values indicate a better performance.

(a) **Root mean squared error of C4.5 decision tree:** The prediction of the performance have similar value of root mean squared error when the process pass through feature selection and when it did not pass through feature selection for C4.5 as shown in Fig. 8.

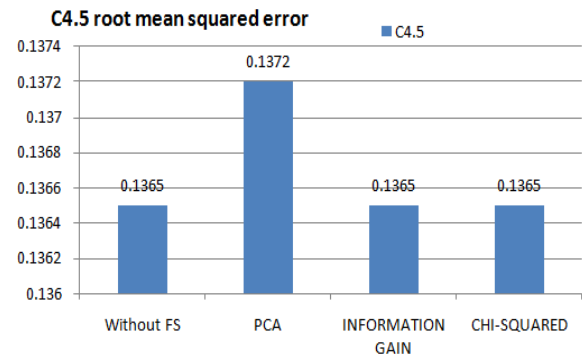


Fig. 8: Root mean squared error for C4.5

(b) **Root mean squared error of functional tree:** Root mean squared error is reduced with Feature Selection during classification using functional tree, especially information gain which reduced to less than half, indicating a better predictive performance as shown in Fig. 9.

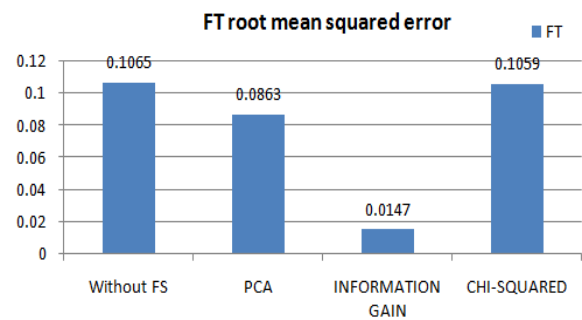


Fig. 9: FT root mean squared error

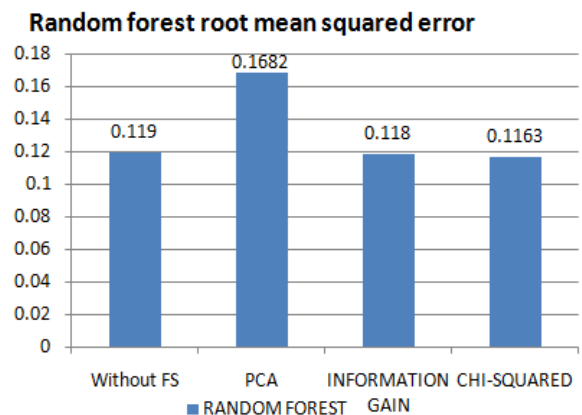


Fig. 10: Random forest root mean squared error

With respect to Random forest classification, Chi-Square and Information gain made the root mean squared error to be reduced, but it was increased in PCA as shown in Fig. 10. Also the three classifiers have the same result; while the value of root mean squared error in information gain is smaller than that of PCA means Information gain is more dependable in predicting skin disease than PCA.

Comparative analysis of skin diseases predictive model

The comparative analysis of the different evaluation metrics: accuracy, sensitivity and root mean squared are described in Tables 1, 2 and 3.

From Table 1, the highest accuracy of 96.99% was obtained from feature selection using PCA with Functional tree (FT). The lowest accuracy of 93.99% was achieved in all feature selection techniques and without feature selection with C4.5.

Table 1: Comparative analysis of accuracy for skin diseases predictive model

Feature selection techniques	Classifier (%)		
	RT	FT	C4.5
No Feature Selection	95.90	95.90	93.99
PCA	96.72	96.99	93.99
Information Gain	96.72	95.90	93.99
Chi-Squared	96.18	95.90	93.99

Table 2: Comparative analysis of sensitivity for skin diseases predictive model

Feature selection techniques	Classifier (%)		
	RT	FT	C4.5
No Feature Selection	95.90	95.90	94.00
PCA	96.70	97.00	94.00
Information Gain	96.70	95.90	94.00
Chi-Squared	96.20	95.90	94.00

Table 3: Comparative analysis of root mean squared for skin diseases predictive model

Feature selection techniques	Classifier (%)		
	RT	FT	C4.5
No Feature Selection	0.1190	0.1065	0.1365
PCA	0.1682	0.0863	0.1372
Information Gain	0.1180	0.0147	0.1365
Chi-Squared	0.1163	0.1059	0.1365

From Table 2, the highest sensitivity of 97.00% was obtained from feature selection using PCA with Functional tree (FT). The lowest sensitivity of 94.00% was achieved in all feature selection techniques and without feature selection using C4.5. In Table 3, the highest root mean squared value of 0.1682 was obtained from feature selection using PCA with random forest tree (RT). The lowest sensitivity of 0.0863 was achieved in PCA feature selection technique with functional tree.

Conclusion

Feature selection on decision tree classifiers in predicting skin disease has been shown not to improve accuracy or sensitivity significantly, the study also observed that the time taken to build the model and classify when feature selection algorithms were used increased significantly. Therefore, the use of feature selection algorithms for tree-based algorithms on skin disease didn't improve accuracy significantly and considering the increased time taken and the engagement of the processor, it is advisable not to be used unless in situations where any little increase in accuracy or sensitivity will make a significant difference.

References

Adeyemo OO & Adeyeye TO 2015. Comparative study of ID3/C4.5 decision tree and multilayer perception algorithms for the prediction of typhoid fever. *Afri. J. Computing & ICT*, 8(1): 103–112.

Ali, Jehad, Rehanullah Khan, Nasir Ahmad & Imran Maqsood 2012. Random forests and decision trees. *Int. J. Comp. Sci. Issues*, 9(5): 272–278.

Aneeshkumar AS & Venkateswaran CJ 2012. An approach of data mining for predicting the chances of liver disease in ectopic pregnant groups. *The Int. Conf. on Commun., Computing and Information Techn.*, pp. 19–22.

Aneeshkumar AS & C Jothi Venkateswaran 2015. Relevance study of data mining for the identification of negatively influenced factors in sick groups." *Procedia-Procedia Comp. Sci.*, 47: 101–108. <http://dx.doi.org/10.1016/j.procs.2015.03.188>.

Biau G. 2012. Analysis of a random forests model. *J. Machine Learning Res.* 13: 1063–1095.

Bindushree DC 2016. Prediction of cardiovascular risk analysis and performance evaluation using various data mining techniques: A review. *Int. J. Engr. Res.*, 5013(5): 796–800.

Breiman L 2001. No Title. *Machine Learning*, 45(1): 5–32.

Çataloluk, Hatice & Metin Kesler 2012. A diagnostic software tool for skin diseases with basic and weighted K-NN. *Int. Sympo. on Innov. in Intell. Sys. and Applic.*, pp. 0–3.

Chandna D. 2014. Diagnosis of heart disease using data mining algorithm. *Int. J. Comp. Sci. and Information Techn.*, 5(2): 1678–1680.

Durairaj M & Ranjani V 2013. Data mining applications in health sector: A study. *Int. J. Sci. & Techn. Res.*, 10.

Fayyad, Usama, Gregory Piatetsky-Shapiro & Padhraic Smyth 1996. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3): 37–54.

Gama J. 2004. Functional trees. *Machine Learning*, 55: 219–250.

Ginghi, Er, Ansari Nausheen, Ansari Zoya & Shaikh Ruhin 2017. An innovative approach for skin disease detection using image processing and data mining. *Int. J. Innovative Res. in Comp. and Commun. Engr.*, 5(4): 8135–1841.

Hastie, Trevor, Robert Tibshirani & Jerome Friedman 2008. *The Elements of Statistical Learning*.

Jin, Hoon, Seoungcheon Kim & Jinhong Kim 2014. Decision factors on effective liver patient data prediction. *Int. J. Biosci. and Biotech.*, 6(4): 167–177.

Karatzoglou, Alexandros, David Meyer & Kurt Hornik 2005. Support vector machines in R. *Journal of Statistical Software*, 21(9): 1–26.

Karim, Hamid & Khani Zand 2015. A comparative survey on data mining techniques for breast cancer diagnosis and prediction. *Indian J. Fundam. and Appl. Life Sci.*, 5(1): 4330–4339.

Karthik, Sistla, Bhimana Rahul, Mukesh Vibhudi, and Tarun Tej 2017. Prediction of dermatological condition using naive Bayesian classification. *Int. J. Pharmacy & Techn.*, 9(1): 28988–28994.

Kaur, Parminder & Aditya Khamparia 2015. Classification of liver based diseases using random tree. *Int. J. Advances in Engr. & Techn.*, 8(3): 306–313.

Lebbe, Ahamed, Sayeth Saabith, Elankovan Sundararajan & Azuraliza Abu Bakar 2014. Comparative study on different classification techniques for breast cancer dataset. *Int. J. Comp. Sci. and Mobile Computing*, 3(10): 185–191.

Parikh KS, Trupti PS, Rahulkrishna K & Rita V 2015. Diagnosing common skin diseases using soft computing techniques. *Int. J. Biosci. and Biotechn.*, 7(6): 275–286.

Patel, Sheenal & Hardik Patel 2016. Survey of data mining techniques used in healthcare domain. *Int. J. Information Sci. and Techn.*, 6(1): 53–60.

Patil RV, Sannakki SS & Rajpurohit VS 2017. A survey on classification of liver diseases using image processing and data mining techniques. *Int. J. Comp. Sci. and Engr.*, 5(3): 29–34.

Priya V & Sathya N 2019. Classification and prediction of dermatitis dataset using naive Bayes and value weighted naive Bayes algorithms. *Int. Res. J. Engr. and Techn.*, 6(2): 1077–1081.

Rajeswari K, Vaithianathan V & Shailaja VP 2013. Feature selection for classification in medical data mining. *Int. J. Emerging Trends & Techn. in Comp. Sci.* 2(2): 492–97.

Rajeswari P & G Sophia Reena 2010. Analysis of liver disorder using data mining algorithm. *Global J. Comp. Sci. and Techn.* 10(14): 48–52.

Seema, Monika Rathi, & Mamta 2012. Decision tree: Data mining techniques. *Int. J. Latest Trends in Engr. and Techn. (IJLTET)*, 1(3): 150–55.