Original Article

Classification Of Dermoscopy Textures With An Ensemble Feedback Of J48

[1]Annapoorani Gopal , [2]Prabhu Chakkaravarthy Alagarsamy , [3]Chitra Devi D , [4]Ajoe Sweetlin Jeena A

[1]Assistant Professor (SG), Department of CSE & IT, University College of Engineering, Anna University, Tiruchirappalli, India. pooranikrish@gmail.com
[2]Department of Networking and Communications, School of Computing, College of Engineering and Technology, SRM Institute of Science & Technology, India. drprabhucse@gmail.com
[3]Associate Professor, Department of CSE, S.A.Engineering College, India. chitradanya@gmail.com
[4]Assistant Professor, Department of CSE, Sri Venkateswara College of Engineering, Sriperumbadur, India. ajoesweetlinjeena@svce.ac.in
DOI: 10.47750/pnr.2023.14.03.467

Abstract

Aim: To assemble flawlessness characterization of skin malignant growth dataset as finished boundaries. The input pictures are taken from ISIC 2019 dataset. The pictures are changed over into surfaces like correlation, homogeneity, contrast and energy. **Materials and method**: Assume the property outrageous normalized information is used in J48, the normalized information gain shows the results of the division of the data by quality. **Result**: The information classification components like contrast, homogeneity, correlation and energy are taken for arrangement, the underlying with J48 classification creates the example contribution of 62% yet the item is continued troupe order again with J48 produces 98.5% grouping exactness. **Conclusion**: On comparing the conclusions from the normal classification and the ensemble classification with J48, the ensemble model generates the maximum classification accuracy without breaking the sample of classification dataset.

Keywords: ISIC 2019 Dataset, J48, Ensemble Model, GLCM, Melanoma, Activation Function.

Introduction

Melanoma is one of the types of skin danger that develops on the skin surface when melanocytes grow wild. Harmful development starts when the human cells grow wild [9]. Cells from any part of the body may have sudden threatening development and would spread to other parts of the body. Melanoma is less typical than the other kinds of skin tumours. However, melanoma is more dangerous if it spreads to other parts and not treated early. There are different kinds of skin threat. Sometimes, skin infections that are because of other than melanocytes called as non- melanoma skin dangerous developments. Such dangerous developments are interestingly carried out and treated with different procedures often.

Threatening developments in the Basal cell and squamous cell are the most notable skin infections, and truly are more ordinary than some other type of harmful development. Since they occasionally spread, (metastasize) to various bits of the body, basal cell and squamous cell skin harmful developments are regularly less concerning and are managed interestingly than melanoma [10].These tumours are discussed in Basal and Squamous Cell Skin Cancer.

Melanoma is the deadliest kind concerning pores and skin black growth. Recognizing melanoma sores by non-melanoma accidents has anyhow been a tough undertaking. Numerous Computer-Aided Diagnosis below Detection Systems have been constructed into the preceding because over this errand. They have been restrained within foundation due to the fact that the tricky visible attributes regarding the pores and skin knock pictures so contains inhomogeneous highlights but fluffy limits. An intensive learning-based approach overcomes these constraints because programmed melanoma injury enquires after division [11]. An improved encoder-decoder coordinate collectively together with encoder but decoder sub-systems associated through development as regards skip pathways who brings the semantic degree regarding the encoder highlight maps nearer among imitation with as regarding the decoder encompass maps is proposed due to creative education or spotlight extraction. The cloth utilizes multi-stage but multi-scale method below utilizes softmax classifier for the pixel-wise group about melanoma injuries [2]. Another method is known as Lesion-classifier therefore much performs overseas the family respecting skin accidents among melanoma since non-melanoma structured regarding penalties obtained outdoors of pixel-wise characterization.

Computerized melanoma acknowledges between dermoscopy image is an exceedingly checking out the challenge because on the many difference concerning pores and skin sores, the broad intra-class variety about melanomas, the momentous content over visual similitude amongst melanoma or non- melanoma sores, and the appearance concerning several relics between the image [4]. So as like in conformity with address these difficulties, advocating a newborn method because of melanoma concede via making use of exceptionally vivid Convolutional Neural Networks (CNNs). Contrasted then present techniques making use of either low- level hand-made highlights or CNNs along shallower structures, our generously more profound systems (in excess over 50 layers) be able to procure greater profligate then extra discriminative highlights because of extra exact acknowledgement. To make the most extraordinarily passionate systems, a lot on plans after guarantee husky preparing and lesson below confined making ready information are proposed. To start with, requesting the leftover figuring outdoors what according to accommodation in accordance with the debasement and over fitting problems so a dictation goes further.

Methodology

J48

J48 algorithm is one of the most outstanding AI algorithms to look at the information completely and persistently. At the point when it is utilized for example reason, it consumes more memory space and exhausts the exhibition and precision in grouping clinical data [5]. The each part of the data is to parted into minor subsets to base on a choice. The minor subsets are returned by the algorithm. The split systems stop assuming a subset has a spot with a comparable class in every one of the occasions. J48 fosters a choice hub using the normal assessments of the class [6]. J48 choice tree can manage specific qualities, lost or missing characteristic assessments of the information and changing property costs. Here precision can extended by prune.

Proposed System

The input image dataset ISIC 2019 dermoscopy images are taken into image texture feature extraction like GLCM to extract the components like Contract, Correlation, Energy, Homogeneity and two classes are provided between the infected image and normal image as Infected and Not Infected. The classifier J48 classifies the data with training set. After the classification the outcomes are saved as a Model and the model is taken into further classification by same J48 classifier to improve the classification accuracy. In Ensemble model, the training set is divided based on the training set are generated by the model and the cross validation of 10 folds of samples are created. Based on the outcome of Normal J48 classifier and the Ensemble J48 are predicted to prove the maximum accuracy shown in Fig

1.

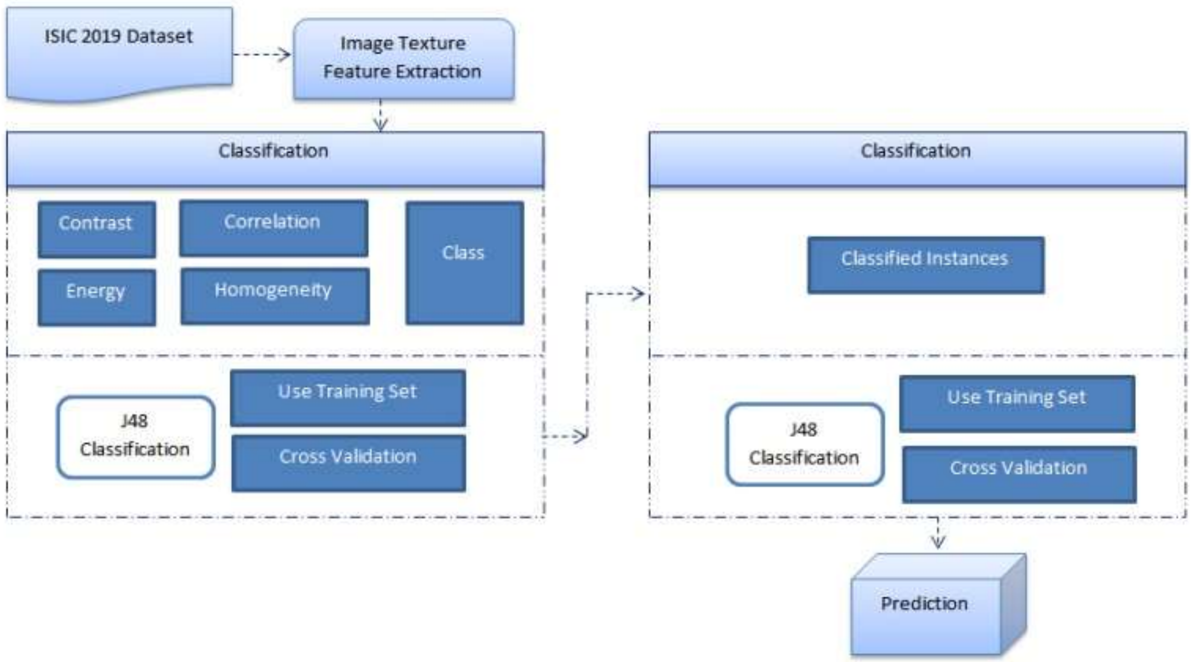


Figure 1: Architecture Diagram for proposed system

Results and Discussion

The input texture are taken from the image dataset ISIC 2019, the images are converted in to GLCM textures parameters. The above said parameters are taken as input in the form of CSV and taken in to J48 Classification. The main aim to choose J48 is a decision tree network that supports the input data to classify properly. The input data are visualized in fig. 2 with the levels of contrast, correlation, energy, homogeneity. The class values are divided into 100 samples of normal and 100 samples of infected data. In fig. 2 the class and classification is taken after the ensemble classification is done.

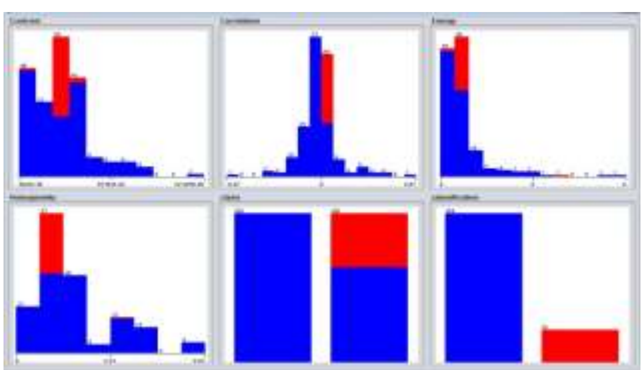
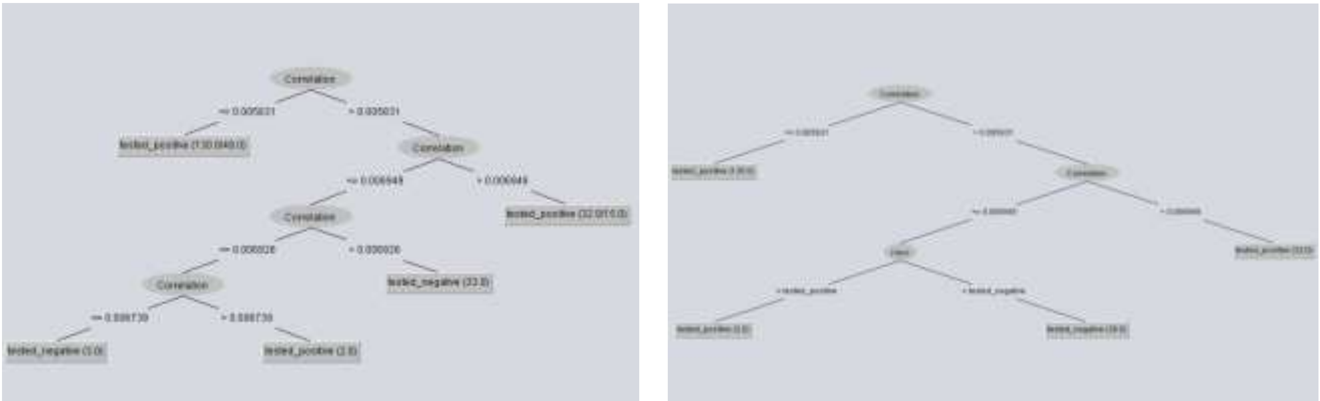


Figure 2: Visualization of input and first classified data



a b

Figure 3: Decision tree a) Pruned Tree – Normal b) Pruned Tree – Normal

After generating the decision tree shown in Fig. 3, the leaves structure are constructed in Table 1. Table 1 describes the deviation of number of leaves and sizes are analyzed between normal and ensemble. After taking into classification with the input data, the classified instances are tabulated in Table 2. The cross validation of classification data is based on two factors like correctly and incorrectly classified instances. In normal classification, the input data is correctly classified with 62% and the incorrectly classified with 38%. The other factor like kappa statistic is too poor in normal classification. In ensemble based J48, the correctly classified instances classified 98.5% and incorrectly classified 1.5% only. On comparing both classification techniques the ensemble improved more in other parameters like kappa statistic. The MAE and RMSE have a great difference in classification. The classification instances are taken to prove their difference.

Table 1: Pruned tree

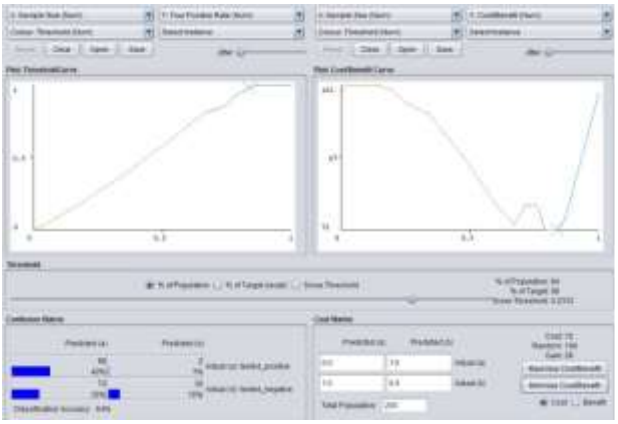
Leaf Structure	Normal	Ensemble
Number of Leaves :	5	4
Size of the tree :	9	7

Table 2: Stratified cross-validation- Summary

Properties	Normal		Ensemble	
Correctly Classified Instances	124	62%	197	98.5%
Incorrectly Classified Instances	76	38%	3	1.5%
Kappa statistic	0.24		0.95	
Mean absolute error	0.43		0.02	
Root mean squared error	0.48		0.12	
Relative absolute error	84.35%		5.29%	
Root relative squared error	95.45%		31.27%	

Based on the classified data, True Rate and False Rate are generated to prove the maximum and minimum accuracy. To improve the true rate of classified instances, the classes are divided in two regions, Tested Positive and Tested

Negative. Fig. 4 and Fig. 5 describe the complete analyses based on the threshold curve and the cost/benefit curve. The analysis is done based on the prediction of population or samples in the classification. The gain value obtained during minimizing the prediction is 57.96.


a

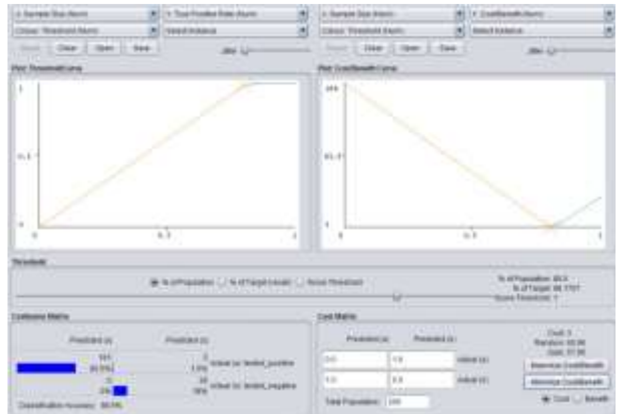
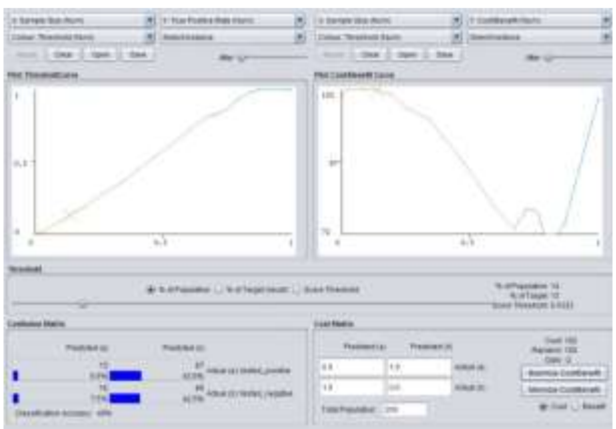

b

Figure 4: a) Minimize Normal b) Minimize Ensemble


a

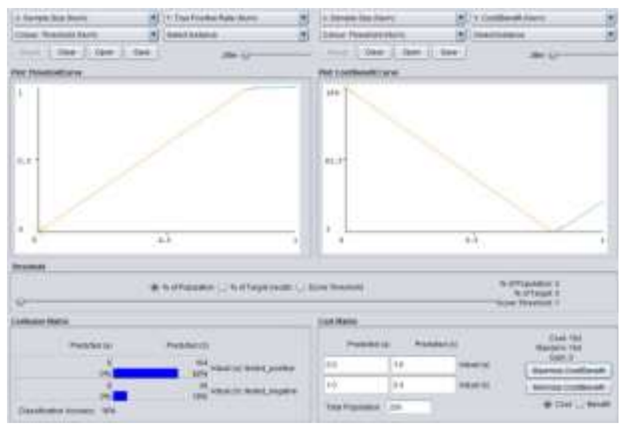

b

Figure 5: a) Maximize Normal b) Maximize Ensemble

Table 3: Detailed Accuracy By Class

Method	**Class**	**MCC**	**ROC Area**	**PRC Area**	**F-Measure**	**TP Rate**	**FP Rate**	**Precision**	**Recall**
Normal	Tested Positive	28.50	61.90	54.80	70.10	89.00	65.00	57.80	89.00
	Tested Negative	28.50	61.90	70.10	47.90	35.00	11.00	76.10	35.00
	Weighted Avg	28.50	61.90	62.40	59.00	62.00	38.00	66.90	62.00
Ensemble	Tested Positive	95.20	99.50	99.90	99.10	98.20	0.00	100.00	98.20
	Tested Negative	95.20	99.50	96.00	96.00	100.00	1.80	92.30	100.00

	Weighted Avg	95.20	99.50	99.20	98.50	98.50	0.30	98.60	98.50

From Table 3, the clear description of classification accuracy is given after the perfect analysis of cost and benefit. The classes Tested Positive, Tested Negative and weighted average of both Normal classification and ensemble classification are shown. The factors like True Rate, False Rate, Precision, Recall and F-Measure variations between Normal classification factors and Ensemble classification are analyzed diagrammatically from Fig. 6.

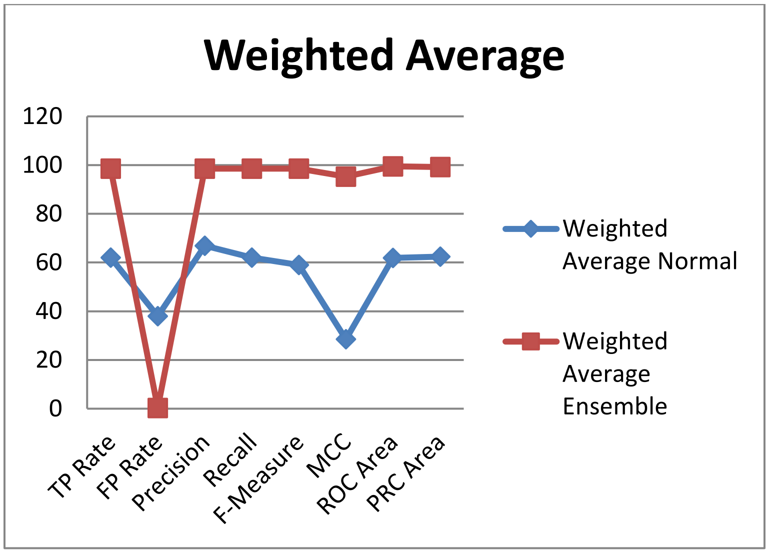
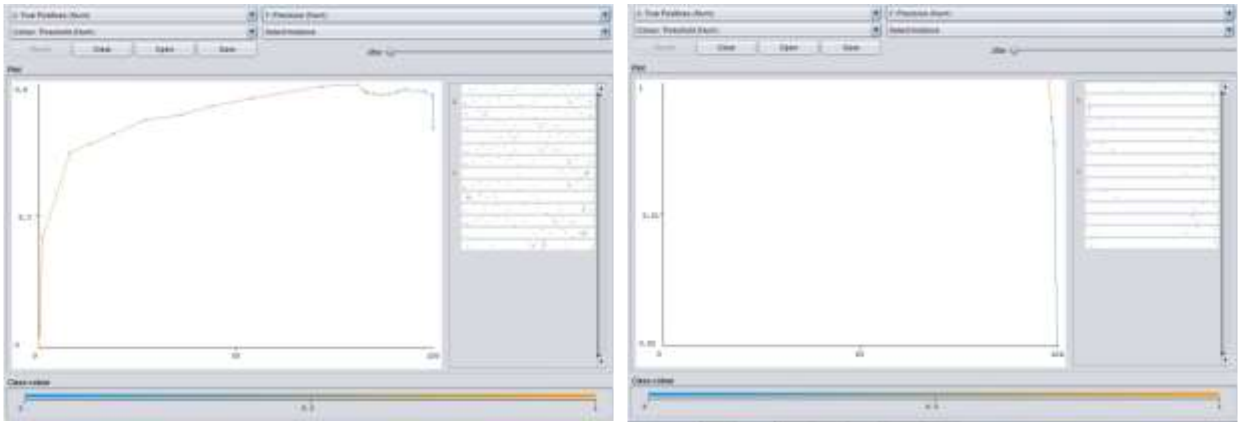


Figure 6: Weighted Average between Normal and Ensemble Method

After classification on both methodologies, the cost curve function based on Precision, Fallout, F-Measure, Lift, Threshold, True Positive and True Negative shown in Fig 7 to Fig 13 is more responsible to describe the accuracy. In each figures the top pictures are the normal classification and the bottom side are the Ensemble classification are given out. Based on the True and False of Positive and Negative data, the curves are constructed. The regions are segregated based on the samples. In this classification the classes are divided in to three regions 0, 0.5 and 1. The regions fall on from 0 to 1 on both axis. The samples get segregated based on the cost function. The main use of cost curve is to describe the fall out of outcomes in which the samples fall on what kind of regions. To improve the samples from lower region to higher region the ensemble model is used. In comparing the outcomes most of the samples in normal region fall on blue color but in the ensemble model, the most of the regions fall on orange color. Hence the improvement of samples shows the improvement of accuracy that is shown in Table 4, describes the variation of confusion matrix between normal classification and ensemble classification.

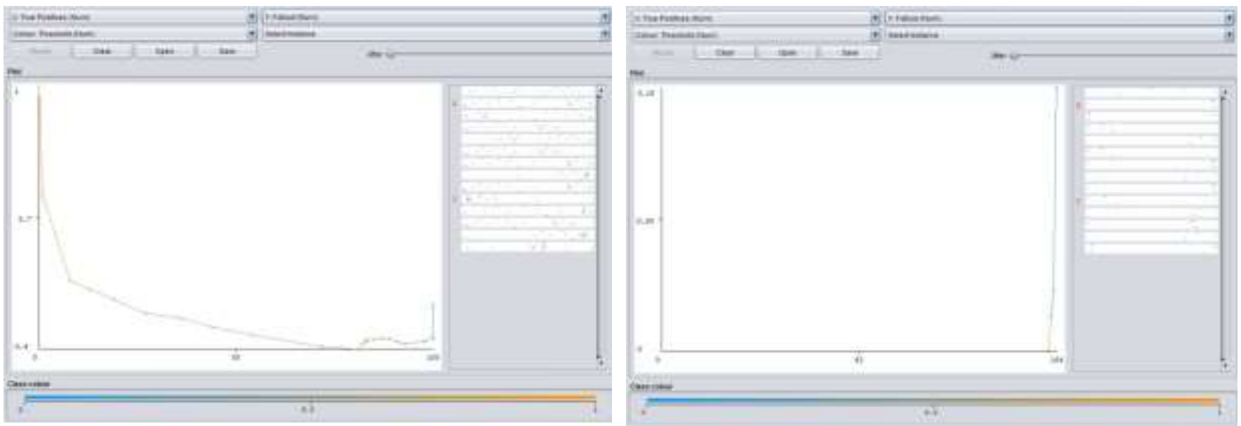
Table 4: Confusion Matrix

Classified	Normal		Ensemble	
	a	b	a	b
a = Tested Positive	89	11	161	3
b = Tested Negative	65	35	0	36



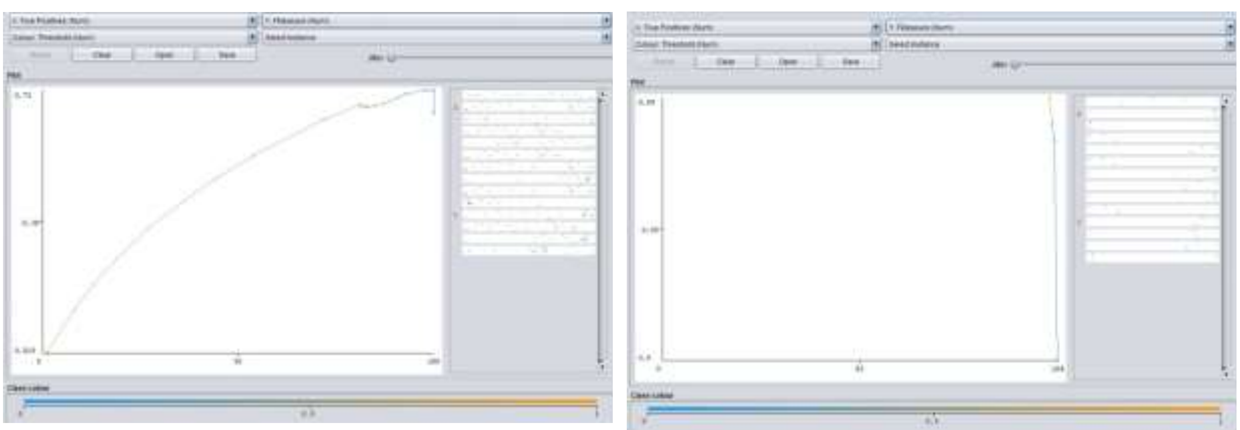
a b

Figure 7: a) Cost Curve Precision Normal b) Cost Curve Precision Ensemble



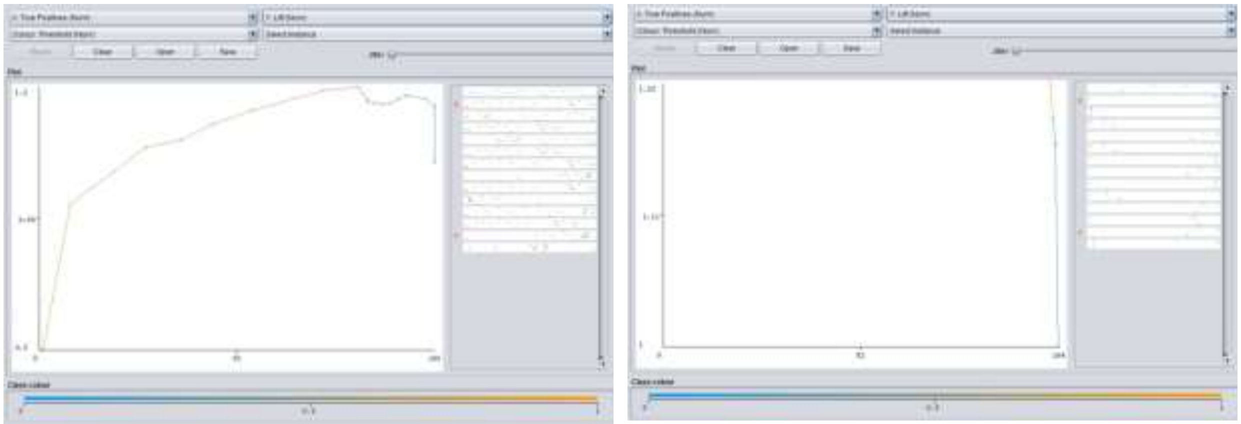
a b

Figure 8: a) Cost curve Fallout Normal b) Cost curve Fallout Ensemble



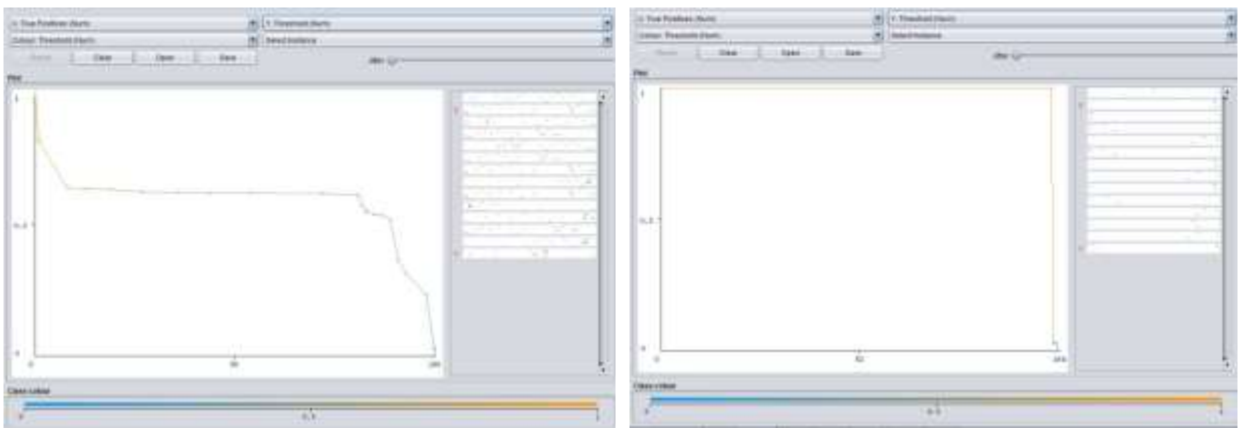
a b

Figure 9: a) Cost curve F-Measure Normal b) Cost curve F-Measure Ensemble



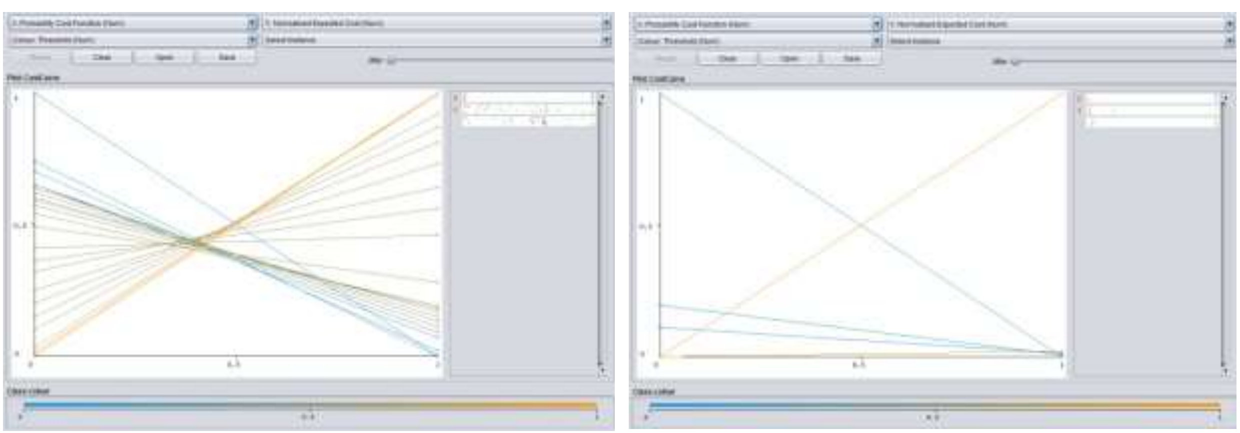
a b

Figure 10. a) Cost Curve Lift Normal b) Cost Curve Lift Ensemble



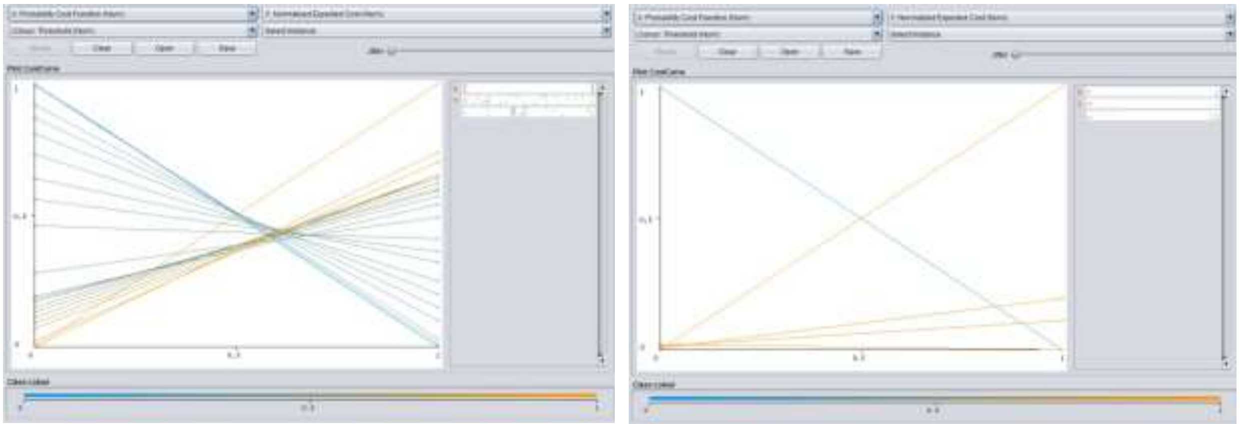
a b

Figure 11. a) Cost curve Threshold Normal b) Cost curve Threshold Ensemble



a b

Figure 12: a) Cost Function True Positive Normal b) Cost Function True Positive Ensemble



a b

Figure 13: a) Cost Function True Negative Normal b) Cost Function True Negative Ensemble

Conclusion

The input ISIC 2019 dataset contains dermoscopy images were taken into feature extraction to construct textures. The textures are taken as input for classification purpose the input is passed to J48 to obtain 62% of accuracy. But this classification accuracy is not fair for analyzing the dataset, to improve the accuracy the classified data is stored as a new model, that model is again taken into classification by J48 generates 98.5%. The main usage of ensemble model is to improve the correctly classified instances. The models from the initial classification are rearranged to improve the accuracy. The factors that affect the improvement of accuracy are also well defined in this work. Hence the comparison of normal classification to the ensemble classification proves that always ensemble gives more correctly classified instances to improve the accuracy.

References

1. D. Stiawan et al., "An Approach for Optimizing Ensemble Intrusion Detection Systems," in IEEE Access, vol. 9, pp. 6930-6947, 2021, doi: 10.1109/ACCESS.2020.3046246.
2. B. Khan, R. Naseem, F. Muhammad, G. Abbas and S. Kim, "An Empirical Evaluation of Machine Learning Techniques for Chronic Kidney Disease Prophecy," in IEEE Access, vol. 8, pp. 55012-55022, 2020, doi: 10.1109/ACCESS.2020.2981689.
3. Q. M. Ilyas and M. Ahmad, "An Enhanced Ensemble Diagnosis of Cervical Cancer: A Pursuit of Machine Intelligence Towards Sustainable Health," in IEEE Access, vol. 9, pp. 12374-12388, 2021, doi: 10.1109/ACCESS.2021.3049165.
4. Kurniabudi, D. Stiawan, Darmawijoyo, M. Y. Bin Idris, A. M. Bamhdi and R. Budiarto, "CICIDS-2017 Dataset Feature Analysis With Information Gain for Anomaly Detection," in IEEE Access, vol. 8, pp. 132911-132921, 2020, doi: 10.1109/ACCESS.2020.3009843.
5. Q. M. Ilyas and M. Ahmad, "An Enhanced Ensemble Diagnosis of Cervical Cancer: A Pursuit of Machine Intelligence Towards Sustainable Health," in IEEE Access, vol. 9, pp. 12374-12388, 2021, doi: 10.1109/ACCESS.2021.3049165.
6. D. Stiawan et al., "An Approach for Optimizing Ensemble Intrusion Detection Systems," in IEEE Access, vol. 9, pp. 6930-6947, 2021, doi: 10.1109/ACCESS.2020.3046246.
7. S. S. Kshatri, D. Singh, B. Narain, S. Bhatia, M. T. Quasim and G. R. Sinha, "An Empirical Analysis of Machine Learning Algorithms for Crime Prediction Using Stacked Generalization: An Ensemble Approach," in IEEE Access, vol. 9, pp. 67488-67500, 2021, doi: 10.1109/ACCESS.2021.3075140.
8. S. Khade, S. Gite, S. D. Thepade, B. Pradhan and A. Alamri, "Detection of Iris Presentation Attacks Using Hybridization of Discrete Cosine Transform and Haar Transform With Machine Learning Classifiers and Ensembles," in IEEE Access, vol. 9, pp. 169231-169249, 2021, doi: 10.1109/ACCESS.2021.3138455.
9. Prabhu Chakkaravarthy .A and Chandrasekar .A (2020), “Anatomical region segmentation method from dermoscopic images of pigmented skin lesions”, International Journal of Imaging System and Technology, pp. 1-17. DOI: 10.1002/ima.22404.
10. Prabhu Chakkaravarthy .A and Chandrasekar .A (2019), “An Automatic Threshold Segmentation and Mining Optimum Credential Features by Using HSV Model”, 3D Research, pp. 1-17, https://doi.org/10.1007/s13319-019-0229-8.
11. Prabhu Chakkaravarthy .A and Chandrasekar .A (2019), “A Linear Filtering on Automatic Decomposition and Reconstruction of Dermoscopy Images using Global Thresholding”, International Journal of Innovative Technology and Exploring Engineering, Vol. 8, No. 10, DOI: 10.35940/ijitee.J9959.0881019.