Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
2326-9865

Extended Feature SpaceBased Automatic Melanoma Detection System

Shakti Kumar[1], Anuj Kumar [2]

[1] Research Scholar, DCSA, Panjab University, Chandigarh

[2] Assistant Professor, DCSA, Panjab University, Chandigarh

[1]shaktibajpai@pu.ac.in[2]anuj_gupta@pu.ac.in

Article Info
Page Number: 8684-8697
Publication Issue:
Vol. 71 No. 4 (2022)

Abstract - Melanoma is the deadliest form of skin cancer. The uncontrollable growth of melanocytes leads to melanoma. Melanoma has been growing wildly in the last few decades. In recent years, the detection of melanoma using image-processing techniques has become a dominant research field. The Automatic Melanoma Detection System (AMDS) helps to detect melanoma based on image processing techniques by accepting infected skin area images as input. A single lesion image is a source of multiple features. Therefore, It is crucial to select the appropriate features from the image of the lesion in order to increase the accuracy of AMDS. For melanoma detection, all extracted features are not important. Some of the extracted features are complex and require more computation tasks, which impacts the classification accuracy of AMDS. The feature extraction phase of AMDS exhibits more variability, therefore it is important to study the behavior of AMDS using individual and extended feature extraction approaches. A novel algorithm ExtFvAMDS is proposed for the calculation of Extended Feature Vector Space. The six models proposed in the comparative study revealed that the HSV feature vector space for automatic detection of melanoma using the Ensemble Bagged Tree classifier on the Med-Node Dataset provided 99% AUC, 95.30% accuracy, 94.23% sensitivity, and 96.96% specificity.

Article History
Article Received: 15 September 2022
Revised: 25 October 2022
Accepted: 14 November 2022

Keywords: Melanoma Detection, Feature Extraction, Machine Learning, Feature Selection.

1Introduction

In India, melanoma of the skin was estimated to have 3916 new cases as reported by the International Agency for Research on Cancer (IARC) [1]. Melanoma is a treacherous form of skin cancer from which many individuals have lost their lives. Melanoma is caused by the abnormal and uncontrollable growth of melanocytes. Melanocytes are responsible for skin pigmentation. There are different types of cancer, like lung cancer, blood cancer, brain cancer, skin cancer, and many more. Skin is the biggest organ in the human body. Skin covers muscles, bones, and other parts of the body. The human skin is exposed to external surroundings. Contact of skin with UV radiation and tanning are the major cause of melanoma skin cancer.

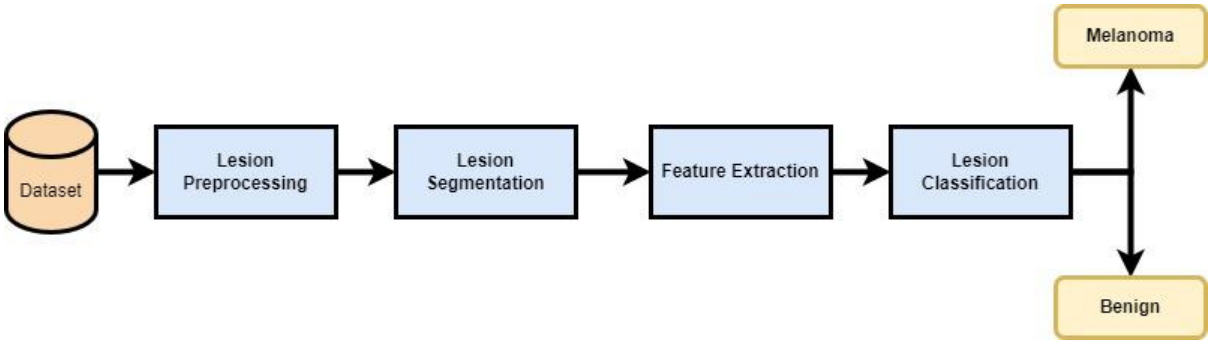

Fig.1 Stages in Automatic Melanoma Detection System

Detection of melanoma can be done by a dermatologist using manual techniques like biopsy or naked eye investigation. These manual techniques are invasive in nature. Alternatively, an automatic melanoma detection system is a non-invasive way to detect melanoma based on image processing techniques [4]-[6]. In [16] The automatic melanoma detection system goes through four basic stages namely the pre-processing stage, the segmentation stage, the feature extraction stage, and the classification stage, as shown in Fig.1.

i) Preprocessing Stageis the first stage of AMDS. This step is vital to deal with images having low analysis quality. In [15,16] Quality Factor for the AMDS is a major concern. The qualityfactor is affected due to the presence of various artifacts like Hairs, colournormalisation, illumination conditions, unreliable colour information, and unequal distribution of intensities. According to [2], these artifacts are the main challenges for the preprocessing step. The pre-processing is required as the images which are given as input to the automatic melanoma detection system contain noise and artifacts.Fig.2 shows the unprocessed images of the Med-Node Data dataset.

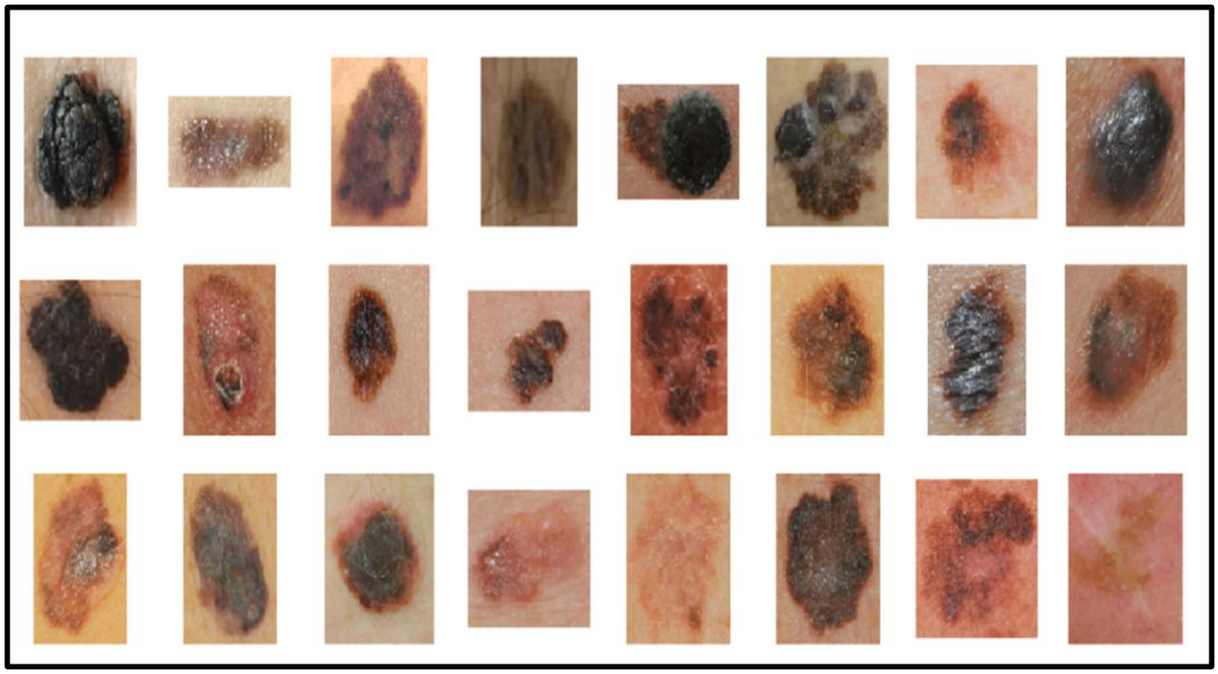

Fig.2 Unprocessed Melanoma Images from Med-Node

The lesion that is used for the study is generally surrounded by hair and contains noise. Different types of marks, illumination, and non-uniform intensities present in the input image may lead to misclassification. Therefore, a pre-processing step is required to remove all these unwanted components. Fig.3 Shows images from the Med-Node Data set after pre-processing.

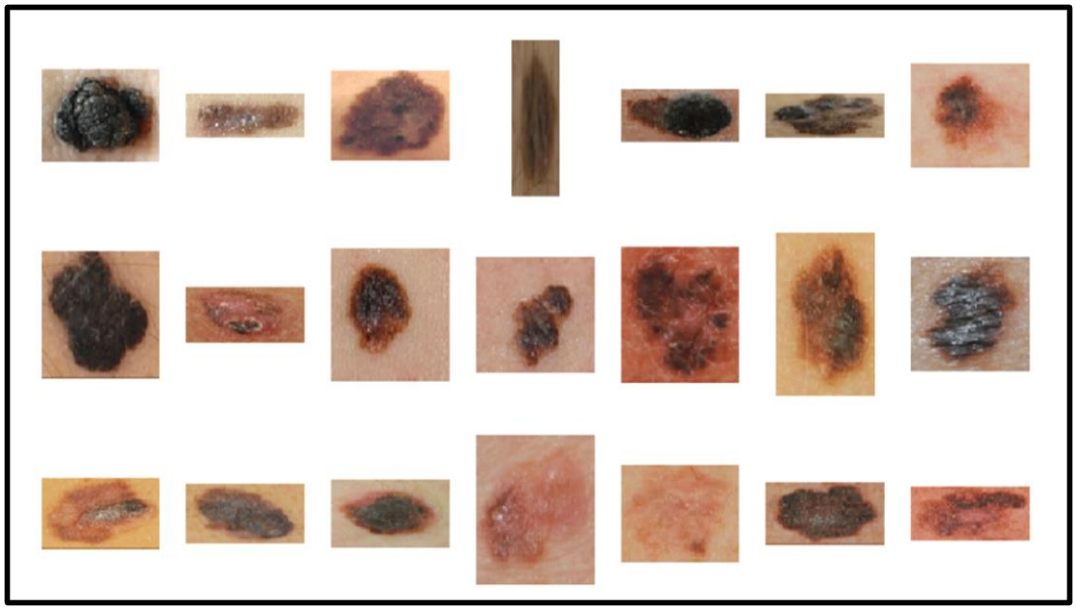

Fig.3 Resized and Denoised Melanoma Images from Med-Node

ii) lesion segmentation, which is the second stage in AMDS, the region of interest is obtained from the overall image. Is a challenging task. According to [3], for a given lesion variety of shapes, texture sizes, and colors make the segmentation process strenuous. The segmentation task becomes easier when there is a good difference between background and foreground intensities. It is believed that accurate lesion segmentation leads to the good extraction of features from the lesion. Fig.4 shows how the segmentation process results in the lesion.

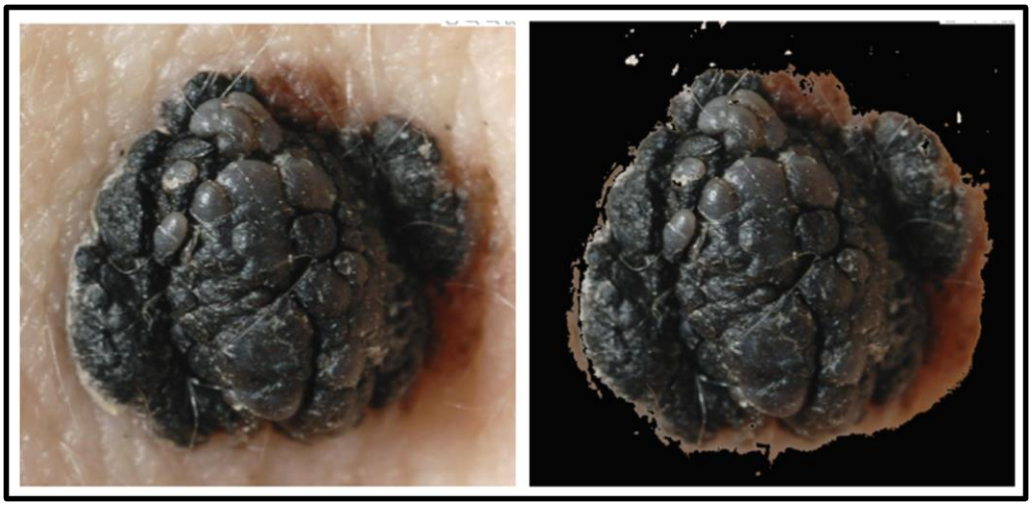

Fig.4 Lesion Image Before and After the Segmentation Process

iii) Feature Extraction After the segmentation of the lesion, the most important phase of the automatic melanoma detection system is feature extraction. The feature extraction face generally takes the segmented lesion as an input and extracts various features from it. According to [13, 14] all features obtained from the feature extraction phase are not important for melanoma detection. The selection of features from the given feature vector space affects the accuracy of the overall automatic melanoma detection system. According to [12] PCA is a good approach for feature selection. Our study aims to provide novelty in terms of comparative studyto understand the behavior of AMDS under 6 different proposed models.

iv) Classificationthe outputs of the feature extraction stage are given as input to the classification stage. The classification stage produces the result of the suspected lesion as benign or melanoma. Simple or hybrid classification techniques can be used in the classification stage.

The proposed method in our experiment, which classifies the lesion as benign or melanoma, was studied using six different models on the publicly available data set Med-Node and summarises the findings by in-depth evaluation. The following points are the key endowments of our present study:

- A novel algorithm ExtFvAMDS is given to compute the Extended Feature Vector Space for lesion images from Med-Node Dataset.
- The proposed model is able to identify the best AUC, classifier, and feature extractor for the Med-Node publicly available dataset.
- The proposed system is executed on an individual as well as a hybrid approach for different stages of AMDS.
- Using the proposed model 6, we have achieved 99% AUC, 95.30% Accuracy, 94.23% sensitivity, and 96.96% specificity using Extended Feature Vector Space on the Med-Node dataset.

The ensuing sections of the paper are described as follows: Section II covers materials and methods, and discussion and analysis of the results are carried out in Section III. Finally, Section IV includes conclusions and future work.

II Material and Methods

In this study, a novel technique is proposed To extract the lesion from the infected area of the skin. The first stage is used for noise removal. the second stage is used for extracting the features, the third stage is used for classification and The fourth stage is used for the analysis of the result. The experimentation is carried out on Med-Node data set consisting of images in both the categories Nevus and Melanoma. A novel algorithm namely ExtFvAMDS is proposed.

A) Dataset

To perform the experiments, MED-Node publically available datasets are selected. The datasets contain lesion images in the category of melanoma and benign. According to [7], **MedNode**-Dataset consists of 70 melanoma and 100 naevus images from the digital image archive of the Department of Dermatology of the University Medical Center Groningen (UMCG) used for the development and testing of the MED-NODE system for skin cancerdetection from macroscopic images. To determine the effectiveness of dataset size cross-validation technique is used.

B) Methodology

To evaluate the classification procedure our proposed methodology goes through 4 steps:

1. Lesion Pre-processing
2. Feature Extraction
3. Lesion Classification
4. Analysis of Result

The comparative study is evaluated using six proposed models as depicted using Fig.s numbered from 5 to 8.

Proposed Model 1 considers unprocessed Med-Node Data with HSV Feature Extractor, Ensemble Bagged Tree Classifier for the result evaluation as shown in Fig.5.

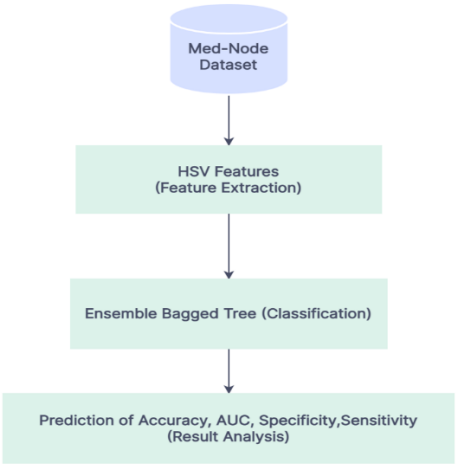

Fig.5 Proposed Model 1

Proposed Model 2 considers pre-processed Med-Node Data with HSV Feature Extractor, Ensemble Bagged Tree Classifier for the result evaluation as shown in Fig.6.

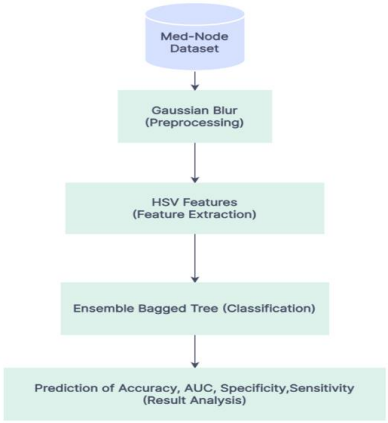

Fig.6 Proposed Model 2

Proposed Model 3 considers unprocessed Med-Node Data with LBP Feature Extractor, Ensemble Bagged Tree Classifier for the result evaluation as shown in Fig.7.

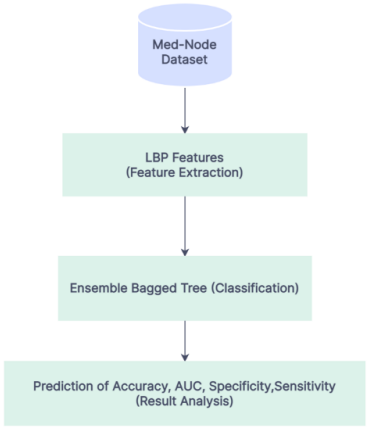

Fig.7 Proposed Model 3

Proposed Model 4 considers pre-processed Med-Node Data with LBP Feature Extractor, Ensemble Bagged Tree Classifier for the result evaluation as shown in Fig.8.

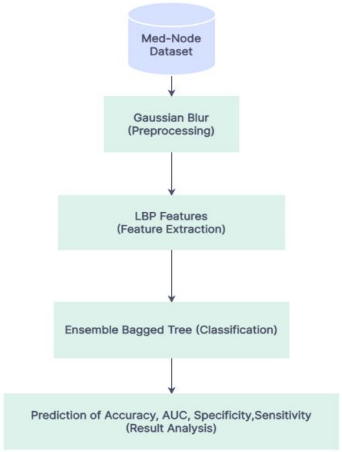

Fig.8 Proposed Model 4

Proposed Model 5 considers unprocessed Med-Node Data with both HSV and LBP Feature Extractor to create an extended feature vector space, and Ensemble Bagged Tree Classifier for the result evaluation as shown in Fig.9.

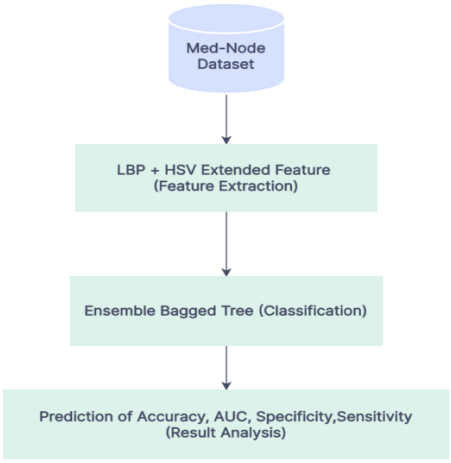

Fig.9 Proposed Model 5

Proposed Model 6 considers pre-processed Med-Node Data with both HSV and LBP Feature Extractor to create an extended feature vector space, and Ensemble Bagged Tree Classifier for the result evaluation as shown in Fig.10.

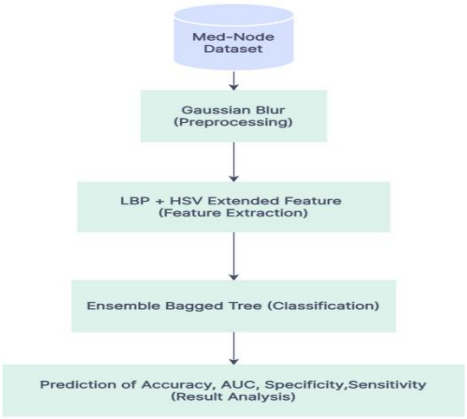

Fig.10 Proposed Model 6

Stage 1 - Lesion Preprocessing

The input images on the basis of which the whole experiment is to be carried out are of different resolutions and sizes. So images must be resized to a proportional scale (500 pixels in width). After the scaling operation, Gaussian Blur is carried out on the scaled images for noise removal.

Stage 2 - Feature Extraction

The feature extraction step is generally carried out in many machine learning applications. In an automatic melanoma detection system,It is the second stage after the pre-processing step.

The output of pre-processing is considered as the input of the feature extraction step. Each extracted feature is a vector representation of the input image.HSV (Hue - Saturation - Value) [8] values are used to represent the colour features of the input image. The RGB input image of the lesion is converted to HSV colour as shown in Fig.**11**. Then a 3-D histogram of all channels (H, S, V) is calculated, each with 8 bin count. The proposed Algorithm **ExtFvAMDS** is used for pre-processing as well as for HSV, LBP and Extended feature vector space computation in the first two stages of AMDS.

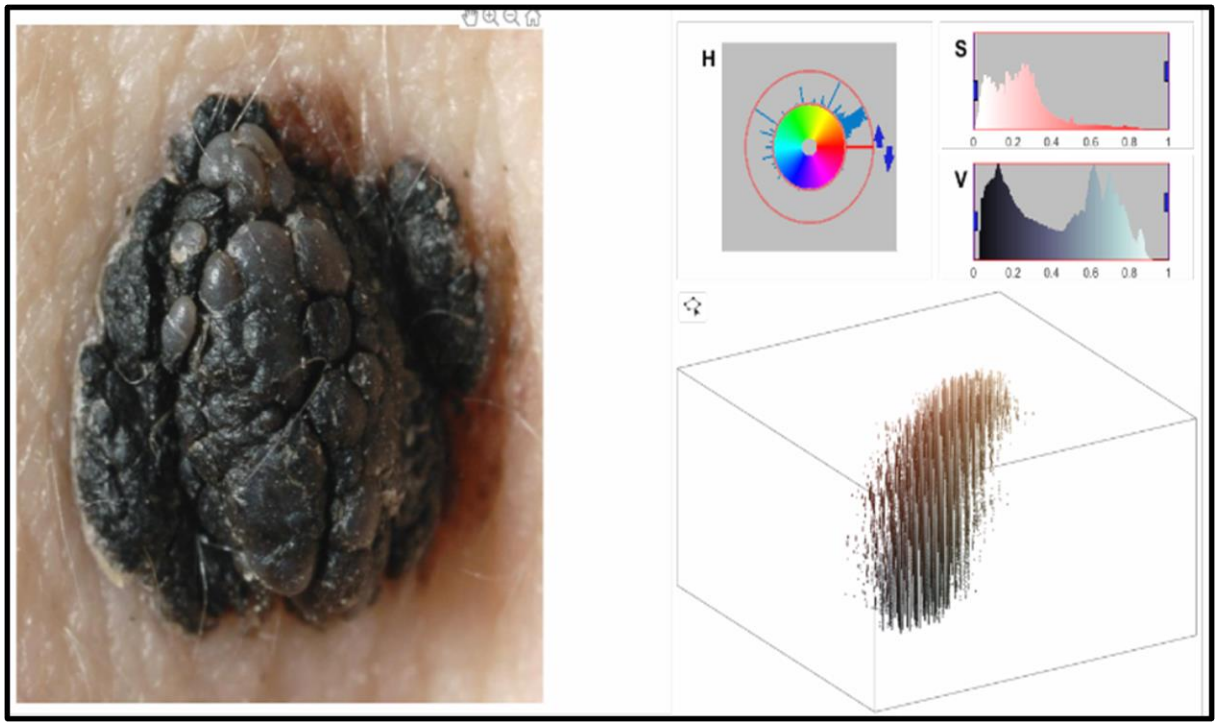

Fig.11HSV Representation of Lesion in 3-D Space

Algorithm : ExtFvAMDS

Input: Lesion Images I_1, I_2,... I_n

Result: Feature Vector (FV) Space

1. Set ImageWidth=500 pixels, SD(σ)=1.1, Filtersize=[7,7]
2. **For** I= 1 to n**do**
3. NewImg= ImgResize(I_i , 500); // **Proportional Scale of Each Input Image**
4. DeNoiseImg=GaussainBlur(NewImg); // **Noise Reduction**
5. FV(I_i)$_{HSV}$=FeatureExtractorHSV(I_i);

6. $FV(I_i)_{Denoised\_HSV}$=FeatureExtractorHSV($DeNoiseImg(I)_i$);
7. $FV(I_i)_{LBP}$=FeatureExtractorLBP(I_i);
8. $FV(I_i)_{Denoised\_LBP}$ =FeatureExtractorLBP($DeNoiseImg(I)_i$);
9. $ExtendedFV(I_i) = FV(I_i)_{HSV} + FV(I_i)_{LBP}$
10. $ExtendedFV_{Denoised}(I_i) = FV(I_i)_{Denoised\_HSV} + FV(I_i)_{Denoised\_LBP}$
11. **End**

LBP(Local Binary Pattern) [9,10] is used to extract the texture features of the input image. The lesion image is divided into 8 x 8 cells. Each cell's pixel is compared with its neighbour. A histogram for each cell is calculated. Then all cells are combined for normalization.

Firstly, we extract features from HSV and LBP. Then we Combine the features of HSV and LBP to extend the feature vector space. HSV unprocessed features are used with proposed Model 1. HSV features with Gaussian Preprocessing are used with the proposed Model 2. LBP unprocessed features are used with the proposed Model 3. LBP features with Gaussian Preprocessing are used with the proposed Model 4. HSV+LBP extended unprocessed features are used with the proposed Model 5. HSV+LBP extended features with Gaussian Preprocessing are used with the proposed Model 6.

Stage 3 - **Classification**

To classify the lesion as melanoma or benign our study has used an ensemble bagged tree classifier.According to [11] ensemble,the bagged tree performs well for imbalanced datasets. The classifier is selected based on the characteristic that the ensemblemeta estimator fits the base classifier on a random subset of the Original dataset. After that, the results are obtained by aggregating the predictions made from the individual classifier.So the classifier provides robust and accurate approach for the classification of the lesion as benign or melanoma.

Stage 4 - Analysis of Result

To compare the effectiveness of each combination (preprocessing, feature extraction, and classification) in the proposed system, we use the following metrics:

Accuracy is measured as the ability to differentiate the Melanoma and benign cases correctly, defined as:

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad \text{.....(1)}$$

Sensitivity is measured as the ability to determine the Melanoma cases correctly, defined as:

$$\text{Sensitivity} = \frac{(\text{T P})}{(\text{T P} + \text{F N})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2)$$

Specificity is measured as the ability to determine the benign cases correctly, defined as:

$$\text{Specificity} = \frac{(\text{T N})}{(\text{T N} + \text{F P})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

In these formulas, TP (true positive) represents the number of cases correctly identified as Melanoma; FP (false positive) represents the number of cases incorrectly identified as Melanoma; TN (true negative) represents the number of cases correctly identified as benign, and FN (false negative) represents the number of cases incorrectly identified as benign.

AUC (Area Under Curve) is a higher level metric, combining the true positive rate (T P R, same as sensitivity) and false positive rate (F P R =F P /(F P +T N)) , indicating how well a classification system distinguishes between the positive classand the negative class. When using only specificity and sensitivity separately, it is difficult to determine whether or not a classification method is overfitted with positive samples (high Sensitivity) or overfit with negative samples (high specificity). Therefore, AUC is proposed in many classification systems as the ultimate metric to measure their effectiveness, both for positive class and negative class. AUC is calculated as the area below a Receiver Operating Characteristic curve, composed of different combinations of TPR and FPR when the classification threshold varies. A higher AUC value denotes theclassification method is closer to a perfect prediction system

III. Results and Discussion

The experiments are carried out on Intel(R) Core(TM) i7-10870H CPU @ 2.20GHz GeForce RTX 3060 Processor using 16.0 GB RAM, The results obtained from the proposed system are listed below inTable 1.

Table 1 Comparative Analysis of Proposed Models in terms of Individual and Extended Feature Vector Space

Proposed Model	**Pre-processing Technique**	**Feature Space**	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**
Model 1	Unprocessed	HSV	77.6	78.70	75.80	87
Model 2	Gaussian Blur	HSV	94.10	94.11	94.11	99
Model 3	Unprocessed	LBP	70.6	71.55	68.51	70

Model 4	Gaussian Blur	LBP	81.8	81.65	81.96	90
Model 5	Unprocessed	HSV+LBP	80.6	80.73	80.32	85
Model 6	**Gaussian Blur**	**HSV+LBP**	**95.3**	**94.23**	**96.96**	**99**

Model 1 has provided 77.6% Accuracy with an AUC value of 87, Sensitivity of 78.70 %, and Specificity of 75.80% shown in Fig.12.

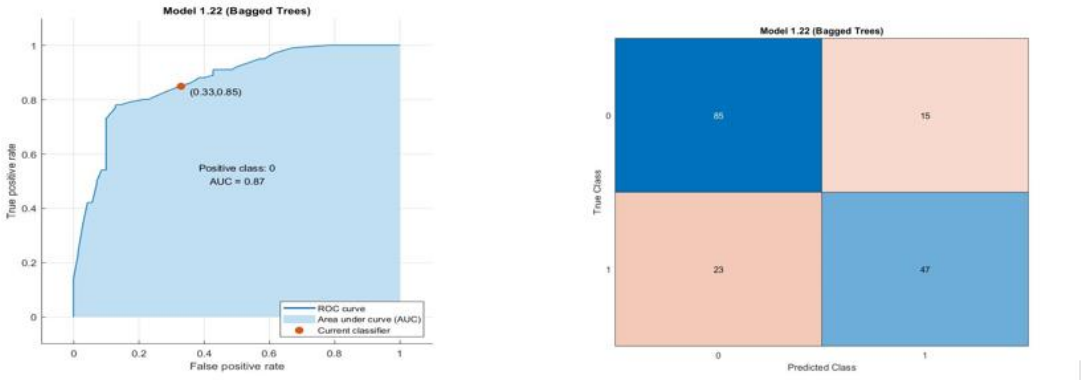

Fig.12 AUC and Confusion Matrix for Proposed Model 1

Model 2 has provided 94.10% Accuracy with an AUC value 99, Sensitivity of 94.11 %, and Specificity of 94.11%.shown in Fig.13.

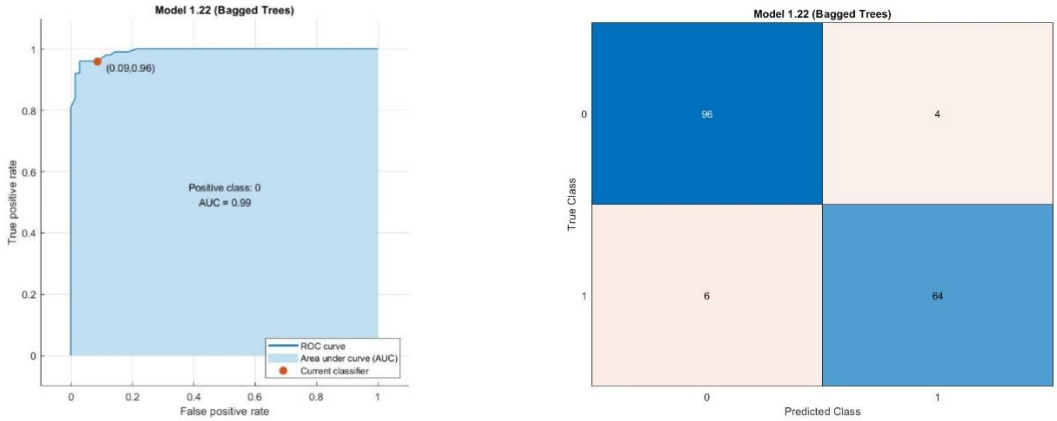

Fig.13 AUC and Confusion Matrix for Proposed Model 2

Model 3 has provided 70.6% Accuracy with an AUC value of 70, Sensitivity of 71.55 %, and Specificity of 68.51% shown in Fig.14.

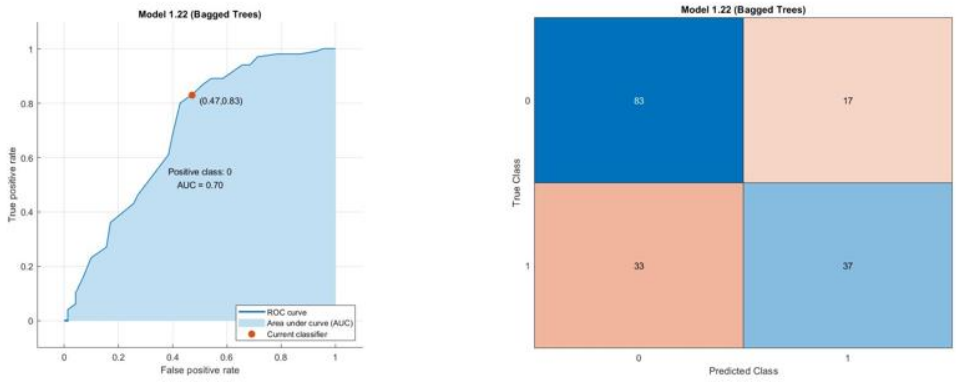

Fig.14 AUC and Confusion Matrix for Proposed Model 3

Model 4 has provided 81.8% Accuracy with an AUC value of 90, Sensitivity of 81.65%, and Specificity of 81.96 % shown in Fig.15.

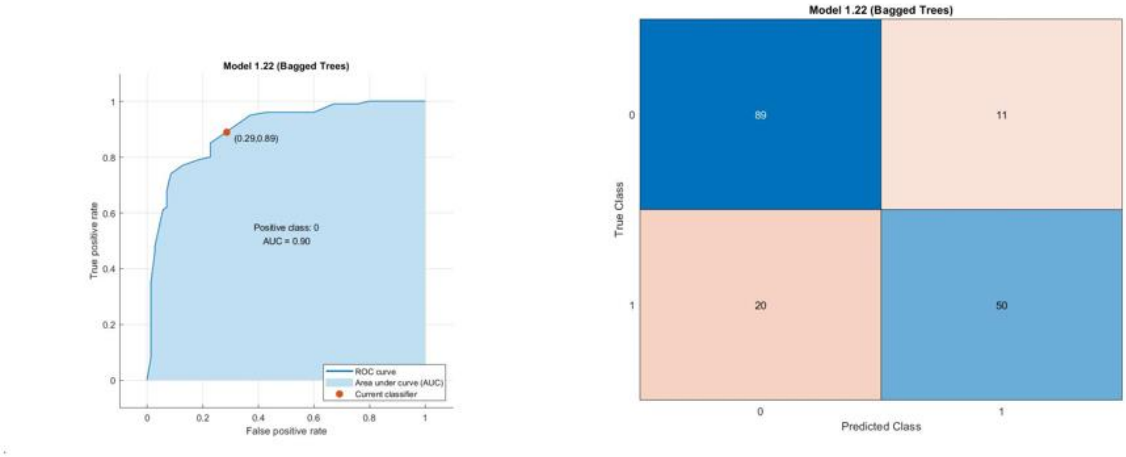

Fig.15AUC and Confusion Matrix for Proposed Model 4

Model 5 has provided **80.6**% Accuracy with an AUC value of 85, Sensitivity of 80.73 %, and Specificity of 80.32% shown in Fig.16.

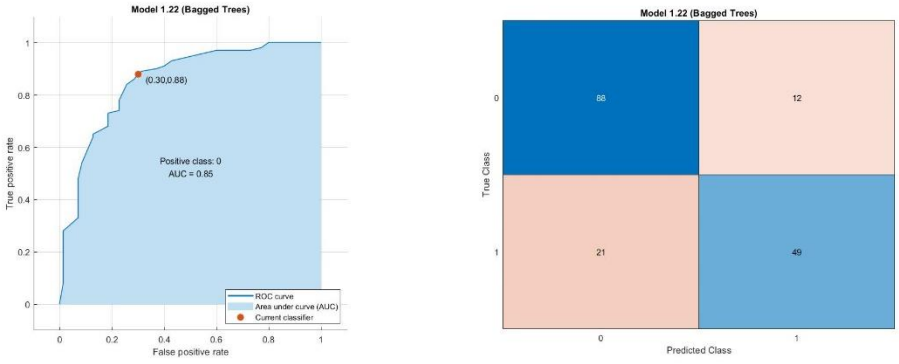

Fig.16 AUC and Confusion Matrix for Proposed Model 5

Model 6 has provided **95.3**% Accuracy with an AUC value of 99, Sensitivity of 94.23 %, and Specificity of 96.96% shown in Fig.17.

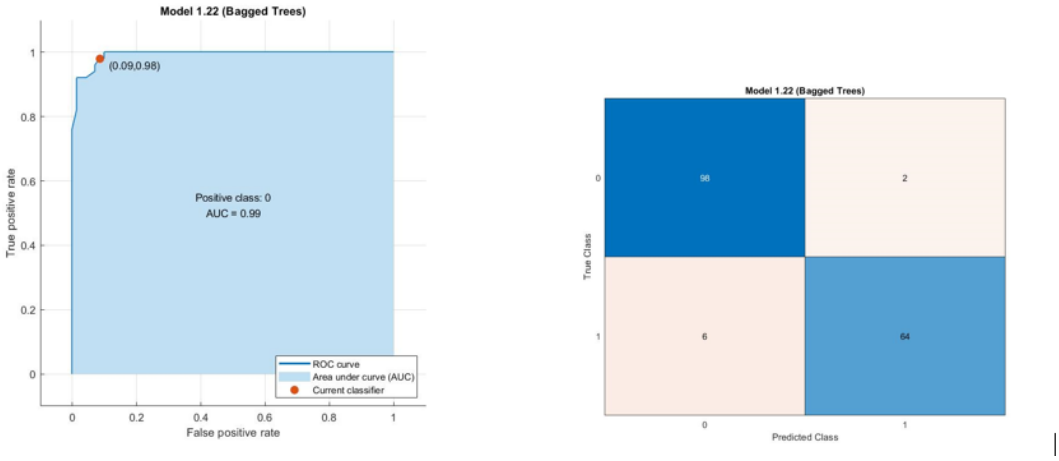

Fig.17 AUC and Confusion Matrix for Proposed Model 6

IV. Conclusion

The proposed study performs an in-depth evaluation of the lesion classification using six different proposed models. The proposed system includes preprocessing feature extraction and classification techniques, and result evaluation for Melanoma detection. Experimentation is done with 6 different models out of which Model 1,3 and 5 skips the preprocessing step, while Model 2,4, and 6 Considers the pre-processing step.Two feature extraction techniques namely HSV and LBP are used with an ensemble bagged tree classifier. Med-Node publicly available data set is used for the study. From the experimental setup, it was observed that Using the Gaussian blur preprocessing technique, using extended feature vector technique, and using ensemble bagged tree classifier the best results are obtained on the benchmark data sets using proposed model 6. The proposed system can be further used by the researchers as a future scope by continuing with more modern approaches for different stages of automatic melanoma detection systems, including CNN and DNN.

References

[1] International Agency for Research on Cancer. https://gco.iarc.fr/today/online-analysis-treemap

[2] Wang, Shuihua, et al. "Advances in data preprocessing for biomedical data fusion: an overview of the methods, challenges, and prospects." Information Fusion 76 (2021): 376-421.

[3] Joseph, Seena, and Oludayo O. Olugbara. "Preprocessing effects on performance of skin lesion saliency segmentation." Diagnostics 12.2 (2022): 344.

[4] Alizadeh, Seyed Mohammad, and Ali Mahloojifar. "Automatic skin cancer detection in dermoscopy images by combining convolutional neural networks and texture features." International Journal of Imaging Systems and Technology 31.2 (2021): 695-707.

[5] Rahman, MdMahbubur, et al. "Hybrid Feature Fusion and Machine Learning Approaches for Melanoma Skin Cancer Detection." (2022).

[6] D. Gutman, N. C. F. Codella, M. E. Celebi, B. Helba, M. A.Marchetti, N. K. Mishra, and A. Halpern, "Skin lesion analysis towardmelanoma detection: A challenge at the international symposium onbiomedical imaging (ISBI) 2016, hosted by the international skinimaging collaboration (ISIC)," CoRR, vol. abs/1605.01397, (2016).[Online]. Available: http://arxiv.org/abs/1605.01397

[7] Giotis I., N. Molders, S. Land, M. Biehl, M.F. Jonkman, N. Petkov "MED-NODE : A computer-assisted melanoma diagnosis system using non-dermoscopic images." Expert Systems With Applications, 42 (19) (2015), pp. 6578-6585doi: 10.1016/j.eswa.2015.04.034

[8] J.-q. Ma, "Content-based image retrieval with HSV color space and texturefeatures," in Web Information Systems and Mining, 2009. WISM 2009.International Conference on. IEEE, (2009), pp. 61–63.

[9] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binarypattern operator for texture classification," IEEE transactions on imageprocessing, vol. 19, no. 6, pp. 1657–1663, (2010).

[10] Abdulsattar, Fatimah Shamsulddin. "On the Effectiveness of Using Wavelet-based LBP Features for Melanoma Recognition in Dermoscopic Images." 2021 International Conference on Information Technology (ICIT). IEEE, 2021.

[11] Galar, Mikel, et al. "A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 42.4 (2011): 463-484.

[12] I. Jolliffe, Principal component analysis. Springer, 2011.

[13] He, Sihui, Zheyang Huang, and XinjieZhong. "Melanoma Classification Using Feature Extraction Methods and Machine Learning Approaches." (2022).

[14] Das, Kinnor, et al. "Machine learning and its application in skin Cancer." International Journal of Environmental Research and Public Health 18.24 (2021): 13409.

[15] Kumar, S., Kumar, A. (2022). Lesion Preprocessing Techniques in Automatic Melanoma Detection System—A Comparative Study. In: Smys, S., Bestak, R., Palanisamy, R., Kotuliak, I. (eds) Computer Networks and Inventive Communication Technologies . Lecture Notes on Data Engineering and Communications Technologies, vol 75. Springer, Singapore. https://doi.org/10.1007/978-981-16-3728-5_27

[16] Kumar, S., Kumar, A. (2020). Automatic Melanoma Detection System (AMDS): A State-of-the-Art Review. In: Gupta, S., Sarvaiya, J. (eds) Emerging Technology Trends in Electronics, Communication and Networking. ET2ECN 2020. Communications in Computer and Information Science, vol 1214. Springer, Singapore. https://doi.org/10.1007/978-981-15-7219-7_17