Haar Wavelet Pyramid-Based Melanoma Skin Cancer Identification With Ensemble of Machine Learning Algorithms

Sudeep D. Thepade, Pimpri Chinchwad College of Engineering, SPPU, Pune, India
https://orcid.org/0000-0001-7809-4148
Gaurav Ramnani, Pimpri Chinchwad College of Engineering, SPPU, Pune, India

ABSTRACT
Melanoma is a mortal type of skin cancer. Early detection of melanoma significantly improves the patient’s chances of survival. Detection of melanoma at an early juncture demands expert doctors. The scarcity of such expert doctors is a major issue with healthcare systems globally. Computer-assisted diagnostics may prove helpful in this case. This paper proposes a health informatics system for melanoma identification using machine learning with dermoscopy skin images. In the proposed method, the features of dermoscopy skin images are extracted using the Haar wavelet pyramid various levels. These features are employed to train machine learning algorithms and ensembles for melanoma identification. The consideration of higher levels of Haar wavelet pyramid helps speed up the identification process. It is observed that the performance gradually improves from the Haar wavelet pyramid level 4x4 to 16x16 and shows marginal improvement further. The ensembles of machine learning algorithms have shown a boost in performance metrics compared to the use of individual machine learning algorithms.

KEYWORDS
Computer-Assisted Diagnostics, Haar Wavelet, Machine Learning, Melanoma Skin Cancer

INTRODUCTION
In clinical informatics, the use of image processing techniques has increased steadily with the rising use of computer-aided diagnoses (CAD) technologies such as X-rays, ultrasounds, MRIs, and the like. Today, as these tools proliferate, an increasing amount of imaging data is being generated and used to create readily available and accessible data in the public domain, leading further to an increased focus on generating machine learning (ML) models for medical diagnosis and classification.

Each year, more people are diagnosed with melanoma than all other skin cancers combined. Melanoma, the deadliest type of skin cancer responsible for the vast number of deaths, is the third most common cancer among people aged between 20 and 39 (American Cancer Society, 2020). Based on the World Health Organization (WHO) statistics, around 132,000 fresh cases of melanoma are diagnosed annually across the globe (American Cancer Society, 2020). In the US alone, the annual cost for the treatment of melanoma has been estimated to be about 3.3 billion USD. (Guy et al., 2015)
For patients with early-stage melanoma detection, the estimated 5-year survival rate is about 98% (American Cancer Society, 2020); thus, prompt detection of melanoma is vital in saving human lives. In conventional diagnosis, a medical expert manually inspects the dermoscopy images. Not only is this method of diagnosis time consuming, but it is also mostly subjective and dependent on the expert’s interpretation of the imaging system. Another disadvantage is the dire shortage of experienced doctors, which is especially evident in rural areas. To alleviate these challenges, CAD-based diagnostics using advanced image processing techniques and ML algorithms may be used. These techniques can reduce the time required by doctors for analyzing the dermoscopy images while providing a less invasive way of achieving a preliminary diagnosis. These techniques leverage the fact that ML algorithms can be well trained and robust to an increasing amount of available and accessible data to aid the medical experts and reduce the subjectivity commonly associated with traditional melanoma diagnoses.

This work proposes the use of a feature extraction technique based on the Haar wavelet pyramid to derive features from dermoscopy skin images and uses the derived features to train ML algorithms for melanoma identification. The presentation is arranged as follows. After the introduction in Section 1, Section 2 offers the background, encompassing a review of extant literature. Section 3 then discusses the proposed method and Section 4 elucidates the environment in which the experimentation is carried out. Section 5 elaborates on the study results and the relevant interpretations. Finally, Section 6 offers the concluding remarks with insights on study limitation, practical implications, and future works.

**BACKGROUND**

Despite only accounting for 1% of skin tumors globally, malignant melanoma is the cause of 60% of deaths from skin cancer (Khazaei et al., 2019). This is primarily because melanoma has been found to be highly likely to expand to other areas of the human body (American Cancer Society, 2020). While early detection does prove incredibly successful in ensuring high survival rates, this can only be facilitated by the presence of medical professionals, a scarcity of whom is seen across the globe, including countries like the US (Pham, 2019), the UK (Taylor, 2020), India (Hazarika 2013). To overcome these challenges, computer-assisted health informatics can prove to be a promising source of support (Kareh and Thoumy, 2018) (Jain and Singh, 2020).

In health informatics based melanoma detection, the approaches proposed can be broadly classified into three categories as methods that use machine learning-based algorithms with crafted dermoscopy skin image features, deep learning-based methods with skin images, and DNA profiling based methods. (Leachman et al., 2016)

Diagnosis methods based on DNA profiling have been explored by Calapre et al. (2017) and Yeh et al. (2019) among others. These approaches necessitate extensive knowledge of the human DNA as well as the genomic structure in order to devise an acceptable solution. Due to the complexity of data involved in such methods, they also tend to require special computing power to obtain a diagnosis in a reasonable time.

More recently, the creation of deep convolutional neural networks (CNNs) (Yu et al., 2016) (Xie et al., 2016) has allowed for more accurate melanoma identification. While these deep networks prove to be highly accurate in differentiating between benign and malignant melanoma, they have some major drawbacks. Training these networks can only be facilitated by a humongous sized dataset, with balanced data between all classes. Neural network training is also computationally expensive, requiring a lot of time (Chen et al., 2018), with rework often needed regarding the values of weights, network design, and parameter tuning (Cheng et al., 2018).

Comparatively, a simpler and less resource-intensive approach is devising lighter handcrafted features of dermoscopy skin images and deploying them to train machine learning algorithms to identify melanoma. Previous attempts along these lines involve the use of texture analysis features (Juntu et al., 2010), GLCM (Korchiyne et al., 2014), local binary patterns (Shan, 2012), and
Experiments involving the use of different transforms on images for feature extraction have also been attempted across various domains. These include the use of Radon transforms (Jadhav and Holambe, 2009), wavelet transforms (Baby et al., 2017) (Redhouane et al., 2014) and discrete wavelet transforms (Hamad et al., 2016). These attempts show that there is a need of devising optimal size features and efficient feature extraction methods for better performance of the machine learning algorithms, while also ensuring robustness against variations in scaling and illumination.

**PROPOSED METHOD**

Initially, the dermoscopy skin images are fetched from the ISIC dataset (International Skin Imaging Collaboration, 2020) in the RGB color space. These images are then rescaled to a size of 512 x 512 for uniformity. The Haar Wavelet Transform is then utilized to obtain a low-frequency area of this 512x512 sized image. This is done by matrix multiplying each color plane of the image with the Haar matrix and then matrix multiplying the result with the transpose of the Haar matrix. The Haar matrix size is kept equivalent to the size of the image that is being transformed. This results in the image being demarcated into four areas, with the top left containing the low-frequency part of the image.

The inverse Haar Wavelet Transform is implemented on this low-frequency part of the image, in a similar manner to the Haar Wavelet Transform. The pixel values from alternate rows and columns are then selected after the inverse Haar Wavelet Transform is performed to obtain the output of the first level of the Haar Pyramid with size 256x256. This process is then repeated for seven levels of the Haar Pyramid, resulting in output sizes like 128x128, 64x64, 32x32, 16x16, 8x8, and 4x4. The pixel values from the Haar Pyramid sizes ranging from 64x64 to 4x4 are availed to create the feature set acquired from the transformed image. Figure 1 describes the feature extraction from the dermoscopy images in the proposed method.

Once the feature sets are obtained for all the specified levels of the Haar Pyramid, these feature sets are employed to train the machine learning algorithms as shown in Figure 2. The same feature extraction methods are used for the feature extraction process as well.
sets are also used to train the ensembles created using multiple machine learning algorithms. These ensembles are formulated using the formula of majority voting. These algorithms are also tested by making use of the n-fold cross-validation technique, keeping the value of n as 10, to ensure uniform stratification of data. The general procedure of n-fold cross-validation is as given herewith:

- The random shuffling of data samples.
- Stratifying the data into n lots.
- For each group:
  - Consider the group as a testing set.
  - Assume the leftover groups as the training set.
  - Use this training set to fit the model and gauge its performance on the test set.
  - Save the performance score and scrap the model.
- Outline the capability of the model using the obtained evaluation scores.

After completing the training and testing stages, the algorithms are then brought into play to detect the dermoscopy images as malignant or benign.

**Haar Wavelet Transform (Kekre and Thepade, 2010) (Shimpi and Thepade, 2018)**

To obtain a Haar Matrix of size NxN, initially, the basic Haar matrix of size 2x2 is considered as follows:

$$H_{2x2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$ (1)
Let $I_{0.5N \times 0.5N}$ be the identity matrix of size $0.5N \times 0.5N$ and $\otimes$ be the Kronecker product. The Haar Matrix of size $N \times N$ is obtained as given in equation 2:

$$H_{N \times N} = \begin{bmatrix} H_{0.5N \times 0.5N} \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \\ I_{0.5N \times 0.5N} \otimes \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \end{bmatrix}$$  \hspace{1cm} (2)

Let $H$ be the Haar Matrix of size $N \times N$. Let $I$ be the dermoscopy skin image of size $N \times N$. Let $IR$, $IG$, $IB$ be the color planes of the dermoscopy skin image. The transformed color planes of the dermoscopy skin image are obtained as given in equations 3 to 5:

$$tIR = H \ast IR \ast H^t$$  \hspace{1cm} (3)

$$tIG = H \ast IG \ast H^t$$  \hspace{1cm} (4)

$$tIB = H \ast IB \ast H^t$$  \hspace{1cm} (5)

The low-frequency areas from the transformed color planes are taken by suppressing other regions to get the modified transformed color planes ($tIR_m$, $tIG_m$, $tIB_m$) and the inverse Haar Transform is applied as shown in equations 6 to 8:

$$itIR = H^t \ast tIR_m \ast H$$  \hspace{1cm} (6)

$$itIG = H^t \ast tIG_m \ast H$$  \hspace{1cm} (7)

$$itIB = H^t \ast tIB_m \ast H$$  \hspace{1cm} (8)

After applying the inverse Haar Transform, the obtained outputs of all color planes $itIR$, $itIG$, $itIB$ are combined to form the transformed image $tI$ of size $0.5N \times 0.5N$, as shown in equation 9:

$$tI = \begin{bmatrix} itIR & itIG & itIB \end{bmatrix}$$  \hspace{1cm} (9)

The Haar Wavelet level 1 image $tI$ provides the first level of the Haar Pyramid. The above equations 1 to 9 are applied using the Haar Wavelet level 1 image $tI$ to obtain a feature set of seven levels of the Haar Wavelet Pyramid in the proposed method. The Haar wavelet pyramid for sample melanoma image is shown in Figure 3.

The proposed method utilizes a feature set of 7 levels of the Haar Pyramid. All images are resized to 512x512. The first level provides an image of 256x256, the second level provides an image of 128x128, the third level provides an image of 64x64, the fourth level provides an image of 32x32, the fifth level provides an image of 16x16, the sixth level provides an image of 8x8 and the seventh level provides an image of 4x4. Henceforth, level 3 of the Haar Pyramid is referred to as ‘64x64’,
level 4 is referred to as ‘32x32’, level 5 is referred to as ‘16x16’, level 6 is referred to as ‘8x8’ and level 7 is referred as ‘4x4’.

If the original image were to be used as a feature vector for training the machine learning algorithms, the time required for processing would be high, as well as being constrained by the memory limitations of the system. The feature vectors with reduced sizes obtained from the Haar wavelet pyramid play a vital role in alleviating this issue. To reduce the computational requirements, the first and second levels are discarded from the feature set. As the levels are increased from level 3, the feature vector size is consistently reduced in half, thus allowing for faster processing.

Machine Learning Algorithms Explored in Proposed Melanoma Identification Method

The five machine learning algorithms explored in the proposed method of detection of melanoma skin cancer are briefly introduced as follows:

- **Naive Bayes**: Naive Bayes is a cluster of classification algorithms. Bayes’ Theorem forms the basis for this algorithm. The algorithm is called naive since it believes that every feature is of equal importance and there is no dependence between features. The algorithm calculates probabilities for every data point to be in a specific class, where the highest probability is believed as the most likely class.
- **Support Vector Machine (SVM)**: SVM is an algorithm primarily deployed for the purpose of classification. The algorithm aims to create a dividing hyperplane farthest from training observations. It attempts to keep the distance between the hyperplane and observations maximum.
• **Random Tree:** A decision tree where the features are selected randomly from the feature collection is defined as a random tree. The tuples are constantly segregated based on certain criteria obtained from the features in a decision tree. A random tree is an example of a supervised learning algorithm.

• **Random Forest:** This is an aggregation of numerous random decision trees. Every decision tree in the ensemble gives an estimate for the category and the category having the highest count of estimates becomes the model’s resulting decision.

• **Alternating Decision Tree (AD Tree):** These are the generalization of decision trees. They constitute decision nodes and prediction nodes, with the decision nodes stating a predicate condition, and prediction nodes containing a single numerical value. The root and leaf nodes of an ADTree are always prediction nodes. ADTree classifies an instance by traversing every path where the decision nodes are true and adding the values stored in any prediction nodes encountered along that path.

**Ensembling Method**

An ensemble is created by having multiple models working parallel on a single feature set. This is called ensemble learning. In ensemble learning, the models are pitted against each other and selected based on majority voting logic. There are 3 methods of majority voting for choosing the class as follows:

1. **Unanimous Voting:** Where all the considered algorithms give the same prediction.
2. **Simple Majority:** Predicted by at least one more than 50% of the number of algorithms.
3. **Plurality Voting or Majority Voting:** Which receives the highest count of votes, irrespective of that number exceeds half of the number of algorithms.

Majority voting is used in the proposed method to create ensembles of machine learning algorithms.

**EXPERIMENTATION ENVIRONMENT**

The data used for evaluation of the proposed method is provided by the International Skin Imaging Collaboration (ISIC) Archive (https://www.isic-archive.com), which is considered the largest collection of dermoscopy images, whose quality is monitored, available in the public domain. The dataset comprises 2000 images (1000 images each of benign melanoma and malignant melanoma). A few of the image examples from the dataset are given in Figure 4. The actual diagnoses of these images are also supplied by ISIC as metadata. The images are provided in the JPG format.

The experiments are conducted using the Python programming language, with the assistance of the Waikato Environment for Knowledge Analysis (WEKA) tool. The N-fold cross-validation method is utilized for training and testing of the machine learning algorithms, keeping the value of the folds as 10.

Measures like accuracy (AC), sensitivity (SE), and specificity (SP) are calculated to analyze the performance shown by the proposed technique of identifying melanoma skin cancer. These are clarified as follows in equations 10 to 12.

Let \( C_{fp}, C_{tp}, C_{fn}, \) and \( C_{tn} \) respectively denote the tally of false positive, true positive, false negative, and true negative. Then:

\[
\text{Accuracy} = \frac{C_{tp} + C_{tn}}{C_{tp} + C_{fp} + C_{tn} + C_{fn}}
\]  

(10)
Specificity = \frac{C_{tn}}{C_{fp} + C_{tn}} \quad (11)

Sensitivity = \frac{C_{tp}}{C_{tp} + C_{fn}} \quad (12)

Sensitivity signifies the proportion of positives that the algorithm recognizes accurately, specificity is the proportion of negatives which are accurately recognized, and accuracy is the proportion of the sum of positives and negatives accurately recognized.

RESULTS AND DISCUSSION

The experiments are performed using the images in the default color space of RGB. All images are resized to 512x512 to maintain uniformity. The features of the images are extracted using multiple levels of Haar Wavelet Transform to obtain Haar Pyramid levels for each image with sizes as 4x4 (Level 7), 8x8 (Level 6), 16x16 (Level 5), 32x32 (Level 4), and 64x64 (Level 3). Three machine learning algorithms are used during the experiments along with four other ensembles of machine learning algorithms aggregated through majority voting. The final results of the investigations are elaborated in this section.

The performance juxtaposition of the examined machine learning algorithms along with the ensembles of several machine learning algorithms collated together is depicted in Figure 5 with the help of percent accuracy for melanoma skin cancer identification with the dermoscopy images from the dataset. From Figure 5, it can be inferred that there is a steady rise in accuracy as the sizes of the Haar pyramid levels increase from 4x4 (Level 7) to 64x64 (Level 3). Overall, better accuracy is observed with Random Forest and Haar Pyramid level with size 64x64 (Level 3).

Figure 6 describes the evaluation of the machine learning algorithms, individually and many machine learning algorithms considered jointly as ensembles, with the help of percentage sensitivity of identifying melanoma skin cancer from the dermoscopy images present in the ISIC dataset. Overall, the ensemble of SVM-AD Tree-Random Forest gives better sensitivity performance with the 4x4 Haar Pyramid level.
Figure 7 depicts the performance contrast of the regarded machine learning algorithms and their ensembles, created using majority voting, with the help of percentage specificity of identification of melanoma skin cancer in the dermoscopy images of the ISIC dataset. From Figure 7, it is seen that the Haar Pyramid level of size 64x64 (Level 3) usually outperforms the other Haar Pyramid sizes. Overall, better specificity is observed with Random Forest using the Haar Pyramid level of size 32x32 (Level 4).

Figure 8 displays the performance differentiation of the scrutinized machine learning algorithms along with the ensembles of algorithms in advocated feature vector extraction of Haar Pyramid-based melanoma skin cancer detection strategy for the individual level variations with help of percentage sensitivity.

Figure 7 and Figure 8 illustrate the performance of machine learning algorithms and their ensembles for melanoma skin cancer identification using Haar Pyramid-based methods. Figure 7 shows the accuracy of different classifiers at various pyramid levels, with Random Forest performing well at Level 3. Figure 8 compares multiple metrics including accuracy and sensitivity, with Random Forest again showing superior performance at Level 3.
Figure 9 gives the performance comparison, using accuracy, sensitivity, and specificity, of the different classification algorithms and the different levels of the Haar Wavelet Pyramid used in the advocated method for melanoma identification. The Random Forest gives better melanoma identification in all performance metrics alias sensitivity, specificity, and accuracy.

Figure 10 gives a performance comparison, using the average of all three performance metrics used (accuracy, sensitivity, and specificity) of the different classification algorithms and the different levels of the Haar Wavelet Pyramid used in the posited technique of melanoma identification. Here the performance of Haar Wavelet features of size 128x128 is almost similar to those of 64x64 and 32x32.

Figure 11 shows the performance comparison of the different levels of the Haar Wavelet Pyramid with the help of accuracy, sensitivity, specificity, and the mean of all three criteria (accuracy, sensitivity, specificity) of melanoma skin cancer detection for the dataset’s dermoscopy images. It can be inferred from Figure 11 that considering the average of all three metrics with equal importance, the Haar
Pyramid levels of sizes 8x8 (Level 6), 16x16 (Level 5), 32x32 (Level 4), and 64x64 (Level 3) give approximately similar performance, with slight deflections observed as the levels increase.

Table 1 denotes the percentage improvement of the levels used in the Haar Wavelet Pyramid in the proposed method. The mean of the three metrics (accuracy, sensitivity, specificity) is used to evaluate performance and the highest level of 4x4 is considered as the base for evaluating the
performance improvement. There is a considerable increase in performance when going from Haar Pyramid level with size 4x4 to Haar Pyramid level with size 8x8, and the subsequent improvements are comparatively smaller in magnitude, resulting in approximately equal performance.

From the results of experimentation done on images from the ISIC dataset for the proposed melanoma identification method, it can be observed that the computationally lighter melanoma detection with reduced feature vector size is attempted by the proposed method with help of Haar Wavelet pyramid. The validation of results is done using three performance metrics alias accuracy, sensitivity, and specificity.

Though the time of feature vector generation is a little more in the proposed method as it needs to extract the levels of Haar wavelet pyramid for melanoma images is the concern currently, this can be made lighter by using hybridization of transform matrices.

The reduced feature vector size will result in faster identification of melanoma skin cancer. In the future, for the speedy preliminary diagnosis of melanoma in the absence of medical expertise, the proposed method may get incorporated even in mobile-based health informatics applications.

Table 1. Percentage improvement of the different levels used in the Haar Wavelet Pyramid in the proposed method of melanoma identification

<table>
<thead>
<tr>
<th>Haar Pyramid Level</th>
<th>Average of all metrics</th>
<th>Percentage improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x4</td>
<td>72.06266667</td>
<td>0%</td>
</tr>
<tr>
<td>8x8</td>
<td>75.66233333</td>
<td>5%</td>
</tr>
<tr>
<td>16x16</td>
<td>76.595</td>
<td>6.29%</td>
</tr>
<tr>
<td>32x32</td>
<td>76.603</td>
<td>6.3%</td>
</tr>
<tr>
<td>64x64</td>
<td>77.27333333</td>
<td>7.23%</td>
</tr>
</tbody>
</table>

CONCLUSION

In health informatics, computer-assisted disease diagnostic methods are proving to be a great blessing by providing faster diagnosis, leading to higher survival rates. With the rise in the availability of
technology, these techniques are also providing a way of diagnosis for people in rural areas, where medical experts are scarce in number. Many attempts are being made over the past few years to enhance the performance of these computer-assisted diagnostic methods. This paper portrays a novel proposition to feature extraction in dermoscopy images using Haar Wavelet Pyramid levels for better space complexity with optimum performance of melanoma skin cancer identification. Overall, 3 machine learning algorithms with 4 ensembles are deployed during the experimentation of the proposed method. The validation of the proposed method is done with square sized images sized 512x512 of the ISIC dataset. The best performance among these variants is observed in the case of Random Forest with Haar Pyramid size 64x64, followed closely by Random Forest with Haar Pyramid size 32x32. In other cases, the ensembles of machine learning algorithms generally have shown an improvement in performance over the functioning of individual machine learning algorithms. Different levels of the Haar Wavelet Pyramid are also compared for their performance for melanoma skin cancer identification and it is observed that the Haar Pyramid levels 64x64, 32x32, 16x16, and 8x8 give roughly similar performance, with minor deviations as the levels increase. However, extensive improvements are obtained concerning the speed at which the algorithms are trained at higher levels of the Haar Wavelet Pyramid, as the image sizes are reduced.

The proposed method attempts computationally lighter melanoma detection with reduced size of features extracted from dermoscopy skin images using the Haar Wavelet Pyramid. This method can be incorporated in mobile-based health informatics applications for faster preliminary diagnosis of melanoma in the absence of medical expertise. Further, the Haar wavelet can be made even lighter in computations using the hybridization of orthogonal transforms for feature extraction. Though the proposed experimentation considers accuracy, sensitivity, and specificity for evaluation, the performance could be improved further for the superior diagnostic ability of the health informatics system.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
REFERENCES


Sudeep Thepade, PhD., has received Ph.D. Computer Engineering from SVKM’s NMIMS in 2011, M.E. in Computer Engineering from University of Mumbai in 2008 with Distinction, B.E.(Computer) degree from North Maharashtra University with Distinction in 2003. He has about 16 years of experience in teaching and industry. Currently He is Professor, Computer Engineering at PCCoE, Pune. He is member of International Advisory Committee for many International Conferences, acting as reviewer for many referred international journals/transactions including IEEE and IET. He has more than 320 papers in National/International Conferences/Journals to his credit.

Gaurav Ramnani is currently pursuing his B.E. Computer Engineering from Pimpri Chinchwad College of Engineering, affiliated to Savitribai Phule Pune University. His areas of interest are Computer Vision, Artificial Intelligence, Machine Learning and Image Processing. He has completed several projects in these fields. He is currently working under Dr. Sudeep D. Thepade.