Melanoma Detection Using HSV with SVM Classifier and De-duplication Technique to Increase Efficiency

Mayura Patil[✉] and Nilima Dongre

Ramrao Adik Institute of Technology, Nerul, Navi Mumbai, India
mayurapatil.b4u@gmail.com, nilima.dongre@rait.ac.in

Abstract. Around one third of all recorded cancers account worldwide for skin cancer, according to the World Health Organization. Every year in the USA there are over 5 million non-melanoma, although about 13,000 cases of melanoma are reported in the UK and Australia. Over the last few decades, occurrences of skin cancer have also risen by 119%, from 1990 to 91270, from 27,600 in 2018. Melanoma has risen 119 million in nations such as the United Kingdom. Not only has the ozone layer reduced ultraviolet radiation safety but the misuse of the atmosphere and heat and tanning [2] has explained this trend. The medical fraternity has spent enormous time and energy on sensitizing people by awareness initiatives. Human skin cancer is the most dangerous variety, with its effects growing rapidly. Early detection of melanoma in dermoscopic photos is extremely important as they are useful for early diagnosis and treatment of ailment. Computer-aided diagnostic tools may promote the detection of cancer early for dermatologists. In this method pre-processing shall take place by applying a list of filters for removing hair, spots and assorted noises from pictures and the methodology of photographs painting shall then be used to fill unspecified areas. In this paper system uses PH2 dataset for evaluation. Also proposed de-duplication method will help in image preprocessing time which will also help in detection of melanoma. KNN, Naïve Bayes and SVM classifier are used for training and testing purpose also SVM shows the highest accuracy of classifier with de-duplication techniques.

Keywords: Melanoma · SVM · Image processing · Data de duplication

1 Introduction

Around one third of all recorded cancers account worldwide for skin cancer, according to the World Health Organization. Every year in the USA there are over 5 million non-melanoma, although about 13,000 cases of melanoma are reported in the UK and Australia. Over the last few decades, occurrences of skin cancer have also risen by 119%, from 1990 to 91270, from 27,600 in 2018. Melanoma has risen 119 million in nations such as the United Kingdom. Not only has the ozone layer reduced ultraviolet radiation safety but the misuse of the atmosphere and heat and tanning [2] has explained this trend. The medical fraternity has spent time and energy on sensitizing people by

awareness initiatives. Nevertheless, irresponsible behavior may not guarantee safety, as there is also a possibility that the number of people with sunburns will develop skin cancer over their lives. Hence there is need to invest in development of technologies for the early diagnosis of skin cancer. Two more popular methods to obtain color images of skin wounds are the numerous, non-invasive dermatologist techniques. The images under observation can be of two types namely macroscopic or dermoscopic based on the collection system. Macroscopic images or clinical image that are taken by regular or smart phone cameras, while dermoscopic images are collected by means of a special lens system with an oil/gel interface (disposal touch dermoscopy). In this article [9], the images are not just dermoscopic, they allow the visualization of additional colors and patterns that make skin injuries more accurate.

Another problem to address is the principle of photo repetition, which is useful for performance. From a visible angle, however, completely different codes can be observed from images with consistent visual perceptions. Therefore, the successful removal of duplicate photo copies by large data center storages and data clouds would increase the processing usage. The optimization of storage may have a critical sensitivity. Manual process extracted redundant objects. The downside was that the de-duplication method was totally dependent on human intervention. In the immense knowledge, tons of human resources would be necessary and vulnerable in order to produce subjective judgment errors. For backup systems and database schemes, replication has been commonly used to significantly improve space usage. Nevertheless, the standard de-duplication software just removes exactly the same things, but is out of inventory for replicate pictures which nevertheless have completely different codes with a constant visual perception. This paper provides a high-precision duplicate de-duplication method to tackle the higher than problem. Duplicate photos are eliminated as the most plan in the proposed approach.

2   Litreature Survey

During the last decade many research have been undertaken in the area of the detection of melanoma, covering a large array of computer vision and patterns. The most widely used techniques found in literature are the segmentation and classification of images. Techniques for segmentation include assignment to the affected region and the non-affected region of threshold binary values. The technique proposed provides a good outcome, but the exactness is near the border of skin color and color of the affected area overlaps. The technique proposed provides good results. N. The hybrid or mix of functional extraction methods used by Moura et al. [5] were the ABCD rule (A is for asymmetry, B is for borderline, C is for color and D is for the diameter of affected area or diameter of mole) and textural characteristics. A classification system with a support vector machine was set and SVM was implemented after a Hybrid feature, which results in a performance rate of 75%. In various classification schemes proposed by many authors, ABCD rule features key role. Shape, colour, edge and texture characteristics are mostly used. The majority of works found in the literature are also based on computer vision and techniques of segmentation. Whereas pattern recognition algorithms are used for the detection of skin lesions due to melanoma with different characteristics. In contrast to the proposed methods, this method is focused on the extraction and use of SVM

classifier for classifying Melanoma Images from all the dermoscopic images of new statistical colors and texture characteristics.

In its planned model, Mogensen, Jemec [3] uses background neural networks. The methodology envisaged, however, has some disadvantages. Slow convergence rates and native minimum trapping are the major disadvantages of this method.

Whereas Rehman, Mobeen [6] has projected that CNN does not need any additional classifier, KNN is used for classification model training since three fully connected layers have been used. Specific advantages of such a classification are its own, as a back propagation algorithm can be used to adjust neuron parameters entirely in layers so as to achieve higher classification patterns. In the approach proposed for TS also Pennisi [7] is very appropriate once benign lesions are handled, whereas when malignant melanoma pictures are divided the detection accuracy significantly decreases. Moreover, this rule is extremely sensitive to images which contain irregular boundaries, a variety of recalls and a range structure, and therefore has a space for a lesion that is less than that of the specific area.

Jacob and Rekha [18] explored double recognition algorithms to identify objects in a series of visually duplicated images. In many applications for large image collections it is important to find visually identical objects. For each image, the Kbits hash code is calculated first, i.e. each picture is converted to a k-bit hash code based on its contents and then only the hash codes can be used to detect the double image.

Prathilothamai and Nair [14] provided a method for almost duplicated object detection and recovery. The issue of near-duplicate detection and image recovery is solved by rigorous interest point detection (DoG detector), local layout (PCA-SIFT) and an active similitude analysis (LSH) of high dimensional images. The photograph representation focuses on sections which use local descriptors and offer high-quality matches under various transformations. Sensitive locality hazard used for local descriptors indexing. A limited solution is a technology vulnerability and a potential drawback that the system requires hundreds or thousands of apps at a time that can be inefficient.

Majumdar and Ullah [16] provided an approach to hierarchical object matching centered on an area. The two pictures are given to identify the largest part of Fig. 1 and the match for Fig. 2 with the most similarity of areas (e.g., surfaces, boundary shapes and colors) defined. The authors [15] suggest sorting criteria to classify the latter into two sets: a collection of edges probable to be in the same category and a sets of edges that are highly likely in different groups to be statistically significant. Any particular edge name as l is selected as a background element when the ratio of a particular group edge to no group edge is higher than a certain threshold. This procedure will lead to asymmetric labeling procedure in which the figure is separated from background discrimination. The main elements are considered edges with highly prominent features, and separated from the background elements. We are not uniquely different if there is overlap in it.

Table 1. Comparative analysis

Paper Name	Author	Technique
Combining ABCD Rule, Texture Features and Transfer Learning in Automatic Diagnosis of Melanoma	N. Moura et al. [5]	ABCD rule of dermoscopy is used for extraction of skin lesion. Feature extraction is utilised on the pre-processed image inorder to find – Area (A), Border (B), Colour(C), and D (Diameter) square measure extraction
Diagnosis of non melanoma skin cancer karatirocyte melanoma: A review of diagnosis and its accuracy of non melanoma skin cancer diagnostic tests and technologies	Mogensen M, Jemec GBE [3]	In this project, model is making use of Backpropagation neural networks. The major drawbacks of this methodology is square measure slow convergence rates and trapping in native minima
Convolution Neural Network based identification	Rehman, Mobeen [17]	The proposed system is advantageous as it does not need extra classifiers as SVM, KNN since three full-connected layers are used for training purpose of classification. This classification is feasible to use back-propagation method which can adjusts various parameters of neurons for further classification
Detection using Delaunay Triangulation	A. Pennisi [7]	The proposed approach is handling benign lesions efficiently, whereas the detection of melanoma accuracy considerably decreases once malignant melanoma pictures are divided. This rule is sensitive to images with irregular shapes and borders, multiple layer of pigmentation, and structure and so, it presents a lesion space that's smaller compared to particular space

3   Proposed Method

Early detection in dermoscopic pictures of malignant melanoma is very important and critical as they can be useful for early treatment. Diagnostic code machine assisted can be valuable to support dermatologists in early detection of cancers. Cancer is the most deadly disease in the world today and the detection and diagnosis are an important area for image processing science. An efficient machine-learning method for the identification
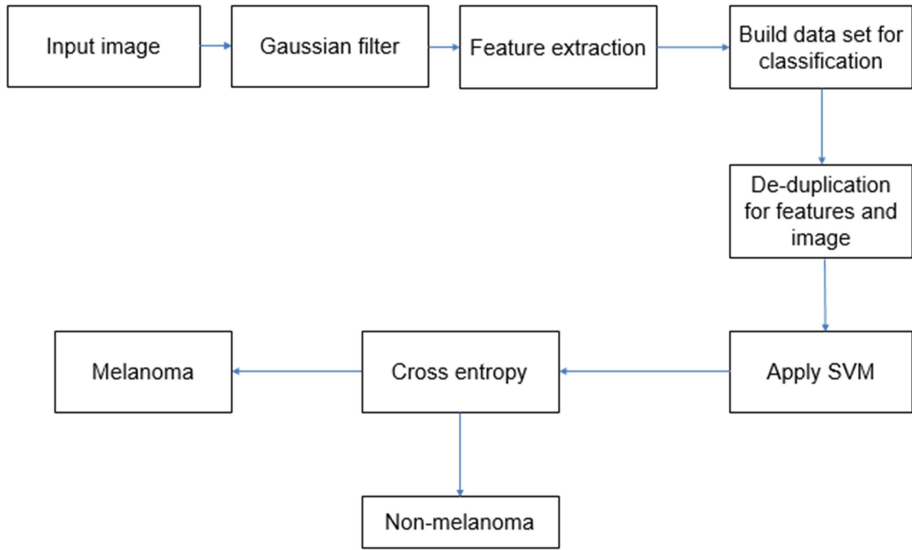
Fig. 1. System Architecture

of dermoscopic objects of melanoma that identifies lesions based on skin melanoma, e.g. different types of color and texture. For the elimination of the same object and function during the train/test phase, software replication definition is used to improve reliability and performance. Several steps are listed below (Fig. 1).

3.1 Pre-processing

It is very important and necessary to rescale the lesion images in order to perform deep learning network. As directly resizing image may distort the shape and size of skin lesion, it requires cropping the center area of lesion image and then needs to proportionally resize the area to a lower resolution.

3.2 Dermoscopic Images

The skin specialist uses optical enlargement and polarized light to magnify the photos and make out the segmented portion. The photos are not easy to identify the segmented part as feathers, nuclei, objects and replacement are part of the photograph. For the purpose of implementation we use ABCDE, the array of 7 points to identify artefacts accurately.

3.3 Gaussian Filter

Gaussian noise is statistical noise, the normal density function probability. Instead of noise patterns, the filter is used to remove noise over edges. The black and white image usually has a natural appearance called the sounds of salt popper.

3.4 Feature Extraction

They have deleted features while keeping the maximum amount of data with large image data. Various methods of extracting color, text and type from an image are used. The efficiency and effectiveness of the current challenges to achieve functionality and recovery. The implementations use gray color matrices to provide a non-redundant texture classification which is immune to the alteration of dynamics and the removal of color. The Gray Level Co-occurrence Matrix (GLCM) is one of the statistical method for extracting textural characteristics from images.

3.5 Gradient Vector Flow

To look in a picture for smoothness. GVF is automatically formatted at low levels. A cycle is placed on a picture with given radius to reach the core of a photo. The snake algorithmic implementation of the process is used:

```
1) Calculate max-min part of a group
2) Take union of 2 set
3) Increase/decrease parts during a set
4) Decrease the key value of an element
```

$$(x + a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k} \tag{1}$$

3.6 Classification

The serial move is to separate the malignant structures from their equivalents after an important stage in an appropriate set of choices. During this step, at least one in each class of cancerous, benign, or healthy is assigned a region of interest to the lesion image. The malignancy level of the tissue (i.e. grading) can be classified as a field of diagnosis. The groups are the potential degrees of cancer of interest in this case. For nomination, a research cluster takes a look at the options. To investigate whether or not a massive difference occurs for several groups at the cost of a minimum of 1 value element. For the photos, however, it is important, for the following reason, to understand the results of the set tests with additional caution. Unit evaluations conclude that the specimens are independent and therefore result in assumptions.

On the other side, the data set consisted of separate cloth footage taken from the same, non-freelancing client, which could contribute to misleading and ambiguous tests. Another research cluster uses algorithms for computer teaching to be informed (from data) by discriminating between different categories.

3.7   Image De-duplication

The client has to calculate the value (hash and feature) of the image I, and upload stands to the server. Then D-phash will 1ist "Duplicate Check" (key part).

```
Phase-I: Duplicate Check
   Input Image;
   1) Read imageI
   imgbgr = inread(I);
   2) Convert the image in greyscale
   Imgray = cvtColor(imgbgr);
   3) Resize the image
   imgdst = resize(imgray);
   4) Compute the DCT matrix F
   F = dct(imgdsr);
   5) Select low frequency DCT matrix keep the loop 8x8ofF
   6) Compute the mean value of F
   mean = (ΣF(i,j) − F(0,0))/63 Where0 ≤ i≤7,0 ≤ j≤7
   F (0,0)is DCT Coefficient
   7) Normalize into Binary form
   8) Construct the Features value p = p(i,j)
   9) Return P;
```

Phase-II (Proof of Ownership): Challenge:

```
   1: The server S randomly select an auxiliary image Ia, it
send id of Aa to the client and request the client C to provide
the proof.
   2: The server S reads the auxiliary image Ia and the image I',
where I' is the image saved on the server which will be similar
to the image I.
   3: The server S resize the auxiliary image Ia: size (width
Ia, Height Ia) = size(width Ia,Height Ia).
   4: Let the blending parameter α = 0.5, the server S generates
the blended image I = Blend (I,Iα,α),which means the server S
computes the feature value.
```

Response:

```
   1: After receiving id of the auxiliary image Ia, the client
C creates the auxiliary image Ia, corresponding to the id and
reads the image I.
   2: The client C resize the size of Ia.
   3: Let α = 0.5, the client C generates the blended image,
which means, the client C computes the feature value, then it
send feature value to the server.
```

Mathematical system

S = {I, O, F}
Where,
S = the system for device specific image processing.
I is set of inputs.
O is set of outputs.
F is set of functions

I = {M1, M2}

Where, M1 = Input Images for Test. M2 = Input Images for Training

O = {O1, O2}

Where, O1 = Melanoma detected Images. O2 = Non-Melanoma Images.

F = {F1, F2, F3}
Where,
F1 = rgb (RGB) To HSV
F2 = de-duplication(set(I))
F3 = feature vector ()

The evaluation of system uses metrics such as accuracy, sensitivity, specificity. Here several terms are commonly used to calculate Specificity, Sensitivity and Accuracy. These terms are True positive (TP), True negative (TN), False negative (FP), False positive (FP). These metrics are given mathematically as follows.

Accuracy - ((T N + T P)/(T N + T P + F N + F P))

4   Results and Dataset

We also mentioned a prototype that is applied in a high specification machine using the Help Vector. Intel Core i5 with 4 GB of RAM was the system setup. For project execution, we used the java open source platform and cv repository.

In a public PH2 database are included 437 image studies consisting of 80 standard nevi, eighty atypical nevi and 40 melanomas. Depending on their picture type, the PH2 data set was split into two groups. There were 160 general cases: 152 with healthy photos were detected and 8 melanoma images recorded a whole skin region.

Table 1 for time required for execution has shown below. Which clearly shows the required time for two method which are system with de-duplication and without de-duplication. As de-duplication system neglect duplicate image and features so it requires less time for final execution. Figure 2 graph shows a comparison without the DE reproduction method of the existing system. This finding shows the time needed for the image test. Clearly, we can say that using image processing duplication technology results better.
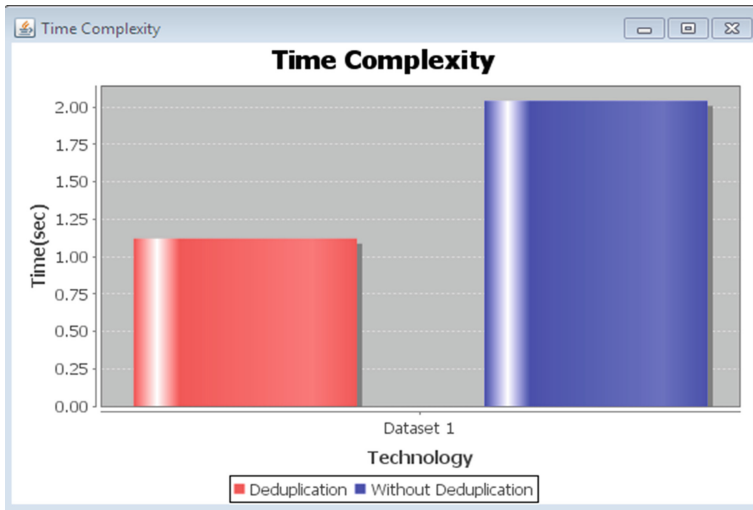
Fig. 2. Time required for execution

Table 2. Time required for execution

Method Used	Required time
De-duplication	1.12 s
Without De-duplication	2.04 s

As system provides better result in respect of accuracy and time by using duplication techniques. So classification also works better in such scenario. System has evaluated three classifier SVM, KNN and Naïve Bayes. So The better result we get through SVM classification. Table 2 contains the accuracy in percent for these three classifier. Also in Fig. 3 shows the graph of the increasing classifiers. The system is assessed by three different KNN, Naive Bayes and SVM classifier classifiers. Between these data sets, SVM gives the best results. Accuracy of classifiers given in table above (Table 4).

Table 3. Classification accuracy

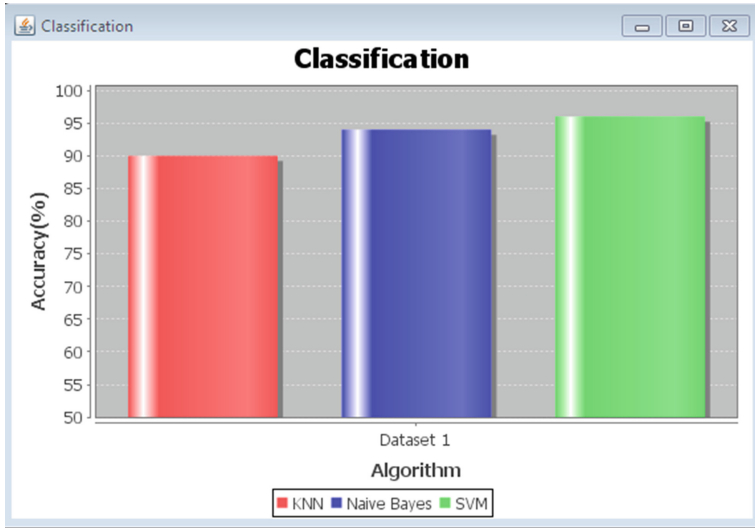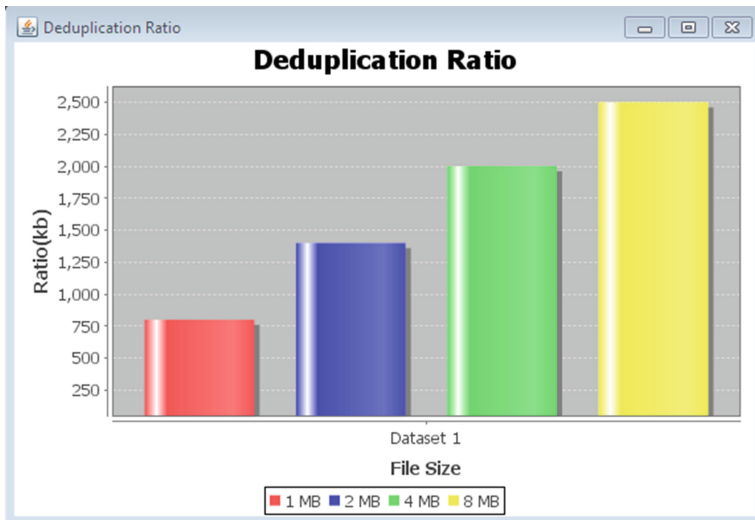| Classifier	Accuracy (in %)
KNN	90
Naïve Bayes	94
SVM	96

Fig. 3. Comparison of classifiers

Fig. 4. Deduplication ratio graph

Third evaluation is checked on de-duplication ratio graph. As Table 3 and Fig. 4 shows ratio graph and changing size of image. After this approach has been used a deduction graph shows that some objects are smaller and have decreased their volume by almost 50% once implementation of techniques. Objects have various sizes in PH2 dataset. Upon implementing replication technique, the table displayed certain objects with their original size.

Table 4. Change in size of image before and after duplication

Before duplication	After duplication
1 MB	750 kb
2 MB	1.4 MB
4 MB	2 MB
8 MB	2.5 MB

5   Conclusion

The system has presented an effective machine-based method for the early detection of melanoma with dermoscopic images based on distinctive effects of skin lesions on melanoma. First, the dermoscopic images will extract new characteristics of color and texture. The vector feature is stored to display all objects. Use of SVM classifier to clinch melanoma images from a set of dermoscopic images of PH2, with the feature vectors stored in the database has worked efficiently. System has evaluated and gave the accuracy of 96% with less time. There is huge scope of further development in this project. This includes detection of various other types of melanoma as we have only focused on visible skin moles, complexity of diagnosis can vary as per the melanoma types, preprocessing time can further be reduced with new techniques, more work can be done towards the security of data.

References

1. Auxilia, L.A.: Accuracy prediction using machine learning techniques for indian patient liver disease. In: 2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, pp. 45–50 (2018)
2. Zhang, W., Gao, L., Tang, Z., Ran, M., Lin, Z.: A benchmark for automatic acral melanoma preliminary screening. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain, pp. 2829–2831 (2018)
3. Mogensen, M., Jemec, G.B.E.: Diagnosis of non melanoma skin cancer karatirocyte melanoma: A review of diagnosis accuracy of non melanoma skin cancer diagnostic tests anud technologies. Dermatol. Surg. **33**, 115874 (2007)
4. Roy, S.S., Haque, A.U., Neubert, J.: Automatic diagnosis of melanoma from dermoscopic image using real-time object detection. In: 2018 52nd Annual Conference on Information Sciences and Systems (CISS), Princeton, NJ, pp. 1–5 (2018)
5. Moura, N., et al.: Combining ABCD rule, texture features and transfer learning in automatic diagnosis of Melanoma. In: 2018 IEEE Symposium on Computers and Communications (ISCC), Natal, pp. 00508–00513 (2018)
6. Rehman, M., et al.: Classification of skin lesion by interference of segmentation and convolotion neural network. In: 2018 2nd International Conference on Engineering Innovation (ICEI). IEEE (2018)
7. Pennisi, A., et al.: Melanoma detection using delaunay triangulation. In: 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), IEEE (2015)

8. Sabbaghi, S., Aldeen, M., Garnavi, R.: A deep bag-of-features model for the classification of melanomas in dermoscopy images. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, pp. 1369–1372 (2016)

9. Adjed, F., Gardezi, S.J.S.: Fusion of structural and textural features formelanoma recognition, March 2018. https://doi.org/10.1049/iet-cvi.2017.0193

10. Dhinagar, N.J., Wilson, M., Celenk, M.: Standardizing pre-processing for digital skin-lesion image analysis. In: 2017 2nd International Conference on Bio-engineering for SmartTechnologies (BioSMART), Paris, pp. 1–4 (2017)

11. Joseph, S., Panicker, J.R.: Skin lesion analysis system for melanoma detection with an effective hair segmentation method. In: 2016 International Conference on Information Science (ICIS), Kochi, pp. 91–96 (2016)

12. Thaiyalnayaki, S., Sasikala, J., Ponraj, R.: Indexing near-duplicate images in web searchusing minhash algorithm. Mater. Today: Proc. **5**(1), 1 (2018)

13. Jacob, C., Rekha, V.R.: Secured and reliable file sharing system with de-duplication using erasure correction code. In: 2017 International Conference on Networks & Advancesin Computational Technologies (NetACT), Thiruvanthapuram, pp. 221–228 (2017)

14. Prathilothamai, M., Nair, P.S.: De-duplication of passports using Aadhaar. In: 2017 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, pp. 1–5 (2017). https://doi.org/10.1109/iccci.2017.8117744

15. Guo, Y., Liu, Y.: CNN-RNN: a large-scale hierarchical image classification framework. Multimedia Tools Appl. **77**(8), 1025110271 (2018)

16. Majumder, S., Ullah, M.A.: Feature extraction from dermoscopy images for an effective diagnosis of melanoma skin cancer. In: 2018 10th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, pp. 185–188 (2018)

17. Rehman, M., et al.: Classification of skin lesion by interference of segmentation and convolution neural network. In: 2nd International Conference on Engineering Innovation (ICEI). IEEE (2018)

18. Jacob, C., Rekha, V.R.: Secured and reliable file sharing system with de-duplication using erasure correction code. In: 2017 International Conference on Networks & Advances in Computational Technologies (NetACT), Thiruvanthapuram, pp. 221–228 (2017)