Skin Cancer Prediction using Enhanced Genetic Algorithm with Extreme Learning Machine

P. Ramya [1], B. Sathiyabhama[2]

[1, 2]Department of Computer Science and Engineering, Sona College of Technology, Salem, Tamil Nadu, India

E-mail: [1]ramyapannerselvam.cse@sonatech.ac.in, [2]sathiyabhama@sonatech.ac.in

Abstract

In the current scenario, the death rate due to the cause of skin cancer is increasing enormously. Diagnosis and prediction of Skin Cancer (SC) have become vital at an earlier stage. The main objective of this research is ensemble machine learning with enhanced genetic algorithm technique to achieve higher accuracy in the prediction of skin cancer at an earlier stage compared to other existing techniques. Although many machine learning and deep learning approaches implemented in detecting skin cancer at an earlier stage still there are few limitations. To overcome these problems in our proposed work, the CNN model, ResNet-16 usually produces successful results in extracting the features automatically and classifying the images very accurately. Therefore, the ResNet model used in our work obtains the deep features with the help of a fully connected layer. Later the feature selection is performed with the help of an Enhanced Genetic Algorithm (EGA) that produces optimized solutions by implementing operations like mutations, crossover, and ensemble with Extreme Learning Machine (EGA-ELM) to classify the images as either melanoma or non-melanoma. The proposed model certainly achieved higher accuracy and effective performance. Finally, the obtained results are to be compared with other popular classifying algorithms like Support Vector Machine (SVM) and various other models.

Keywords: Skin Cancer, Genetic Algorithm, Extreme Learning Machine (ELM), CNN, Optimization

DOI: https://doi.org/10.36548/jtcsst.2023.1.001
Received: 3.12.2022, received in revised form: 2.01.2023, accepted: 17.02.2023, published: 27.02.2023

1. Introduction

Many cancerous diseases affect the day-to-day life activities of the present population and converts into the infectious disease at a faster rate. Among all cancers available in the world today, skin cancer is one of the most threatening diseases all over the world. The major factors like exposure to sun or ultraviolet radiation, different working hazards, etc. are the cause's skin cancers [18]. The abnormal growth of cells in the human body leads to SC which falls broadly into two categories called malignant or benign. The malignant type of SC is very dangerous since once affected it grows faster, and spreads to other parts of the body, which very soon leads to death. The early prediction and detection of SC help medical practitioners to decrease the mortality rate. One of the conventional techniques in diagnosing SC is a Dermoscopic examination. The SC is determined using the advanced illumination system where the upper layers of the skin lesion can be determined and examined at an early stage. The other technique involved is a histopathological examination where the lump lesions are removed and examined by pathologists in the laboratory and confirm it. Between these two methods, the latter is the most significant method in diagnosing SC. Since it is a time-consuming process of examining in the laboratories and waiting for a couple of weeks for the biopsy test to confirm the result. By then if SC affects the person it starts to proliferate to the other parts of the body faster and turns into a malignant disease.

To overcome all these problems and make a faster diagnosis of SC disease, the CAD system plays a vital role in the medical field in diagnosing diseases at a faster rate using many AI techniques like Machine Learning [7], Deep Learning [10] and Optimization methods [11] as an ensemble technique.

In our proposed, work the deep feature extraction is done with the help of a convolution Neural Network (CNN) combined with a Genetic Algorithm (GA) Extreme Learning Machine (ELM) is used for the Effective diagnosis and classification of SC disease with higher accuracy.

2. Literature Review

As discussed earlier, AI techniques are used in almost all fields in the current scenario, and it outperforms almost all fields. Especially it provides promising results in the medical field with greater accuracy and efficiency with the help of many machine learning and deep learning models. This in turn helps the medical practitioners in diagnosing the disease faster and has a societal impact of saving human life evidently.

Erdal basaran et.al [2] proposed a model for the classification of SC at an earlier stage using the EfficentNetB0 CNN model for automatic feature extraction combined with the KNN algorithm for effective diagnosis. Genetic algorithm (GA) with PSO that provided optimized results and further hybridized the model with SVM classifier for better classification with an accuracy of 89.17% obtained. However, since SVM used parameters tuning done for each iteration.

Ulzii-Orshikh Dorj et.al [3] proposed a model for the detection of skin disease using the deep convolution network. The pre-trained AlextNet a CNN model used for feature extraction and ECOC SVM used as a classifier for the classification of SC disease. Four kinds of skin disease classified squamous cell carcinoma with an accuracy of 95%, actinic keratosis at 94.17%, basal cell carcinoma at 91.8%, and melanoma at 90.74%. Even though four types of skin cancer were detected, the learning process consumes more time since SVM.

Albadr, Musatafa et.al. [4] have proposed a model using the CNN algorithm implemented in four transfer learning techniques like Alex Net, Reset 34, Res Net 50, and VGG 16 for image classification as malignant or Benign. The CNN model was trained using 3700 clinical images and tested over 600 images. Among the four techniques, Google Net provides a reliable result and a better approach to the early detection of skin cancer and treatment [7]. Further, the research can be extended with a multiclass classifier and classify more classes by increasing the convolution neural network layers.

Xiaowei Xue et.al. [5] Proposed an ensemble model GA-ELM that overcomes the problem of ELM overtraining the entire dataset moreover since it has the random determination of parameters at times there might be un optimal parameters generated to achieve the generalization performance and stability. So, the Genetic algorithm is used to provide an optimized solution to the random parameter and achieve higher accuracy.

Guang-Bin Huang et.al [13] proved that ELM is best suited for binary and multiclass classification and regression when compared to traditional LS-SVM and PSVM that is widely used in binary classification applications but when SVM is used for regression and multiclass classification, it would be a time-consuming process with higher computational complexity.

From the above discussion, it is noticeable that in skin cancer detection, the SVM classifier used requires parameter tuning after each iteration, at times it is a consuming process. Moreover, it is best suited as a binary classifier only. In our proposed work we ensemble the Genetic Algorithm with ELM and achieved a good result in the classification of skin cancer. Further, the paper comprises various sections as in section 2. Materials and Methods section 3. Experimental Results, Discussion, and section 4. Conclusion.

3. Proposed Methodology

This section discusses the process and methods carried on for the better classification and detection of skin cancer.

3.1 Extraction of Features

Compared to all other techniques the CNN pre-trained models developed by the google AI team produce promising results in the image classification process. The ResNet-50 Architecture is much faster than AlexNet, and EfficientNetB0. The ResNet-50 used in our proposed work to overcome the Exploding /Vanishing gradient problem more effectively. ResNet -50 architecture is like ResNet-34 the only difference here it is stacked of 3 layers instead of 2 layers. This is the standard feature extraction technique used in many vision applications. In our proposed work, we fed the image to the pre-trained neural network and represented the image in the intermediate layer of the neural network. The features are extracted as a feature vector and for ResNet-50 the vector size is 2048.

3.2 Image Classification using EGA-ELM

The Extreme Learning Machine (ELM) is a popular machine-learning model that focuses on classification and prediction areas successfully. ELM is considered the best alternative for Feedforward Neural networks that fix the issues in gradient-based machine learning algorithms. ELM does not fine-tune all inner parameters as in the case of gradient-based algorithms.

The motivation to focus on ELM is that it requires substantially lesser training time than the backpropagation (BP) algorithm and support vector machine (SVM). Although ELM possesses more hidden neurons it is better than the BP algorithm and SVM, it is sufficient that it tunes only a few parameters randomly. Moreover, the ELM is the best multi-class classifier compared to other models. In our proposed model we chose ELM with Genetic algorithm as a hybrid model to optimize the classification even though the tuning is done in random using ELM.

ELM uses a wide type of feature mapping that includes kernels and hidden nodes. In the proposed model ELM, not only provide a unified solution but also has the classification capability of classifying disjoint regions. The output function of ELM for generalized SLFN is denoted as

$$\mathrm{FL}(\mathrm{X}) = \beta\mathrm{i}\,\mathrm{hi} = \mathrm{h}(\mathrm{x})\,\beta \qquad (1)$$

Where β = [β1 ………… βL] T is the vector output weights between the hidden layer node and the output node. h (x) indicates the output vector and maps the d-dimensional input space to L dimensional feature space. H(x) indicated for feature mapping. In general, the ELM not only reaches out smallest training error but also the smallest norm of output weights.

ELM Objectives

- High Learning accuracy
- Least Human invention
- Fast learning speed at lower costs
- Simple Learning Algorithm
- Approximation of generalization performance
- Generalization capability to manage high-dimensional data.

Many researchers had focused on the random parameter initialization that is relevant to the optimal performance of the ELM classifier.

The Genetic Algorithm is the adaptive heuristic search algorithm that evolves from the evolutionary algorithms. In common genetic algorithms are usually used to generate effective solutions for many optimization and search problems. The main strategy involved in the analogy of genetic algorithms is genetic structure and chromosome behavior. Some criteria are,

i. Individuals compete for resources and mate
ii. The successful individual mate creates more offspring compared to others.
iii. The genes from the fittest reproduce throughout the generation.

The above-mentioned is the process of genetic algorithms and finding the fittest at random. The GAs has some of the specific operators to perform the searching process and determine the fitness score. The three operators are the selection operator, crossover operator, and mutation operator. The selection operator is used to find the best individual with good fitness score for successive generations then later the crossover operator is responsible for the mating among the individuals in random with the help of the selection operator to create a new offspring. Finally, to maintain diversity and avoid premature convergence a random gene is inserted in offspring. The above-mentioned are the operations performed with GAs to find out the optimal solutions for many search problems in random that well suits with ELM to make it more optimized.

Although Genetic algorithm provides optimized solution the convergence would be slow so here in our proposed model we had used Enhanced Genetic algorithm (EGA) that converges faster to fitness group than classical Genetic algorithm. Moreover, the EGA-ELM proves to have a better positioning performance than other algorithms.

The EGA-ELM was used to classify the ISIC dataset into Malignant and benign types of SC. The proposed model utilizes three different functions, which help in optimizing the input values of hidden nodes using mutation, crossover, and selection operations.

Here the input values (extracted features) are represented as x_i and t_i as the expected output. Initially, the extracted features are taken as input x_i and E_i as the expected output. The EGA-ELM ensemble model obtains the input values, defines the hidden nodes in random and characterize them as chromosomes, and is represented as,

C = {w11, w12 … w1n, w21, w22, …. w2n, wL1, wL2…. wLn, b1…. bL}

Where w refers to the weight of input and hidden node varies from 1 to -1, b the bias value and L refers to the total number of hidden nodes, n represents the number of input nodes.

The Fitness function used in our proposed model to enhance the accuracy largely is

$$F(c) = \frac{\sqrt{\sum_i^N || \sum_k^l \rho_k g\ (w_k x_i + b_k) - t_i ||_2^2}}{N} \qquad \text{-----Eq (1)}$$

N represents the training samples, ρ output weight, and t_i expected output.

$$H_\rho = T \qquad \text{----Eq (2)}$$

$$H = \begin{bmatrix} g(w1.x1 + b1) & \cdots & g(w1.x1 + bL) \\ \vdots & \ldots & \vdots \\ g(w1.xN + b1) & \cdots & g(w1.xN + bL) \end{bmatrix}_{N*L} \quad \text{----Eq (3)}$$

In our proposed model, the CNN model ResNet-50 a pre-trained model used to extract the exact features, and later along with the Enhanced Genetic Algorithm (EGA), the Extreme Learning Machine (ELM) is used to optimize the features and classify the dermoscopic images from the ISIC repository as either malignant or benign.

When applying the Enhanced genetic algorithm, the crossover/mutation used as a genetic operator to determine the varying chromosomes from one generation to the next. In our proposed work, we had implemented the uniform crossover technique among different types like single point crossover and two-point crossover available for genetic algorithm. Using the uniform crossover, the gene is selected in random from one of the parent chromosomes.

Parent:

000000000000000000000

111111111111111111111

Children:

100011010100100111101

011100101011011000010

The dermoscopic images were initially trained using the CNN model ResNet-50 and then the extracted 1000 features were obtained from the fully connected layer (FC). The EGA-ELM algorithm is depicted as follows which mainly focuses on generating the g initial population and then calculates the fitness value for each chromosome of the population using the fitness function eq-1. In the next step, the 3 selection criteria are used random, K-tournament, and Roulette methods and perform crossover and mutation operations.

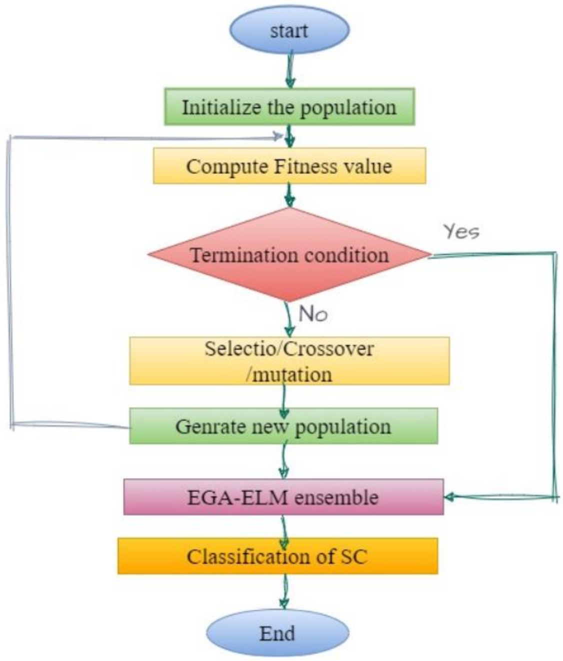

Figure 1. EGA-ELM Workflow Model

Algorithm EGA-ELM

Input

Given Dataset split for Training and Testing, Hidden nodes, total number of iterations total number of neurons for ensemble.

Training Stage

1. Initial generation of population in random

2. Compute the fitness value by eq.1 and preserve the output weights of individuals. Validation data is chosen from training data for each iteration.

3. For k= 1,2……. k

(i) Using Uniform crossover and mutation breed individuals as new population.

(ii) From the existing population choose the population for next generation

(iii) k =k+1

4. End for

5. Based on fitness value select the superior individuals then again sort that population using the norm output weights.

Ensemble Stage

1. For r =1,2…. M
2. Ensemble phases optimize the weights and bias using the ELM function and predict the result.
3. r = r+1

Table 1. Parameters used in our proposed model for the classification process

ELM
Input weights and bias - W
Output weight matrix - Wo
Hidden node numbers
(with increment of 25) - h
Input Nodes - input attributes
Activation function - Sigmoid
EGA
No.of Iternations - 150
Size of population - 70
Crossover and mutation - Arithmetic and uniform

4. Experiments and Results

4.1 Dataset

The dataset implemented for the proposed work of skin cancer detection is the ISIC (International Skin Imaging Collaboration) dataset from kaggle, which facilitates the application of digital skin imaging to help in reducing the mortality rate. The Dataset comprises 3297 color images which are segregated as 1497 as malignant and 1800 as benign and the classification of these images is carried on with the ensembling of a Genetic Algorithm with Extreme Learning Machine (ELM).

The dataset taken from ISIC archive [16] is involved with redundant images and irrelevant images also so the preprocessing of the images also done in order to remove the redundant images and make the dataset best suited for the classification process with EGA-ELM ensemble model to predict the SC as either Malignant or benign.

Figure 2. SC benign and malignant images from the ISIC archive

4.2 Results and Discussion

The ensemble model EGA-ELM process the classification with varying hidden neurons in between the range 100-300 with an increment of 25. So, the total number of iterations was 100 for each experiment. Python was used for implementation with a configuration of 3.20GHz and 16 GB RAM, SSD 1TB (windows 10). The performance of the learning algorithm is evaluated using various metrics like Sensitivity, Specificity, Accuracy, Precision, and F-Score. Along with all these a confusion matrix is calculated based on the error rate to enhance the classification accuracy.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Sensitivity\ (TPR) = \frac{TP}{TP + FN}$$

$$Specificity\ (TNR) = \frac{TN}{FP + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F - Score = 2 \times \frac{Prescision\ \times Sensitivity}{Precision + Sensitivity}$$

Training of the dataset performed by splitting the dataset into 70% of data for training and 30% of the dataset for testing is used. The SC images effectively classified with an accuracy of 89.19%. Here in the work the EGA-ELM implemented and used for detection of skin cancer later the same model would be compared with other relevant techniques to see the effective performance of the model.

Table 2. Compared with other model used for disease prediction

Model	**Accuracy**	**Sensitivity**	**Specificity**	**Precision**
ResNet-50 + EGA+ELM (Proposed model)	**89.19**	85.27	87.21	89.01
EfficientNetB0+GA+ELM (Covid)	**89.17**	85.12	87.11	88.86
SVM+ GA	**87.21**	85.01	86.95	87.10

Performance measures of Selection criteria with different performance metrics in figure 3 and 4.

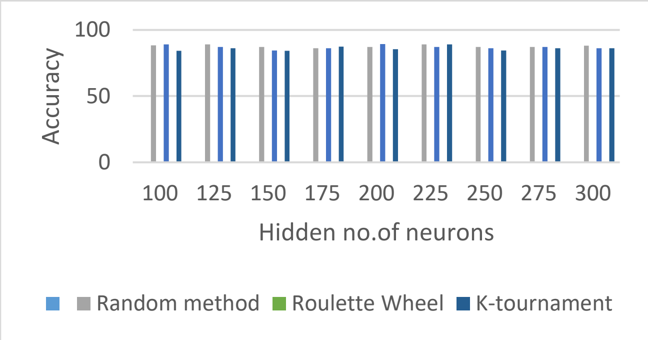

Figure 3. Performance comparison in terms of accuracy

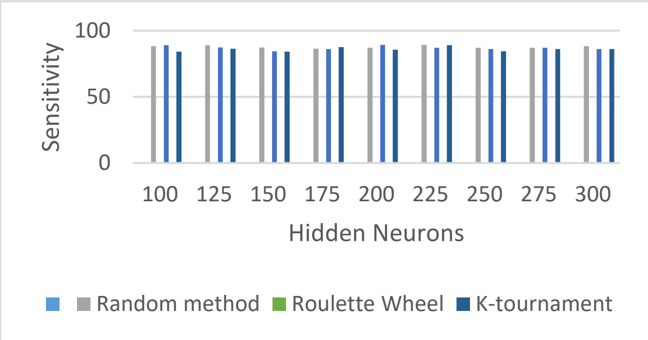

Figure 4. Performance comparison in terms of sensitivity

Skin Cancer (SC) is the most threatening disease in today’s world. The early diagnosis of SC has become more vital. Medical practitioners can diagnose but it would be a time-consuming process so in our proposal, we developed a CAD model to detect skin cancer at an early stage that assists the practitioners. Our model comprises a CNN model ResNet-50 that is used to train the model and extract the features. Later the extracted features are taken as input to the EGA-ELM model to optimize the random selection of parameters and the classification is done at higher accuracy of 89.19%. In addition, this model will be compared with other State-of-art-Techniques.

5. Conclusion and Future Work

This work clearly depicts the detection of Skin Cancer using the ISIC dataset as either malignant or benign. Using the EGA-ELM model the error rate optimized, and the random parameter chosen is optimal for the classification process. Further, the research can be extended with microarray data consisting of gene expression and use Big Data to analyse the huge amount of data and perform the classification with other DL models..

References

[1] Bai, Lixia, et al. "A cooperative genetic algorithm based on extreme learning machine for data classification." Soft Computing (2022): 1-17.

[2] BASARAN, Erdal, and Yuksel ÇELIK. "Skin Cancer Diagnosis Using CNN Features with Genetic Algorithm and Particle Swarm Optimization Methods." (2022).

[3] Dorj, Ulzii-Orshikh, et al. "The skin cancer classification using deep convolutional neural network." Multimedia Tools and Applications 77.8 (2018): 9909-9924.

[4] Albadr, Musatafa Abbas Abbood, et al. "Optimised genetic algorithm-extreme learning machine approach for automatic COVID-19 detection." PloS one 15.12 (2020): e0242899.

[5] Xue, Xiaowei, et al. "Genetic ensemble of an extreme learning machine." Neurocomputing 129 (2014): 175-184.

[6] Nisha Oommachen, Vismi V, Soumya S, Jeena C D. “Melanoma Skin Cancer Detection Based on Skin Lesions Characterization”, IOSR Journal of Engineering (IOSRJEN) e-ISSN: 2250-3021, pISSN: 2278-8719 Vol. 3, Issue 2 (Feb. 2013), V1, PP 52-59.

[7] M. Goyal, T. Knackstedt, S. Yan, and S. Hassanpour, “Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities,” Computers in Biology and Medicine, vol. 127. Elsevier Ltd, p. 104065, Dec. 2020, doi: 10.1016/j.compbiomed.2020.104065.

[8] Murugan, S. A. H. Nair, A. A. P. Preethi, and K. P. S. Kumar, “Diagnosis of skin cancer using machine learning techniques,” Microprocess. Microsyst., vol. 81, p. 103727, Mar. 2021, doi: 10.1016/j.micpro.2020.103727.

[9] X. Du-Harpur et al., “Clinically Relevant Vulnerabilities of Deep Machine Learning Systems for Skin Cancer Diagnosis,” J. Invest. Dermatol., vol. 141, no. 4, pp. 916–920, Apr. 2021, doi: 10.1016/j.jid.2020.07.034.

[10] M. Toğaçar, B. Ergen, and Z. Cömert, “Classification of white blood cells using deep features obtained from Convolutional Neural Network models based on the combination of feature selection methods,” Appl. Soft Comput., vol. 97, p. 106810, 2020, doi: https://doi.org/10.1016/j.asoc.2020.106810.

[11] D. Bi, D. Zhu, F. R. Sheykhahmad, and M. Qiao, “Computer-aided skin cancer diagnosis based on a New meta-heuristic algorithm combined with support vector method,” Biomed. Signal Process. Control, vol. 68, p. 102631, Jul. 2021, doi: 10.1016/j.bspc.2021.102631.

[12] Y. Celik, “An enhanced artificial bee colony algorithm based on fitness weighted search strategy,” https://doi.org/10.1080/00051144.2021.1938477, vol. 62, no. 3, pp. 300–310, 2021, DOI: 10.1080/00051144.2021.1938477.

[13] Huang G.B et.al. (2006) ‘Extreme Learning Machine: theory and applications. Neurocomputing, 70(1), 489-501.

[14] Guang Bin Huang et.al. (2012) ‘Extreme Learning Machine for Regression and Multiclass classification’, IEEE Transactions on Systems Man and Cybernetics-Part-B: Cybernetics vol-42. No.2.

[15] Sasikala et.al. (2020) ‘Towards Improving Skin Cancer Detection Using Transfer Learning’, Biosci.Biotech.Res. Comm. Special Issue Vol 13. no.11 pp 55-60.

[16] G.-B. Huang, Q. Zhu, C. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (1) (2006) 489–501.