# Skin Disease Identification and Classification Optimization Study Using Random Forest Boosted Deep Learning Neural Networks

**A. Kalaivani[1*], S. Karpagavalli[2]**

**Abstract**

Human skin disease is a relatively typical disorder. Detecting skin illness and determining its kind is a difficult challenge in the medical field. Damage and unrepaired skin cells result in genetic flaws or mutations on the skin that can also result in skin cancer. It is essential to identify skin cancer early since it spreads slowly to other parts of the body and is more curable including its early stages. Medical Imaging is concerned with the development of automated methods that assist physicians in their diagnoses. We obtain a clear examination of certain categorization strategies utilized in medical imaging in this research. This paper proposes a unique neural network (Convolutional neural) for exceptional classification models to categorize dermoscopy data using the random forest (RF) classifier. Given the severity of these challenges, researchers have designed several skin cancer early detection tools. It assists patients and dermatologists in determining the kind of disease from a digital image of the afflicted region during the early stages of skin disease. These data show that the suggested method might assist general practitioners in quickly and accurately diagnosing skin disorders, decreasing complications, and fatalities. Artificial intelligence (AI) is quickly expanding in medical disciplines in a modern context. Numerous machine learning (ML) and based on deep neural networks (DL) approaches are used in medical diagnosis. These methods greatly speed up the diagnosis process while simultaneously improving it. The suggested technique outperformed existing methods with over 94.2 percent accuracy using the HAM10000 dataset. The results of the study observed that the architecture produces results that are equivalent to other traditional art models. A system with all the required qualities was successfully created in this study using the recommended predictive deep learning-based classifier model, which showed a considerable gain in high precision of 94.2 percent.



**Key Words:** Artificial intelligence, Dermoscopy, Digital image, Neural network, Random forest, Skin disease

**DOI Number: 10.14704/nq.2022.20.8.NQ44021** **NeuroQuantology 2022; 20(8):197-206**

## 1. Introduction

As helps to protect our inside organs from dangerous germs, air, and radiation, the epidermis perceives the external environment. Several internal and environmental variables can have an impact on the skin. Three-quarters- quarters of all cancer diagnoses worldwide, according to the Department Of Health, are due to skin cancer. When healthy malignant cells develop into a malignant tumor as a result of uncontrolled growth, skin tissue is harmed. It has an impact on every part of the body. It frequently appears on sun-exposed areas like the lips, upper back, forearms, and so on. Only melanoma tumors that are detected early enough can be treated. The person will suffer a painful death as a result as they are distributed to several organ regions. The standard method was a biopsy for identifying skin cancer. Recent developments in the medical field, namely the incorporation of digital technologies, have led to the introduction of a revolutionary diagnosis mechanism called medical image processing [1]. Images are gathered using a specialized tool called a dermatoscopy in order to classify them.

**Corresponding author:** A. Kalaivani
**Address:** [1]Assistant Professor, Department of Computer Science, Nallamuthu Gounder Mahalingam College, Pollachi Coimbatore, Tamilnadu, India, Associate Professor, Department of Computer Science, [2]PSGR Krishnammal College for Women Coimbatore, Tamilnadu, India
E-mail: kalaivanimathsca@gmail.com[1], karpagavalli@psgrkcw.ac. In[2]

Recent advancements in deep learning (DL) have become strong enough to overcome the aforementioned drawbacks. The disadvantage of having engineers hand-craft features may be eliminated as it can also continuously learn abstract qualities from the source data and classify them accurately. As a result, numerous DL techniques, including sparse auto-encoders, multilayer perceptron, and low-resolution filtering, have become more widely used in defect diagnosis. These are some of the best DL types in a neural network.

As a result, diagnosing these ailments has gained significance since, in contrast to professional analysis using testing methods, it can yield precise findings quickly. Skin color, lesion contour, body dimensions dispersion, and the distribution of skin lesions are all visual signs that may be used to diagnose skin syndrome. The complexity of categorization is enhanced when each human cell characteristic is considered separately, and human extraction of characteristics is insufficient for classifying [2]. Supervised learning for disease detection has emerged as a cutting-edge research subject in dermatology. In order to examine the features of skin lesions and the status of imaging techniques, the main discussion objective is to give a general review of skin disease diagnostics utilizing pattern recognition. The training need for these deeper learning-based optimization strategies, however, is a substantial quantity of labeled examples for every group. Dermoscopy appears to be an imaging technique that minimizes reflections from the epidermal layer and also permits detailed, real-time analysis of abnormalities on the inside of the dermal layer. The ability to acquire more sensory information from the layers of the skin at higher resolution and without reflection flaws will allow for the creation of far more accurate computerized disease diagnoses [3].

A voting model is employed by the ensemble way of learning known as random forest (RF) to categorize data. Especially in comparison to some other ML methodologies, the Random Forest method has fewer drawbacks, such as low complexity, quick computation reaction times, great accuracy rates, insensitivity to considerations, no necessity for characteristics regularization, less out-fitting, and other advantages [4], [5]. Importantly, RF is less immune to noise [4]. As a result, when dealing with a large amount of data with decent characteristics, especially in a noisy environment, it is preferable to employ RF. The literature [5], [6] has shown that RF-based sensor problem diagnostics are possible. The use of random forests has a variety of advantages. It has been demonstrated to be on par with other neural network models like boosting and artificial neural networks, with the advantage that random forests are less dependent on the operating parameters and that choosing the operating parameters is straightforward. In contrast to individual decision trees, over-fitting is less of an issue, and thinning the trees would be unnecessary. Finally, the ability to generate accuracy, variable significance, and information about outliers automatically make random forests easier to employ.

Deep learning has drastically changed the pattern recognition landscape during the previous few decades. These are the most sophisticated area of multilayer perceptron research within machine learning. These algorithms were motivated by the structure and operation of the nervous system. Convolutional neural networks are employed in a number of industries, such as computational biology, analytical thinking, and natural language processing [7, 8] [9]. Expert systems had already delivered exceptional results in several areas compared to earlier machine learning approaches. 
Furthermore, deep learning algorithms have been increasingly employed for illness diagnosis in technology. In this paper, we thoroughly investigate and investigate deep learning-based skin cancer screening methods.

The development of medical technology aspects on light and optical data storage will indeed result in a significant improvement in the speed and accuracy of skin condition diagnosis. In addition to this, the expense of such a diagnosis is often high and prone to change. Neural network approaches [10-12] are more efficient than traditional models in classifying visuals and data. There has been a need for exact abnormality detection and illness categorization employing magnetic resonance imaging (Magnetic resonance), optical imaging, and signaling data in the field of healthcare diagnosis.

It will be easier to provide better therapy for patients if the illness category can be accurately identified. Deep neural networks are adaptable to changes in the problem being handled and may resolve critical challenges by automatically detecting input data properties. Using even the most basic computer models, deep learning models will acquire the inferred data to discover and study the characteristics in the unexposed data patterns, producing noticeably high efficiency. With a picture of the afflicted region, the researchers are now considering using a deep learning model to determine the type of skin problem.

As a result, computer-based health diagnosis comes into play, as it may deliver a result in less time and with more accuracy than human analysis using laboratory processes. The most frequently applied technique for skin disease prediction is deep learning. Deep learning models will uncover and investigate characteristics in uncovered data patterns using inferred data, yielding substantial efficiency even for modest computational models. This work demonstrates a reliable technique for reliably diagnosing skin disorders while reducing testing expenses using supervising approaches. Because of this, the researchers are thinking of using a classification model to diagnose common ailments from an image of the affected region.

In this study report, numerous skin disorders were predicted using computational methods and neural networks. The findings demonstrate the efficiency of classification algorithms on a dataset, as well as the kind and importance of the dataset in the discussion. The chapter is divided into Section I, which is an overview, Section II, which is a review of the literature study, Section III, which is the proposed model, Section IV, which gives a detailed explanation of results and discussions, and Section V, which is a ultimate conclusion with performance evaluation, and shows the proposed framework performance metrics through the random forest with neural network with sources of information.

## 2. Literature Survey

Several researchers have proposed strategies for detecting skin problems based on imaging investigations. Go over some of the techniques that have been documented in the literature at this point.

Data analysis is a rapidly expanding field that converts data into meaningful knowledge. This method aids the authorized person in making well-informed choices and making the best results for their benefit [13]. It makes use of concealed patterns discovered in huge databases to assess, anticipate, and guide subsequent behavior. It results in the creation of tools for automatically recognizing past records and building models to forecast occurrences in the future. Multivariate, decision trees, and Gradient boosting [14] are just a few examples of the tools available for data mining machine learning algorithms that may be used to identify and predict different illnesses. Depending on the patient, different test results in various contexts were required for the diagnosis of illness. The number of tests required for data analysis will be minimized if data mining is used. Data analysis needs to increase performance and save resources [15].

For the prediction and diagnosis of various illnesses, a range of classification and classification methods play an important role. Ensemble learning classifiers and stochastic forests classifiers are used to identify diabetes risk [16, 17, 18]. The k-means algorithm's prediction accuracy is improved by employing together class and group approaches and adapting this one to diverse information [19]. To diagnose diabetes risk, a suite of classification methods, omitting the random forest technique, is useful to the records. When the outcomes of each approach are associated, it is clear that Random Forest outperformed the others in terms of accuracy and ROC curve [20, 21].

An approach for building a categorization of an integrated collective framework using decision trees that develop in arbitrarily chosen data feature spaces is termed a random forest [22]. According to experimental findings, stochastic forest learners may classify data in domains with numerous classes with great accuracy [23]. Random forests have recently sparked attention in image classification and bioinformatics [24]. Image classification presents particular problems with the ever-growing amount of image data and information, systems, and archives because the dimensions of image data are quite high, data dispersion, and multi-class labels. Numerous characteristics found in picture data are useless as class features. Relevant traits are far more likely to be overlooked if, throughout this forest-building process, researchers use another tiny subdomain from the high dimensional data at random. [25].

Chang et al. [26] recommended performing five tests concentrating on six important skin illnesses and constructing the best prediction model in dermatology utilizing some integrated neural network categorization procedures. With this study, several dermatitis were anticipated and investigated. The purpose of this analysis is to create a dermatological categorization model based on important clinical and histological characteristics. To choose important features, a well-known features extraction approach called maximum relevance minimum redundancy (m RMR) is applied.

The model is then subjected to a support vector machine (SVM), a common classification tool. The after-effects show that by selecting features, the classification model's accuracy improves noticeably. Some of the best aspects have been chosen as the most important criteria for classifying skin disorders in patients of various ages [24]. Automatic segmentation of dermoscopic pictures, according to Halil Murat Unver et al., and Enes Ayan et al., [27,28], is a perplexing issue due to the existence of various artifacts such as hairs, gel bubbles, and rule markers. Low contrast and various picture size further complicate categorization. As a consequence, the authors created an effective method based on a grab-cut algorithm and a deep neural network. They de-hair the region, find the lesion, segment the relevant portion, and then post-process using morphological operators for further categorization.

Every year, new cases of skin cancer are discovered, with melanoma being the worst among them. Harangi et al., [29] utilizing a deep learning method, identified skin infections. The efficiency of deep learning-based systems, as per the author, has increased dramatically over the past ten years, and their results seem to beat those of conventional image dealing techniques in sorting investigations. The fact that classification algorithms must first categorize hundreds of photographs in the training set before being assessed is one of their biggest limitations. The researcher created a series of deep learning models to help classify dermoscopy images into three categories: seborrheic keratosis, nevus, and melanoma.

Several methods for constructing random forest models from data subspaces have been presented [30]. The bagging technique is some of the most well-known methods for creating individual trees and was first presented by T.K. Ho. It involves randomly selecting a subdomain of characteristics at each node to develop decision tree branches, then combining all individual trees to create a random forests model.

According to reports, image classification [31], detection of photovoltaic arrays [32], detection of online security breaches [33], motion segmentation [34], recognition system [35], classification of species of trees [36], and ship identification on satellite images [37] have all made use of the benefit of CNN and when it comes to feature extraction, CNN outperforms, while random forest classification keeps improving. A unique bearing fault diagnostic strategy established on a neural network with RF was recently developed and investigational test results show that the suggested methodology is more effective at identifying bearing faults under difficult operational conditions than conventional approaches and traditional neural networks [38]. However, until recently, the unique technology combining CNN and RF has been employed very infrequently for gas sensor failure diagnostics.

Provided the above mentioned, it would be advantageous to develop an in-deep learning classical approach that includes components of several approaches, such as separation, data expansion, and integrated learning. To choose the best CNN model arrangement, with due attention all provinces in all models, we suggested using evolutionary approaches.

## 3. Procedure And Resources

In this part, the CNNet-RF model is briefly described. This involved in the main goal is to develop a suitable performance technique for classifying skin lesion images into seven groups. 
This section explains how the given approach for identifying, retrieving, and evaluating images of skin problems works. The method will be incredibly effective in diagnosing seven different types of skin issues. The architecture may be divided into preprocessing, extraction of features, and categorization. On the other hand, extensive neural networks pose significant difficulties in the extraction of medical image features. While DCNN needs huge data of image information to produce additional useful features, medical image data can occasionally be hard to get by whenever a dataset is small. The model will be over-fitted if a little dataset is used for the deep learning step of the process. It could be difficult to train a model that performs well across all domains, and a single network might not be able to extract all of the crucial attributes. We provide a machine learning-based effective framework to collect model findings and improve performance. Figure 1 displays the recommended diagram architecture.

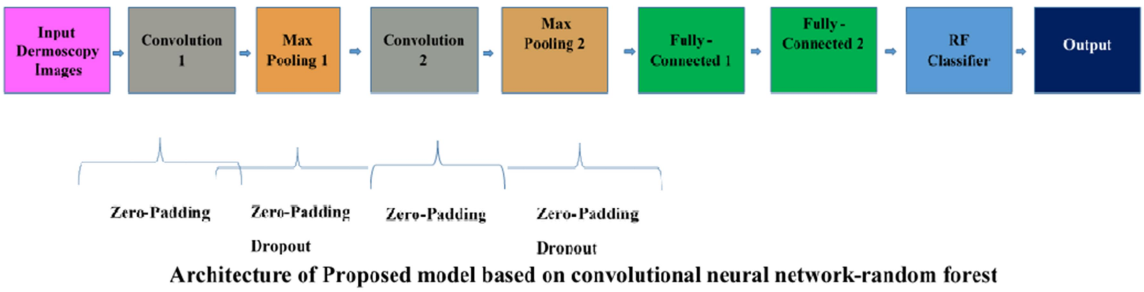


**Figure 1. The Architecture of CNNet-RF Classifier based on CNN**

## 3.1 Image Data Collection for Lesion classification

In this investigation, the data stated below were gathered to evaluate the presented approach. Dermoscopy images image that has been examined by dermatologists make up the vast majority of skin lesion collections. Furthermore, the bulk of research does not use a traditional exploratory technique and only uses a few datasets. The ISIC archive provided the HAM10000 data set, which may be found at https://isic-archive.com/. This includes 10015 skin photos of seven different types and only skin images with 505 lesions that have been verified by pathology. Figure 2 depicts a selection of dermoscopy images from the collection.

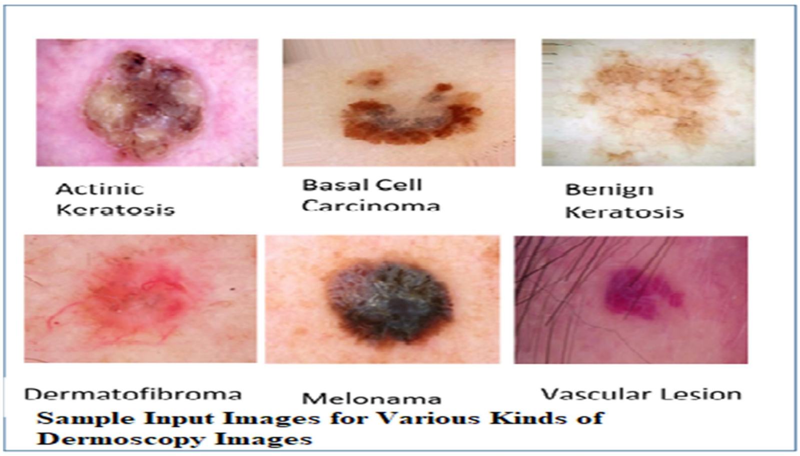


**Figure 2. Sample Dermoscopy Images for Skin Lesions of HAM Dataset**



## 3.2 Data Preprocessing for Skin Lesion Images

To more closely approximate the real distribution of the data, the deep network model requires a lot of training instances, and lack of adequate data may cause over-fitting and other issues that limit the capacity of model for classification. Amongst the most challenging parts of medical image analysis is the low quality of the majority of medical image collections. A common strategy for boosting the volume of training data and enhancing the generalizability of the model is data augmentation. We employ a number of image processing techniques on the entire data, which included a 180-degree translation transformation, turning the photographs horizontally and vertically, and changing the image size and breadth directions by 10%, in order to generate five new instances for each image acquired.

## 3.3 Random forest classifier

In this article, we suggest a fundamentally alternative method for creating the neural network's basic architecture, so that it builds a model with characteristics comparable to those of a randomized decision tree without simulating a particular random forest. Numerous trees that were all growing at random rates make up a random forest multi-way classifiers. The posterior distribution across the pattern classes is used to estimate the names of each massive tree leaf node.

The results of classification are significantly influenced by a classifier decision. Depending on the local base level, gradient descent, and over-fitting issues in the training procedure, it is difficult for the usual CNN, which is constructed on the nonlinear activation, to obtain the greatest generalization capacity. Random forests have a number of benefits over image categorization techniques. In addition to producing auxiliary statistics like classification error and variable relevance, it is non-parametric, equipped to handle both numerical and categorical large datasets, simple to parameterize, resistant to over-fitting, and skilled at managing outliers in training data.

Neural networks and random forests have several features in common. Both can simulate arbitrary decision limits, however, it may be argued that neural networks are slightly more potent in this regard when a lot of data is provided. However, because of its randomized ensemble method, random forests are incredibly resistant to over-fitting, in contrast to neural networks, which are highly susceptible to it. An effect of the numerous parameters employed to build the model is the over-fitting of neural networks. Utilizing particular knowledge about the data domain, techniques like convolutional neural networks substantially minimize the number of parameters to be learned (e.g., images). This clearly shows that the use of a neural network design that dramatically decreases the number of parameters with data will aid in enhancing accuracy.

The relatively homogenous supervised learning group uses random forest as one of its ensemble learning techniques. A random subset of the feature vectors is available to each decision tree or base learner in the random forest. The feature vector so fits the following definition:

$$v = (v1, v2, v3, \ldots. vp) \quad (1)$$

Here p is the available vector dimensions attribute for the training sample. The foremost objective stands to identify the forecast task, denoted as f (v), which calculates the X constraint. The estimate of the function is well-defined as follows:

$$L\,(X\,,f\,(v)) \quad (2)$$

The objective is to reduce the work was intended of the loss, and L is referred known as the great gradient descent. Mean square error loss and zero-one loss are common solutions for applications in the analysis and categorization sectors, respectively. Accordingly, equations (3) and (4) define these two primary functions as follows.

$$L\,(X\,, f(v)) = (X - f\,(v)2) \quad (3)$$

$$L(X\,, f(v)) = \grave{I}\,(X \neq f(v)) = \begin{cases} \text{Zero,} & \text{If } X = f\,(v), \\ \text{One,} & \text{Else.} \end{cases} \quad (4)$$

A group of base learners joins together to establish an integration with the base framework. If followers are considered basic learners,

$$h1\,(v), h2\,(v), h3\,(v) \ldots.., hk\,(v), \quad (5)$$

Equation (6) will serve as the foundation for both the averaging in statistical applications and the voting in classification applications (7),

$$f(v) = \frac{1}{k}\sum_{k=1}^{K} hk(v) \quad (6)$$

$$f(v) = \arg\max \sum_{k=1}^{K} \grave{I}\,(x = hk(v)\,) \quad (7)$$

Using different training datasets and feature adjustments that maximize the output effectiveness and accuracy, the random forest attempts to identify a subset of characteristics. New input is recognized using this set of characteristics.

An ensemble learning method based on decision trees is the random forest approach. Assume that the training dataset has N class and that every tree selects a sub-training set of N training instances at random. If there are M features, select m (m is fewer than M) features, and then select the right feature from all divided. In addition to ensuring that the training results are "integral," this enables each tree to acquire training outcomes based on distinct sub-training sets. The voting results, or the results of numerous feeble classifiers combined to produce a robust classifier, decide the final classification. The suggested framework is suited for huge data sets and input trials with great dimensional properties since it incorporates the RF early in the process.



Consequently, in this learning, RF is used as a classifier. The stability of the model against noise, generalizability, and classification impact are all enhanced and over-fitting is reduced when only a few parameters need to be altered.

## 3.4 Deep CNNet-RF Method for Skin Diseases Classification

When creating integrated multi-leaner models, integration systems are used to boost multi-class problem acquaintance and handle complex multistage difficulties by consuming CNN outline extraction features. In this paper, the CNNet-RF technique is put out. It has several layers: an RF classifier layer, a convolutional layer, a pooling layer, and a fully connected layer. Additionally, it uses dropout and zero-padding techniques to optimize performance and prevent over-fitting. Feature extraction images are analyzed for classification using the CNNet-RF method.

The characteristics are then input into an RF classifier, which has established great presentation in hydrogen sensor fault method arrangement, to get fault diagnosis results. The overall layout of the recommended model based on the CNNet-RF technique described in this study is shown in Figure 1. The CNNet-RF method involves a percentage of data for training and examination samples. After gathering both synthetic and real data, a deep convolutional neural network was trained to classify skin lesions using both the real and synthetic data.

CNN representations include multiple layers to extract characteristics. The network classifier is the fully connected layer, whereas the network extractor mostly consists of convolutional and pooling layers. Convolutional layers use convolutional kernels to execute convolution operations on input images in order to extract information. Kernels take on the features of the entire image by sliding on it like a window. Additionally, the convolution activity of each kernel only pertains to a constrained region known as the receptive field of the input.

The receptive field and weight sharing are key components of a convolution neural network since they can modify the number of training parameters. The pooling operation, which includes the commonly used max pooling and average pooling, is a type of down sampling that aims to shorten training time, extend the receptive field, and prevent over-fitting. In addition, for classification, the fully connected convolution neural network receptive field and weight sharing are essential elements because they can change the number of training parameters. The pooling operation is a sort of down sampling that tries to save training time, widen the approachable field, and avoid over-fitting. It comprises the frequently used max pooling and average pooling. Additionally, the fully connected layer converts the learned feature representation to the label space for categorization. If you need to divide the data into n classes, there are n neurons in the final entirely linked layer.

The network will produce more accurate classification results because the fully connected layer extracts features for each class separately. This is because only a portion of the neural cable network output will be determined by a given input, and the prediction about one class will not be influenced by the predictions of the other six classes. This new network, which employs a random classifier as the classification layer, is used to integrate the outputs of the whole network, as shown in Figure 1. The results are better than the mean of other state-of-the-art techniques. All dermoscopy data are reduced in size to 224 by 224, and a model uses those images as input. The learning rate is initially set at 0.0001 and the Adam approach is utilized as the optimizer. The initial value of our timeframe is 100. To prevent over-fitting, we use an early halting strategy with patience of 10 iterations.

The Deep Convolutional Neural Networks with Random Forest Classifier (CNNet-RF Classifier), which was created by fine-tuning the parameters of the model, was used to test the proposed deep learning model for classification. Computer algorithms called classifiers may place fresh observations into one of several subcategories.

In CNN training, over-fitting is a problem that commonly occurs. Reducing the connection between feature detectors in a neural network helps prevent over-fitting and boosts performance. By ignoring half of the feature detectors in each training batch, the over-fitting problem is significantly reduced in dropout, a deep learning models training method. As a result, during dropout, the weights of half of the hidden layer nodes becomes zero. In this work, over-fitting is successfully avoided by using dropout. We set the preservation probability to $p = 0.5$ during dropout, which suggested that the result for every neuron was zero with a probability of 0.5.



When using a CNN to process image data, the convolutional kernel only performs operations on the bulk of the input image edge pixels once, while pixels in the image center are scanned several times. This lessens the reference degree of border information to some extent. After using zero padding, the new boundary, on the other hand, affects a fraction of the actual processing. This problem can be remedied to some extent. Input photos of different sizes can be complemented simultaneously to make them the same size.

## 4. Experimental Results

The empirical findings demonstrate that for the non-linear issue of dermoscopy image classification for the HAM Dataset, the CNNet-RF model has greater resilience and higher generalization capacity. Mean correctness, accuracy, recall, and f1-score are the performance metrics of the recommended models that are used to assess CNNet-RF model performance. In this part, the CNNet-RF model's performance is assessed using Python 3.6 and the HAM10, 000 databases, with dataset details from every class being arbitrarily selected for preparation. The remaining 30% of HAM images taken in each class are kept for testing purposes. Table 1 compares the identification performance of individual deep learners from the original CNN with that of the suggested ensemble classifier. Figure 3 compares the effectiveness of the traditional network model to the recently proposed technique.

**Table 1. Deep Neural Network through Random Forest Classifier for Dermoscopy Classification: Fine-grained Experimental Results**

| Metric-based Performance Percentage | Traditional CNN-Based Model | CNNet-RF Framework |
|---|---|---|
| Precision | 90. 3 | 94. 12 |
| Re-call | 92. 21 | 95. 24 |
| F1-measure | 91. 28 | 94. 13 |
| Accuracy | 91. 68 | 94. 2 |



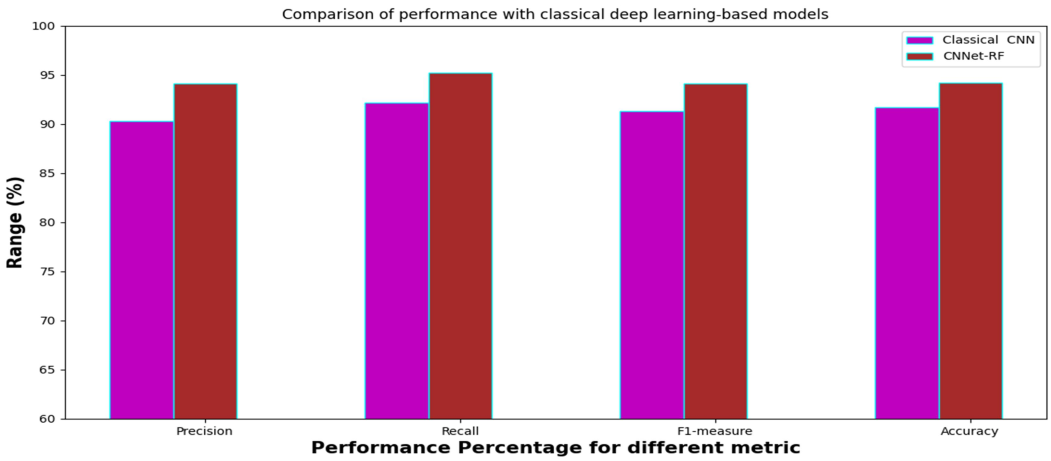


**Figure 3. Classification Performance of Proposed Model with Traditional Model**

Figure 3 shows that, in positions of accuracy, the suggested method achieves improved than the most recent models. The experimental findings show that the recommended model coupled genetically modified CNN with a Random Forest classifier successfully and surpassed current base models in terms of classification accuracy.

## 5. Conclusion

To assist dermatologists in the detection of skin abnormalities, we developed a unique approach in this work called the CNNet-RF classification method. This method accurately classifies the many types of skin problems. Without the assistance of a specialist, the suggested technique may combine the three key components of conventional methodologies into an only learning build for feature extraction, feature selection, and arrangement. The suggested approach assures that the training model structure has greater generalization ability and stability, compared to the classic CNN method, for the classification process on image datasets, by optimizing the structure of a CNN with an RF classifier using dropout and zero-padding. The results of the experiments demonstrate that the suggested technique can learn features well and produce strong HAM detection results. The suggested approach outperformed the CNN by itself and other approaches, with a prediction accuracy of 94.2 percentile on the modes examined. Future developments and additions will be many. A skin injury will first be discovered on the surface of the skin, followed by the process of developing a skin illness in a smartphone system, and lastly, the detection of all skin illnesses in the entire region and their harshness. The investigation shows that even while convolutional neural networks are effective at discriminating between skin conditions, the precision of individual detectors may still be raised by employing a fresh strategy.

## References

Mussarat Yasmin, Muhammad Sharif and Sajjad Mohsin "Survey Paper on Diagnosis of Breast Cancer Using Image Processing Techniques", Research Journal of Recent Sciences, vol:2(10), 88-98, October 2013.

Balazs Harangi. (2018). Skin lesion classification with ensembles of deep convolutional neural networks. Journal of Biomedical Informatics, 86 ,25-32. https://doi.org/10.1016/j.jbi.2018.08.006.

Rashid, H.; Tanveer, M.A.; Aqeel Khan, H. Skin Lesion Classification Using GAN Based Data Augmentation. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 916–919. [CrossRef]

L. Breiman, ''Random forests,'' Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

D. Mohapatra, B. Subudhi, and R. Daniel, ''Real-time sensor fault detection in tokamak using different machine learning algorithms,'' Fusion Eng. Des., vol. 151, Feb. 2020, Art. no. 111401.

Z. Noshad, N. Javaid, T. Saba, Z. Wadud, M. Saleem, M. Alzahrani, and O. Sheta, ''Fault detection in wireless sensor networks through the random forest classifier,'' Sensors, vol. 19, no. 7, p. 1568, 2019.

Bisla, D.; Choromanska, A.; Stein, J.A.; Polsky, D.; Berman, R. Towards Automated Melanoma Detection with Deep Learning: Data Purification and Augmentation. arXiv 2019, arXiv:1902.06061. Available online: http://arxiv.org/abs/1902.06061 (accessed on 10 February 2021).

Farag, A.; Lu, L.; Roth, H.R.; Liu, J.; Turkbey, E.; Summers, R.M. A Bottom-Up Approach for Pancreas Segmentation Using Cascaded Superpixels and (Deep) Image Patch Labeling. IEEE Trans. Image Process. 2017, 26, 386–399. [CrossRef]

Sae-lim, W.; Wettayaprasit, W.; Aiyarak, P. Convolutional neural networks using MobileNet for skin lesion classification. In Proceedings of the 16th International Joint Conference on Computer Science and Software Engineering, Chonburi, Thailand, 10–12 July 2019; pp. 242–247.

Castillo, D.; Lakshminarayanan, V.; Rodríguez-Álvarez, M.J. MR Images, Brain Lesions, and Deep Learning. Appl. Sci. 2021, 11, 1675. [CrossRef]

SivaSai, J.G.; Srinivasu, P.N.; Sindhuri, M.N.; Rohitha, K.; Deepika, S. An Automated Segmentation of Brain MR Image through Fuzzy Recurrent Neural Network. In Bio-Inspired Neurocomputing; Bhoi, A., Mallick, P., Liu, C.M., Balas, V., Eds.; Studies in Computational Intelligence; Springer: Singapore, 2021; Volume 903.

Babu S, Vivek EM, Famina KP, Fida K, AswathiP, Shanid M, Hena M (2017) Heart disease diagnosis using data mining technique. In: International Conference on Electronics, Communication, and Aerospace Technology, ICECA2017.

Sampaul TGA, Robinson YH, Julie EG, Shanmuganathan V, Nam Y, Rho S (2020) Diabetic retinopathy diagnostics from retinal images based on deep convolutional networks. Preprints. https://doi.org/10.20944/preprints202005.0493.v1

Vimal S et al (2020) Deep learning-based decision-making with WoT for smart city development. In: Jain A, Crespo R, Khari M (eds) Smart innovation of web of things, CRC Press, Boca Raton, pp 51–62. https://doi.org/10.1201/9780429298462

Kalaiselvi C, Nasira GM (2014) A new approach of diagnosis of diabetes and prediction of cancer using ANFIS. In: IEEE Computing and Communicating Technologies, 2014, pp 188–190

Kumari M, Vohra R, Arora A (2014) Prediction of diabetes using Bayesian network. Int J Comput Sci Inf Technol (IJCSIT) 5(4):5174–5178

Krishnaiah V, Narsimha G, Chandra NS (2013) Diagnosis of lung cancer prediction system using data mining classifcation techniques. Int J Comput Sci Inf Technol 4(1):39–45

Long NC, Meesad P, Unger H (2015) A highly accurate frefy-based algorithm for heart disease prediction. Expert Syst Appl 42:8221–8231

Esteghamati A, Hafezi-Nejad N, Zandieh A, Sheikhbahaei S, Ebadi M, Nakhjavani M (2014) Homocysteine and metabolic syndrome: from clustering to additional utility in prediction of coronary heart disease. J Cardiol 64:290–296



Lee BJ, Kim JY (2016) Identifcation of type 2 diabetes risk factors using phenotypes consisting of anthropometry and triglycerides based on machine learning. IEEE J Biomed Health Inform 20(1):39–46

L. Breiman, "Random forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001.

R.E. Banfield, L.O. Hall, K.W. Bowyer, and W.P. Kegelmeyer, "A Comparison of Decision Tree Ensemble Creation Techniques," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 1, 2007, pp. 173-180.

A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns" In ICCV, pp. 1-8, 2007.

D. Amaratunga, J. Cabrera, and Y.S. Lee, "Enriched Random Forests," Bioinformatics, vol. 24, no. 18, pp. 2010-2014, 2008.

Chang, Chih-Hao Chen 2009. Applying decision tree and neural network to increase quality of dermatological diagnosis, Expert System with Applications 3, 4035 – 4041.

Xie J, Wang C. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases. Expert Systems with Applications. 2011 May 1;38(5): 5809-15.

Unver HM, Ayan E (2019) Skin lesion segmentation in dermoscopic images with combination of YOLO and GrabCut algorithm. Diagnost J 9:3

Basha AJ, Balaji BS, Poornima A (2020) Support vector machine and simple recurrent network based automatic sleep stagclassificationon of fuzzy kernel. J Ambient Intell Human Comput 2020:1–9.

Harangi B (2018) Skin lesion classification with ensembles of deep convolutional neural networks. J Biomed Informatics 86:25–32

T.K. Ho, "The Random Subspace Method for Constructing Decision Forests," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, no. 8, pp. 832-844, 1998

T. Li, J. Leng, L. Kong, S. Guo, G. Bai, and K. Wang, ''DCNR: Deep cube CNN with random forest for hyperspectral image classification,'' Multimedia Tools Appl., vol. 78, no. 3, pp. 3411–3433, Feb. 2019.

J. M. Malof, L. M. Collins, K. Bradbury, and R. G. Newell, ''A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery,'' in Proc. IEEE Int. Conf. Renew. Energy Res. Appl. (ICRERA), Nov. 2016, pp. 650–654.

E. Min, J. Long, Q. Liu, J. Cui, and W. Chen, ''TR-IDS: Anomaly-based intrusion detection through text-convolutional neural network and random forest,'' Secur. Commun. Netw., vol. 2018, pp. 1–9, Jul. 2018.

S. Bai, ''Growing random forest on deep convolutional neural networks for scene categorization,'' Expert Syst. Appl., vol. 71, pp. 279–287, Apr. 2017.

Y. Liu, X. Yuan, X. Gong, Z. Xie, F. Fang, and Z. Luo, ''Conditional convolution neural network enhanced random forest for facial expression recognition,'' Pattern Recognit., vol. 84, pp. 251–261, Dec. 2018.

U. Knauer, C. S. von Rekowski, M. Stecklina, T. Krokotsch, T. P. Minh, V. Hauffe, D. Kilias, I. Ehrhardt, H. Sagischewski, S. Chmara, and U. Seiffert, ''Tree species classification based on hybrid ensembles of a convolutional neural network (CNN) and random forest classifiers,'' Remote Sens., vol. 11, no. 23, p. 2788, 2019.

E. Anggiratih and A. E. Putra, ''Ship identification on satellite image using convolutional neural network and random forest,'' Indonesian J. Comput. Cybern. Syst., vol. 13, no. 2, pp. 117–126, 2019.

G. Xu, M. Liu, Z. Jiang, D. Soffker, and W. Shen, ''Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning,'' Sensors, vol. 19, no. 5, p. 1088, 2019.