International Journal of Computers and Informatics

Journal Homepage: https://www.ijci.zu.edu.eg

Int. j. Comp. Info. Vol. 4 (2024) 1–19

Paper Type: Original Article

Towards Higher Accuracy in Diagnosis of Skin Cancer: An Adaptive CNN-RF Model for Diagnosis of Skin Cancer based-on Different Oversampling Methods

Salwa Elsayed [1,*], Mahmoud M. Ismail [1], Amal F. Abdel-Gawad [1] and Israa Mohamed [1,2]

[1] Department of Decision Support, Faculty of Computers and Informatics, Zagazig University, Zagazig, 44519, Egypt.
Emails: salwa.elsayed@fci.zu.edu.eg, mmsabe@zu.edu.eg, amgawad2001@yahoo.com, israasalem@zu.edu.eg.
[2] Faculty of Engineering and Computer Sciences, King Salman International University, South Sinai, Egypt.

Received: 12 Mar 2024 **Revised:** 15 Jun 2024 **Accepted:** 12 Jul 2024 **Published:** 15 Jul 2024

Abstract

One of the most hazardous diseases in the world is skin cancer. Convolutional neural networks have recently received further interest for their use in spotting skin malignancies in dermoscopy images. In this paper, a new hybrid model based on CNN and random forest algorithm is proposed for detecting and classifying skin cancer images. The dataset used for this study is based on the HAM10000 dataset. Dataset was first preprocessed and different oversampling methods were applied to overcome class imbalance. Then, features of the preprocessed dataset were extracted using customized CNN. Finally, these features were classified using RF algorithm. The effective hyper-parameters for CNN and RF were detected besides different batch sizes and image sizes were implemented to ensure consistency of the proposed model. The proposed model has the ability to achieve 98.88 % for accuracy, 0.99 for precision, 0.99 for recall and 0.9999 for AUC.

Keywords: Skin Cancer; CNN; Machine Learning; Features Extraction; Skin Diagnosis; SMOTE-ENN; ADASYN.

1 | Introduction

The increased incidence of skin cancer makes it imperative to create a reliable system for automatically classifying skin cancer. The most deadly and rapidly spreading type of skin cancer is melanoma as it more dangerous and aggressive type. The brain, liver, and lungs are just a few of the human organs where it can quickly spread. Melanoma can lead to a wide range of problems, after receiving therapy for skin cancer, some of these problems include lymph node involvement, which causes lymphedema or swelling in your arm or leg. The Surveillance, Epidemiology, and End Results database reveals a rise in the number of cases of melanoma per 100,000 people in the US between 1990 and 2018, from 13.8 in 1990 to 22.6 in 2018 [1]. The American Cancer Society predicts that there will be 97,610 new cases of melanoma in the US in 2023 and that there will be about 7,990 deaths, including 2,570 females and 5,420 males [2]. These statistics demonstrate the need for quick and precise early diagnosis procedures to assist clinicians in identifying skin cancer at an early stage and enhancing the likelihood of a cure before the disease progresses [3]. Convolutional neural networks (CNNs), which make up the majority of deep learning techniques, have recently received a lot of interest for their use in spotting skin malignancies in dermoscopy images. The need for early detection and the increased

Corresponding Author: salwa.elsayed@fci.zu.edu.eg

incidence of skin cancer make it imperative to create a reliable system for automatically classifying skin cancer. Numerous well-known networks, such as GoogleNet [4], VGGNet [5], and ResNet [6], have been successfully applied to the classification of skin cancer since the advent of deep learning. The authors of this study [7], proposed a hybrid method for classifying malignant and benign skin cancer. They used deep features of deep neural networks and the stacked cross-validation algorithm (SCV-DF), and it has a 90.9% accuracy rate. A CNN model was also presented in [8], along with its seven distinct architectures, including ResNet50, VGG16, InceptionV3, VGG19, Xception, and MobileNetV2. The authors discovered that Xception provided an accuracy of about 85.303%. As well, this paper [9] proposed a highly efficient hybrid detection and segmentation method that combined RetinaNet and MaskRCNN. The ISIC-2018 and PH2 datasets were used to train and validate the suggested technique, and the results demonstrated that the proposed method outperformed the alternatives. This paper [10], introduced a system for detecting skin cancer that uses support vector machine (SVM), probabilistic neural networks, random forest (RF), and coupled SVM+RF classifiers, among other classification methods. The outcomes demonstrated that the SVM+RF classifier outperformed other classifiers.

Due to the variety of skin lesions it contains, the HAM10000 dataset is frequently used by researchers. Figure 1 shows a subset of seven classes of the HAM10000 dataset. [11], suggested using adversarial learning and transfer learning to classify skin diseases to increase the generalizability of models to new samples and decrease cross-domain shift with an accuracy (0.909) and AUC (0.967). [12], developed a weight pruning technique for thin neural networks to compensate for accuracy loss, enhance model performance, and increase model dependability in the classification of skin cancer with an accuracy of 0.975. [13], proposed a technique for classifying skin diseases that combines CNN with one-versus-all (OVA) and this method gave an accuracy of 0.929. To get domain-dependent noise vectors, the authors of this paper [14], used a variational autoencoder network. In addition, a student-like distribution was utilized to increase the diversity of the images, and an auxiliary classifier was applied to create images of certain classes with an accuracy of 0.925. Inception-v3, InceptionResNet-v2, ResNet-152, and DenseNet-201 are four brand-new deep CNN models that [15] used to classify eight different forms of skin tumors on the HAM10000 and PH2 datasets. Finally, experimental findings showed that, in terms of ROC AUC score, these CNN models' diagnostic level surpasses that of dermatologists. Pre-trained Mobilenetv2 is used in the act of the back pillar of the DeepLabv3+ model and trained on the best parameters that significantly increase the segmentation of infected skin lesions [16]. DesneNet201's pre-trained feature extraction is used to do the multi-classification of the skin lesions. Three models built on Cubic SVM, Weighted KNN, and Fine KNN were suggested by the authors. Finally, the models attained mean ROC of 0.98 and accuracy of 0.9065, 0.8696, and 0.9201 respectively on the HAM10000 dataset.

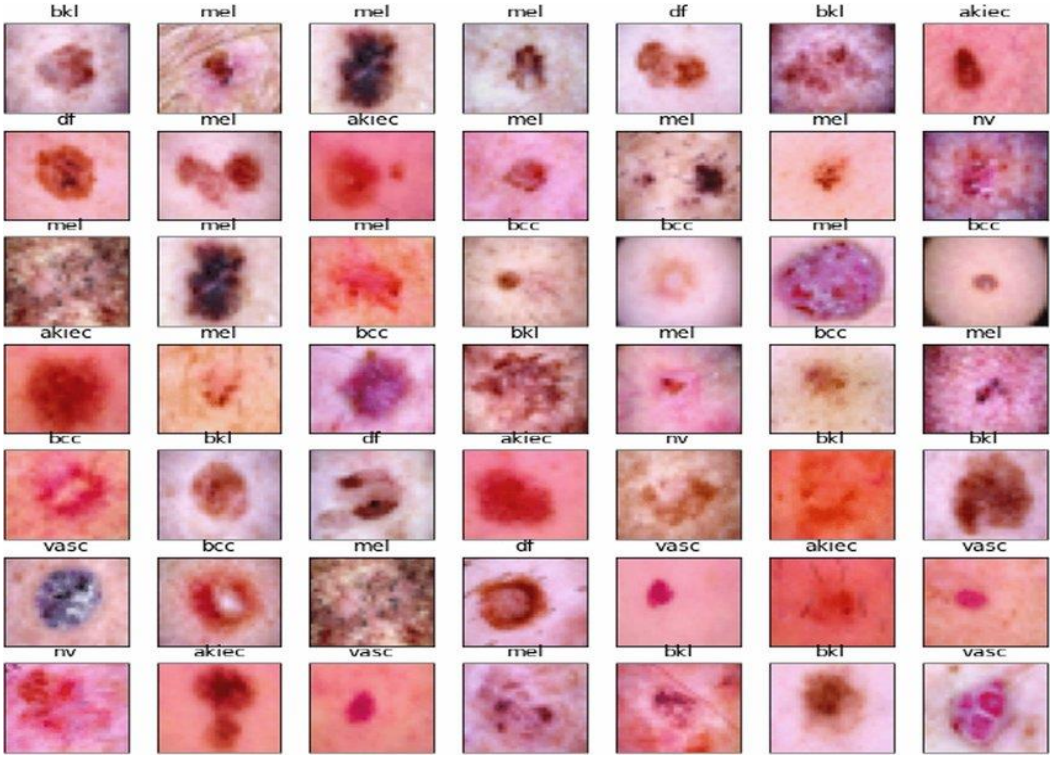

Figure 1. A subset of seven classes of the HAM10000 dataset.

[17], examined how boundary localization and normalization approaches affect the detection of melanoma. Utilizing four comparable datasets, PH², ISIC 2016, ISIC 2017, and HAM10000, the proposed method is assessed. According to experimental results, the DenseNet-121 with Multi-Layer Perceptron (MLP) performs better on the PH², ISIC 2016, ISIC 2017, and HAM10000 datasets with accuracy scores of 0.9833, 0.8047, 0.8116, and 0.81 respectively. On the HAM10000 dataset, [18] combined Inception ResNet-v2 with the Soft-Attention technique to get the best classification result, with an accuracy of 0.934, an average precision of 0.937, and an AUC of 0.984. [19], introduced the Two-Timescale Update Rule to produce features with great fine-grainedness and integrated the attention mechanism with PGGAN to acquire overall features of skin lesions images, all while enhancing the stability of GAN with an AUC 0.793. [20], suggested deep learning as a technique for precisely extracting a lesion zone. The quality of the image is first improved using Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN). Then, regions of interest (ROI) are separated from the entire image using segmentation. To categorize skin lesions, the image is next examined using CNN and an updated version of Resnet-50. On the HAM10000 dataset, the suggested method obtained an F-score of 0.86, accuracy of 0.86, precision of 0.84, and recall of 0.86. AUC of 0.99 was achieved by various deep learning models using the HAM10000 dataset. Two CNN models were used by [21] to separate benign and malignant skin lesions into different categories. According to the data, the model had a better classification accuracy than dermatologists' diagnosis when the results were compared to those.

We summarize our contributions as follows:

- We propose a new hybrid model based on a convolutional neural network and random forest algorithm (CNN-RF model) for detecting and classifying skin cancer images.
- The proposed model (CNN-RF) can overcome class imbalance by utilizing different oversampling methods. The proposed model was validated using the HAM10000 dataset and different batch sizes to show the ability of our proposed model.
- In addition, we implement the proposed model on different image sizes to ensure the consistency and flexibility of our model.

- The proposed model can classify seven classes of the utilized dataset with 98.88 % for accuracy, 0.99 for precision, 0.99 for recall, 0.99 for F1-score, and 0.9999 for AUC.

The outline of this paper is arranged as follows. Our proposed method is presented in detail in Section 2. Results and discussion are shown in Section 3. Sensitivity analysis and comparison are explained in Section 4. The conclusion and future work are summarized in Section 5.

2 | Proposed Methodology

Using machine learning and specialized convolutional neural networks, we created a fully automated method for identifying and classifying skin lesions. The suggested work mainly focused on pre-processing and classification. This study makes use of the standard HAM10000 dataset, which consists of 10015 images of skin lesions classified into seven categories. In this section, first, the utilized methods and algorithms will be reviewed. Then the proposed model will be discussed in detail. We applied five types of oversampling methods to overcome the imbalance of the HAM10000 dataset: random oversampling, SMOTE, SMOTE-ENN, SMOTE-TOMEK, and ADASYN. One of the most frequently chosen methods for dealing with an imbalanced dataset is resampling the data. For this, there are primarily two kinds of methods: under-sampling and oversampling. Oversampling is typically favored over under-sampling methods. The reason is that when we under-sample, we frequently leave out data points that might contain crucial information [22, 23]. Five different types of oversampling techniques will be covered briefly.

Random Over-Sampling: The simplest oversampling method for resolving the dataset's imbalance is random oversampling. By re-creating the minority class samples, it balances the data. While there is no information loss as a result, the dataset is more likely to be over-fit because the same data is being copied [24].

SMOTE (Synthetic Minority Oversampling Technique): SMOTE is an oversampling technique in which made-up samples are created for the minority class. By using this approach, the over-fitting problem caused by random oversampling is mitigated. It focuses on the feature space to create new instances by using interpolation between the positive instances that are close together [25].

SMOTE-ENN: Another hybrid strategy is SMOTE + ENN, which removes additional observations from the sample space. Here, the nearest neighbors of each member of the majority class are calculated using ENN, another under-sampling technique. This method can be used in conjunction with SMOTE's oversampled data to perform thorough data cleaning [26].

SMOTE-TOMEK: Such a hybrid method, SMOTE+TOMEK, seeks to eliminate overlapping data points for each of the classes scattered in the sample space. Under-sampling and oversampling strategies are combined in hybridization techniques. This is done to improve how well classifier models perform on samples produced using these strategies [27].

ADASYN (Adaptive Synthetic Sampling Approach): Another kind of SMOTE is ADASYN. ADASYN generates synthetic data following the data density and the density of the minority class has an inverse relationship with the growth of synthetic data. The ADASYN method would then focus excessively on these feature space regions. Before employing the ADASYN, it could be preferable to eliminate the outlier [28].

Convolutional Neural Network (CNN): CNN is excellent at processing inputs such as images, speech, or audio, which sets them ahead of other neural networks. Convolutional, pooling and fully-connected layers are their three primary types of layers. The fundamental layer of a CNN is the convolutional layer, which is similarly where the most of computation happens. It requires input data, a filter, and a feature map. Pooling layers, also referred to as down-sampling, do dimensionality lessening, which depresses the number of parameters in the input. In the fully connected layer, each node in the output layer is directly associated with a node in the layer above it. Using the features that were repossessed from the previous layers, this layer conducts the classification operation [29, 30].

Random Forest (RF): A popular machine learning algorithm identified as random forest blends the output of various decision trees to yield a single outcome. Its common use is inspired by its flexibility and usability

because it can loosen classification and regression problems. Additional benefits of RF include robustness: RF is a strong algorithm that can deal with erratic data. It can generalize effectively to new data since it is less likely to overfit the data. Accuracy: RF is one of the most reliable and accurate machine learning algorithms. Feature Importance: The assessment of feature relevance provided by RF can aid in feature selection and data comprehension. These factors make RF useful for a variety of tasks, particularly image classification [31, 32].

Figure 2 shows the phases that make up the proposed work. The proposed hybrid classification model consists of three phases, including data pre-processing, features extraction using customized CNN, and classification of skin lesions using the RF algorithm. At first, data were pre-processed before being used by CNN. This operation included oversampling techniques, image resizing, standardization, and splitting data. HAM10000 dataset has seven classes and each class has a no. Image as follows: (akiec: 327, bcc: 514, bkl: 1099, df: 115, nv: 6705, vasc: 142, melanoma: 1113). As seen the dataset is imbalanced so, resampling techniques are required to overcome the class imbalance. We implemented five types of oversampling methods: random oversampling, SMOTE, SMOTE-ENN, SMOTE-TOMEK, and ADASYN. Then we standardized the pixel values of images to a range of 0–1. After that, data was split as 80% for training and 20% for testing.

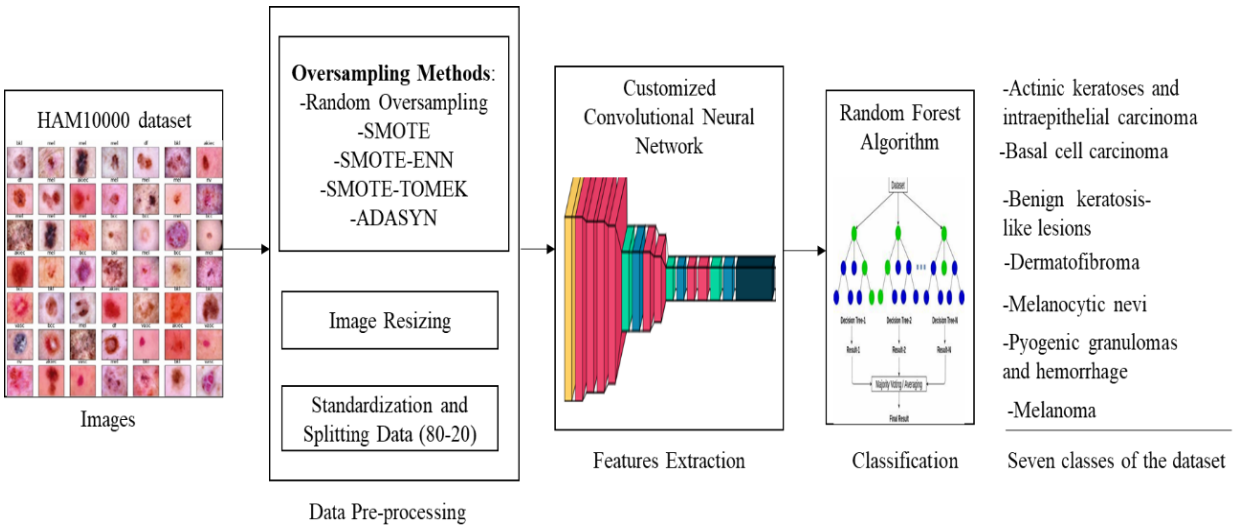

Figure 2. The phases that make up the proposed work.

Features extraction phase: in this phase, features were extracted using a customized CNN. Figure 3 shows a high-level overview of the customized CNN architecture. A CNN model was built, consisting of 19 layers and these layers include convolutional, pooling, and fully connected layers. We used the complete architecture of CNN with dense layers to obtain the results of CNN as a prediction model. However, when applying our proposed model CNN-RF, dense layers were removed from CNN architecture to use CNN as a feature extraction method and RF as a prediction tool. The image size of the HAM10000 dataset is 28 × 28 with a depth of 3 to speed up the process. Table 1 shows the hyperparameters of the proposed hybrid model (CNN-RF). After many experiments, we found that the best Learning rate for CNN is 0.001 and the best optimizer is Adam optimizer. To fit the CNN model, we used different batch sizes (16, 32, 64, 128, and 256), and the results according to each batch size will be reported later in this paper. In the classification phase, features that were extracted from the CNN model are fed to RF as an input. Then RF model makes predictions for these features. We used no. The estimator for RF is 50. The output of the RF model is one of the seven classes of the HAM10000 dataset according to the given image.

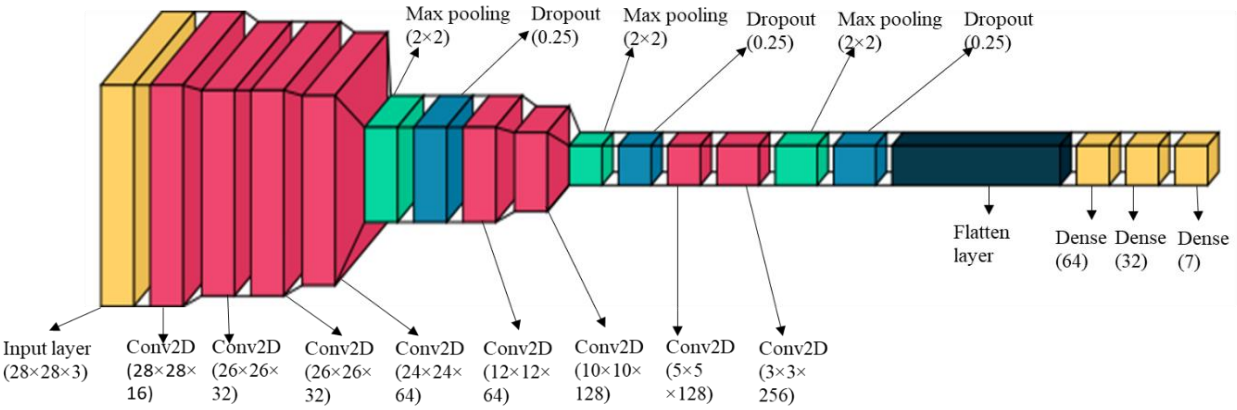

Figure 3. A high-level overview of CNN Architecture.

Table 1. Hyperparameters of the proposed model.

Parameter	value
Filter_size	(3,3)
Pool_size	(2,2)
Loss function for CNN	Categorical Cross-entropy
Epochs	30
Batch-size	16,32,64,128,256
Optimizer	Adam
Learning_rate	0.001
Number of CNN layers	16 layers
Estimators for RF	50 (random_state=0)

3 | Results and Discussion

Table 2 shows the number of images in each class after applying each method of oversampling methods. For simplicity, each class was represented with a digit number as follows: (akiec was represented with 0, bcc was represented with 1, bkl was represented with 2, df was represented with 3, nv was represented with 4, vasc was represented with 5, melanoma was represented with 6). Figure 4 shows the HAM10000 dataset after applying different oversampling methods. The authors faced the problem of memory being filled because the number of images is large and after applying different oversampling methods, the number of images become very large so, a high memory size is required. As shown in Table 2 and Figure 4, oversampling methods can achieve a balance between all classes of the dataset. Tables (3-7) show different accuracy measures of the customized CNN and the proposed hybrid model CNN-RF with different batch sizes. The accuracy measures include accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). The results of CNN were obtained when we used CNN as a prediction method for the HAM10000 dataset and compared these results with the proposed CNN-RF. As shown from Table 3 when the batch size is 16, CNN can achieve the highest accuracy with the SMOTE-ENN method and obtained 0.9728 accuracy and AUC 0.9980. According to CNN-RF can achieve the highest accuracy with the random oversampling method and get 0.9822 accuracy and AUC 0.9998. Figure 5 shows accuracy and loss curves for CNN with SMOTE-ENN when batch size is 16 for 30 epochs. As seen from the Figure, CNN is capable of getting better accuracy in a small number of epochs. Roundly from epoch 5 to epoch 30, CNN can obtain higher accuracy and achieve a small loss for diagnosing seven classes of the dataset which ensures the consistency of our customized CNN. Figure 6 shows the confusion matrix and ROC curve for CNN with SMOTE-ENN when the batch size is 16. Figure 7 shows the confusion matrix and ROC curve for CNN-RF with Random-versampling when the batch size is 16. The figures show the ability of our customized CNN and the hybrid CNN-RF to classify seven classes of the dataset. The confusion matrix shows that CNN-RF surpasses CNN in classifying more no. True images. Also, ROC curves show the ability of CNN and CNN-RF to get better AUC for each class which is nearest

to 1. CNN got 1 of AUC for all classes except for class 4 got 0.99 for AUC but CNN-RF got 1 of AUC for all classes.

Table 2. Number of images in each class for oversampling methods.

class	original	Random-Oversampling	SMOTE	SMOTE-ENN	SMOTE-TOMEK	ADASYN
0	327	6705	6705	6705	6705	6724
1	514	6705	6705	6704	6705	6641
2	1099	6705	6705	6592	6705	6886
3	115	6705	6705	6705	6705	6691
4	6705	6705	6705	3498	6705	6705
5	142	6705	6705	6705	6705	6730
6	1113	6705	6705	6592	6705	6810
Total	10015	46935	46935	43501	46935	47187

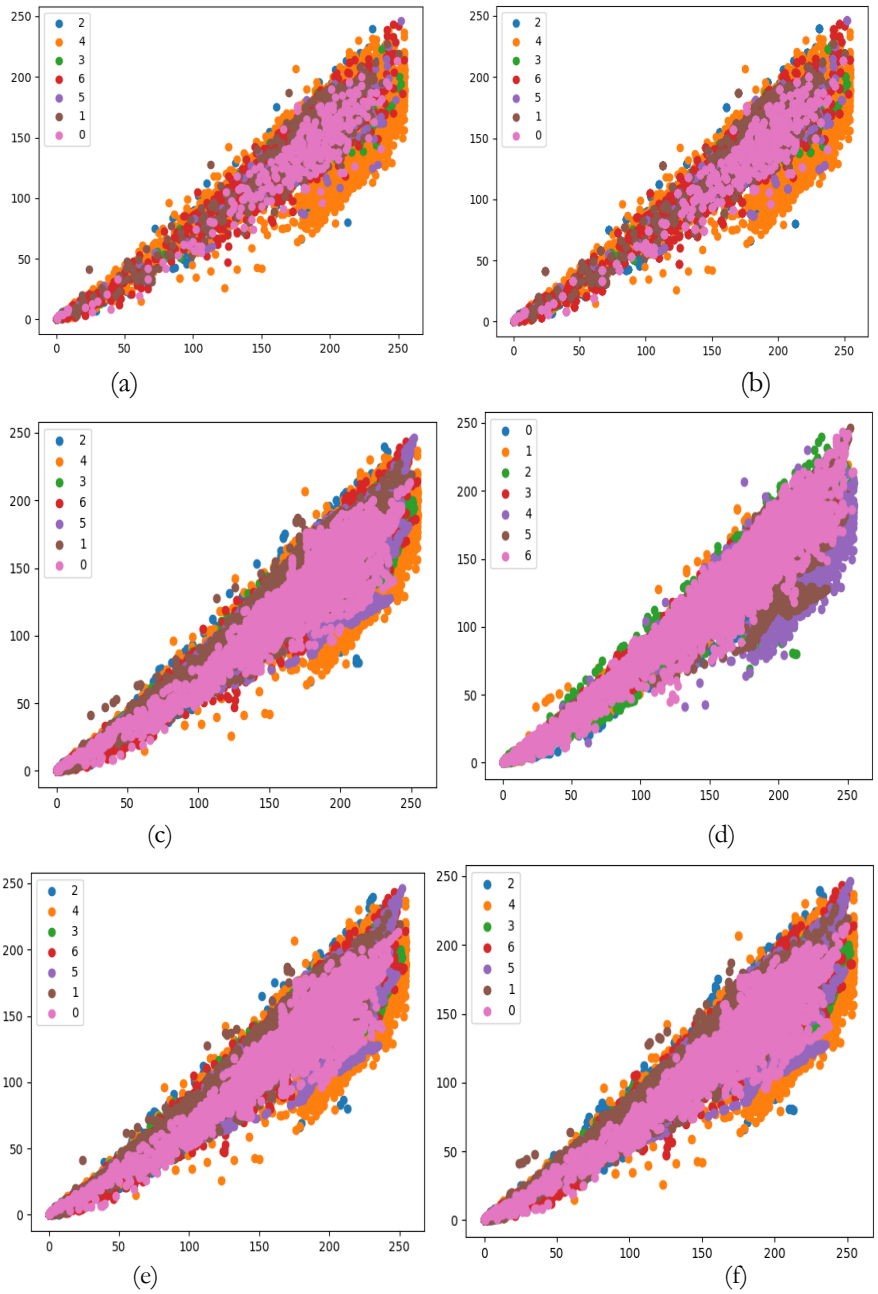

Figure 4. HAM10000 dataset after applying oversampling methods. (a) Original dataset, (b) Random oversampling, (c) SMOTE, (d) SMOTE-ENN, (e) SMOTE-TOMEK, and (f) ADASYN.

Table 3. Accuracy measures of CNN and CNN-RF (Batch Size=16).

Model	Accuracy	Precision	Recall	F1-score	AUC
CNN(Random-Oversampling)	0.9622	0.96	0.96	0.96	0.9972
CNN(SMOTE)	0.9469	0.95	0.95	0.95	0.9949
CNN(SMOTE-ENN)	0.9728	0.97	0.97	0.97	0.9980
CNN(SMOTE-TOMEK)	0.9377	0.94	0.94	0.94	0.9944
CNN(ADASYN)	0.9378	0.94	0.94	0.94	0.9944
CNN-RF(Random-Oversampling)	**0.9822**	**0.98**	**0.98**	**0.98**	**0.9998**
CNN-RF(SMOTE)	0.9354	0.94	0.94	0.93	0.9937
CNN-RF(SMOTE-ENN)	0.9657	0.97	0.97	0.97	0.9976
CNN-RF(SMOTE-TOMEK)	0.9371	0.94	0.94	0.94	0.9943
CNN-RF(ADASYN)	0.9378	0.94	0.94	0.94	0.9943

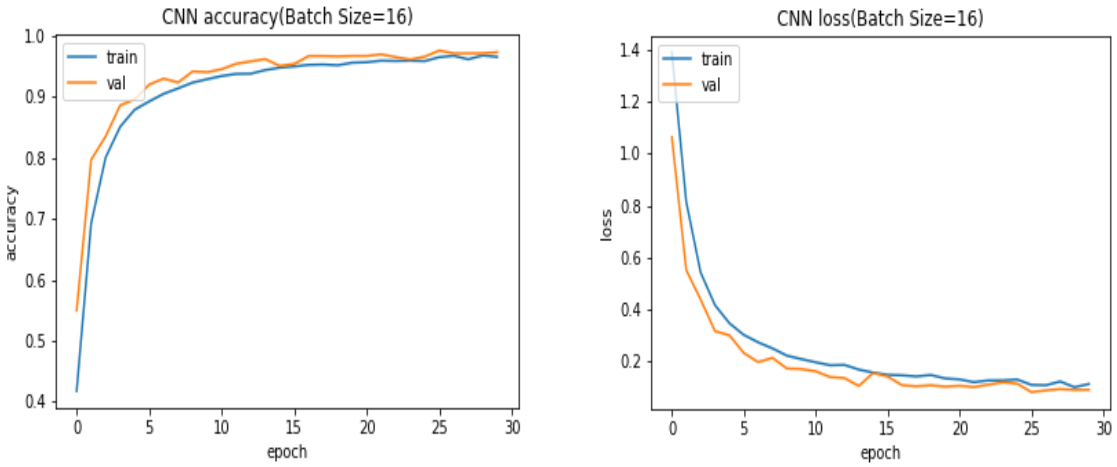

Figure 5. Accuracy and loss curves for CNN with SMOTE-ENN when batch size is 16.

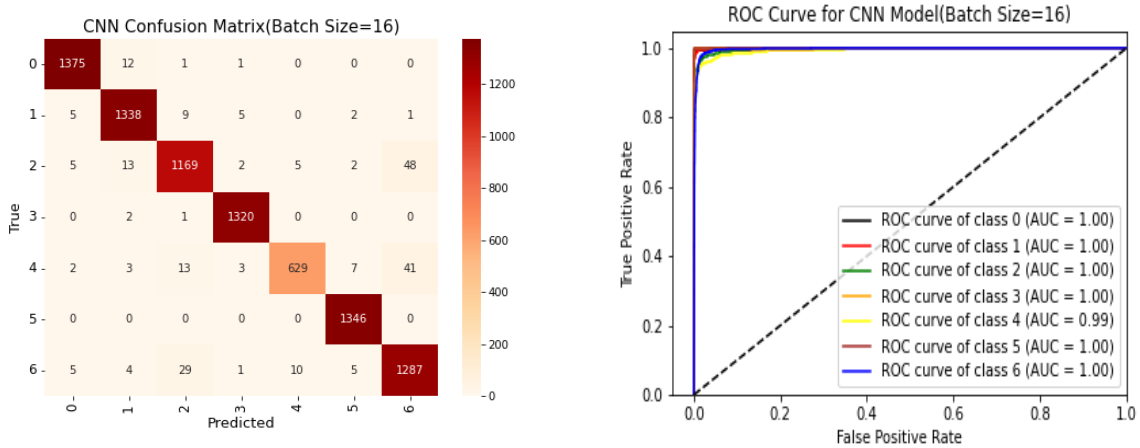

Figure 6. Confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 16.

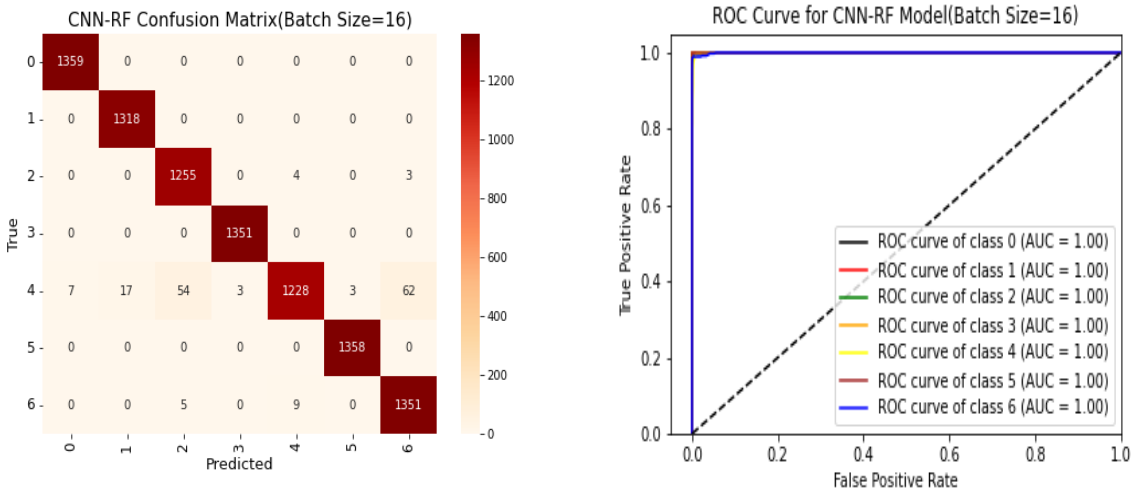

Figure 7. Confusion matrix and ROC curve for CNN-RF with Random-Oversampling when batch size is 16.

Table 4 shows accuracy measures when batch size is 32. As shown from the table, CNN can achieve the highest accuracy with the SMOTE-ENN method and obtained 0.9760 accuracy and AUC 0.9985. According to the hybrid model, CNN-RF can achieve the highest accuracy with the random oversampling method and get 0.9836 accuracy and AUC 0.9998. Figure 8 shows the confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 32. Figure 9 shows the confusion matrix and ROC curve for CNN-RF with Random-Oversampling when batch size is 32. As seen from the figures, our customized CNN and the hybrid CNN-RF can classify seven classes of the dataset. The confusion matrix shows that CNN-RF exceeds CNN in classifying more no. True images. Also, the ROC curve shows the ability of CNN and CNN-RF to get 1 of AUC for all classes.

Table 4. Accuracy measures of CNN and CNN-RF (Batch Size=32).

Model	Accuracy	Precision	Recall	F1-score	AUC
CNN(Random-Oversampling)	0.9655	0.97	0.97	0.96	0.9973
CNN(SMOTE)	0.9473	0.95	0.95	0.95	0.9958
CNN(SMOTE-ENN)	0.9760	0.98	0.98	0.98	0.9985
CNN(SMOTE-TOMEK)	0.9527	0.95	0.95	0.95	0.9960
CNN(ADASYN)	0.9476	0.95	0.95	0.95	0.9955
CNN-RF(Random-Oversampling)	**0.9836**	**0.98**	**0.98**	**0.98**	**0.9998**
CNN-RF(SMOTE)	0.9426	0.94	0.94	0.94	0.9952
CNN-RF(SMOTE-ENN)	0.9690	0.97	0.97	0.97	0.9981
CNN-RF(SMOTE-TOMEK)	0.9441	0.95	0.94	0.94	0.9951
CNN-RF(ADASYN)	0.9485	0.95	0.95	0.95	0.9957

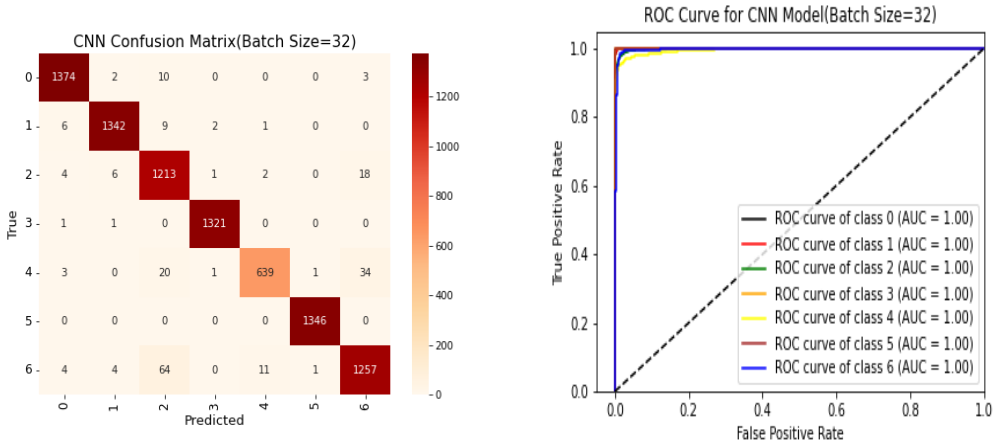

Figure 8. Confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 32.

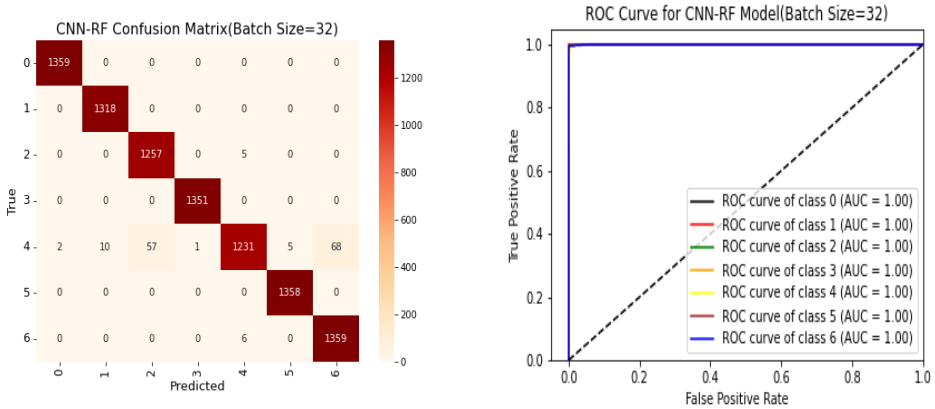

Figure 9. Confusion matrix and ROC curve for CNN-RF with Random-Oversampling when batch size is 32.

Table 5 shows accuracy measures when the batch size is 64, as shown from the table, CNN can achieve the highest accuracy with the SMOTE-ENN method and obtained 0.9767 accuracies and AUC 0.9983. According to the hybrid model, CNN-RF can achieve the highest accuracy with the random oversampling method and get 0.9878 accuracy and 0.9999 AUC. Figure 10 shows the confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 64. Figure 11 shows the confusion matrix and ROC curve for CNN-RF with Random-Oversampling when batch size is 64. The figures show that CNN and CNN-RF can obtain better results for seven classes of the dataset. The confusion matrix shows that CNN-RF surpasses CNN in classifying more no. True images. Also, the ROC curve shows the ability of CNN and CNN-RF to get better AUC for each class which is nearest to 1. CNN got 1 of AUC for all classes except for class 4 got 0.99 for AUC but CNN-RF got 1 of AUC for all classes.

Table 5. Accuracy measures of CNN and CNN-RF (Batch Size=64).

Model	Accuracy	Precision	Recall	F1-score	AUC
CNN(Random-Oversampling)	0.9676	0.97	0.97	0.97	0.9981
CNN(SMOTE)	0.9560	0.96	0.96	0.96	0.9965
CNN(SMOTE-ENN)	0.9767	0.98	0.98	0.98	0.9983
CNN(SMOTE-TOMEK)	0.9554	0.96	0.96	0.95	0.9966
CNN(ADASYN)	0.9549	0.96	0.95	0.95	0.9966
CNN-RF(Random-Oversampling)	**0.9878**	**0.99**	**0.99**	**0.99**	**0.9999**
CNN-RF(SMOTE)	0.9529	0.95	0.95	0.95	0.9959
CNN-RF(SMOTE-ENN)	0.9768	0.98	0.98	0.98	0.9986
CNN-RF(SMOTE-TOMEK)	0.9521	0.95	0.95	0.95	0.9958
CNN-RF(ADASYN)	0.9512	0.95	0.95	0.95	0.9966

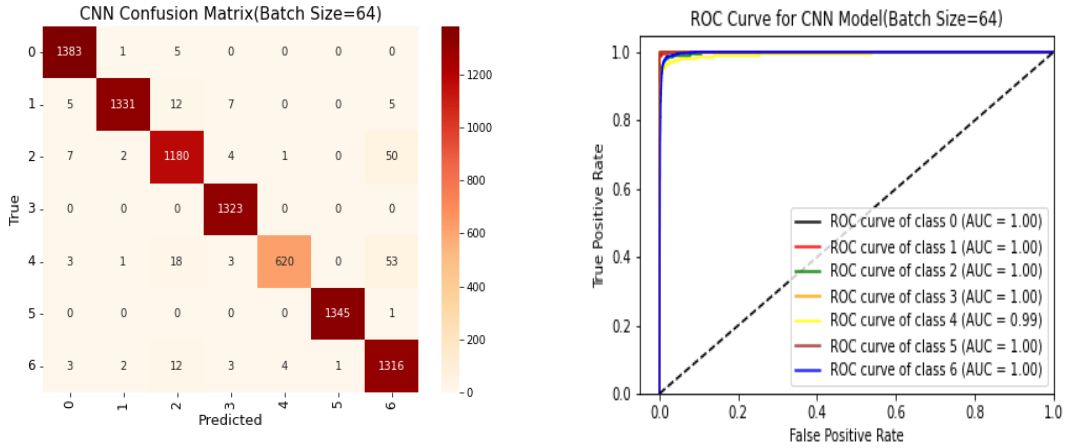

Figure 10. Confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 64.

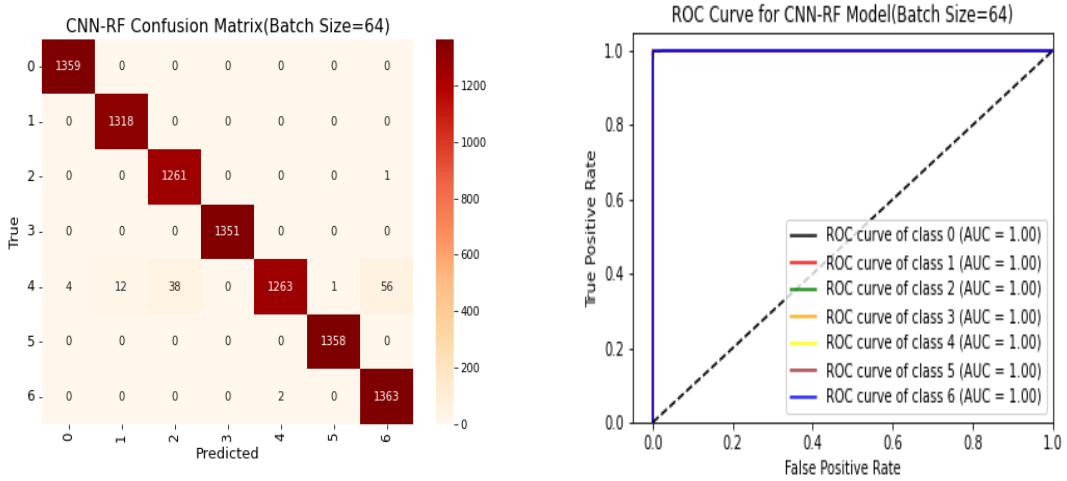

Figure 11. Confusion matrix and ROC curve for CNN-RF with Random-Oversampling when batch size is 64.

Table 6 shows accuracy measures when the batch size is 128, as shown from the table, CNN can achieve the highest accuracy with the SMOTE-ENN method and obtained 0.9782 accuracy and AUC 0.9988. According to the hybrid model, CNN-RF can achieve the highest accuracy with the random oversampling method and get 0.9844 accuracy and AUC 0.9998. Figure 12 shows the confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 128. Figure 13 shows the confusion matrix and ROC curve for CNN-RF with Random-Oversampling when batch size is 128. As seen from the figures, CNN and CNN-RF can diagnose HAM10000 dataset classes effectively. The confusion matrix shows that CNN-RF surpasses CNN in classifying more no. True images. However, the ROC curve shows that CNN and CNN-RF have equal ability in getting 1 of AUC for all classes.

Table 6: Accuracy measures of CNN and CNN-RF (Batch Size=128).

Model	Accuracy	Precision	Recall	F1-score	AUC
CNN(Random-Oversampling)	0.9711	0.97	0.97	0.97	0.9983
CNN(SMOTE)	0.9583	0.96	0.96	0.96	0.9968
CNN(SMOTE-ENN)	0.9782	0.98	0.98	0.98	0.9988
CNN(SMOTE-TOMEK)	0.9510	0.95	0.95	0.95	0.9961
CNN(ADASYN)	0.9566	0.96	0.96	0.96	0.9958
CNN-RF(Random-Oversampling)	**0.9844**	**0.98**	**0.98**	**0.98**	**0.9998**
CNN-RF(SMOTE)	0.9560	0.96	0.96	0.96	0.9968
CNN-RF(SMOTE-ENN)	0.9784	0.98	0.98	0.98	0.9987
CNN-RF(SMOTE-TOMEK)	0.9522	0.95	0.95	0.95	0.9962
CNN-RF(ADASYN)	0.9515	0.95	0.95	0.95	0.9959

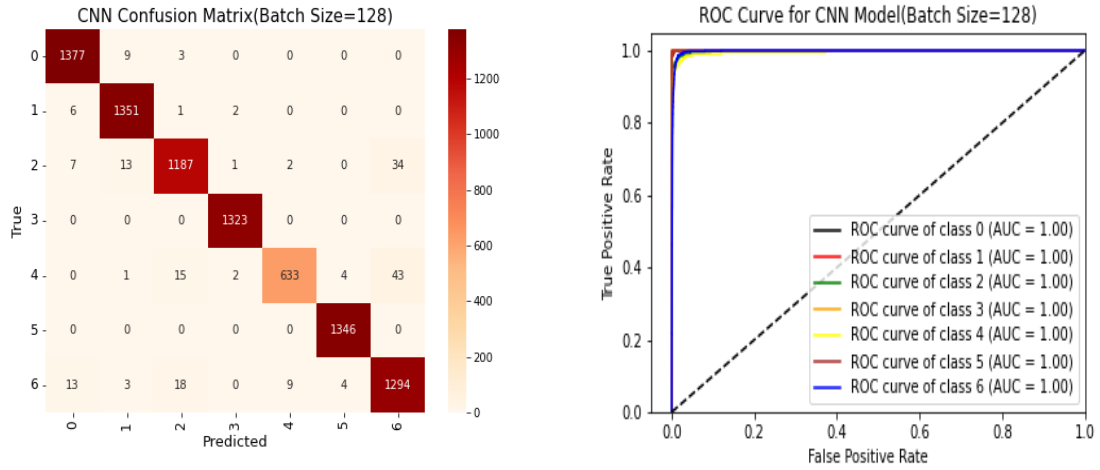

Figure 12. Confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 128.

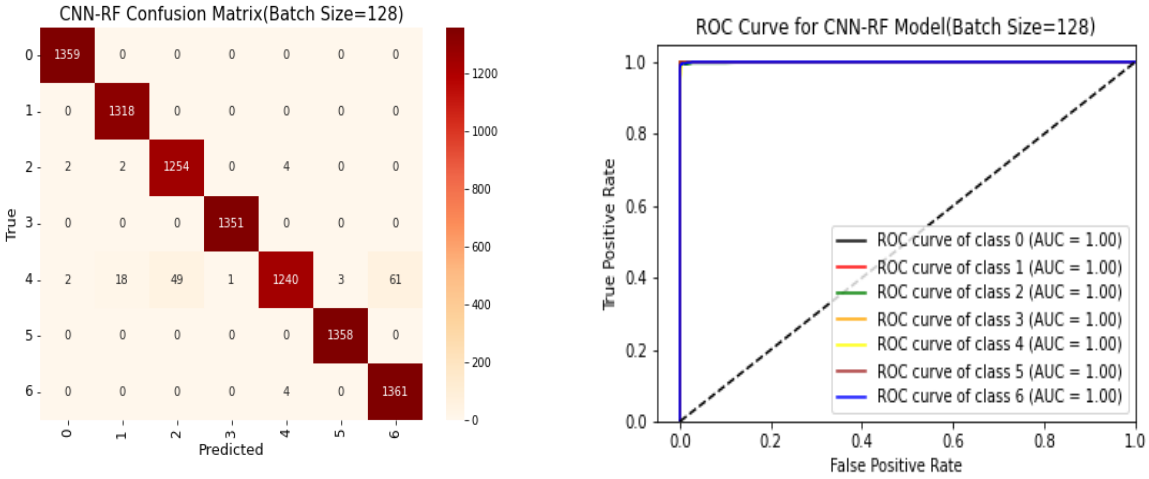

Figure 13. Confusion matrix and ROC curve for CNN-RF with Random-Oversampling when batch size is 128.

Table 7 shows accuracy measures when the batch size is 256, as shown from the table, CNN can achieve the highest accuracy with the SMOTE-ENN method and obtained 0.9820 accuracies and AUC 0.9988. According to the hybrid model, CNN-RF can achieve the highest accuracy with the random oversampling method and get 0.9865 accuracy and AUC 0.9998. Figure 14 shows the confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 256. Figure 15 shows the confusion matrix and ROC curve for CNN-RF with Random-versampling when batch size is 256. The figures show the ability of CNN and CNN-RF to classify seven classes of the dataset. The confusion matrix shows that CNN-RF surpasses CNN in classifying more no. True images. However, the ROC curve shows that CNN and CNN-RF have equal capability in getting 1 of AUC for all classes.

Table 7. Accuracy measures of CNN and CNN-RF (Batch Size=256).

Model	Accuracy	Precision	Recall	F1-score	AUC
CNN(Random-Oversampling)	0.9706	0.97	0.97	0.97	0.9979
CNN(SMOTE)	0.9637	0.96	0.96	0.96	0.9975
CNN(SMOTE-ENN)	0.9820	0.98	0.98	0.98	0.9988
CNN(SMOTE-TOMEK)	0.9630	0.96	0.96	0.96	0.9972
CNN(ADASYN)	0.9593	0.96	0.96	0.96	0.9971
CNN-RF(Random-Oversampling)	**0.9865**	**0.99**	**0.99**	**0.99**	**0.9998**
CNN-RF(SMOTE)	0.9549	0.96	0.95	0.95	0.9966
CNN-RF(SMOTE-ENN)	0.9788	0.98	0.98	0.98	0.9989
CNN-RF(SMOTE-TOMEK)	0.9567	0.96	0.96	0.96	0.9970
CNN-RF(ADASYN)	0.9583	0.96	0.96	0.96	0.9972

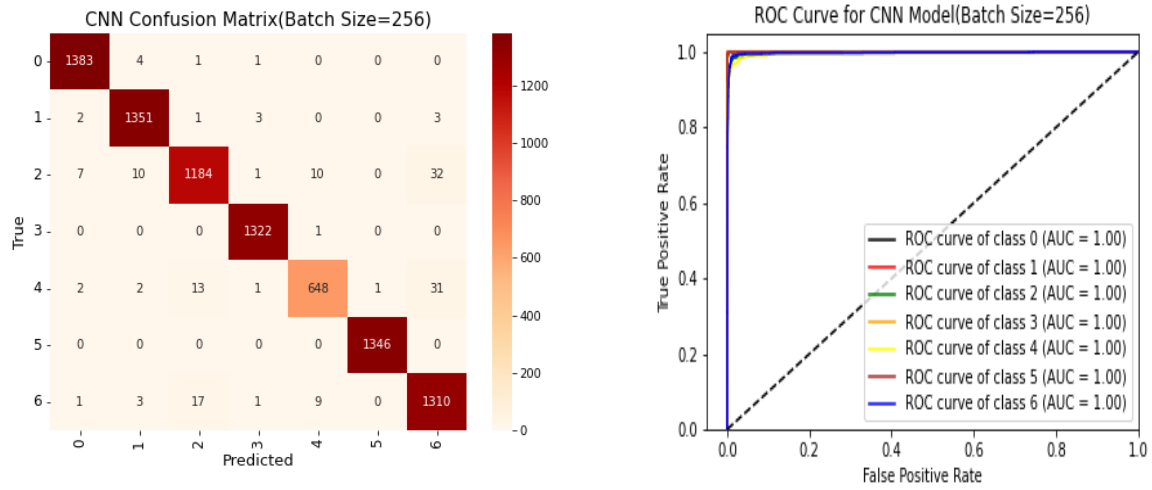

Figure 14. Confusion matrix and ROC curve for CNN with SMOTE-ENN when batch size is 256.

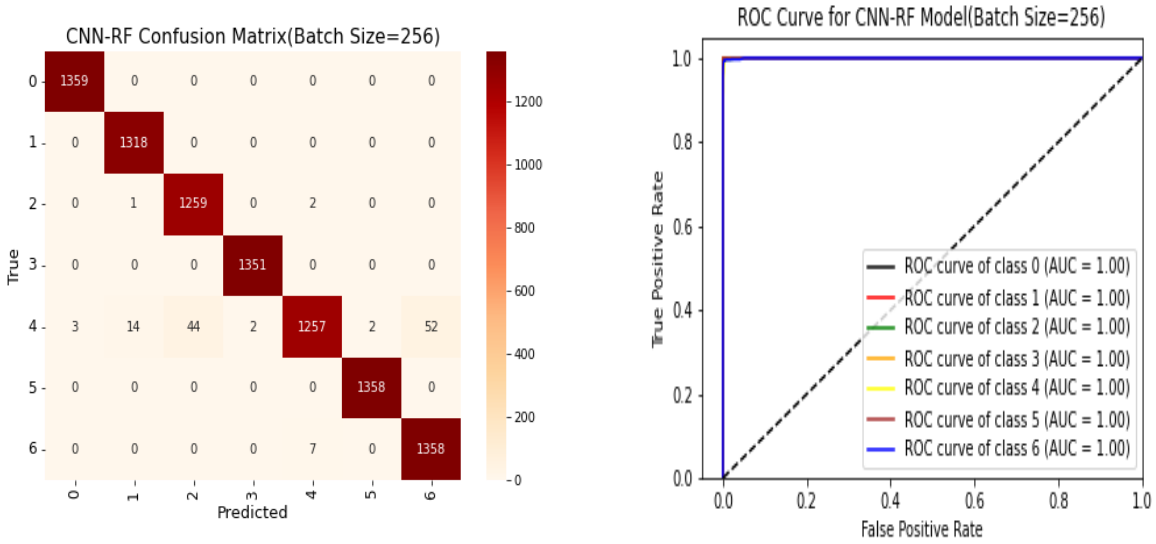

Figure 15. Confusion matrix and ROC curve for CNN-RF with Random-Oversampling when batch size is 256.

As seen from the results, the proposed hybrid model CNN-RF can deliver the best results for the HAM10000 dataset with the random oversampling method and can obtain the value 1 of AUC for all classes with different batch sizes. According to CNN results, the best oversampling method for CNN is SMOTE-ENN which can deliver better results than other methods. For batch size 16, the worst oversampling method for CNN is SMOTE-TOMEK with an accuracy of 0.9377 and AUC of 0.9944 but, for CNN-RF the worst oversampling method is SMOTE with an accuracy of 0.9354 and AUC of 0.9937. For batch size 32, the worst oversampling method for CNN is SMOTE with an accuracy of 0.9473 and AUC of 0.9958, and also for CNN-RF the worst oversampling method is SMOTE with an accuracy of 0.9426 and AUC of 0.9952. For batch size 64, the worst oversampling method for CNN is ADASYN with an accuracy of 0.9549 and AUC 0.9966, and also for CNN-RF the worst oversampling method is ADASYN with an accuracy of 0.9512 and AUC 0.9966. For batch size 128, the worst oversampling method for CNN is SMOTE-TOMEK with an accuracy of 0.9510 and AUC 0.9961 but for CNN-RF the worst oversampling method is ADASYN with an accuracy of 0.9515 and AUC 0.9959. For batch size 256, the worst oversampling method for CNN is ADASYN with an accuracy of 0.9593 and AUC 0.9971 but for CNN-RF the worst oversampling method is SMOTE with an accuracy of 0.9549 and AUC 0.9966. Table 8 and Table 9 summarize all the important previous results for CNN and CNN-RF. As shown from the table, when batch size is increased, the accuracy of the CNN model is increased also which ensures the model stability and how batch size can affect the performance model.

Table 8. CNN with SMOTE-ENN with different batch sizes.

Batch size	Accuracy
16	0.9728
32	0.9760
64	0.9767
128	0.9782
256	0.9820

Table 9. CNN-RF with Random-Oversampling with different batch sizes.

Batch size	Accuracy
16	0.9822
32	0.9836
64	0.9878
128	0.9844
256	0.9865

4 | Sensitivity Analysis and Comparison

Among all the previous results of this paper, the proposed hybrid model CNN-RF is the best with a random oversampling method for different batch sizes. To ensure the consistency of our proposed model, we applied the same model for image size 80 × 80 instead of image size 28 × 28 which was the original image size for our selected dataset. Table 10 shows the accuracy measures of CNN-RF (image size 80 × 80) with different batch sizes. As seen from the table, CNN-RF with batch size 128 gives the highest accuracy 0.9888, AUC 0.9999, and 0.99 for each precision, recall, and F1-score than other models. Figure 16 shows the confusion matrix for CNN-RF with random oversampling with different batch sizes. This figure shows that CNN-RF can predict the most no. True images. Figure 17 shows the ROC curve for CNN-RF with random oversampling with different batch sizes. Also, this figure ensures the ability of our proposed model to achieve AUC 1 for all classes of the dataset for all batch sizes. These results prove that our proposed hybrid model CNN-RF can achieve more satisfactory results for all measures and all batch sizes despite scaling image size to 80 × 80 which improves the ability of our proposed model to obtain the best and more satisfactory results for the HAM10000 dataset.

Table 10: Accuracy measures of CNN-RF with Random-Oversampling (image size 80 × 80) for different batch sizes.

Model		Accuracy	Precision	Recall	F1-score	AUC
CNN-RF	batch size=16	0.9885	0.99	0.99	0.99	0.9998
	batch size=32	0.9847	0.98	0.98	0.98	0.9998
	batch size=64	0.9881	0.99	0.99	0.99	0.9999
	batch size=128	**0.9888**	**0.99**	**0.99**	**0.99**	**0.9999**
	batch size=256	0.9815	0.98	0.98	0.98	0.9999

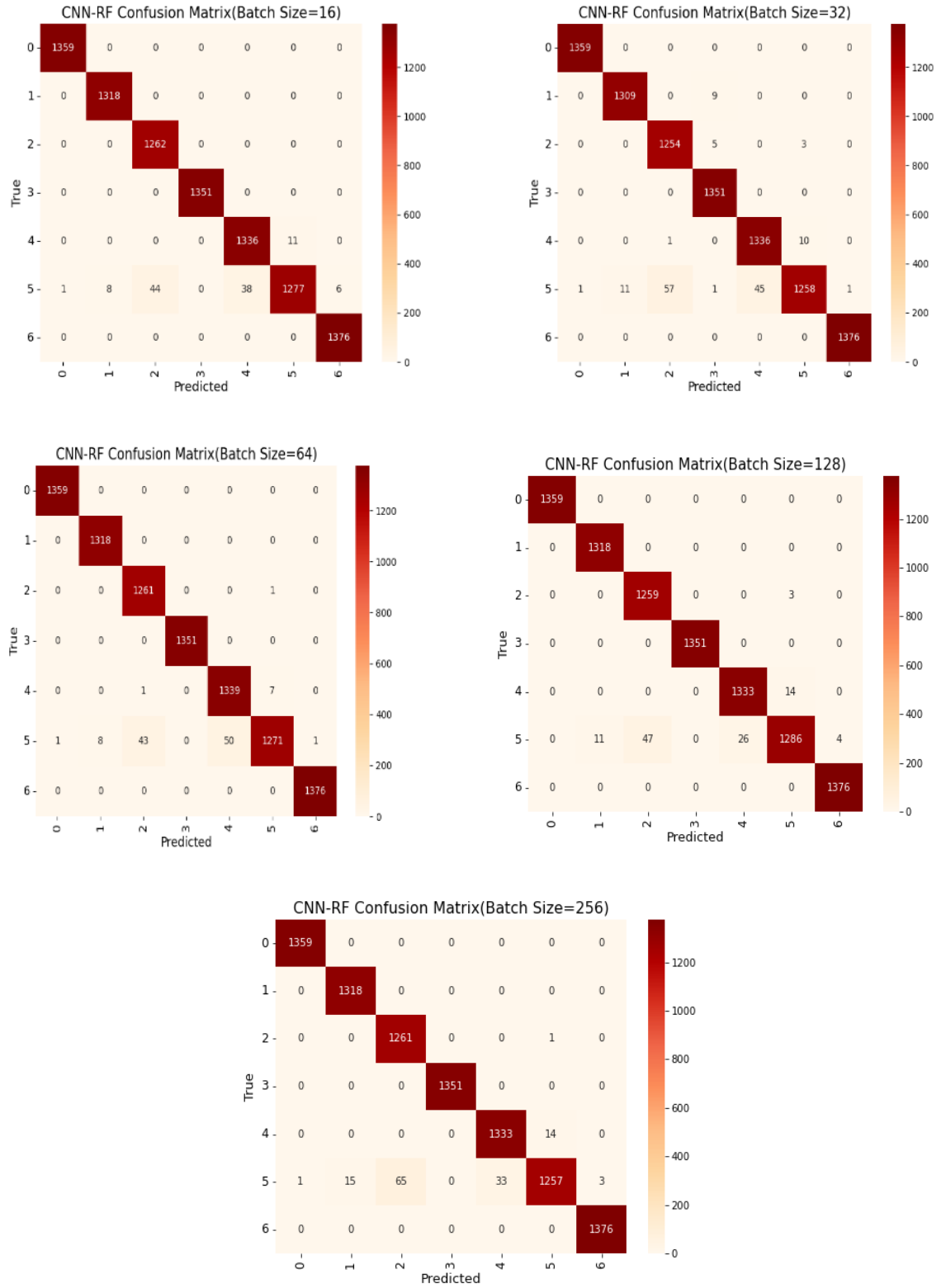

Figure 16. Confusion matrix for CNN-RF with Random-Oversampling with different batch sizes.

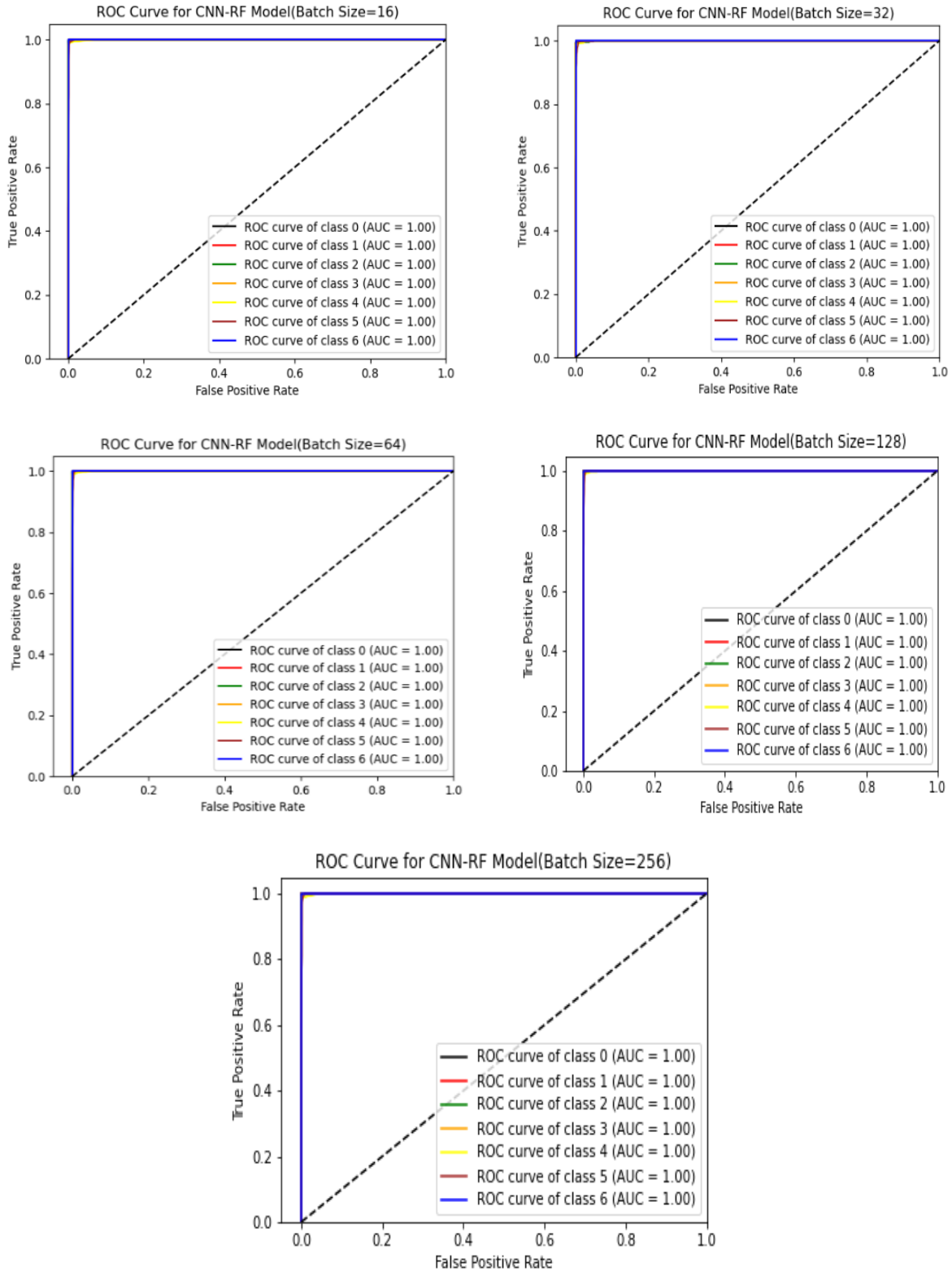

Figure 17. ROC curve for CNN-RF with Random-Oversampling with different batch sizes.

To show the ability of our proposed hybrid model against the literature, we compared it with models of more recent papers. The selected papers for comparison were for the classification of the HAM10000 dataset. Table 11 shows the comparison of the proposed model with recent existing techniques. As seen from the table, the proposed hybrid model CNN-RF can obtain the highest accuracy 98.88%, and beat other models.

Table 11. Comparison of the proposed model with recent existing techniques on the HAM10000 dataset.

Ref#	Year	Method	Accuracy (%)
[33]	2021	Shifed MobileNetV2	81.90
[34]	2021	MASK RCNN + DenseNet + LS-SVM	88.50
[35]	2021	ResNet-50 + Naïve Bayes	85.50
[36]	2022	EW-FCM+wide-shufenet	84.80
[37]	2022	CNN	95.18
[38]	2022	XAI/IML explanation techniques + CNN	80.00
[39]	2022	CNN, AlexNet, Vgg-16, Vgg-19	92.25
[40]	2023	InceptionResnet-V2	91.30
[41]	2023	EfficientNET-B1	84.30
[16]	2023	Mobilenetv2-DeepLabv3+	92.01
[42]	2023	effective data augmentation + Inception-Resnet-v2	95.09
[43]	2023	wavelet transform-based deep residual neural network (WT-DRNNet)	95.73
Proposed	2024	CNN-RF	**98.88**

5 | Conclusion and Future Work

In this paper, a new hybrid model based on a convolutional neural network and random forest algorithm is proposed for detecting and classifying skin cancer images. The dataset utilized for this study is based on the HAM10000 dataset. The dataset was first preprocessed effectively and different oversampling methods were applied to determine the best method. Then, features of the preprocessed dataset were extracted using a customized convolutional neural network. Finally, these features were classified into seven classes using a random forest algorithm. The effective hyper-parameters for CNN and RF were detected besides different batch sizes and image sizes were implemented. The proposed model (CNN-RF) with random oversampling method can achieve 98.88 % accuracy, 0.99 precision, 0.99 for recall, 0.99 for F1-score, and 0.9999 for AUC. The proposed model was compared with our customized CNN as a prediction tool and with the most recent existing work on the same dataset and proved its ability to achieve better accuracy than other models. The results of the proposed model show that it can effectively detect and classify seven classes of the HAM10000 dataset.

According to future trends, there are many algorithms aiming to aid us in understanding how machine learning models make predictions. However, many researchers have applied explainable artificial intelligence and shown its importance in providing effective explanations for predictions. Additionally, different techniques for feature selection can be used to develop models, such as wrapper and filter methods. Also, they can replace the random forest algorithm in our model with other machine learning algorithms and achieve more satisfactory results. In addition, they can use the proposed model for other datasets and obtain more satisfactory results.

Acknowledgments

The author is grateful to the editorial and reviewers, as well as the correspondent author, who offered assistance in the form of advice, assessment, and checking during the study period.

Author Contributions

Salwa Elsayed: Conceptualization, Methodology, Software, Data curation, Writing- Original draft preparation. **Israa Mohamed:** Investigation, Writing- Reviewing and Editing. **Amal F. Abdel-Gawad:** Visualization, Investigation, Writing- Reviewing and Editing. **Mahmoud M. Ismail:** Conceptualization, Supervision, Writing- Reviewing and Editing.

Funding

This research has no funding source.

Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to the privacy-preserving nature of the data but are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Ethical Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

References

[1] P. Aggarwal, P. Knabel, and A. B. J. J. o. t. A. A. o. D. Fleischer Jr, "United States burden of melanoma and non-melanoma skin cancer from 1990 to 2019," vol. 85, no. 2, pp. 388-395, 2021.

[2] K. S. f. M. S. C. A. online:. (2023). Available: https://www.cancer.org/cancer/melanoma-skin-cancer/about/key-statistics.html

[3] Y. Wu, B. Chen, A. Zeng, D. Pan, R. Wang, and S. J. F. i. O. Zhao, "Skin Cancer Classification With Deep Learning: A Systematic Review," vol. 12, 2022.

[4] C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1-9.

[5] K. Simonyan and A. J. a. p. a. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778.

[7] A. J. T. g. Al-Karawi, "Stacked Cross Validation with Deep Features: A Hybrid Method for Skin Cancer Detection," vol. 16, no. 1, pp. 33-39, 2022.

[8] H. A. Hasan and A. A. Ibrahim, "Hybrid detection techniques for skin cancer images," in 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), 2020, pp. 1-8: IEEE.

[9] N. Ahmed, X. Tan, L. J. M. T. Ma, and Applications, "A new method proposed to Melanoma-skin cancer lesion detection and segmentation based on hybrid convolutional neural network," pp. 1-24, 2022.

[10] A. Murugan, S. A. H. Nair, A. A. P. Preethi, K. S. J. M. Kumar, and Microsystems, "Diagnosis of skin cancer using machine learning techniques," vol. 81, p. 103727, 2021.

[11] [1Y. Gu, Z. Ge, C. P. Bonnington, J. J. I. j. o. b. Zhou, and h. informatics, "Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification," vol. 24, no. 5, pp. 1379-1393, 2019.

[12] K. Xiang et al., "A novel weight pruning strategy for light weight neural networks with application to the diagnosis of skin disease," vol. 111, p. 107707, 2021.

[13] K. Polat, K. O. J. J. o. A. I. Koc, and Systems, "Detection of skin diseases from dermoscopy image using the combination of convolutional neural network and one-versus-all," vol. 2, no. 1, pp. 80-97, 2020.

[14] B. Ahmad, S. Jun, V. Palade, Q. You, L. Mao, and M. J. D. Zhongjie, "Improving skin cancer classification using heavy-tailed Student t-distribution in generative adversarial networks (TED-GAN)," vol. 11, no. 11, p. 2147, 2021.

[15] A. Rezvantalab, H. Safigholi, and S. J. a. p. a. Karimijeshni, "Dermatologist level dermoscopy skin cancer classification using different deep learning convolutional neural networks algorithms," 2018.

[16] M. Zafar, J. Amin, M. Sharif, M. A. Anjum, G. A. Mallah, and S. J. M. Kadry, "DeepLabv3+-Based Segmentation and Best Features Selection Using Slime Mould Algorithm for Multi-Class Skin Lesion Classification," vol. 11, no. 2, p. 364, 2023.

[17] H. K. Gajera, D. R. Nayak, M. A. J. B. S. P. Zaveri, and Control, "A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features," vol. 79, p. 104186, 2023.

[18] S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft attention improves skin cancer classification performance," in Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data: 4th International Workshop, iMIMIC 2021, and 1st International Workshop, TDA4MedicalData 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 4, 2021, pp. 13-23: Springer.

[19] I. S. A. Abdelhalim, M. F. Mohamed, and Y. B. J. E. S. w. A. Mahdy, "Data augmentation for skin lesion using self-attention based progressive generative adversarial network," vol. 165, p. 113922, 2021.

[20] G. Alwakid, W. Gouda, M. Humayun, and N. U. Sama, "Melanoma Detection Using Deep Learning-Based Classifications," in Healthcare, 2022, vol. 10, no. 12, p. 2481: MDPI.
[21] M. A. Kassem, K. M. Hosny, and M. M. J. I. A. Fouad, "Skin lesions classification into eight classes for ISIC 2019 using deep convolutional neural network and transfer learning," vol. 8, pp. 114822-114832, 2020.
[22] L. Abdi, S. J. I. t. o. K. Hashemi, and D. Engineering, "To combat multi-class imbalanced problems by means of over-sampling techniques," vol. 28, no. 1, pp. 238-251, 2015.
[23] B. Santoso, H. Wijayanto, K. Notodiputro, and B. Sartono, "Synthetic over sampling methods for handling class imbalanced problems: A review," in IOP conference series: earth and environmental science, 2017, vol. 58, no. 1, p. 012031: IOP Publishing.
[24] M. Hayaty, S. Muthmainah, and S. M. J. I. J. o. A. I. R. Ghufran, "Random and synthetic over-sampling approach to resolve data imbalance in classification," vol. 4, no. 2, pp. 86-94, 2020.
[25] S. Maldonado, C. Vairetti, A. Fernandez, and F. J. P. R. Herrera, "FW-SMOTE: A feature-weighted oversampling approach for imbalanced classification," vol. 124, p. 108511, 2022.
[26] M. Muntasir Nishat et al., "A comprehensive investigation of the performances of different machine learning classifiers with SMOTE-ENN oversampling technique and hyperparameter optimization for imbalanced heart failure dataset," vol. 2022, pp. 1-17, 2022.
[27] C. Rana, N. Chitre, B. Poyekar, and P. Bide, "Stroke prediction using Smote-Tomek and neural network," in 2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2021, pp. 1-5: IEEE.
[28] G. Ahmed et al., "Dad-net: Classification of alzheimer's disease using adasyn oversampling technique and optimized neural network," vol. 27, no. 20, p. 7085, 2022.
[29] A. Radhakrishnan, M. Belkin, and C. J. P. o. t. N. A. o. S. Uhler, "Wide and deep neural networks achieve consistency for classification," vol. 120, no. 14, p. e2208779120, 2023.
[30] Q. Zhang, J. Xiao, C. Tian, J. Chun-Wei Lin, and S. J. C. T. o. I. T. Zhang, "A robust deformed convolutional neural network (CNN) for image denoising," vol. 8, no. 2, pp. 331-342, 2023.
[31] J. Li, J. Shi, J. Chen, Z. Du, and L. J. F. i. O. Huang, "Self-attention random forest for breast cancer image classification," vol. 13, p. 1043463, 2023.
[32] E. J. W. C. Xi and M. Computing, "Image classification and recognition based on deep learning and random forest algorithm," vol. 2022, 2022.
[33] K. Thurnhofer-Hemsi, E. Lopez-Rubio, E. Dominguez, and D. A. J. I. A. Elizondo, "Skin lesion classification by ensembles of deep convolutional networks and regularly spaced shifting," vol. 9, pp. 112193-112205, 2021.
[34] M. A. Khan, T. Akram, Y.-D. Zhang, and M. J. P. R. L. Sharif, "Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework," vol. 143, pp. 58-66, 2021.
[35] F. Afza, M. Sharif, M. Mittal, M. A. Khan, and D. J. J. M. Hemanth, "A hierarchical three-step superpixels and deep learning framework for skin lesion classification," vol. 202, pp. 88-102, 2022.
[36] L. Hoang, S.-H. Lee, E.-J. Lee, and K.-R. J. A. S. Kwon, "Multiclass skin lesion classification using a novel lightweight deep learning framework for smart healthcare," vol. 12, no. 5, p. 2677, 2022.
[37] B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. J. S. R. Lakshmanna, "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," vol. 12, no. 1, p. 18134, 2022.
[38] M. Saarela and L. J. A. S. Geogieva, "Robustness, stability, and fidelity of explanations for a deep skin cancer classification model," vol. 12, no. 19, p. 9545, 2022.
[39] I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. J. E. Virvou, "Deep learning methods for accurate skin cancer recognition and mobile application," vol. 11, no. 9, p. 1294, 2022.
[40] G. Alwakid, W. Gouda, M. Humayun, and N. J. D. Jhanjhi, "Diagnosing Melanomas in Dermoscopy Images Using Deep Learning," vol. 13, no. 10, p. 1815, 2023.
[41] A. Tajerian, M. Kazemian, M. Tajerian, and A. J. P. o. Akhavan Malayeri, "Design and validation of a new machine-learning-based diagnostic tool for the differentiation of dermatoscopic skin cancer images," vol. 18, no. 4, p. e0284437, 2023.
[42] F. J. M. T. Bozkurt and Applications, "Skin lesion classification on dermatoscopic images using effective data augmentation and pre-trained deep learning approach," vol. 82, no. 12, pp. 18985-19003, 2023.
[43] F. Alenezi, A. Armghan, and K. J. E. S. w. A. Polat, "Wavelet transform based deep residual neural network and ReLU based Extreme Learning Machine for skin lesion classification," vol. 213, p. 119064, 2023.