ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 17, Issue 4, 2023

Comparison of the Convolution Neural Network and the Random Forest Classification Approach for the Detection and Classification of Melanoma Skin Cancer Images

[1]Dr.G Mahesh, [2] Dr.Reddeppa, [3]Dr.Anvesh

[1,2,3]Associate professor,

department of CSE,

Sree Institute of Technology,Tirupati

Abstract. The primary goal is to enhance the accuracy of skin disease and cancer diagnosis by replacing Random Forest with a Convolutional Neural Network. An open-access machine learning repository called uci is used to get the datasets. Each of the two groups uses twenty images of melanoma to make a diagnosis: one employs a Convolution Neural Network and the other a Random Forest. The maximum allowable inaccuracy is 0.8, which is half of the G power. With a significance level of 0.001, the suggested technique outperformed the first research utilizing Random forest (RF), which had an accuracy of 86.1%, sensitivity of 94.02%, and specificity of 87.11%. Convolution Neural Networks seem to be much more effective than Random Forests when it comes to skin cancer diagnosis, according to this perspective.

Topics covered include melanoma imaging, convolution neural networks, radiation fields, ML, DL feature extraction, novel approaches to skin cancer detection, and deep learning.

INTRODUCTION

Protecting cells from environmental hazards like heat and viruses is the job of the body's biggest organ, the skin [1]. The regulation of core body temperature is only one of its many functions; it also facilitates the storage of water and fat [2]. An infection with cancerous potential is among the worst skin problems that might occur. It is critical to diagnose and treat this illness early on to prevent it from spreading to other areas of the body [1]. Melanoma is the worst skin cancer, although any skin tumour might be lethal [2, 3]. To differentiate skin illness from melanoma, physicians typically use on clinical tests in addition to visual examinations of the sore's form, surface, and tune to establish the sore's benign or malignant nature [4].

From 62 papers found in Google Scholar and 51 publications provided by the IEEE, it was found that new skin cancer category had greater accuracy, sensitivity, and specificity. built a system that uses specificity and accuracy criteria to identify skin cancer early on [1]. The author used GLCM and the ABCD rule, two feature extraction approaches, after doing some preparatory processing. All three of these models are fed the recovered features: K Nearest Neighbour, Random Forest, and Support Vector Machine. Here, a deep learning-based convolution neural network model is suggested by the writer [5]. The writer may eliminate any extraneous noise by using a kernel or filter. developed an algorithm that, by examining surface, shape, balance, and shading characteristics, may predict the onset of skin cancer [6]. We have been very successful in several areas because we are committed to doing comprehensive research based on strong evidence: [7]. By including more layers into the design and training the network using deep neural networks, we can teach it a more complex classification model. Classifier performance should be improved as a result. Classification accuracy for new skin cancer detection is imprecise due to factors such as a small training sample, environmental impacts during data collecting, and input features that are noisy and uninformative. Because of that, it's crucial. The paper finds that the best way to increase classification accuracy is using supervised learning. The end objective is to make skin cancer detection via analysis more precise, which will allow for the early identification of skin diseases.

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 17, Issue 4, 2023

MATERIALS AND METHODS

In Chennai, India, researchers at SIMATS, an engineering school, conducted the study in the Department of Electronics and Communication Engineering. Information gathered from the publicly available UCI website. Both groups employ CNNs, however one uses a Random Forest (RF) while the other uses twenty melanoma photos for training [6]. The clinic has determined that for the 80% G power, a permitted error of 0.5 is the maximum [7].

Using a picture of the affected area as input data, this cutting-edge skin cancer detection system finds novel skin illnesses using machine learning algorithms. Our data is fed into a bag-of-features model that was trained using CNN as the classifier. This model performs adequately when tasked with object and texture identification. Through the app's user interface, the MATLAB program and a Windows tablet app are able to connect with one another. After taking a picture, the user may use MATLAB's pre-trained classifier to identify the skin condition.

In the first step of pre-processing, the picture is restored, undesirable hair is removed, and the image is enhanced. For improved precision, it's a solid processing strategy to use. Image editing to improve its quality. Improving and modifying the picture to get more accurate results is the first step in automated processing.

One aspect of feature extraction is the discovery of certain visual qualities that may be used for further processing. Following the transition from picture to non-pictorial datasets, this stage is an essential part of image analysis solutions. By feeding the results of a feature extraction step into a pattern recognition or classification strategy, you may label, classify, or label anything. Below, you will find more information on the following features that may be extracted from melanoma pictures.

Equation 1 explains how to get the mean, which is the average of the image input samples. The standard deviation (SD) is determined in Equation 2, which is relevant to the image processing operation, using the input data values (x).To calculate the variance in image processing, we take the average of all the values of the input data, where x is the variable being measured, as per Equation 3.For the EEG test signal, the third-order statistical feature is defined by the scenes feature, which is calculated from the signal's mean function (Equation 4).The symbol "F" represents the standard deviation of the signal.

$$=1 / n\sum_{i=1}^{n} V_1 \qquad (1)$$

$$\sigma = \frac{\sqrt{(X-\mu)}}{N} \qquad (2)$$

$$\sigma^2 = \frac{\sqrt{(X-\mu)^2}}{N} \qquad (3)$$

$$\mu_3 = \sigma^{-2} \sum_{i=0}^{N-1} (i-\mu)^3 p(i) \qquad (4)$$

$$\mu_4 = \sigma^{-3} \sum_{i=0}^{N-1} (i-\mu)^4 [p(i) - 3] \qquad (5)$$

$$m_k = E((x-\mu)^k \qquad (6)$$

Equation 5 provides the mean function for the EEG test signal and determines the fourth-order statistical characteristic, skewness. Using its multivariate properties, Equation 6 [8] shows the moment characteristic of the test EEG signal.

Data Analysis using Statistics

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN 2454-9940

www.ijasem.org

Vol 17, Issue 4, 2023

With the use of SPSS, we generated graphs, descriptive statistics (mean, standard deviation, and standard error), and a t-test for independence [10]. In this context, the values of the retrieved attributes are considered independent variables, whereas accuracy, specificity, and sensitivity are considered dependent variables.

RESULTS

Mean accuracy, standard deviation, as well as standard error rate were measured after 20 iterations in the statistical study, which compared CNN with Random Forest. According to the findings, CNN outperforms RF.

Compare the Random Forest classifier with the Convolution Neural Network classifier; after 20 iterations, the Random Forest classifier exhibited lower mean accuracy, standard deviation, overall standard error rate (Table 1).

Specification	Group.	N	Mean	Std. deviation	Std Error Mean
Accuracy	CNN	20	98.31	.0806	.0182
	RF	20	86.16	.00723	.0162
Specificity	CNN	20	89.11	.00721	.00161
	RF	20	87.24	.0072	.00162
Sensitivity	CNN	20	97.66	.0041	.0020
	RF	20	94.02	.0053	.0096

Table 2 displays the results of the independent t test for Specificity, Sensitivity, and Accuracy. We have computed the significance value for each of the 20 photos using Random Forest and Convolution Neural Network. In order to determine the 95% confidence intervals, a P value lower than 0.05 was considered statistically significant. There is a significant level of less than 0.05.

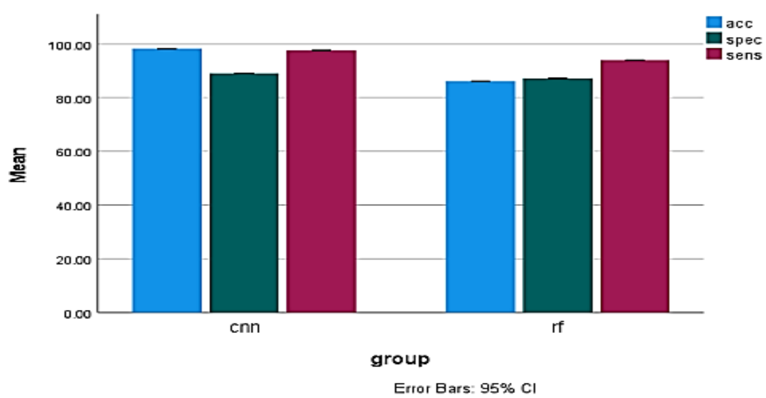

FIGURE. 1. We compare CNN and RF in terms of sensitivity, specificity, and accuracy in melanoma skin cancer images, and we also provide the error bars for the two groups. Blue indicates precision, green indicates specificity, and pink indicates sensitivity in relation to the two datasets. Both Mean Accuracy detecting with 95% CI and 1 SD and CNN versus RF are shown on the left side of the table.

Table 2 shows the results from an independent sample t-test, which were used to assess the sensitivity, specificity, and accuracy. Now you can use the Random Forest with the Convolution Neural Network to compare the significance values of the 20 photos. In order to calculate the 95% confidence intervals, a P value of less than 0.05 was considered statistically significant. The results support our idea, as the significance threshold is less than 0.05.

Levene's test for Equality of variance				T	Df	Sign (2-tailed)	T-test for equality of means		95% confidence interval of the difference	
		F	Sig				Mean	Std. Error Diff	Lower	Upper
Accuracy	Equal variance assumed	.102	.751	50.26	38	.001	12.14	.002	12.14	12.15
	Equal variance not assumed			50.26	37.58	.001	12.14	.002	12.14	12.15
specificity	Equal variance assumed	.109	.892	81.35	38	.001	1.86	.002	1.85	1.86
	Equal variance not assumed			81.35	38	.001	1.86	.002	1.85	1.86
sensitivity	Equal variance assumed		0.18	17.58	38	.001	3.64	.001	3.64	3.64
	Equal Variance not assumed			17.58	19	.001	3.64	.001	3.64	3.64

In Tables 1 and 2, you can see the outcomes of the group statistics. Compared to the typical significance range, the independent sample t test's significance level of 0.001 is lower. So, it's obvious that the first two groups are completely distinct from one other. Mean Convolution Neural Network outperformed Random Forest, with an accuracy of 98% and an error rate of 0.180, as shown in Table 1. Random Forest's accuracy was 86.17% and its error rate was 0.162. The following graph is shown in Figure 1: on the X-axis, it compares CNN and RF; on the Y-axis, it displays the Mean Accuracy detection of ± 1SD.

ISSN 2454-9940
www.ijasem.org
Vol 17, Issue 4, 2023

The outcomes of a statistical analysis conducted on twenty samples are reported in Table 1. In contrast to CNN's 0.80 standard deviations and 0.182 standard error, the RF method resulted in 0.072 standard deviations. Because the p-value is less than 0.001, we cannot reject the null hypothesis. The relationship between the independent variables (the input values) and the dependent variables (the outcomes of those changes) is seen in Table 2. Two algorithms' accuracy was compared using an independent t-test, and a statistically significant difference was found (p 0.001). The CNN model achieved a level of accuracy of 98%.

DISCUSSION

The different classifiers with the same findings have been compared to evaluate the proposed method that is tested with various datasets. The system uses CNN model and obtained the accuracy (98 %), sensitivity (97.6 %) and specificity (89.11 %) in Skin cancer detection.

Automated detection of diabetes using CNN and heart rate signals using machine learning algorithms obtained is 96 % accuracy [9].CNN is used to identify white blood cancer from microscopic pictures of bone marrow in deep learning and attained 97.2% accuracy [9]. The authors have implemented Breast cancer detection using the CAD tool used to diagnose the disease that obtained 99.1 % accuracy [7]. Congestive heart failure detection using Random Forest classifiers has achieved 99.9 % accuracy [8].

Based on the analysis, If the additional data includes extra qualities that can be employed in the present method, the performance level may be enhanced even further. Furthermore, it is assured that the proposed methodology can be developed using modified deep learning architecture.

CONCLUSION

Innovative skin cancer detection and classification of disease is proposed in this article From the result is observed that CNN provides significantly better performance in the detection of disease compared to RF algorithm .The results obtained using CNN have attained greater Accuracy (98 %), Sensitivity (97.46 %) and Specificity (89.11 %) than the Random Forest Accuracy (86.70 %), Sensitivity (93.4 %) and specificity (85.81 %).

REFERENCES

1. A. Murugan, S. A. H. Nair, and K. P. S. Kumar, "Detection of Skin Cancer Using SVM, Random Forest and kNN Classifiers," *J. Med. Syst.*, vol. 43, no. 8, Jul. 2019, doi: 10.1007/s10916-019-1400-8.
2. S. S. Mitra, "Management of skin cancer other than melanoma," *Practical Clinical Oncology*. pp. 485–498. doi: 10.1017/cbo9781107280250.038.
3. K. Lu *et al.*, "Dopamine Prevents Ultraviolet B-induced Development and Progression of Premalignant Cutaneous Lesions through its D Receptors," *Cancer Prev. Res.* , vol. 14, no. 7, pp. 687–696, Jul. 2021.
4. G. Millington, "The first purely open access dermatology journal from the British Association of Dermatologists and Wiley," *Skin Health and Disease*, vol. 1, no. 1. 2021. doi: 10.1002/ski2.2.
5. J.-H. Kim and J. Lee, "Controlling your contents with the breath: Interactive breath interface for VR, games, and animations," *PLoS One*, vol. 15, no. 10, p. e0241498, Oct. 2020.
6. B. Ghosh, S. Chatterjee, R. Dhakar, S. Muley, P. Mitra, and J. Chatterjee, "Arecanut-induced fibrosis display dual phases of reorganising glycans and amides in skin extracellular matrix," *Int. J. Biol. Macromol.*, vol. 185, pp. 251–263, Aug. 2021.
7. S. R. Samuel, S. Acharya, and J. C. Rao, "School Interventions-based Prevention of Early-Childhood Caries among 3-5-year-old children from very low socioeconomic status: Two-year randomized trial," *J. Public Health Dent.*, vol. 80, no. 1, pp. 51–60, Jan. 2020.
8. G. Ramkumar, A. G, S. K. M, M. Ayyadurai and S. C, "An Effectual Underwater Image Enhancement using Deep Learning Algorithm," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1507-1511, doi: 10.1109/ICICCS51141.2021.9432116
9. G. Anitha, P. Nirmala, S. Ramesh, M. Tamilselvi and G. Ramkumar, "A Novel Data Communication with Security Enhancement using Threat Management Scheme over Wireless Mobile Networks," 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2022, pp. 1-6, doi:

ISSN 2454-9940
www.ijasem.org
Vol 17, Issue 4, 2023

10.1109/
10. Kumar, A., Prabu, R.T. & Das, A. Configuration analysis of SnS based solar cells for high-efficiency devices. *Opt Quant Electron* **54**, 521 (2022). https://doi.org/10.1007/s11082-022-03940-0