IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 11, Issue 5, May 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105165|

Skin Disease Detection using Machine Learning

Ankit Sanjyal, Paresh Subedi, Dr S Vijay Kumar

Final Year B.Tech Student, Dept. of CSE, Jain University, Bangalore, Karnataka India

Associate Professor, Dept. of CSE, Jain University, Bangalore, Karnataka India

ABSTRACT: Dermatology is the area of biology that deals with the diagnosis and treatment of conditions that primarily affect the skin. Due to temperature, humidity, and other environmental conditions, the vast spectrum of dermatologic illnesses varies geographically as well as seasonally. Because of its unevenness, tone, hairiness, and other mitigating factors, human skin is one of the most unexpected and difficult surfaces to mechanically synthesise and assess. Only a small number of studies have focused on the medical paradigm of the problem, despite the fact that many studies use PC Vision techniques to identify and simulate human skin victimisation. Patients typically disregard early symptoms because there aren't any medical services in distant places, which could make the condition worse over time. Consequently, there is a growing need for high accuracy automatic skin disease detection systems. In order to distinguish between healthy skin and skin that has a disease, as well as to classify skin diseases into their main classes, such as melanocytic nevi, melanoma, benign keratoses-like lesions, basal cell carcinoma, actinic keratoses, vascular lesions, and dermatofibroma, we develop a multiclass deep learning model. We utilised Deep Learning to train our model. Deep Learning is a subset of machine learning, however unlike machine learning, it makes use of big datasets, which significantly reduces the number of classifiers. The machine self-learns, divides the supplied data into levels of prediction, and provides accurate findings in a very short amount of time, encouraging and supporting the growth of dermatology.Convolutional Neural Network (CNN) is one of the most used algorithms for picture categorization, thus that is the one we utilised.

KEYWORDS*:* Dermatoscopicimages, Deep Learning, Data Enhancement ,Convolutional Neural Network(CNN), Model Training, Testing and Evaluation.

I. INTRODUCTION

Even in the healthcare industry, artificial intelligence is replacing automation in all sectors of application. These diseases have caused worry in recent years because of their abrupt onset and complexity, which has elevated life risks. These skin abnormalities are highly contagious and must be treated early to stop them from spreading. Unprotected exposure to excessive ultraviolet radiation (UR) is a primary contributor to disease. Among all skin lesions, benign kind is thought to be less harmful than malignant melanoma and can be treated effectively, but malignant Melanoma is the most deadly variety.According to the survey's findings, skin cancer is most prevalent on the trunk, lower extremities, and upper extremities. There are many patients who are between the ages of 30 and 60. Additionally, Melanocytic Nevi, Cancer, and Dermatofibroma are not common in those under the age of 20.

II. EXISTING TECHNOLOGY

A. ArtificalNeuralNetwork(ANN).

A statistical nonlinear predictive modelling technique called an artificial neural network (ANN) is used to learn the intricate correlations between input and output. The biological pattern of our brain's neurons served as inspiration for the structure of ANN [2]. Three different types of computation node exist in an ANN. Through back-propagation, ANNs learn how to compute at each node. Two different types of data sets—trained and untrained data sets—are used to achieve accuracy using supervised and unsupervised learning techniques and various neural network topologies, such as the feed-forward and back-propagation methods, which utilise the data set in a unique way. The accuracy of 80% acquired using artificial neural networks in multiple studies is subpar [2]. Additionally, parallel processing-capable CPUs are needed for ANNs. Although ANN generates a probing solution, it does not explain why or how it occurs, which undermines network confidence.

ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105165|

B. BackPropagationNetwork(BPN).

Back propagation is a technique used in Artificial Neural Networks to determine each neuron's contribution to the error after processing a cluster of data (in image recognition, several images). Back propagation is very susceptible to erratic and noisy data. 75%–80% accuracy is attained with the BNN classifier [2]. Although BNN has advantages over other learning algorithms in terms of prediction and classification, its processing time is slower [5] [2].

C.SupportVectorMachine(SVM).

The supervised non-linear classifier SVM creates the best n-dimensional hyperplane to divide all the data points into two groups [2]. Choosing an honest kernel function in SVM is difficult. For large datasets, a lengthy training period is necessary. We are unable to make minor adjustments to the final model due to its difficulty of use, and tuning the SVM parameters becomes challenging. SVMs consistently produce the greatest outcomes when compared to ANNs [3].

III. LITERATURE

The fourth most frequent source of skin burden worldwide is skin diseases. In order to ease this burden and assist patients in doing an early assessment of the skin lesion, a reliable and automated approach has been devised.

This approach, which is mostly found in the literature, solely classifies skin cancer. Early detection and application of skin treatments result in greater efficacy and less disfigurement, but research is difficult since skin disorders share many traits. In this initiative, we try to find skin conditions.The most frequent skin lesions (melanocytic nevi, melanoma, benign keratosis-like lesions, basal cell carcinoma, actinic keratoses, vascular lesion, and dermatofibroma) are diagnosed using a revolutionary system that is provided in this research paper. Pre-processing, a deep learning algorithm, training the model, validation, and classification phases make up the suggested methodology. Convolution Neural Networks (CNN) and the Keras Application API were used in experiments on 10010 photos to achieve a seven-class classification accuracy of 93%.

V. DATASET

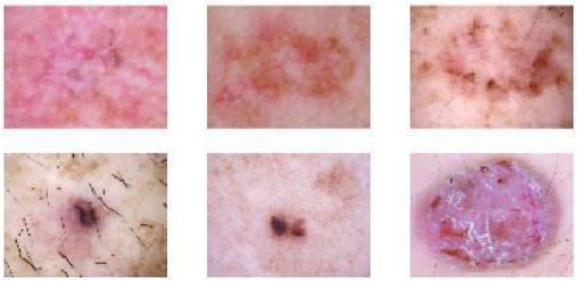

Fig.1.SampleData

The fig. above is the sample data setw hichwehavetrainedandtested.

I. IMPLEMENTATION(METHODOLOGY)

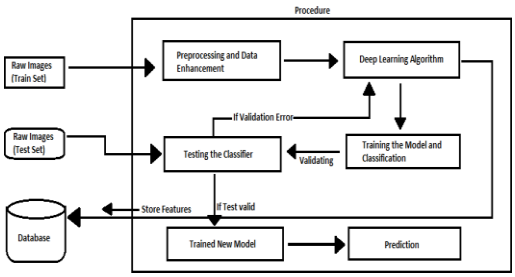

Fig.2.Procedure

ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105165|

Todevelop any ML-AIbasedsystem,beitthissystem;following stepsare tobe followed.

1. DataGathering.

Dermatoscopic pictures from a publically accessible dataset based on Skin-Cancer-MNIST (Modified National Institute of Standards and Technology Database)-HAM10000 were used to evaluate the suggested method. There are countless possibilities. It is possible to use publicly available data to save time and effort.*DataPreprocessing&Enhancement.*

The fundamental tenet of this stage is "Trash In- Good Out" [6]. By eliminating errors and unreliable information, the process can be sped up by validating your dataset using some simple profiling techniques [4]. When working with this data, AI systems don't produce fantastic results.

1.1. DataCleaning.

Unreliable and subpar output can be the result of dirty data, which can also confuse users. Data cleaning is therefore the first step in data pre-processing. Filling in missing numbers, reducing outliers from noisy data, and removing inconsistencies are all methods of cleaning data

.DataTransformation.

Data Transformation involves converting data from one format intoanother.Itinvolvestransformingactualvaluesfromonerepresentationtothetargetrepresentation.

1.2. ExploratoryDataAnalysis(EDA).

In this we explore different features of the dataset, theirdistributionsandactual counts.

1.3. LabelEncoding.

Thedatasetislabelled into7differentcategories:

1. MelanocyticNevi
2. Melanoma
3. Benignkeratosis-likelesions
4. Basalcellcarcinoma
5. ActinicKeratoses
6. Vascularlesions
7. Dermatofibroma

3. Training.

For this we have todivide the data into training set andtesting set. This division can be in any ratio. Also, the batchsizeand number ofepochshasto bedecided beforehand.

4. ModelBuilding.

Convolutional neural network (CNN) technology has been employed. Convolutional neural networks, often known as CNNs or ConvNets, are a subcategory of deep neural networks in which the computer learns on its own, divides the input data into prediction levels, and produces accurate results in a remarkably short amount of time [2]. A convolutional neural network (CNN) is a deep learning technique that consists of fully connected layers at the end, similar to a multilayer neural network, after a mix of convolutional and pooling layers in a predetermined order [2]. Among all other potential algorithms for classifying photos, CNN excels. To extract the best features, Sparse Connectivity, Shared Weights, and Pooling Feature are essential aspects. Additionally, the training time for deep learning techniques has decreased due to the introduction of Graphical Processing Units (GPUs). The public now has access to huge datasets of recently processed data and pre-trained networks.The figure below shows the difference between Sparse and DenseConnectivity.

ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105165|

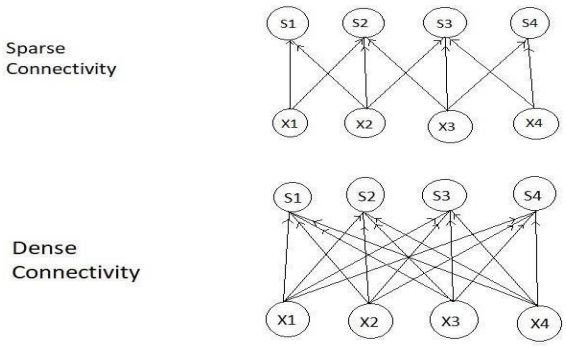

Fig.3.SparseandDenseConnectivity

4.1. Explanation.

We used the Keras Sequential API, which only requires that you add a layer at a time, beginning with the input. Learnable features from the Conv2D layer. There are 32 filters in use in this case. Each filter applies the kernel filter to a portion of the image that is determined by the kernel size. The filter maps are transformed photos. The pooling layer, which merely serves as a down sampling filter, is the following crucial layer. We have max pooling; maxpool() selects the pair of adjacent pixels with the highest value. This layer is used to scale back (to decrease) machine value and to somewhat lessen overfittingCNN may easily mix local features and learn global features by combining the two aforementioned layers. The network is given non-linearity by activation function relu. We employ a regularisation technique called the Dropout function, where a portion of the layer's nodes are arbitrarily ignored (their weights are set to zero) for each training sample. The network can be more broadly applied as a result. Now, we must flatten the final feature maps in order to turn them into a single 1D vector, which is why Flatten Layer is employed. In order to use fully connected layers after some of the levels above, this flattening step is necessary.It includes all of the native alternatives that the earlier convolutional layers have discovered. Dense().. We must create a score function, a loss function, and the appropriate optimisation method after the layers have been introduced. The error rate between the predicted labels and the observed labels is really measured by our loss function, which we describe as binary cross entropy. The optimizer is second in importance. The advantage of the Adam Optimizer is that it also uses the features of other optimizers. The Adam algorithm is well-known and well-liked in the learning models community. The next step is the metric function, which use metric accuracy to assess the system's performance. Another key term is learning rate (LR). It uses the annealing process. In order to lose as little as possible, it is ideal to have a falling learning rate during rate. It is called ReduceLROnPlateau, which literally translates to "reduce the LR to reach the global minimum of the loss function."

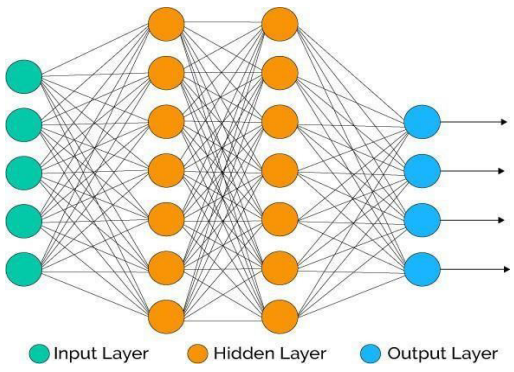

Fig.4.Fully Connected Network

5. ModelEvaluation.

"The more accurate the model, the better it is." The accuracy and loss acquired are used to evaluate each model. There are two levels of accuracy: test accuracy and validation accuracy. Prior to this, the Validation set was distinct from the Train set in that it was used to choose parameters and was not dependent on the Train set. For example, if your model has 90% train accuracy and 89% validation accuracy, it should have 89% accuracy on new data.

ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105165|

6. GraphicalAnalysis.

This involves plotting Histogram and the Confusion Matrix.TP, TN, FP, and FN are involved in the confusion matrix [2]. Correct decisions (true positives) and false positives are both included in the collection of classification algorithm decisions. False negatives are decisions that the classification algorithm declares to be negative but are actually positive. True negative decisions are ones that the classification system accurately classifies as negative.

VI. OVERVIEW

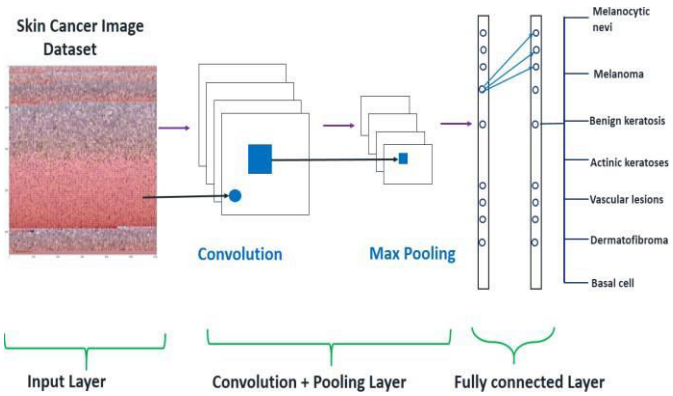

Fig.5.ConvolutionNeuralNetwork(CNN

VII. RESULTS

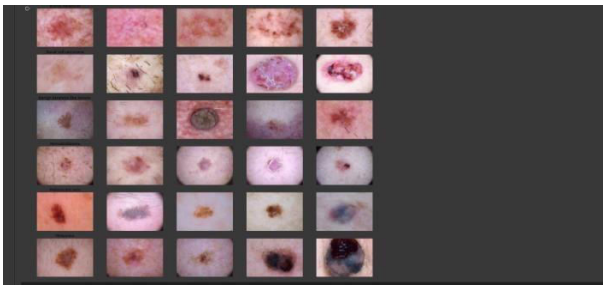

Fig 6. Streamed Images

```
tile_df = pd.read_csv(os.path.join(base_skin_dir, 'HAM10000_metadata.csv'))
tile_df['path'] = tile_df['image_id'].map(imageid_path_dict.get)
tile_df['cell_type'] = tile_df['dx'].map(lesion_type_dict.get)
tile_df['cell_type_idx'] = pd.Categorical(tile_df['cell_type']).codes
tile_df.sample(3)
```

	lesion_id	image_id	dx	dx_type	age	sex	localization	path	cell_type	cell_type_idx
6019	HAM_0000422	ISIC_0027664	nv	follow_up	40.0	male	trunk	/content/HAM10000_images_part_1/ISIC_0027664.jpg	Melanocytic nevi	4
9667	HAM_0005018	ISIC_0028008	nv	consensus	55.0	male	upper extremity	/content/HAM10000_images_part_1/ISIC_0028008.jpg	Melanocytic nevi	4
7009	HAM_0006145	ISIC_0031713	nv	histo	60.0	female	back	/content/ham10000_images_part_2/ISIC_0031713.jpg	Melanocytic nevi	4

Fig.7.ModelSummary

ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105165|

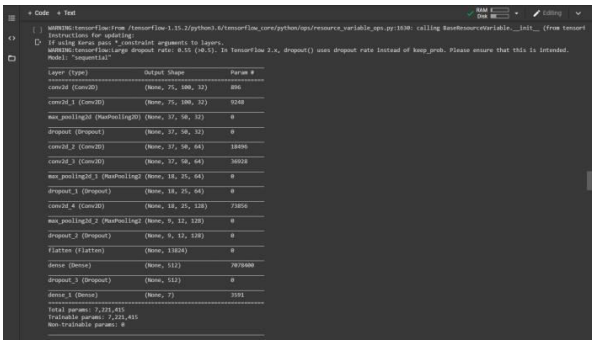

Fig.8.Epoch-50

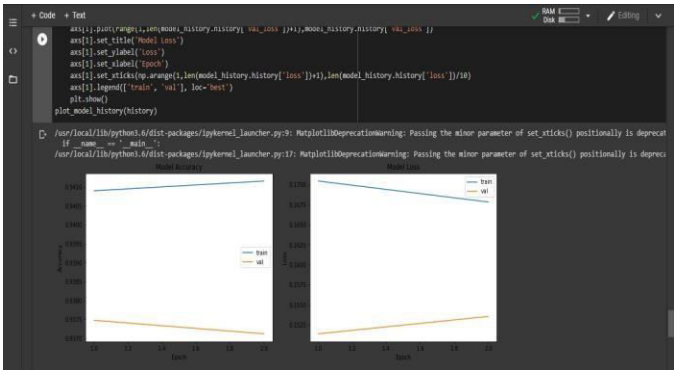

Fig.9.GraphicalPlottingforEpoch-2

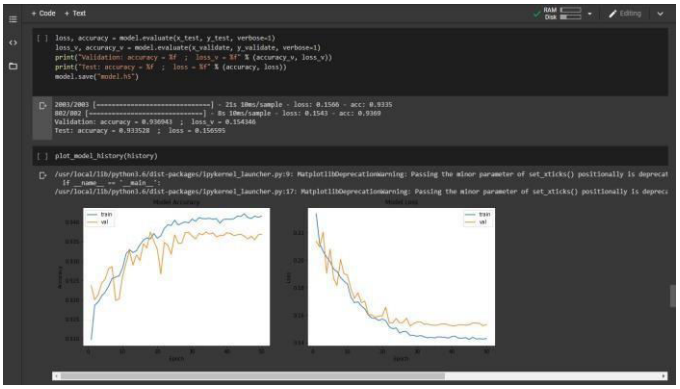

Fig.10.GraphicalPlottingforEpoch-50

Fig.12. Testing for an Image wih DetectionTABLEI.

Sr. No.	Evaluation		
	Metric/Parameter	*Testing*	*Validation*
1.	Accuracy	93.35%	93.35%
2.	Loss	15.65%	15.65%

For epochs – 2

Sr. No.	Evaluation		
	Metric/Parameter	*Testing*	*Validation*
1.	Accuracy	93.28%	93.28%
2.	Loss	16.01%	16.01%

For Epochs-50

ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 |

|| Volume 11, Issue 5, May 2023 ||

| DOI: 10.15680/IJIRCCE.2023.1105165|

VI. DISCUSSION

The proposed system aims in automatic computer-based detection ofSkin diseases so as to reduce life risks. This has been no doubt achallenging task owing to the fine-grained variability in theappearance of skin.

VII. CONCLUSION

Even though skin conditions are the fourth most common cause ofsickness in people, many people still avoid seeing doctors. Weprovided a reliable and automated approach for thedermatological illnesses' diagnosis. When skin conditions aretreated early on, they are less disfiguring andmore successful. Itis important to note that it is intended to replace doctors because nomachine has yet been able to fully replicate humanreasoning andintuition. For the first time, studies at the European Society of MedicalOncology have demonstrated that an AI or ML system mayreplace askilled dermatologist. This gives a quick overview of the system andthe implementation process.

VIII. ACKNOWLEGEMENT

Efforts have been made on this project. However, without the kind support and assistance of many people and organisations, it would not have been feasible. We want to express our sincere gratitude to each and every one of them.
We are very grateful to Dr. S. Vijaykumar, our internal guide, for his moral support and exemplary advice. Their insightful advice and timely suggestions encouraged us to continue working on our idea.
We would like to thank our parents and the other students at Jain University for their wonderful encouragement and cooperation, which enabled us to finish this project.
We also want to express our gratitude and appreciation to our project development team members and everyone else who volunteered their skills in order to assist us.

REFERENCES

[1] Shamsul Arifin, M., Golam Kibria, M., Firoze, A., Ashraful Amini, M., & Hong Yan. (2012). Dermatological disease diagnosis using color- skin images. 2012 International Conference on Machine Learning and Cybernetics. doi:10.1109/icmlc.2012.6359626). [2] Jana, E., Subban, R., & Saraswathi, S. (2017). Research on Skin Cancer Cell Detection Using Image Processing. 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC). doi:10.1109/iccic.2017.8524554.
[3] Mhaske, H. R., & Phalke, D. A. (2013). Melanoma skin cancer detection and classification based on supervised and unsupervised learning. 2013 International Conference on Circuits, Controls and Communications (CCUBE). doi:10.1109/ccube.2013.6718539.
[4] Alfed, N., Khelifi, F., Bouridane, A., & Seker, H. (2015). Pigment network-based skin cancer detection. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2015.7320056.
[5] Lau, H. T., & Al-Jumaily, A. (2009). Automatically Early Detection of Skin Cancer: Study Based on Nueral Netwok Classification. 2009 International Conference of Soft Computing and Pattern Recognition. doi:10.1109/socpar.2009.80.
[6] Dubal, P., Bhatt, S., Joglekar, C., & Patii, S. (2017). Skin cancer detection and classification. 2017 6th International Conference on Electrical Engineering and Informatics (ICEEI). doi:10.1109/iceei.2017.8312419

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor: 8.379

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details