Hybrid Deep Learning Model for Skin Cancer Detection Based on CNN and LSTM

KHYATI H. YADAV
PG Student, Department of Electronics & Communication Engineering
C. G. Patel Institute of Technology,
Uka Tarsadia University, Bardoli, Surat

Dr. MAYANK R. KAPADIA
Assistant Professor, Department of Electronics & Communication Engineering
C. G. Patel Institute of Technology,
Uka Tarsadia University, Bardoli, Surat

Abstract

Skin cancer can be classified into two primary categories: melanoma and non-melanoma. Since melanoma lesions are the most harmful and destructive of all the lesion kinds, they are to blame for the notable rise in mortality and morbidity that has occurred recently. The skin cancer detection techniques have various phases which include pre-processing, segmentation, feature extraction and classification. In the previous year's various transfer learning models are proposed for the skin cancer detection but those techniques are unable to achieve great accuracy. In this research work, hybrid deep learning model is proposed which is the combination of CNN and LSTM. The proposed model is implemented in python and results are analyzed in terms of accuracy, precision and recall. It is analyzed that proposed model achieves great accuracy as compared to existing models.

Keywords

Skin cancer detection, CNN, LSTM, Transfer learning

1. Introduction

The main method for diagnosing skin cancer is a visual examination performed by dermatologists, which usually has an accuracy rate of about 60%. Dermoscopy use raises diagnosis accuracy to 89%, a substantial improvement [1]. With detection rates of 82.6% for melanocytic lesions, 98.6% for basal cell carcinoma, and 86.5% for squamous cell carcinoma, dermoscopy also exhibits great sensitivity. Even though dermoscopy has significantly increased the accuracy of melanoma diagnosis, there are still difficulties in correctly recognizing some lesions, especially early melanomas that lack distinguishing dermoscopic features [2]. Dermoscopy is a highly successful diagnostic tool for cutaneous melanoma; nevertheless, its applicability in detecting featureless melanomas is restricted. This highlights the need for additional improvements in diagnostic accuracy in order to improve patient survival rates [4]. Aiming for greater diagnostic precision in the identification of skin cancer, dermoscopy's shortcomings prompted the creation of computer-aided detection techniques. As seen in Figure 1 [4], the five main processes of computer-aided skin cancer detection are picture acquisition, pre-processing, segmentation, feature extraction, and classification.

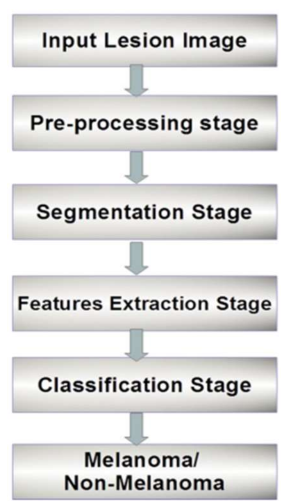

Figure-1: Basic Workflow of Skin Cancer Detection [4]

Pre-processing techniques include morphological procedures, color gray-scaling, binarization, data augmentation, and adjustments to contrast and intensity. In this phase, unwanted components like sounds and artifacts are removed from the pictures. Resizing the photos also contributes to image uniformity and usually lowers computing complexity. Right after pre-processing, there is an image segmentation step that creates regions of interest by separating diseased from healthy tissue. To ensure an accurate diagnosis, the segmentation technique is essential for separating normal tissues before obtaining features from the lesions. Traditional segmentation methods, including thresholding, clustering, edge and region-based techniques, along with the widely recognized ABCDE approach, have traditionally been employed for analyzing skin lesions in melanoma detection [5]. However, these techniques face challenges in effectively handling the intricate visual characteristics of skin lesions and may struggle to accurately segment the melanoma region. In recent times, segmentation methods based on intelligent approaches, such as fuzzy logics, genetic algorithms [6], Artificial Neural Networks (ANN), and more recently, deep learning techniques, have been increasingly utilized for achieving precise and reliable segmentation of skin lesions. Within the field of artificial intelligence, machine learning enables computer systems to acquire knowledge through direct observation of examples, data, and experience. Machine learning systems are capable of carrying out complex processes by learning from data instead of adhering to pre-programmed rules [7], by empowering computers to do certain jobs intelligently. Exciting developments in machine learning over the past few years have increased its potential in a variety of applications [8]. Machine learning systems can now be trained on a vast amount of data thanks to increased data availability, and their analytical skills are reinforced by rising computer processing power. Algorithmic developments within the area itself have also increased the power of machine learning. These developments have allowed systems to operate at levels that were clearly below human capabilities just a few years ago, but which are now capable of outperforming people on certain tasks. Machine learning-based systems are now widely utilized by individuals in their daily lives. Examples of these systems include recommender systems used by online shops, speech recognition systems used by virtual personal assistants, and image recognition systems used on social media. As machine learning continues to grow, it has the potential to facilitate revolutionary developments in a number of fields, with enormous societal and financial implications. Unsupervised, supervised, and semi-supervised learning are the three general categories into which machine learning algorithms fall [9]. Learning is referred to as supervised when cases are provided with known labels, or the matching right outputs, as opposed to unsupervised learning, which uses unlabeled instances. Labeled and unlabeled examples are combined in semi-supervised learning to create a suitable function or classifier.

2. Literature Review

S Maqsood, et.al (2023) outlined how smart healthcare makes clever changes to present medical practices to make them more effective, dependable, and personalized using next-generation technology like artificial intelligence (AI) and the Internet of Things (IoT) [10]. Tele dermatology is one of the most well-known applications of telemedicine and e-health in medical image analysis. Medical information is sent to specialists in this business via telecommunications technologies. Because skin is visually observable, tele dermatology is a good tool for identifying skin lesions, especially in rural areas. Dermoscopy is one of the newest methods for identifying skin cancer. Many computational methods have been developed in the literature to classify skin cancers. Lesions with low contrast, imbalanced datasets, a high degree of memory complexity, and the extraction of redundant features are some of the challenges that still need to be overcome.

K M. Hosny, et.al (2024) suggested a deep, innate learning technique for categorizing seven different kinds of skin lesions [11]. Several explanation strategies were employed to validate the suggested deep intrinsic learning. Both local and global decision-making processes were explained using Explainable AI (X-AI). Furthermore, we offer graphic data to support clinicians' confidence in the suggested approach. The suggested approach was tested on the difficult dataset, HAM10000. With our straightforward, stage-based X-AI architecture, medical professionals can gain a deeper understanding of the workings of black-box AI models. Because the suggested strategy explains the reasoning behind its choices

U. Gangan, et.al (2023) suggested a step-by-step method for detecting skin cancer that uses the VGG16 architecture as its foundation and other layers to classify skin lesions into multiple classes [12]. Our approach seeks to provide a reliable and effective method for detecting skin cancer that can distinguish between seven different kinds of skin lesions. By putting this strategy into practice, skin cancer might be identified early and treated promptly, improving patient outcomes and improving public health. Our research provides a viable way to address the drawbacks of resource-intensive models and manual feature extraction by focusing on multiclass classification and employing the VGG16 architecture.

M Ahammed, et.al (2022) presented a method for digitally eliminating facial hair that relies on morphological filtering, such as the Black-Hat transformation and painting algorithm, and uses Gaussian filtering to de-noise or de-blur the images [13]. Furthermore, we segment the impacted lesions using the automatic Grab cut segmentation method. We use statistical features and the Gray Level Co-occurrence Matrix (GLCM) approaches to uncover underlying input patterns from the skin photos. The skin images are effectively classified as melanoma (MEL), melanocytic nevus (NV), basal cell carcinoma (BCC), actinic keratosis (AK), benign keratosis (BKL), dermatofibroma (DF), vascular lesion (VASC), and squamous cell carcinoma (SCC) using three computationally efficient machine learning techniques: Decision Tree (DT), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) classifiers. These classifiers are applied using the extracted features. ISIC 2019 Challenge and HAM10000 are two common datasets that are used to validate the models. SVM outperforms the other two classifiers by a small margin. We have also made a comparison between our work and the latest techniques.

R R. Babu, et.al (2024) created the Jaya Artificial Ecosystem-based Optimization-LeNet (JAEO-LeNet) strategy for using skin pictures to detect skin cancer early on [14]. The input skin picture is denoised using the bilateral filter, allowing TransUNet to be used for skin lesion segmentation. Segmented skin lesion images are augmented, and multiple features are extracted from the augmented image. LeNet, a deep learning model, is used to identify skin cancer using the attributes that are extracted. The LeNet weights are suggested to be adjusted using the JAEO algorithm in order to improve detection performance. The validation of skin cancer detection is done using the SIIM-ISIC Melanoma categorization database. Furthermore, the JAEO-LeNet achieved its peak performance with 91.99% accuracy, 90.95% sensitivity, and 92.13% specificity.

Karar Ali, et.al (2022) provided Confusion Matrices for all eight models based on the classification results for each class [15]. The Efficient Net B4, our top model, obtained an F1 Score of 87% and a Top-1 Accuracy of 87-91%. We assessed Efficient Net classifiers with metrics accounting for the high-class imbalance. Our results suggest that better classification performance is not necessarily correlated with higher model complexity. Models with intermediate complexity, like Efficient Net B4 and B5, produced the best results. Numerous aspects, including resolution scaling, data augmentation, noise reduction, effective transfer learning of ImageNet weights, and fine-tuning, contributed to the outstanding classification results. Another finding revealed that, when employing Confusion Matrices, some classes of skin cancer performed better at generalization than others.

S. Rasool, et.al (2022) presented a model that uses Bootstrapping Ensembles and Convolutional Neural Networks (BE-CNN) to identify and label skin lesions [16]. The study's authors came up with this notion. This method (improved-SN) is based on the Computer-Intensive Segmentation Network (CI-SN). However, by generating pre-bootstrapping uneven lesion covers, the Compute-Intensive Segmentation Network is able to identify and classify skin lesions. The division and arrangement networks are supposed to collaborate and share knowledge with one another as a result of this strategy. A "bootstrapping" procedure is utilized to do this. We arrived at this conclusion after noticing that the proposed model outclassed the commonly used condition and stages-based approaches in terms of how well skin lesions were classified into their respective stages. The results show that skin lesions in a connected model can be partitioned and classified using a continuous bootstrapping technique.

P Thapar, et.al (2024) proposed a workable method for detecting skin cancers from dermoscopy images, enhancing experts' capacity to discriminate between benign and malignant lesions [17]. Dermoscopy photos were employed in the Swarm Intelligence (SI) technique to identify lesions on the skin areas inside the Region of interest (ROI). The best segmentation results were obtained using the Grasshopper Optimization approach. Based on these results, the Speed-Up Robust Features (SURF) technique is used to extract features. Skin cancers were categorized into two groups using the ISIC-2017, ISIC-2018, and PH-2 databases. The proposed segmentation and classification methodologies have been assessed for classification efficacy, specificity, sensitivity, F-measure, preciseness, the MCC, the dice coefficient, and Jaccard's index. The results show an estimated 98.52% classification accuracy, 96.73% preciseness, and 97.04% Matthews Correlation Coefficient (MCC). The approach we propose outperformed the state-of-the-art approaches in all performance metrics.

3. Research Methodology

The introduced model is the transfer learning approach that is the combination of Efficient net and Bidirectional LSTM. The various phases of proposed model are explained below: -

1. Input image and Pre-process: The skin disease images are employed for input and Gaussian filter utilizes to pre-process it. This filter will reduce noise from the image This filter makes images non-blurry and is also known as a smoothing operator. It eradicates intrinsically available fine image details. Its impulse response refers to a Gaussian function (GF) that is used to outline the probability distribution of the noise. It efficiently removes Gaussian noise. This filter is linear and lower pass having a GF of a given standard deviation.

2. Segmentation: The technique of snake segmentation will be applied for segmenting the region of interest. The Snake segmentation technique is inspired from the raster scan due to which it will cover maximum edges of the image SAC algorithm [6-8] is employed for modelling a parameterized primary contour curve in the image space, and an energy function (EF) is put forward to characterize the shape of the area in accordance with the internal and external power. The features of curve help to determine the first one and the attributes of image assist in describing the external energy including curvature, curve length, etc. EF is diminished to converge the primary contour curve $C(s) = (x(s), y(s), s \in [0,1])$ continuously to the boundary of the destination region in the restraints of both energies:

$$E(C) = \int_0^1 \alpha E_{int}(C(s)) + E_{img}\left(C(s) + \gamma E_{con}(C(s))\right) ds$$

Three portions are included in EF such as E_{int} uses to illustrate the internal energy for ensuring that the curve if smooth and regular; E_{img} is utilized to denote the image energy, assigned in accordance with the desired position attributes like edges; the constrained energy is represented with E_{con}. SAC algorithm is useful as the geometric restraints are taken in account. Avoiding the quality of image, the major focus is on extracting the closed boundaries. However, some limitations are occurred still. The challenging task is of tackling the region due to its dependence on the first contour. The position, shape and number of control points are capable of acquiring the preferred impact only in case of selecting an appropriate primary contour.

3. Classification: To classify the skin disease the approach of Efficient net and Bidirectional LSTM. Efficient Net is a Convolutional Neural Network (CNN) model in which compound coefficients are employed to perform effectual scaling. This algorithm is capable of extracting significant and typical features from images. Moreover, this algorithm is implemented for capturing applicable features from huge-scale datasets. Several computer vision applications employ this algorithm to accomplish diverse tasks such as to recognize an object and segment the data. This algorithm is performed robustly and accurately to segment the images. The architecture of this algorithm plays a significant role in scaling up the dimension of width, resolution and depth of resources which are present in a continuous ratio. This algorithm is composed of Mobile Inverted Bottleneck (MB Conv) layers in which depth-wise separable convolutions are integrated with inverted residual blocks. Moreover, the Squeeze-and-Excitation (SE) optimization is assisted in enhancing performance of this algorithm.

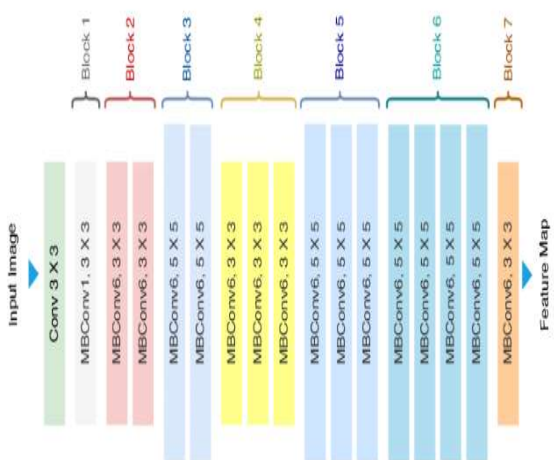

Figure-2: Efficient Net Architecture

The MB Convolution layer, a key element of the Efficient Net framework, draws inspiration from the inverted residual blocks found in MobileNetV2 but with modifications. It begins with a depth-wise convolution, succeeded by a 1x1 convolution to expand channel numbers, and concludes with another 1x1 convolution to reduce channels back to their original count. This bottleneck approach preserves model expressiveness while facilitating efficient training. In addition to the MB Convolution layers, Efficient Net incorporates the SE block, which helps the model prioritize crucial features while suppressing less relevant ones. The SE block employs global average pooling to condense the feature map's spatial dimensions to a single channel, followed by two fully connected layers. These layers enable the model to grasp channel-wise feature relationships, generating attention weights that amplify significant information by applying them to the original feature map through multiplication. Efficient Net offers multiple versions, such as B0, B1, etc., with different scaling coefficients. Users can select the most suitable variant by weighing the trade-offs between model size and accuracy that each version offers. Moreover, Efficient Nets aim to strike a balance between performance and model size, showcasing superior efficiency in memory usage and computational resources when compared to other deep learning architectures. The output of efficient net will be given as input to bidirectional for the classification. While theoretically capable of processing sequences of any length, practical limitations arise due to issues like vanishing or exploding gradients, causing RNNs to effectively look back only a few steps. Long Short-Term Memory (LSTM) networks address these challenges by introducing purpose-built memory cells, known as LSTM cells. In LSTM networks, these memory units consist of an input gate, a forget gate, and an output gate, which replace the conventional hidden layer units found in RNNs. These memory units enable the network to decide whether to retain or forget the information in the output, facilitating the retention of long sequences of computational information. Furthermore, LSTM introduces these three gates to mitigate gradient vanishing problems, enhancing its ability to remember information over extended periods. In contrast to the traditional LSTM, a Bidirectional LSTM (Bi-LSTM) is an LSTM variant that processes input data in two directions: forwards and backwards. This bidirectional approach harnesses information from both directions, enabling the model to capture and learn from the input sequence in a more comprehensive manner. In standard LSTM, information is only learned in a unidirectional manner, moving sequentially from one part of the sequence to another. However, with Bi-LSTM, the LSTM model has the capability to simultaneously learn from the input sequence in both the forward and backward directions, resulting in a richer understanding of the data.

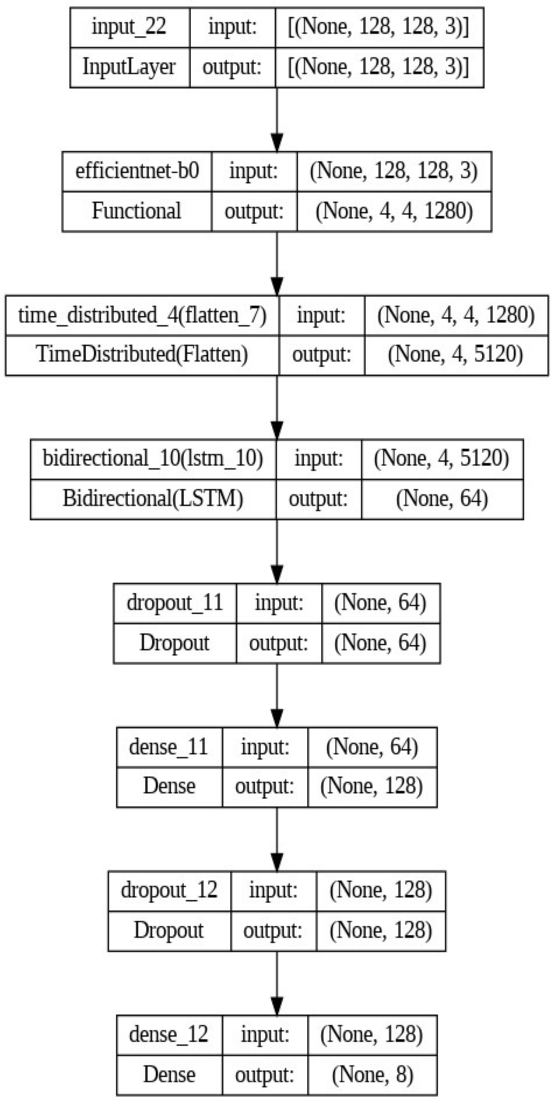

Figure-3: Proposed Methodology

As shown in figure 3, the proposed model is illustrated which is the combination of efficient net and bidirectional LSTM.

4. Result and Discussion

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python's elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

4.1 Performance Analysis Parameters

The detail of performance analysis parameters is given below:

- **Accuracy:** Accuracy quantifies the efficacy of evidence field recovery and data processing. It represents the proportion of correctly classified results, expressed by the following equation:

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

- **Precision:** Precision is an evaluation metric that gauges the ratio of accurately identified positives to the total number of identified positives. This can be represented as follows:

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

- **Recall:** Recall, also termed sensitivity, indicates the proportion of pertinent cases retrieved out of the total number of pertinent cases. It can be calculated in the following manner

$$Recall = \frac{TP}{TP + FN}$$

4.2 Results

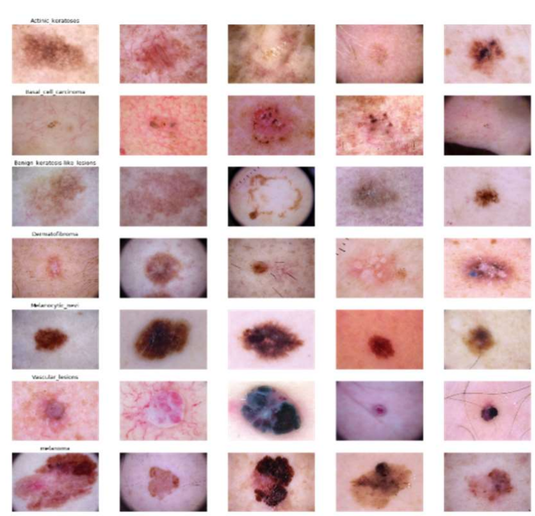

Figure-4: Dataset Images

As shown in figure 4, the dataset is collected is from MNIT and it is related to skin cancer. The dataset has two classes which is cancer or non-cancer. The dataset images are illustrated for the classification.

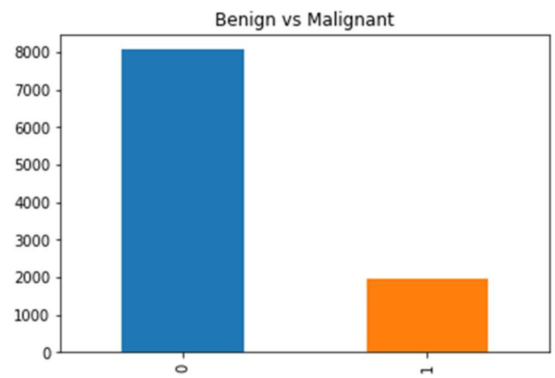

Figure-5: Dataset Class Distribution

As shown in figure 5, the dataset has two classes which are cancer or non-cancer it can also be called as benign and malignant. In the dataset non-cancer class has large data as compared to cancer class.

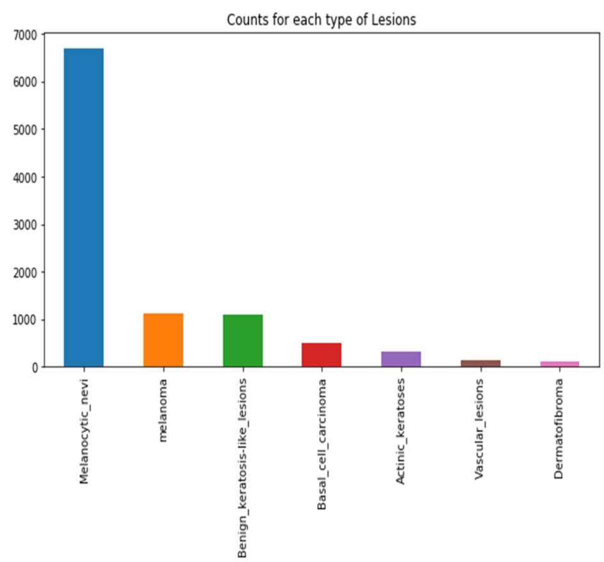

Figure-6: Type of Lesions

As shown in figure 6, in the dataset there are seven lessons in the dataset. The number of instances of each lesion is illustrated in the dataset.

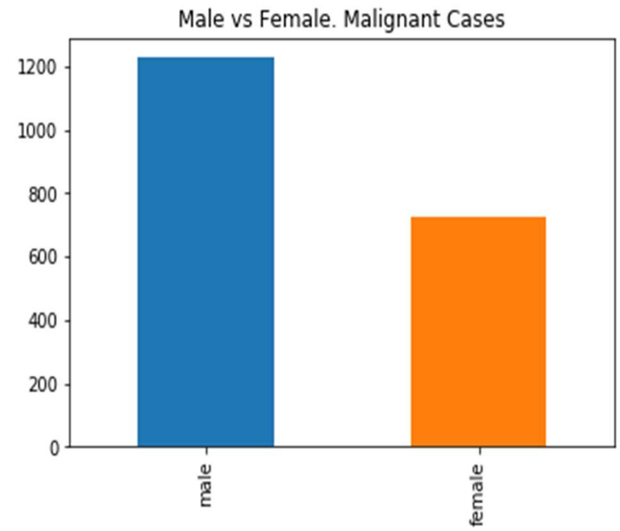

Figure-7: Male and Female Data Distribution

As shown in figure 7, the dataset is of males and females for the skin cancer detection. It is analysed from the data that male data is more in number as compared to female.

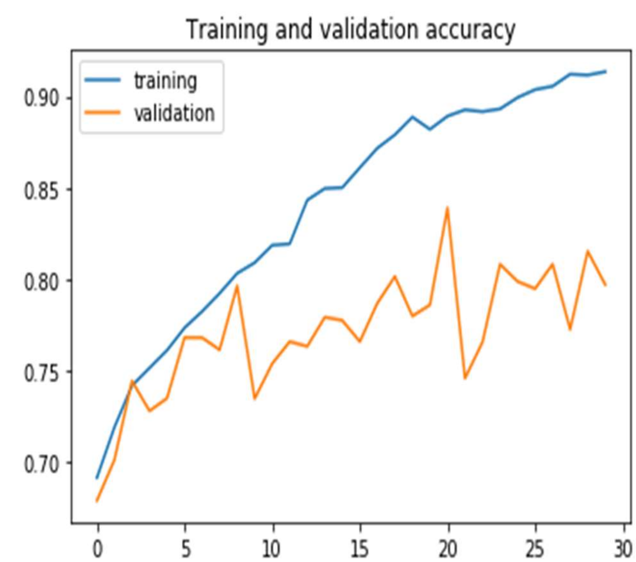

Figure-8: Training and validation accuracy

As shown in figure 8, the training and validation accuracy of the proposed model is show which is above 93 percent.

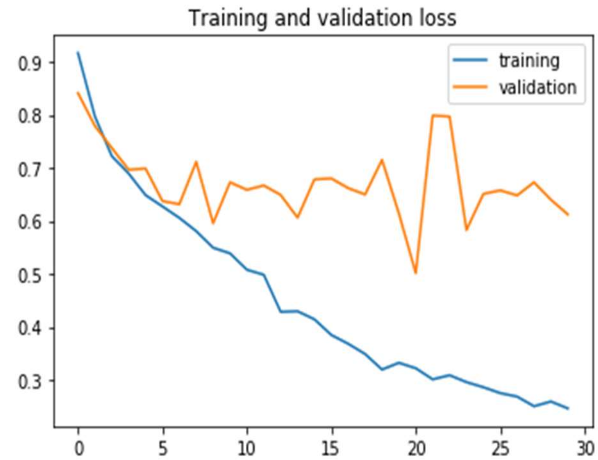

Figure-9: Training and validation loss

As shown in figure 9, the training and validation loss of the proposed model is shown. It is analysed that is below 8 percent

Table-1: Performance Analysis

Model	Accuracy (%)	Precision (%)	Recall (%)
Fine Tuning with VGG 16	83.30	83	83
Fine Tuning Inception	81.10	81	81
Dense Net	77.89	77	77
Proposed Model	92.10	92	92

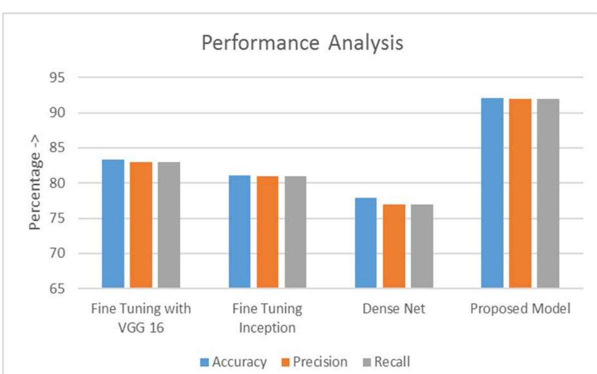

Figure-10: Performance Analysis

As shown in figure 10, the performance analysis of proposed model is compared with VGG16, Inception and dense net. The proposed model has maximum accuracy, precision and recall as compared to existing models.

Conclusion

Skin cancer is a severe health issue. Computer-based technology provides a relaxed, inexpensive and quick diagnosis of skin cancer symptoms. Several techniques, non-invasive in nature, have been proposed to investigate the symptoms of skin cancer, whether they represent melanoma or non-melanoma. The general process applied in skin cancer detection is image acquisition, pre-processing, segmentation of the acquired pre-processed image, extraction of the required features and classification of disease based on features extracted. In the previous year many machine learning algorithms has been proposed which can detect skin cancer but the accuracy of the models is less. The proposed model is the hybrid deep learning model for skin cancer detection. The proposed model is the combination of CNN and LSTM and it achieved accuracy of upto 92 percent which is approx. 6 percent high than existing models.

References

[1] M. Hossain, K. Sadik, M. M. Rahman, F. Ahmed, M. N. Hossain Bhuiyan and M. M. Khan, "Convolutional Neural Network Based Skin Cancer Detection (Malignant vs Benign)," 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2021, pp. 0141-0147

[2] S. Singh and J. J. Zhao, "A deep skin cancer classification approach using image and structured information," 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Niagara Falls, ON, Canada, 2022, pp. 930-933

[3] H. L. Gururaj, N. Manju, A. Nagarjun, V. N. M. Aradhya and F. Flammini, "DeepSkin: A Deep Learning Approach for Skin Cancer Classification," in IEEE Access, vol. 11, pp. 50205-50214, 2023

[4] M. Shah, "LRNet: Skin Cancer Classification using Low-Resolution Images," 2021 International Conference on Communication information and Computing Technology (ICCICT), Mumbai, India, 2021, pp. 1-5

[5] N. P, S. K. S, V. Saijagadeeshkumar, G. P. Kumar, K. Yerriswamyreddy and K. J. Krishna, "Skin Cancer Detection and Control Techniques Using Hybrid Deep Learning Techniques," 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, 2023, pp. 442-446

[6] G. S. Jayalakshmi and V. S. Kumar, "Performance analysis of Convolutional Neural Network (CNN) based Cancerous Skin Lesion Detection System," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-6

[7] L. Song, J. P. Lin, Z. J. Wang, and H. Wang, "An end-to-end multi-task deep learning framework for skin lesion analysis," IEEE J. Biomed. Health Inform., vol. 24, pp. 2912–2921, 2020.

[8] S. Aziz, M. Bilal, M. U. Khan, and F. Amjad, "Deep learning-based automatic morphological classification of leukocytes using blood smears," in Proc. 2020 Int. Conf. Elect. Commun. Comput. Eng. (ICECCE), Istanbul, Turkey, Apr. 2020, pp. 1–5.

[9] Y. Ren, L. Yu, S. Tian, J. Cheng, Z. Guo, and Y. Zhang, "Serial attention network for skin lesion segmentation," J. Ambient Intell. Humaniz. Comput., pp. 1–12, 2021.

[10] S Maqsood, Robertas Damaševičius' "Multiclass skin lesion localization and classification using deep learning-based features fusion and selection framework for smart healthcare, Neural Networks", 2023, Vol. 160, Pages 238-258, sciencedirect.com/science/article/pii/S0893608023000229

[11] K M. Hosny, Wael Said, Mahmoud Elmezain, Mohamed A. Kassem, "Explainable deep inherent learning for multi-classes skin lesion classification, Applied Soft Computing", 2024, Vol. 159, 111624, doi.org/10.1016/j.asoc.2024.111624.

[12] U. Gangan, M. V and R. T. Rajasekaran, "Skin Cancer Detection using Machine Learning and Deep Learning*," 2023 International Conference on System, Computation, Automation and Networking (ICSCAN),* PUDUCHERRY, India, 2023, pp. 1-6, doi: 10.1109/ICSCAN58655.2023.10394907.

[13] M Ahammed, Md. Al Mamun, Mohammad Shorif Uddin, "a machine learning approach for skin disease detection and classification using image segmentatio, Healthcare Analytics", 2022, Vol. 2, 100122, doi.org/10.1016/j.health.2022.100122.

[14] R R. Babu, Felix M. Philip, "Optimized deep learning for skin lesion segmentation and skin cancer detection, Biomedical Signal Processing and Control", 2024, Vol. 95, Part B, 106292, doi.org/10.1016/j.bspc.2024.106292.

[15] Karar Ali, Zaffar Ahmed Shaikh, Abdullah Ayub Khan, Asif Ali Laghari, "Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer, Neuroscience Informatics", 2022, Vol. 2, Issue 4, 100034, doi.org/10.1016/j.neuri.2021.100034.

[16] S. Rasool, U. N. Dulhare, M. N. Khan, D. Gangodkar, A. Rana and R. Kalra, "Automated Multiclass Classification Using Deep Convolution Neural Network on Dermoscopy Images," *2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS),* Tashkent, Uzbekistan, 2022, pp. 711-716, doi: 10.1109/ICTACS56270.2022.9988394.

[17] P Thapar, Manik Rakhra, Mahmood Alsaadi, Aadam Quraishi, Aniruddha Deka, Janjhyam Venkata Naga Ramesh, "A hybrid Grasshopper optimization algorithm for skin lesion segmentation and melanoma classification using deep learning, Healthcare Analytics", 2024, Vol. 5, 100326,