International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Computer-Aided Detection of Skin Cancer Detection from Lesion Images Via Deep Learning Techniques: 3d CNN Integrated Inception V3 Networks

Manju C. P.[1],Jeslin P. Jo[2] ,Vinitha V.[3]

Submitted: 26/04/2023 **Revised**: 27/06/2023 **Accepted**: 06/07/2023

Abstract: A rising number of hereditary and metabolic peculiarities leads to malignant growth, in most parts of the body. The spread of destructive cells might be risky in any area of the body. One of the most well-known diseases is skin malignant development, and its recurrence is spreading around the globe. The three primary subtypes of skin cancer are squamous, basal cell, and melanoma, with melanoma being the most deadly and clinically aggressive. In this way, it is essential to do skin cancer screenings. One of the most impressive techniques for quickly and effectively identifying the development of skin cancer is the use of Deep Learning (DL). As a result, some steps are employed in this research to identify harmful skin: a) Information gathering using the HAM10000 dataset which includes of 10015 lesion images, b) Preprocessing of raw Inge to avoid anomalies using techniques like CLAHE, FMM and Frangi vessel, c) feature extraction using convolutional Autoencoder and finally d) a 3D CNN network for classifications and transfer learning with an Inception V3-trained model to increase the network's knowledge. Experimental evidence demonstrates that, on a variety of metrics, the suggested algorithm performs more effectively than other state-of-the-art systems (accuracy:0.96, sensitivity:0.97, specificity:0.97).

Keywords*: Autoencoder,CNN, CLAHE, FMM, Frangi vessel filter, Inception v3 Lesion, Melanoma, Transfer learning*

1. Introduction

Melanoma, the far more dangerous variety of skin cancer, spreads rapidly if neglected. Melanosomes, the cells that create the pigments known as melanin, which gives skin its color, are where it all begins[1]. When it reaches the dermis, it can circulate and disseminate to different body parts (lower layer of the skin). Percutaneous melanomas, or melanomas that develop on the skin, constitute the majority typical type of melanomas. Melanoma can occasionally begin as a mole, and if caught early enough, it can be efficiently cured [2]. Melanoma, a deadly kind of skin cancer that arises when melanocytes experience a malignant transformation, spreads fast throughout the body in contrast to neural crest neoplasia. There were 18.1 million new instances of cancer in 2018, making it one of the deadliest diseases, and 9.5 million of these are appropriate for evaluating fatalities. The National Cancer Institute (NCI) predicts that by 2040, there would be 29.5 million new instances of cancer, leading to 16.5 million deaths. In keeping with the National Cancer Institute's latest SEER (“Surveillance, Epidemiology, and End Result”) research, the five-year survival rate for cutaneous melanoma from 2010 to 2016 was 92.7 per cent. Melanoma skin cancer was primarily observed more in adults ages from 21 and 39, based on a newly released SEER study. Figure 1 shows the various skin cancer variants.

[1]*Assistant professor, Department of Electronics and Communication, Federal Institute of Science and Technology, email:manju.peetamber@gmail.com*
[2]*Assistant Professor, Department of Electronics and Communication, Federal Institute of Science and Technology,email: jeslin@fisat.ac.in*
[3]*Assistant Professor, Department of Electronics and Communication, Federal Institute of Science and Technology,email: vinithav@fisat.ac.in*

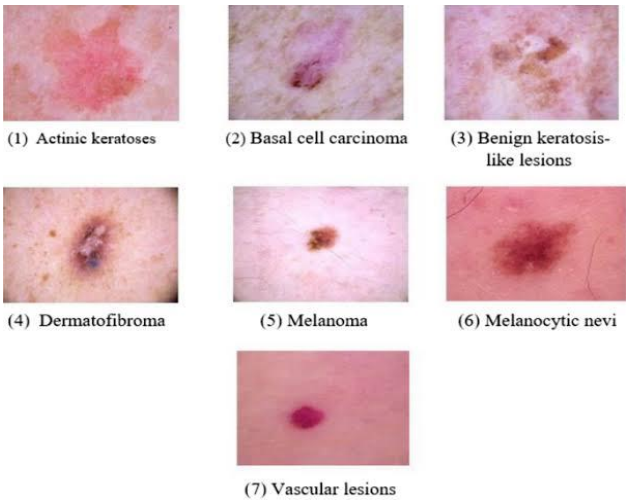


Fig 1. Variants of Skin cancer and its dermatology images

But, age is not a factor in melanoma skin cancer. Age, gender, area, and ethnicity all significantly affect the frequency of melanomas. The most prevalent form of cancer globally, also known as skin cancer or dermoscopic cancer, affects a substantial portion of the population. The American Cancer Society (ACS) said that, as of January 1, 2019, there were 16.9 million malignant cases in the country, with 8.8 million women and 8.1 million men affected. By the beginning of January 2030, 684470 cases of cutaneous melanoma were predicted to occur, bringing the total to 22.1 million. [3] [4]. When creating data sets for the dermato-pathological diagnoses, the patient's age, the lesion's topography, and any colour changes should all be taken into consideration. [5].

Generally harmless, dark spots, tiny moles, or skin surface rashes should not be ignored. Use the ABCDE rule to undertake an initial skin assessment to look for any indications that a skin lesion may turn into melanoma [6, 7]. A magnified and illuminating picture of a slice of skin is created using dermoscopy, a secure skin-scanning procedure, in an effort to even more precisely locate cancer. A visual examination is substantially less accurate than a dermoscopy, which is routinely used to diagnose melanoma [8–10]. The classification of the melanoma warning signs as A (Asymmetry) indicates that the two parts of a melanoma lesion shouldn't be the same. B (Border) claims that the majority of melanomas have crooked surfaces and notch margins. C (Color) shows that a mole that exhibits a variety of colours, such as blue, red, or tan, denotes a concern. Larger melanoma lesions with darker tones are indicated by the letters D (Diameter/Dark).

Seldom is a melanotic melanoma found to be colourless. The E (evolving) symbol indicates a step towards fatality with each alteration in the size, shape, colour, or texture of the skin lesion, which may or may not output in itching.

The technique of diagnosing skin cancer from dermoscopy images requires an effective multiclass classification [11]. The methods depend on faster ROI choice and categorization methods using CNN to increase the effectiveness and execution speed [12]. One choice for effective lesion categorization using the U-net method is deep neural networks (CNNs). Fully connected networks and deconvolutional networks are combined (FCNs). The segmented images were effectively used to retrieve a large number of colour, texture, and structure attributes. For texture analysis, one uses the Local Binary Pattern (LBP) approach. This has been shown to be a very effective completing operation. The edge handbook, gabor, and histogram (HOG) approaches are utilised to retrieve form features [10].

The medical health service has undergone a complete revolution due to the emergence of various networked medical tools and apps resulting from technological advancements. This significantly promotes the sharing of medical knowledge and the delivery of essential medical consultations through the Internet between doctors, other healthcare professionals, and patients. Despite an increase in interest in skin cancer screening, a targeted monitoring method that is widely used has been made possible by the identification of people with skin lesions who are more likely to acquire skin cancer. For the investigation of pigment lesions, automated diagnostic tools are needed. Over the past several decades, it has caught the attention of scientists. Feature extraction, separation, classification, and postprocessing are some of the methods included in these systems. It is still necessary to correctly classify and identify the dermatological lesion. Recent developments in dermoscopic techniques and machine learning algorithms that have reduced the incidence of misinterpretation have led to a major increase in the emphasis on desktop systems.

1.1 Key Highlights

The goals of this research study are to develop effective DL for identifying skin cancers.

- The present state-of-the-art models require more time to identify skin cancer.
- To enhance the efficiency of the existing systems, we can improve accuracy while reducing search time. This can be accomplished by adjusting hyperparameters such as batch size, input image size, and network parameters including the number of hidden layers, hidden nodes, and dropout layer. By fine-tuning these elements, we can elevate the system's efficiency.
- Also, classifiers are used to detect as a standalone system which certainly brings much lower efficiency, so the use of the feature extraction stage as prior and also increasing quality of the image will improve even more.
- A large number of current methods don't use transfer learning to improve classifier expertise.
- In order to combat this, work has been done on developing an efficient transfer learning-based classifier for skin cancer diagnosis.
- To boost the network, use of Autoencoder technique which acts like a double neural network.

Organization of paper; As we already come across skin cancer and its variants information in the introduction part 1, The article also contains the parts that follow: part 2 depicts a literature review, part 3 says the general methodology; part 4 depicts performance analysis and A conclusion is provided in part 5 to wrap up the paper.

2. Literature Review

Sharma et al [11] found that Skin cancer can be detected and classified accurately with reduced time and an accuracy of 97.43%. The study can be expanded to assess the relationship between skin burns and the detection of additional skin cancer disorders.Barman et al. [12] devised a technology that gives an image to the input layer, and then to the hidden layer of CNN for extraction and training. The output layer undergoes classification by applying softmax. Recall 84.1% Precision 83.2% F1 Score 83.2%

In order to improve the functionalities, Mohamed A. Kassem et al. [13] suggested changing the Google Net design by adding layers. And use a bootstrap multi-class support vector machine images that the viewer is unfamiliar with are categorized (outliers).it achieves precision, sensitivity, specificity, the accuracy of 80.36%, 79.8%, 97%, and 94.92% respectively. A computer-aided system using the Gaussian filter, Median Filter, K-mean clustering, ABCD, and MSVM components was proposed by M. Krishna Monika et al.

Long et al. [15]. used the CNN, BP, Whale, and ReLu algorithms to achieve an accuracy of 91,2% and an NPV of 95.2%. Specification 92%, PPV 84% 85% of the time. The platform's efficiency should be increased by expanding feature extraction to include more characteristics and more accurately.

In order to achieve an accuracy of 87.2%, Hwang et al. [16] suggested a system with CNN with ResNet152 and InceptionResNet-V2. Reliability Specificity: 97.04% 97.23% Block-wise tuning of pre-train deep CNN architectures is replaced with layer-wise fine-tuning to increase the effectiveness of the end-to-end learning method.

3. Methodology

This part describes the overall framework's process, which includes the phases as listed below: For any deep learning framework, feeding of enough Dataset is required to make a classifier to learn and understand the image instances. In order to make the model train with such a dataset, the use of a lesion-based image dataset is taken for this work (HAM10000). Upon collection, these datasets go through a preprocessing step where anomalies and noise are removed from the photos in order to enhance their quality. Then with help of Autoencoder, features are extracted where the classifier is. Finally, with these extracted features, the classifier will detect the skin lesion with help of 3D CNN. Also, to further increase the knowledge of the model, the use of Inception v3 as a pre-trained model has also been carried out (Figure 2).

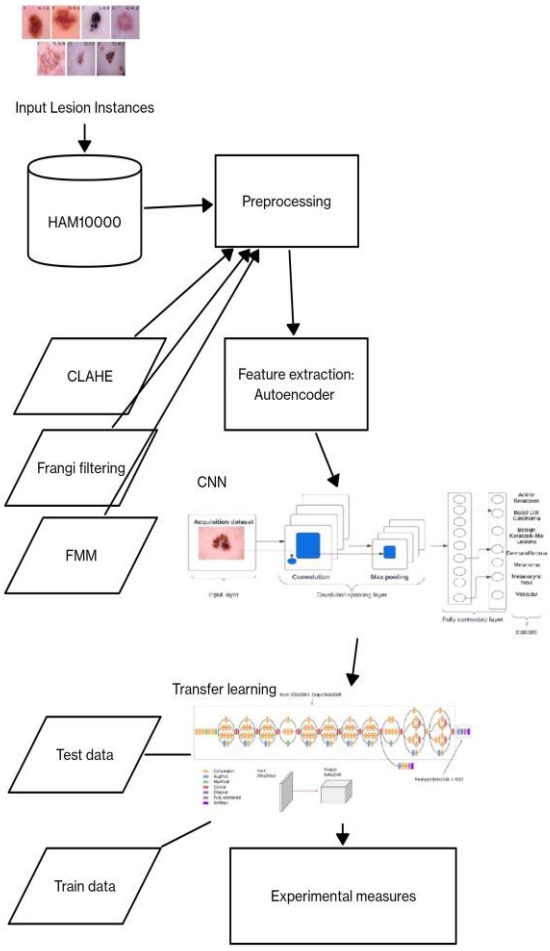


Fig 2. The suggested systems overall structure

3.1 Data Collection

This dataset contains binary segmentation masks as PNG files of all HAM10000 dataset images. The area segments lesion area as evaluated by a single dermatologist. They were initiated with an FCN lesion segmentation model, where afterwards all of them were verified and either approved or corrected/redrew with the free-hand selection tool in FIJI. Figure 3 shows some instances from this dataset and its categories [17,18].

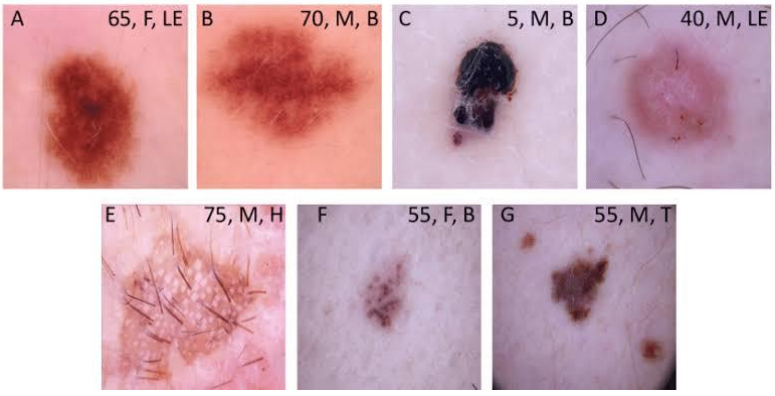


Fig 3. Image instances from HAM10000 Dataset

The 10015 dermatoscopic pictures used in the HAM10000 training dataset came from "Cliff Rosendahl's skin cancer clinic in Queensland, Australia, and the Department of Dermatology at the Medical University of Vienna, Austria", and were saved over the course of 20 years. Images and meta-data were kept up to date on the Australian website using Excel datasets and PowerPoint. The Austrian website began gathering photographs before digital cameras were invented, and it has maintained images and metadata in a variety of ways over time (Table 1).

Table 1. General detailing about the HAM10000 dataset

Dataset	License	Total images	Pathological verification	Akiec	Bcc
HAM10000	CC BY-NC 4.0	10016	53.30%	327	514

3.2 Preprocessing

Enhancing the accuracy and precision of attributes retrieved from the image is the goal of preprocessing. In order to base algorithm acquired images during preprocessing, curvelet features are used, then CLAHE for contrast enhancement, the Frangi vessel filter for artefact identification and elimination, and FMM for inpainting.

3.2.1 Denosing technique

On images, a variety of denoising methods can be applied, including median filtering, frequency domain filtering, wavelet transform, and curvelet. The edge features are preserved while the noise in the images is reduced using a nonlinear method called median filtering. Spatial smoothing and frequency domain filtering is utilized to take the random noise out of the image. Edge and texture data are lost while using these methods. By restricting the noisy wavelet coefficients, the picture can be denoised using the wavelet transform. The image's edges must still be reconstructed using a significant number of wavelet coefficients. The proposed methodology uses a curvelet to remove the undesirable noise from the dermoscopic image yet simultaneously maintaining regional properties like edges, texturing, and minute details seen in conventional images.

The geometric transform known as the curvelet has multiple scales and resolutions. The foundation of the curvelet is a multiscale ridgelet coupled with a spatial bandpass filtering process. With different lengths and widths, curvelets appear at all scales, locations, and orientations. To represent edges with a smooth function, the curvelet transform was created. Because curved edges can appear on images just as they might on ones with straight edges, a transformation is required to represent the curved edges[19]. The ridgelet is employed to specifically record the curved edges since, at small sizes, a curved edge is virtually straight. The most effective technique for segmenting object edges is curvelet. The mean square error of the image data is reduced by the curvelet transform, which is also very precise and straightforward. Image denoising is an application for it [28,29]. Setiawan et al. [19] proposed curvelet-based multiscale denoising with non-local means and uses an image filter. In the 2D domain $\mathbb{R}2$, all curvelets at scale 2-j are found by rotation and translations of mother curvelet ψj with different factors. [30] Curvelets at scale 2-j, orientation θl and position xj,lζ is described in Equation (1)

$$\Psi_{j,l,\zeta}(x) = \Psi_j\left(R_{\theta_l}\left(x - x_\zeta^{j,l}\right)\right), \quad (1)$$

Where position

$$x_\zeta^{j,l} = R_{\theta_l}^{-1}\left(\zeta_1.2^{-j}, \zeta_2.2^{-\frac{j}{2}}\right), \quad (2)$$

The rotation of radians is represented by Rθ, while its inverse is represented by R1θ. If θ is positive, the vector rotation's direction is counterclockwise. If θ is negative, it is clockwise.

$$R_\theta = \begin{pmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{pmatrix} \quad (3)$$

Rotation angle

$$\theta_l = 2\pi.2^{-\lfloor\frac{j}{2}\rfloor}.l, \quad (4)$$

l = 0, 1, …, such that 0 ≤θl ≤2π, and translation parameter

$$\zeta = (\zeta_1, \zeta_2)\epsilon\mathbb{Z}^2. \quad (5)$$

The continuous curvelet transform of a function f ∈L2 ℝ2 ð Þ is defined as:

$$cvt(j,l,\zeta) = \int_{R^2} f(x)\overline{\Psi_{j,l,\zeta}(x)}dx. \quad (6)$$

While the input consists of cartesian arrays, including an image of form f t½ 1, t2, 0 ≤t1, t2 ≤p, where t1, t2, and p are real integers, digital curvelet transforms can be specified as follows:

$$cvt^D(j,l,\zeta) = \sum_{0\le t_1,t_2\le P} f[t_1,t_2]\overline{\Psi^D_{j,l,\zeta}[t_1,t_2]}. \quad (7)$$

A single image pixel's measurements (t1, t2) and values (p), respectively, are represented by the elements t1 and t2. The method for denoising that is being suggested includes curvelet decomposition, parameterization, and curvelet reconstructing from a subset of variables.

3.2.2 CLAHE Algorithm

It is essential to change an image's color space in order to decrease the number of channels, decouple luminance and chrominance information, give sufficient perceptual homogeneity, and provide invariance to various imaging conditions including viewing angles, illumination intensities, and highlights. [34] After denoising, a grayscale image using the L*a*b colour scheme was produced. Uneven lighting caused by external factors and sensor problems must be removed during image

acquisition. Contrast enhancement is required even though low-contrast skin images make locating lesion borders challenging. A collection of spectral bands that increase the contrasting between the lesion and healthy skin are found using the Independent Histogram Pursuit (IHP) approach [35]. Apply this method to separate the lesion from the healthy tissue.

The extra bands were additionally approximation in order to highlight the multiband image's distinctive appearance. All multispectral and dermoscopic images can be properly processed with IHP. Contrast enhancement is accomplished using the Adaptive Histogram Equalization (AHE) algorithm36, which is upgraded using AHE variations. The CLAHE technique is used for lighting correction. 14,22 The intensity presented at each pixel is set to represent the rank of that pixel's intensity in the histogram as part of the CLAHE image optimisation process. It looks at the relative intensity histogram of the each pixels. By modifying a traditional histogram, the method restricts the contrast enhancement to the user-selectable maximal at each intensity level. The entire image is subjected to histogram equalization. Contrast limitation is employed to stop the noise from intensifying.

CLAHE employs adaptive thresholding equalization. The L channel of the image is subjected to the CLAHE method using the "L*a*b" colour space form. This colour change has a consistently perceptible colour space. It functions in restricted areas with increased contrast. To get rid of the artificially created boundaries, the neighbouring areas are merged utilizing bilinear interpolation. To prevent boosting the noise, contrast is restricted. For artefact removal and identifiers, a contrast-enhanced image is given. Using the CLAHE method, the image is scaled and the illumination is improved.

3.2.3 Frangi Filtering

In volumetric displays and maximum intensity projectors used in medical imaging, the Frangi vesselness filtration method is employed to minimize noise and background and accentuate vessels. Many artefacts, including skin lines, and blood vessel ,hair can be seen in dermoscopic images, which makes segmentation and feature extraction for the identification of lesions more difficult. The detection of hair-occluded information is especially difficult because of the tiny curvature architecture of hairs, the variation of colours of hairs, the thick and thin architecture, and the volume of the hairs which makes the elimination of these artefacts is hard. Using several fundamental characteristics like magnitude, thickness, length, and direction, pixels are distinguished from the pixels around them in order to find hair-like structures.

Lesion characteristics are utilized to differentiate lesion pixels from hair pixels since they have some unique characteristics. The image is smoothed using a Gaussian filter (GF) wherein g represents standard deviation. A Hessian matrix, that's a partial second-order derivative of the smoothed image, serves as the basis for the detection of hair lines. The 2-D kernel G that you mentioned is not necessarily the second-order derivative of the GF G. Instead, it is the Gaussian kernel itself that is convolved with the image to smooth it. The second-order derivative of the GF is then applied to the smoothed image to compute the Hessian matrix.

Convolution is used to create directional smoothing functions.

$$G_\sigma = \begin{bmatrix} \frac{\partial^2 g_\sigma}{\partial x^2} & \frac{\partial^2 g_\sigma}{\partial x \partial y} \\ \frac{\partial^2 g_\sigma}{\partial x \partial y} & \frac{\partial^2 g_\sigma}{\partial y^2} \end{bmatrix} \qquad (8)$$

The equation is a representation of the Hessian matrix.

$$H\sigma = G_\sigma * f = \begin{bmatrix} H_{xx} & H_{xy} \\ H_{xy} & H_{yy} \end{bmatrix} \qquad (9)$$

The Frangi vessels filter is produced via eigenvalue evaluation of the hessian matrix on multiple Gaussian scales. The Frangi filter determines whether an image area has vessels by using the eigenvectors of the Hessian. Eigenvalues are denoted by λ1 and λ2 (j j λ2 ≥j jÞ λ1), which is obtained using Equation (13).

$$\lambda_{1,2} = \frac{(H_{xx}+H_{yy}) \pm \sqrt{(H_{xx}-H_{yy})^2 - 4H_{xy}^2}}{2} \qquad (10)$$

Vesselness response V0ð Þ σ 23 is given by the mathematical Equation:

$$V_0(\sigma) = \begin{cases} 0 \; for \; dark \; tubes \; if \; \lambda_2 < 0, \\ e^{\left(\frac{R^2}{2v^2}\right)}\left(1 - e^{\left(\frac{g^2}{2\sigma^2}\right)}\right) \; other \; wise \end{cases} \qquad (11)$$

Here S ¼ λ2 1 þλ2 2q is second order structure-ness, R ¼ ð Þ k k λ1 =k k λ2 is a blob-ness measure and V0 is vesselness value, v, c is image-dependent parameters for blob-ness and structure-ness terms. For the identification of hairs, which is produced by an iterative technique, variables v and c are tuned to 0.5 and 1, correspondingly. The filter's capability to measure R and S. Blob-like structure and background noise is suppressed by setting the values of v and c. The vesselness function is examined at various scales, and the filter responses are then added to produce the final vessel-enhanced image. When some settings are altered, distortions result and hairs become misidentified. The primary paths into which the local second-order architecture of the image might be divided were retrieved using the Hessian eigenvalues after the hairs' varying growth directions had been identified. After extracting the principal directions, they were rotated to reveal the direction with the least amount of curvature.

The Frangi filter is used to the image for enhancement of hairs, edges, and specific details that are required to identify hair pixel and replace. This filter utilizes the threshold value to normalise the filtered image. The Frangi threshold is compared to these normalized values to determine the appropriate threshold for normalizing an image and removing undesirable curves and tumor forms. The suggested methodology offers a hair-repaired method. The method involves locating hair to use the Frangi vesselness filter and honing this using morphological dilation methods. It also filters blood vessels, ruler markings, and hair curvature lines, effective hair-developed techniques employing line detection are used. There is a need to refine hair segmented lines since weak parts of hair occluded lines may be disrupted owing to hairline recognition. These segmentation hairlines also feature curvature-like features, which are connected by the pixel linking function in different directions. These curves could indicate a tumor's region or microscopic gels that are eliminated using morphological surgery. Larger gaps between lines and the borders of hair curves are filled using a morphologically based image closure process. To fill in the spaces and maintain the structure's natural curves, use a three-radius square-shaped structuring element. Using a connected-component labelling technique, the curve-like objects are retrieved, removing any undesired small and circular shapes. Four connected objects can be combined with the morphological area opening and circularity requirement in the connected-component labelling algorithm to eliminate extraneous objects. Utilizing morphological area opening, small objects that are less than 130 pixels are eliminated. The radial distances' mean and variance serve as the foundation for the circularity condition.

3.2.4 FMM for noise removal

The removal of hair areas can remove the impact of hairs on diagnosis assessment. To increase segmentation, however, the hair-obscured melanoma texture needs to be restored. The process of inpainting involves filling in missing information in areas of an image that are already known while reconstructing the fundamentals of pictorial art. Lesions caused by the hair removal effect are repaired using the FMM. Fast marching is recommended due to its durability and non-iterative nature over other hair inpainting techniques. FMM is a texture-focused methodology.

Melanoma texture retrieval employs a method that repairs the hair-occluded region based on a fast-marching technique. By utilizing the FMM inpainting algorithm to replace the recognized hairlines with the background, the lesion is restored without changing the texture. Non-iterative PDEs are the fast inpainting techniques employed in this case, and they operate the structure tensor to strongly define the consistency path that alternates between diffusion and directed passage. For texture features, a quick inpainting technique is performed so as not to alter the tumor's patterns. Faster than the alternative technologies, FMM is simple to use.

3.3 Feature extraction: Autoencoder

Machines can learn from enormous datasets from deep learning algorithms. These algorithms automatically assign explicit categories to input datasets (medical images). They make it possible to provide accurate classifications that, in some situations, can outperform human evaluation. Deep learning models are powerful models for finding complex information correlations and spotting high-level abstractions because they produce previously unheard-of results in many fields. The network depth enables the extraction of relevant traits with some inherent value for the classification task.

An autoencoder, a kind of unsupervised learning framework, is composed of three stages: the input layer, the hidden layer, and the output layer, according to Figure 4. The base of the autoencoder's training stage consists of the encoder and the decoder. The first one is used to map the input knowledge into the hidden layer, while the second is regarded as the reconstruction of the input data from the hidden layer [32].

Considering the unlabeled input dataset xn € Rm+1, the hidden encoder vector h_n, which is derived from x_n, and the output layer's decoder, denoted by y_n. The encoding and decoding processes are represented, accordingly, by the following formula:

$$h_n = fe(\omega_1 x_n + b1) \quad (12)$$

$$y_n = fd(\omega_2 h_2 + b2) \quad (13)$$

where fe and fd stand for the encoding and decoding functions, respectively. The weight matrices for the encoder and decoder are designated as W1 and W2, respectively. The bias vectors B1 and B2 are.

As shown by the following expression, the autoencoder's parameters are tuned to reduce the reconstruction error.

$$\emptyset(\theta) = arg\min_{\theta} \frac{1}{\theta^n} \sum_{i=1}^{n} L(x^i, y^i) \quad (14)$$

where the loss function L is shown below;

$$is \quad by\ L(x, y) = \| x - y \|^2 \quad (15)$$

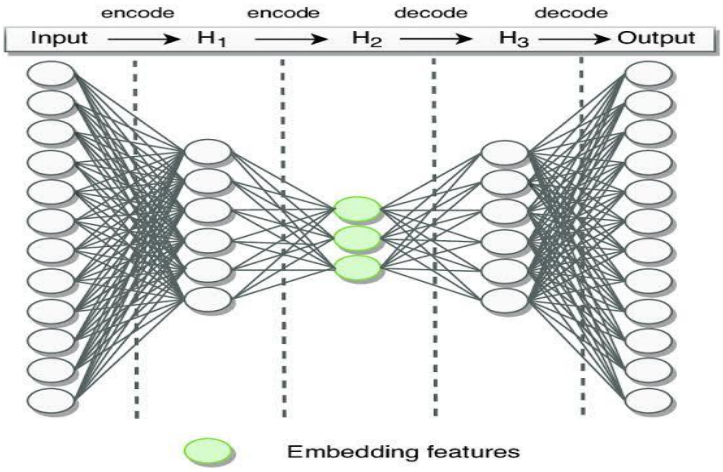


Fig 4. Autoencoder structure for feature extraction process

3.5 Classification: 3D CNN

In opposition to a conventional neural network, a CNN develops interesting structures by applying techniques to an image's raw pixels. The CNN model was created and built with Python 3.7.9 using the Tensorflow and Keras modules. Figure 5 shows the CNN Model from a high-level perspective. The layers and hyperparameters employed in the network are listed in Table 2.

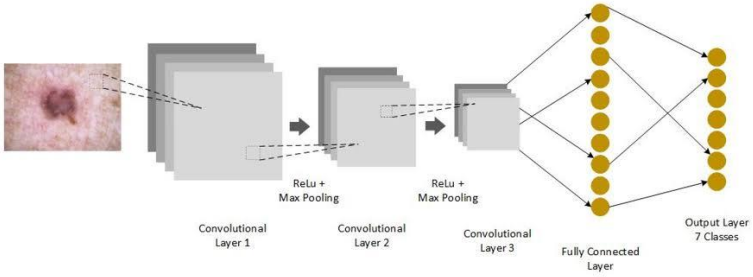


Fig 5. The architecture of skin lesion classification of CNN

Table 2. Summarised architecture layers of CNN

Layer	Hyperparameters
Conv2D	32 filters, 3 * 3 filter size, ReLU activation, same padding, followed by batch normalization
MaxPool2D	3*3 pool size is used for reducing image spatial dimensions fastly from 96*96 to 32*32
Dropout (Core layer)	0.25 Neurons
Conv2D	64 filters,3*3 filter size, ReLU , same padding
Conv2D	64 filters,3*3 filter size, ReLU , same padding, batch normalization is done
MaxPool2D	2*2 pool size
Dropout (Core layer)	0.25 Neurons
Conv2D	128 filters, 3*3 filter, ReLU , same padding, batch normalization is done
Conv2D	128 filters, 3*3 filter, ReLU , same padding , batch normalization
MaxPool2D	2*2 pool size
Dropout (Core Layer)	0.25 Neurons
Flatten (core Layer)	-
Dense	1024 Units, ReLU , and batch normalization
Dropout (core Layer)	0.5 Neurons
Dense	7 Units, softmax activation

The values of some common hyperparameters are utilized to improve model evaluation. The CNN model's selected hyperparameter values are highlighted. The rationale behind the hyperparameter values used in the suggested work is explained in the next section. Because to Adam's simplicity, computational efficiency, and success when dealing with enormous quantities of information and elements, it has become the most popular optimisation problem for the creation of deep neural networks. Loss Function: The Multi-Class uses the "categorical cross-entropy" loss function to calculate the loss value. Epochs: There are 150 epochs in total. Experiments showed that a model with 150 epochs produces low loss and does not overfit to the training set (or not overfitted as best as we can). Batch Size: The 32-piece batch size was found to produce the greatest results after early testing with batches of 5, 10, 20, and 40 carried out. Learning Rate: A beginning rate of 0.001 is used for learning. The learning rate determines how far up the gradient we "step" each time. The step increases as the value increases and decreases as the value increases.

3.5.1 Transfer learning

Due to the high computational cost, training a CNN with numerous layers from scratch might take a long period. Transfer learning has been created as a solution to this problem. The procedure cuts down on the time needed to train a network for a certain set of object classes. The goal of using a fully trained network initially and then retraining it on a fresh collection of objects is to create general networks that can be modified to meet various demands without requiring extensive retraining and restructuring. It is usual practice to accomplish this using the ImageNet [20] training set, a benchmark set created by academics, especially for computer vision that consists of 1.26 million images and 1000 generic object classes. Conventional machine learning assumes that the information used for testing would have the same distribution as the information used to train the network, however, this is mostly false [6]. Yet, the pre-information training from the source task may be transmitted and used to help the networks in a domain it hasn't yet experienced. The network can be fine-tuned by changing the weights of the adequate number of layers, or it can serve as a feature extractor by replacing the last fully-connected layer with one that is more relevant for the new domain, based on the quantity of information available in the new domain and the variability between the new and old domain [27].

Googlenet Inception v3: is a design that offers high-performance networks at a cheap computational cost. There are two main metrics used in the literature [30] to gauge accuracy: top-5 error rate and top-1 error rate. They gauge how frequently, the network fails to incorporate the right class in the top-5 and top-1 output. 5.6% top-5 and 21.2% top-1 error rates are achieved with Inception v3. To put these numbers into perspective, the top-5 error rate for AlexNet [19] is 15.3%, while the original Inception (GoogLeNet) [29] reaches 6.67% in the same category. Google's Inception v3's architecture places a strong emphasis on the value of memory management and the limitations of processing power. This may imply that Inception v3 is more suited to mobile applications than the alternative, more computationally intensive models (Figure 6).

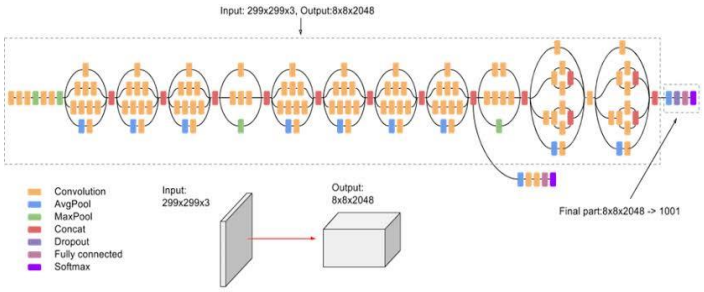


Fig 6. The architecture of Inception V3 as the transfer learning process

4. Performance Analysis

Software requirements include Tensorflow, an open-source Python library for generating DL models, and Google Collaboratory, for producing DL models. Hardware requirements include Ryzen 5/6 series CPU, NV GTX CPU, 1 TB HDD, and Windows 10 OS. Measures including accuracy, sensitivity, specificity, recall, precision, F1-score, detection rate, TPR, and FPR are used in experimental analysis when comparing other algorithms, expressed as L.

The overall evaluation is shown in Table 3 under the headings of accuracy, sensitivity, and specificity. The suggested model outperforms as seen in Figure 7(a,b,c graphical)'s representations of accuracy, sensitivity, and specificity because of its intricate architecture and consideration of training hyperparameters.

Table 3. Detailed examination of accuracy, sensitivity, and specificity

Models	Accuracy	Specificity	Sensitivity
L1[11]	93	90	94
L2[12]	84	89	93
L3[13]	90	94	96
L4[14]	89	93	95
L5[15]	85	91	94
L6[16]	88	95	96
CNN-Inception v3	96	97	97

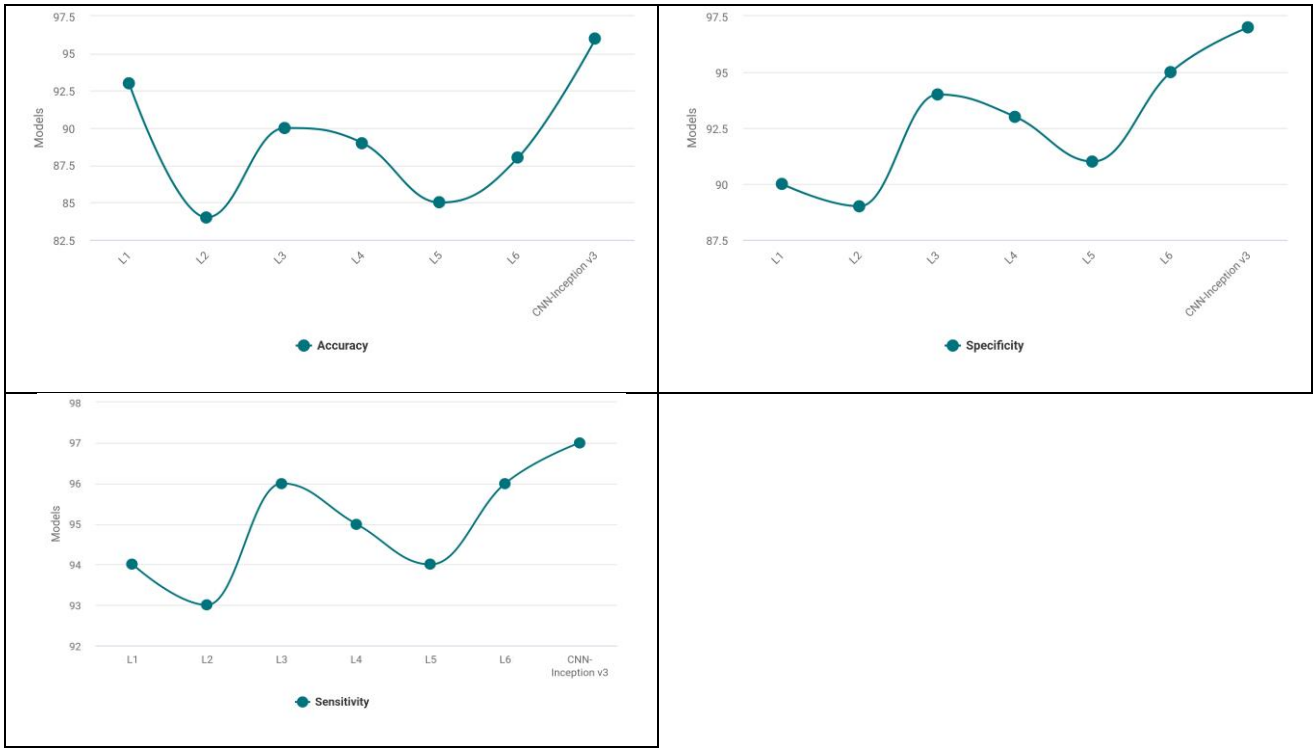


Fig 7. Models vs a) Accuracy, b) Specificity, c) Sensitivity over HAM10000 dataset instances.

Table 4 compares precision, recall, and F1-score analyses. Figure 8(a,b,c) displays graphical data from Table 4 showing that the proposed model outperforms other models due to a more precise preprocessing step that results in significantly better accuracy for enhancing performance and identification.

Table 4. Detailed examination of precision, recall, F1-score

Models	Precision	Recall	F1-score
L1[11]	82	78	83
L2[12]	78	71	75
L3[13]	85	77	81
L4[14]	83	76	80
L5[15]	80	73	79
L6[16]	84	78	84
CNN-Inception v3	89	84	92

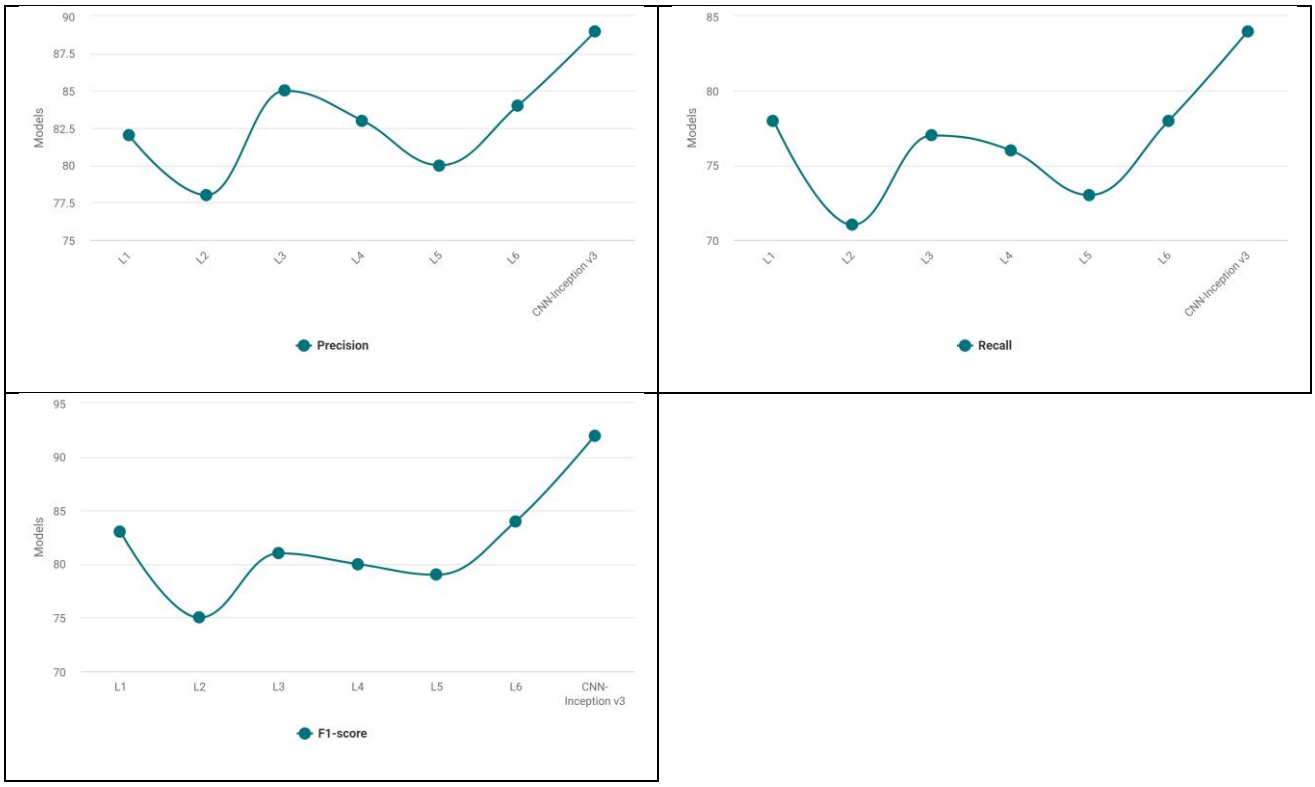
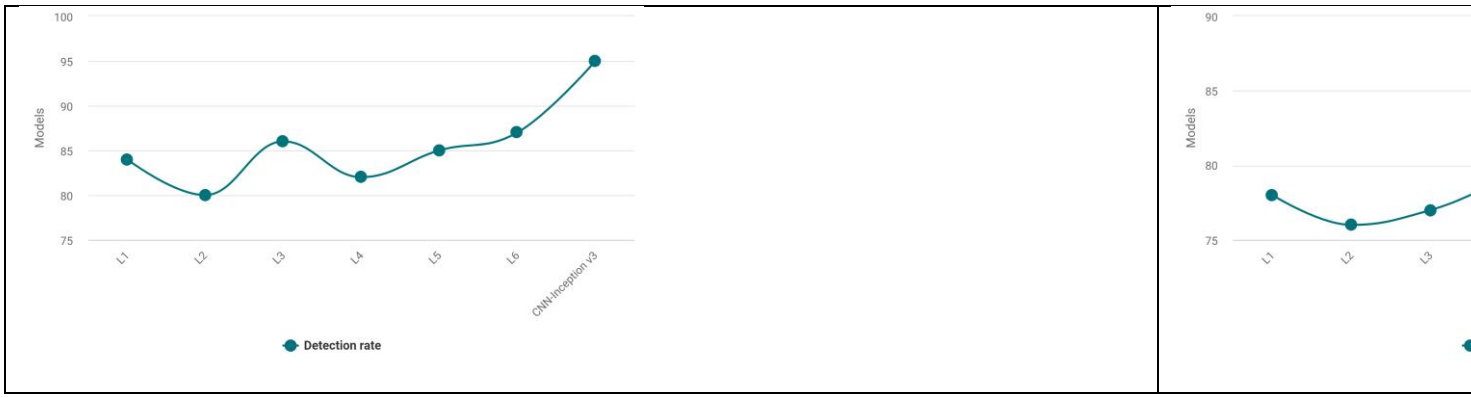


Fig 8. Models vs a) Precision, b) Recall, c) F1-score for HAM10000 dataset instances

Table 5 compares the analyses of the detection rate, TPR, and FPR. Figure 9(a,b,c) displays a graphical representation of Table 5 in which the suggested model outperformed other models due to an increase in classifier information learned by transfer learning utilizing Inception V3. Also, the use of the Autoencoder technique as feature extraction brings an additional boostage.

Table 5. Detailed examination of detection rate, TPR, FPR

Models	**Detection rate**	**TPR**	**FPR**
L1[11]	84	78	22
L2[12]	80	76	24
L3[13]	86	77	23
L4[14]	82	79	21
L5[15]	85	80	20
L6[16]	87	84	16
CNN-Inception v3	95	89	11



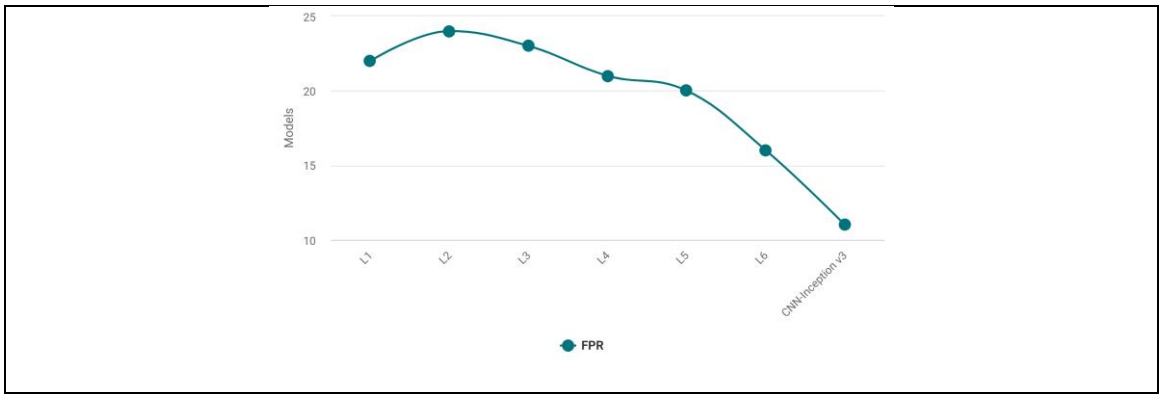


Fig 9. Models vs a) detection rate, b) TPR, c) FPR over HAM10000 dataset instances

5. Conclusion

This paper presents an instrument in which a non-programming foundation individual can foster complex profound learning models. It opened the choices of adaptability in planning profound learning classifiers by alluding to the overall techniques and the circling designs in the improvement of profound learning models. Therefore, this paper brings effective deep learning using CNN with the transfer learning process of the pre-trained model (Inception V3). The help of Autoencoder as feature extraction brings more boostage to the model in terms of efficiency and performance which brings an accuracy of 0.96. For future work, the use of hybrid models as transfer learning will be of high value for the paper as well.

References

[1] Ali, S. N., Ahmed, M., Paul, J., Jahan, T., Sani, S. M., Noor, N., & Hasan, T. (2022). Monkeypox skin lesion detection using deep learning models: A feasibility study. arXiv preprint arXiv:2207.03342.

[2] Varma, P. B. S., Paturu, S., Mishra, S., Rao, B. S., Kumar, P. M., & Krishna, N. V. (2022). SLDCNet: Skin lesion detection and classification using full resolution convolutional network-based deep learning CNN with transfer learning. Expert Systems, 39(9), e12944.

[3] ul Haq, I., Amin, J., Sharif, M., & Almas Anjum, M. (2022). Skin lesion detection using recent machine learning approaches. In Prognostic Models in Healthcare: AI and Statistical Approaches (pp. 193-211). Singapore: Springer Nature Singapore.

[4] Choudhary, P., Singhai, J., & Yadav, J. S. (2022). Skin lesion detection based on deep neural networks. Chemometrics and Intelligent Laboratory Systems, 230, 104659.

[5] Nawaz, M., Mehmood, Z., Nazir, T., Naqvi, R. A., Rehman, A., Iqbal, M., & Saba, T. (2022). Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering. Microscopy research and technique, 85(1), 339-351.

[6] Hum, Y. C., Tan, H. R., Tee, Y. K., Yap, W. S., Tan, T. S., Salim, M. I. M., & Lai, K. W. (2022). The development of skin lesion detection applications in smart handheld devices using deep neural networks. Multimedia Tools and Applications, 81(29), 41579-41610.

[7] Shetty, B., Fernandes, R., Rodrigues, A. P., Chengoden, R., Bhattacharya, S., & Lakshmana, K. (2022). Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. Scientific Reports, 12(1), 18134.

[8] Gouda, W., Sama, N. U., Al-Waakid, G., Humayun, M., & Jhanjhi, N. Z. (2022, June). Detection of skin cancer based on skin lesion images using deep learning. In Healthcare (Vol. 10, No. 7, p. 1183). MDPI.

[9] Chandra, R. (2022). Skin lesion detection using deep learning (Doctoral dissertation, Purdue University Graduate School).

[10] Shahsavari, A., Khatibi, T., & Ranjbari, S. (2022). Skin lesion detection using an ensemble of deep models: SLDED. Multimedia Tools and Applications, 1-20.

[11] M. Kumar, M. Alshehri, R. AlGhamdi, P. Sharma, and V. Deep, "A DE-ANN Inspired Skin Cancer Detection Approach Using Fuzzy C-Means Clustering," Mob. Networks Appl., vol. 25, no. 4, pp. 1319–1329, 2020, doi: 10.1007/s11036-020-01550-2.

[12] M. Hasan, S. Das Barman, S. Islam, and A. W. Reza, "Skin cancer detection using convolutional neural network," ACM Int. Conf. Proceeding Ser., pp. 254–258, 2019, doi: 10.1145/3330482.3330525.

[13] M. A. Kassem, K. M. Hosny, and M. M. Fouad, "Skin Lesions Classification into Eight Classes for ISIC 2019 Using Deep Convolutional Neural Network and Transfer Learning," IEEE Access, vol. 8, pp. 114822–114832, 2020, doi: 10.1109/ACCESS.2020.3003890.

[14] M. Krishna Monika, N. Arun Vignesh, C. Usha Kumari, M. N. V. S. S. Kumar, and E. Laxmi Lydia, "Skin cancer detection and classification using

machine learning," Mater. Today Proc., vol. 33, no. xxxx, pp. 4266–4270, 2020, doi: 10.1016/j.matpr.2020.07.366.

[15] L. Zhang, H. J. Gao, J. Zhang, and B. Badami, "Optimization of the Convolutional Neural Networks for Automatic Detection of Skin Cancer," Open Med., vol. 15, no. 1, pp. 27–37, 2020, doi: 10.1515/med-2020-0006.

[16] B. Ahmad, M. Usama, C. M. Huang, K. Hwang, M. S. Hossain, and G. Muhammad, "Discriminative Feature Learning for Skin Disease Classification Using Deep Convolutional Neural Network," IEEE Access, vol. 8, pp. 39025–39033, 2020, doi: 10.1109/ACCESS.2020.2975198.

[17] Nugroho, A. A., Slamet, I., & Sugiyanto. (2019, December). Skins cancer identification system of HAMl0000 skin cancer dataset using convolutional neural network. In AIP conference proceedings (Vol. 2202, No. 1, p. 020039). AIP Publishing LLC.

[18] Garg, R., Maheshwari, S., & Shukla, A. (2021). Decision support system for detection and classification of skin cancer using CNN. In Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020 (pp. 578-586). Springer Singapore.

[19] Setiawan, A. W. (2020, February). Effect of colour enhancement on early detection of skin cancer using convolutional neural network. In 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT) (pp. 100-103). IEEE.

[20] Jaisakthi, S. M., Mirunalini, P., & Aravindan, C. (2018). Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms. IET Computer Vision, 12(8), 1088-1095.

[21] Choudhary, P., Singhai, J., & Yadav, J. S. (2021). Curvelet and fast marching method-based technique for efficient artefact detection and removal in dermoscopic images. International Journal of Imaging Systems and Technology, 31(4), 2334-2345.

[22] Jaisakthi, S. M., Mirunalini, P., & Aravindan, C. (2018). Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms. IET Computer Vision, 12(8), 1088-1095.

[23] Hammouche, R., Attia, A., Akhrouf, S., & Akhtar, Z. (2022). Gabor filter bank with deep autoencoder-based face recognition system. Expert Systems with Applications, 116743.

[24] Toğaçar, M., Cömert, Z., & Ergen, B. (2021). Intelligent skin cancer detection applying autoencoder, MobileNetV2 and spiking neural networks. Chaos, Solitons & Fractals, 144, 110714.

[25] Majji, R., Om Prakash, P. G., Cristin, R., & Parthasarathy, G. (2020). Social bat optimisation dependent deep stacked auto-encoder for skin cancer detection. IET Image Processing, 14(16), 4122-4131.

[26] Dhar, P., & Guha, S. (2021). Skin lesion detection using fuzzy approach and classification with CNN. International Journal of Engineering and Manufacturing (IJEM), 11(1), 11-18.

[27] Khan, M. A., Muhammad, K., Sharif, M., Akram, T., & de Albuquerque, V. H. C. (2021). Multi-class skin lesion detection and classification via teledermatology. IEEE journal of biomedical and health informatics, 25(12), 4267-4275.

[28] Mahbod, A., Schaefer, G., Wang, C., Dorffner, G., Ecker, R., & Ellinger, I. (2020). Transfer learning using a multi-scale and multi-network ensemble for skin lesion classification. Computer methods and programs in biomedicine, 193, 105475.

[29] Miglani, V., & Bhatia, M. P. S. (2020). Skin lesion classification: A transfer learning approach using efficient nets. In Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020 (pp. 315-324). Singapore: Springer Singapore.

[30] Mahbod, A., Schaefer, G., Wang, C., Ecker, R., Dorffner, G., & Ellinger, I. (2021, January). Investigating and exploiting image resolution for transfer learning-based skin lesion classification. In 2020 25th international conference on pattern recognition (ICPR) (pp. 4047-4053). IEEE.

[31] Singh, J. ., Mani, A. ., Singh, H. ., & Rana, D. S. . (2023). Quantum Inspired Evolutionary Algorithm with a Novel Elitist Local Search Method for Scheduling of Thermal Units. International Journal on Recent and Innovation Trends in Computing and Communication, 11(3s), 144–158. https://doi.org/10.17762/ijritcc.v11i3s.6175

[32] Mwangi, J., Cohen, D., Costa, R., Min-ji, K., & Suzuki, H. Optimizing Neural Network Architecture for Time Series Forecasting. Kuwait Journal of Machine Learning, 1(3). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/13

[33] Dhablia, D., & Timande, S. (n.d.). Ensuring Data Integrity and Security in Cloud Storage.