RESEARCH

Open Access

Multi-class segmentation skin diseases using improved tuna swarm-based U-EfficientNet



Manikandan Rajagopal^{1*}, Shubhangi N. Ghate², Rajeswari P³ and E. N. Ganesh⁴

*Correspondence: manikandan. rajagopal@christuniversity.in

 ¹ School of Business and Management, Christ University, Bangalore 560029, India
 ² Ramrao Adik Institute of Technology, Navi Mumbai 400706, India
 ³ Department of AI & DS, Panimalar Engineering College, Chennai, Tamil Nadu 600123, India
 ⁴ Department of ECE, St. Peter's Institute of Higher Education

and Research, Chennai, India

Abstract

Early location of melanoma, a dangerous shape of skin cancer, is basic for patients. Indeed, for master dermatologists, separating between threatening and generous melanoma could be a troublesome errand. Surgical extraction taken after early determination of melanoma is at its way to dispense with the malady that will result in passing. Extraction of generous injuries, on the other hand, will result in expanded dismalness and superfluous wellbeing care costs. Given the complexity and likeness of skin injuries, it can be troublesome to create an accurate determination. The proposed EfficientNet and UNet are combined and arrange to extend division exactness. Also, to decrease data misfortune amid the learning stage, adjusted fish swarm advancement (IMSO) is utilized to fine-tune the U-EfficientNet's movable parameters. In this paper, a ViT-based design able to classify melanoma versus noncancerous injuries is displayed. On the HAM1000 and ISIC-2018 datasets, the proposed ViT demonstrated accomplished the normal precision of 99.78% and 10.43% FNR with computation time of 134.4632s of ISIC-2018 datasets. The proposed ViT show accomplished the normal exactness of 99.16% and 9.38% FNR in with computation time of 133.4782s of HAM1000 dataset.

Keywords: Skin diseases, Modified tuna swarm optimization, GrabCut-stacked convolutional neural network

Introduction

Skin injuries are developments on the skin that contrast from the encompassing skin in appearance. The two sorts of skin injuries are essential and auxiliary. Essential skin injuries are irregular skin developments that can create some time recently birth or all through adulthood. Auxiliary skin injuries can create from treated or aroused unique skin injuries. Patches, papules, knobs, and tumors are illustrations of essential skin injuries. Scales, scabs, scraped spots, disintegrations, and ulcers are cases of auxiliary skin injuries. Some of the time, threatening tumors can be deadly. The foremost unsafe kind of skin cancer is melanoma [1]. The fundamental establishment for the judgments around skin issues is the subjective assessment of dermatologists, dermatologists, or a long time of specialized encounter, which is always warranted to be error-free. To partitioned the disarranges, a really tall degree of information and competence is required. Treatment is as often as possible done without or is costly for numerous due to the tall



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, and indicate if changes were made. The images or other third party material is not included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativecommons.org/licenses/by/4.0/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdo-main/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

dependence on masters and tall costs of restorative exams and interviews. Skin issues go unnoticed universally in spite of stressing measurements [2]. Skin issues may be recognized by computer-supported computer program, and the results are more reliable and unsurprising.

Deep learning aids in the resolution of challenging learning issues that cannot be handled by the current rule-based approaches. Deep learning-based algorithms perform well in a range of difficult computer vision and image classification applications. As a result, deep learning algorithms are frequently used to analyze medical images for a variety of applications, including illness identification [3]. To get useful representations, nevertheless, elaborate structures need learning several training cases. Unfortunately, building extensive medical images datasets for map learning is more difficult than for other uses. Being expensive and time-consuming, acquisition and labelling call for specialized equipment and trained medical personnel. One of the biggest problems with current deep learning and computer vision frameworks is the lack of data. Everybody has several skin mutations and lesions. As a result, there is a constant absence of data on the skin.

The two most frequently utilized imaging procedures for the distinguishing proof of progressed skin injuries are dermoscopic and plainly visible (clinical) imaging [4]. Dermatologists need the capacity to characterize injuries utilizing dermoscopic imaging, which might regularly be intangible to the human eye. Clinical pictures captured with ordinary cameras, in any case, are more broadly accessible but of lesser quality [5, 6]. Dermatologists can distinguish skin injuries by utilizing dermoscopy, a noninvasive skin imaging strategy, to see past the skin's surface. In any case, depending on the information and encounter of the specialist, dermoscopic prognostic esteem might run significantly, from 24 to 77% [7]. Besides, utilizing dermoscopy by untrained dermatologists really makes it harder to analyze injuries. In this manner, it is fundamental to construct a CAD (computer-aided conclusion) framework in arrange to diminish the trouble of visual assessment and symptomatic botches caused by subjectivity, as well as to reduce the burden of dermatology and get to dermatologists [4, 8].

CAD is a useful tool to assist doctors in routine theranostics by analyzing clinical images. Deep learning (DL) offers a strong basis for computer vision work, and design is no exception [4]. Existing digital dermoscopic equipment can be used to store and retrieve dermoscopic images in a processor [9]. As dermoscopic diagnoses depend constantly on the expertise of the dermatologist, CAD systems are beneficial for less experienced dermatologists and have been demonstrated to be less vulnerable to intersubject variability. Lesion categorization, feature extraction, and images segmentation are often included procedures in conventional approaches for automated dermoscopic image assessment [10, 11]. Accuracy is one of segmentation's key characteristics, making it the most crucial phase in the skin cancer diagnosis [12]. The various skin kinds and textures, as well as the various lesion colors, sizes, and types, make segmentation challenging [13]. The study's primary contributions are as follows:

• To improve performance, train a ViT (Vision Transformer) architecture with both synthetic and raw data utilizing the proper data augmentation approaches.

• Design of the ITSO algorithm: The improved tuna swarm optimization (ITSO) technique, which aims to accelerate convergence and minimize motion that causes the

solution to stray from the optimal region, combines TSO with Pelican's effective solution updating behavior.

• ITSU-EfficientNet's design: ITSU-EfficientNet is a hybrid architecture that combines EfficientNet and UNET and replaces the encoder portion of UNET with EfficientNet to increase granularity accuracy. Additionally, in order to minimize information loss during the information learning stage, we apply the ITSO method to modify the hybrid deep learning model's adjustable parameters.

The rest of the paper is organized as takes after the following: related works to this investigation are detailed in Segment 2. In Area 3, the subtle elements approximately the proposed strategies are displayed. Area 4 appears the test points of interest and execution assessment of the comparative strategies and the benchmarking models. At last, Segment 5 concludes the paper.

Motivation

Skin diseases can be difficult to diagnose accurately, necessitating training and precise methods that increase the diagnostic precision. As a result, dermatologists place a high value on the availability of accurate instruments because they enable better patient care, accurate diagnosis, and fewer patient biopsies. Therefore, this issue has taken into consideration the role of deep learning and its various techniques. Deep learning can assist medical professionals in making diagnoses and facilitate patient outcomes when diagnosing certain diseases. As can be seen from above, applying these techniques in these kinds of situations can help doctors save a great deal of time and effort while still enabling accurate diagnosis—which is ultimately what matters most.

Related work

Dermoscopic images have been the main focus of the majority of melanoma categorization research since they offer more visual information and are often employed by dermatologists in practise. Deep learning was used in a recent work on a CAD system for classifying skin lesions. We discovered that for the majority of approaches, the greater images size necessitates a longer training period for the model. Additionally, the performance of models is hampered by the fact that most existing generic datasets for skin lesion categorization are unbalanced. It [14] employed a deep learning technique based on Vision Converter to categorize skin lesions. This study used a two-layer system to classify skin cancer precisely. In order to determine categories with 96.14% accuracy, the sensor divides the enhanced data into discrete pieces and sends each chunk to a multilayer perceptron classifier. The HAM10000 database is preprocessed in research [15] utilizing a variety of morphological techniques. Utilizing manually created feature extraction methods, features are sought. EfficientNet-B0 and ResNet50V2 are two transfer learning models that were successful in classifying skin lesions with an accuracy of 94.9%. The HAM10000 database was altered in this study [16] utilizing Multi-Focus Segmentation Network (MFS-Net), which is based on deep learning methods. Segmentation maps are produced by Deep Feature Search using parallel partial decoder algorithms. In order to provide segmentation outputs, two distinct attention modules are finally constructed. The authors used the provided algorithm to arrive at a dice score of 90.6%.

A hybrid transformer encoder and CNN were used in literature [17] to provide skin scale segmentation using semi-supervised learning, obtaining both global characteristics and localized features. The developed model is the most effective way to handle problems like overfitting and instability. Here, a semi-supervised model learns semantic properties, enhancing the model's capacity for learning. The developed model performs better than conventional skin lesion segmentation techniques. Noisy annotation criteria reduce performance while still enriching data samples.

It [18] created a GrabCut approach to segment skin lesions, as well as edge boundary removal from images preprocessing and skin hair removal to enhance model performance. In the hair removal project, we create a mask for eliminating hair from input images after building a hair contour detecting system. An equalization approach is used to enhance the contrast once the hair has been removed. In terms of dice and jack-card coefficients, segmentation utilizing the grab-and-cut method performs better. However, segmenting tiny lesions is not a good use of this technique.

The process of identifying skin issues is being automated by researchers utilizing a computer-supported framework. Perform the fundamental tasks of image capture, preprocessing, segmentation, feature extraction, and images classification with this method. They struggle with the problem of their being insufficient training data for deep learning models. The color and texture characteristics of paintings were employed [19] to detect skin conditions. First, median filtering is used to preprocess the images. In order to acquire image segments, rotate the denoised image. After utilizing the GLCM tool to extract text characteristics, categorize skin disorders into herpes, dermatitis, and psoriasis using SVM.

A system uses computer algorithms and image processing to recognize and gauge the severity of eczema [20]. The framework can recognize and gauge the severity of eczema by enabling users to upload images of afflicted skin. To discriminate between mild and severe eczema, our system employs images segmentation, feature extraction, and statistical classification. An intensity index is assigned to the images in order to identify the kind of eczema. The cross-channel correlation was carried out by the suggested deep residual network model, which disregarded the spatial dimensions. When handling individual channels in an input feature map, cross-correlation is employed. Because it is less susceptible to background oscillations, the cross-channel correlation will decrease. In order to address the imbalance in the dataset, the images and labels were replaced with a vector of images and weights [21].

The initial dataset pictures are utilized to prepare and assess the recommended remaining profound residual *deep* convolutional neural network (RDCNN) organize without any division or preprocessing. Another, sectioned pictures are utilized to test the proposed RDCNN. In conclusion, the trained model that was utilized within the moment experiment is put away and utilized as a pre-trained show within the last try. It is at that point prepared once more with an elective dataset. The recommended RDCNN performs recognizably way better than the current profound convolutional systems [22].

There are three essential steps within the proposed method. The preprocessing step includes portioning the locale of intrigued (ROI) within the input color skin pictures. Moment, revolution, and interpretation changes are connected to improve the ROI pictures that have been sectioned. Third, a few profound convolutional neural organize (DCNN) structures are utilized, counting GoogleNet, ResNet101, and Alex-net. With altered GoogleNet, where the classification exactness was 99.29%, the proposed strategy significantly improved the classification handle [23].

Problem statement

To anticipate these weakening medicines and get effective treatment, melanoma treatment incorporates chemotherapy and radiotherapy. Early conclusion is among the foremost efficient arrangements. A few CAD frameworks are right now accessible for the identification of color skin injuries, such as Dell'Eva-Burroni Melanoma Picture Handling Computer program, which gives a moo execution in genuine usage. Be that as it may, common conclusions around the execution of these frameworks are difficult to define. In sorted out investigation, the different picture securing methods such as dermoscopic, clinical, and ordinary camera pictures encourage to complicate the classification assignment in one worldwide technique. Subsequently, the unused CAD programs are moreover distant from perfect and require advance propels to improve melanoma location and conclusion. In expansion, two significant problems are postured within the classification of skin color injuries into dangerous and generous cancer.

Proposed method

These are segmenting skin lesions to identify cancer and classifying them according to disease severity. With effective segmentation techniques, diseases may be identified more precisely. Deep learning algorithms are frequently employed in image processing for several additional computer vision applications as a consequence of the encouraging results. As a result, in this work, we use deep learning to create a skin lesion detection system. In order to achieve this, the UNetEfficientNet (ITSU-EfficientNet) based on enhanced tuna shoal optimization is presented. Here, input skin images are taken from the dataset and first go through a preprocessing step called median filtering to get rid of any image artefacts. The skin is then constructed using the newly created U-EfficientNet, which we created by integrating UNet and EfficientNet. Here, the suggested ITSO technique is used to fine-tune the classifier's ideal parameters. It is created by fusing TSO with the efficient Pelican optimization algorithm solution updating technique.

Reflective noise and artifacts removal

There are different sorts of procedures for picture smoothing like Gaussian obscuring and middle obscuring to expel indeed a little amount of commotion. For the evacuation of clamors and artifacts reflection from the pictures, a basic thresholding calculation is utilized. When the regions of artifacts are recognized, an engraving movement is connected moreover to the information of their distinguished pixels range, remaining absent from the effect of the artifacts.

Each of the single pixel (x, y) can be identified and classified as a reflection artifact agreeing to the taking after condition:

$$I(x, y) > T_{R1} \text{ and } I(x, y) - I_{avg}(x, y) > T_{R2}$$

where *I* denotes the image with single pixel I(x, y) and $I_{avg}(x, y)$ signifies the normal concentration of the pixel's neighborhood which is computed by the neighborhood cruel

channel and the edge values which are gotten by test, i.e., TR1 = 0.85 and TR2 = 0.097, individually.

Proposed ITSU-EfficientNet-based skin lesion segmentation

According to Fig. 1, the suggested ITSU-EfficientNet skin lesion segmentation was created by combining UNet with EfficientNet. In this case, adjustable parameters are modified using ITSO. Encoder and decoder are the two components that make up the UNet architecture. The decoder is comparable to the UNet decoder in the proposed model, which substitutes EfficientNet B7 for the UNeT encoder. The 3×3 convolutional blocks that make up the UNet decoder do not sample the feature maps after each block in order to concatenate and transmit the encoder's features to the subsequent 3×3 convolutional block. To create the segmentation map, a 1×1 transformation is used after the last block.

Improved tuna swarm optimization (ITSO)

Globally, optimum solutions to optimization issues can be found with the use of metaheuristic algorithms that strike a balance between diversity and richness. Tuna school optimization (TSO) is a balanced diversification and improvement strategy centered on the tuna, a large marine predator. This marine predator consistently swims quickly and in the shape of a fish's tail, giving it away. The predator can swim, but it moves much more slowly than the quick little fish. Predators in the ocean hunt in packs to catch their prey. As a result, we employ two separate techniques to capture the intended fish.

Mathematical modelling

Initialization, expansion, and enhancement are the three partitioned stages of ITSO scientific demonstrating. Fair the candidate (fish) and target (in this case) are initialized within the highlight space. Amid the moment step, known as expansion, candidates look at the useful space for the best worldwide arrangements. The candidate includes space employments and a spiral jumping approach to realize this. The improvement stage's last objective is to completely utilize the highlight space that was as of now found amid the enhancement arrange [5]. Here, in arrangement to diminish moving arrangements absent from the perfect put, we include Pelican's compelling arrangement overhauling behavior into the Pelican optimization calculation. A strategy that fine-tunes the perfect injury division strategy parameters yields the result of protest procurement.

Initialization

Candidates are displayed in the feature space and distributed at random as follows:

$$S_m^{\tau} = k \cdot (U - V) + V, \quad m = 1, 2, \dots Z$$
 (1)

The lower and upper bounds of the highlight space are indicated by V and U individually, the irregular number chosen from the uniform dissemination within the run [0, 1] is indicated by k, and the mth candidate redundancy within the include space is S_m^{τ} .



Fig. 1 Architecture of proposed U-EfficientNet

Wellness work

After initializing the candidates within the include space, the wellness is assessed based on cruel square blunder and is communicated as follows:

$$F_{ITSO} = \frac{1}{T_{sol}} \sum_{i=1}^{T_{sol}} (SS_{obser} - SS_{tar})^2$$
(2)

where the fitness of the solution is F_{ITSO} , the total number of samples is T_{sol} , the observed solution is SS_{obser} , and the target solution is SS_{tar} .

Broadening

Within the broadening stage, the winding diving-based scavenging is formulated by the candidates for capturing the target. The prey alters its swimming course to undertake to urge absent from the candidate. In doing so, they make minor spirals that trap the trap and ruin its capacity to elude. Moreover, data trade among candidates empowers them to dive advance into the utilitarian space. Candidate position overhauls are made within the enhancement stage within the way depicted underneath.

$$S_{m}^{\tau+1} = \begin{cases} \delta_{1}.(S_{G}^{\tau} + \gamma.|S_{G}^{\tau} - S_{m}^{\tau}|) + \delta_{2} \cdot S_{m}^{\tau}, & m = 1\\ \delta_{1}.(S_{G}^{\tau} + \gamma.|S_{G}^{\tau} - S_{m}^{\tau}|) + \delta_{2} \cdot S_{m-1}^{\tau}, & m = 2, 3, \dots Z \end{cases}$$
(3)

$$\delta_1 = q + (1 - q) \cdot \frac{\tau}{\tau^{maxx}} \tag{4}$$

$$\delta_2 = (1-q) - (1-q) \cdot \frac{\tau}{\tau^{maxx}} \tag{5}$$

$$\gamma = e^{pn} \cdot \cos(2\pi p) \tag{6}$$

$$n = e^{3\cos(((\tau^{maxx} + 1/\tau) - 1)\pi)}$$
(7)

The location of the mth candidate in the feature space is represented by $S_m^{\tau+1}$, and the optimal best candidate in the τ^{th} iteration is indicated as S_G^{τ} . The movement of the individual candidate in the feature space is controlled by the weight coefficients δ_1 and δ_2 , wherein the movement extention is decided by the constant q. Display the maximum number of iterations as τ^{maxx} and the range of random values between 0 and 1 as p.

When the ideal candidate fizzled to capture the target, at that point, the arbitrary candidate is chosen from the swarm, and subsequently, the investigation criteria are upgraded, and the position overhauling is defined as follows:

$$S_{m}^{\tau+1} = \begin{cases} \delta_{1}.(S_{k}^{\tau}+\gamma.|S_{k}^{\tau}-S_{m}^{\tau}|) + \delta_{2} \cdot S_{m}^{\tau}, & m = 1\\ \delta_{1}.(S_{k}^{\tau}+\gamma.|S_{k}^{\tau}-S_{m}^{\tau}|) + \delta_{2} \cdot S_{m-1}^{\tau}, & m = 2, 3, \dots Z \end{cases}$$
(8)

Candidates chosen at random are identified by S_k^{τ} . To prevent the solution from diverging from the ideal solution, the effective location update operation of Pelican optimization is combined with that of TSO in this case. The modified solution of the ITSO is expressed as follows:

$$S_m(\tau) = \begin{cases} S_i(\tau+1) & F_{ITSO}(\tau) < F_{ITSO}(\tau+1) \\ S_i(\tau+1) & Otherwise \end{cases}$$
(9)

In order to direct the solution to the ideal location, the candidate update solution uses the effective solution update criterion, which increases the convergence rate of Algorithm 1.

When tackling optimization issues, a metaheuristic calculation with adjusted heightened and broadening can help in finding the worldwide best arrangement. Based on the marine predator known as fish, the fish swarm optimization (TSO) show is the one with adjusted expansion and heightened.

Computational modeling

There are three diverse steps within the numerical modeling of the ITSO: initialization, broadening, and escalated.

Initialization in this setting alludes as it were to setting up targets and candidates (fish) within the highlight space. The moment stage is called broadening, amid which the candidates look the feature space in an exertion to discover the most excellent arrangement universally. To do this, the candidates within the include space utilize the winding plunging procedure. Eventually, the reason of the heightened stage is to completely utilize the highlight space that was discovered amid the expansion stage. Here, the Pelican optimization calculation joins the Pelican's viable arrangement upgrading behavior to play down the development of the arrangement absent from the ideal area. The calculation for altering the injury division technique's perfect parameters accomplishes the objective of capturing the target.

Diversification

The candidates devise a spiral diving-based foraging strategy to capture the target during the diversification phase. The prey attempts to evade the candidate by swimming in a different direction. As a result, the candidates create a tight spiral to ensnare the prey, decreasing its ability to flee. Furthermore, the candidates' exchange of information facilitates the exploration of a larger area within the feature space. The candidates' updated positions during the diversification phase are developed.

The investigation criteria are made strides, and the position upgrading is defined when the ideal candidate is incapable to capture the target and an arbitrary candidate is chosen from the swarm.

As a result, the arrangement overhauled by the candidates utilizing the proficient arrangement upgrading criteria coordinates the arrangement to the perfect area that quickens the algorithm's rate of meeting.

Termination

The algorithm iteration ends upon the acquisition of the global best solution or the achievement of r^{max} . Algorithm 1 presents pseudo-code for the ITSO algorithm.

Algorithm 1. Pseudo-code for ITSO algorithm

Input: The values τ_{max} and Z are initialized Output: Best solution S_G^{τ} Locate the candidates in the feature space $S_m(m = 1, 2, ..., Z)$ The factor q is defined while ($\tau < \tau^{max}$) Estimate the fitness Update S_G^{τ} For (all candidates) do Update δ_1 , δ_2 and a End for $\tau=\tau++$ End while Return the best solution

Vision transformers for image classification

A network of transformers is used by visual transformer, a deep learning architecture, to carry out images categorization tasks. Vision Transformer employs self-attention, like other transformer networks, to process the input images as a collection of patches rather than a grid of pixels. As a result, the network is better equipped to handle image classification tasks and capture long-range relationships of images. It has been demonstrated that Vision Converter can perform at the cutting edge on a number of images categorization benchmarks, including ImageNet. It is capable of carrying out a variety of images categorization tasks, including scene and object detection [6]. Vision Converter provides a number of benefits over existing deep learning architectures for images categorization in addition to its excellent performance. For instance, it can take input images of any size without human scaling or cropping and may be trained with a modest quantity of data. As a result, it is appropriate for problems involving images categorization in real-world settings.

Vision Modern approaches employ transformers to appropriately categorize images. The ViT framework utilizes a transformer model with several image blots in place of the framework's convolutional layers. The ViT architecture receives a linear embedding series of separated patches from the input frame that has been reduced and converted into N patches.

$$I \in \mathbb{R}^{H \times R \times C} \Rightarrow Ip \in \mathbb{R}^{N \times (P^2 \times C)}$$
(10)

$$N = H \times \frac{W}{p^2} \tag{11}$$

where I stand for the image, H, W, and C for the original image height, width, and channels of the original image, the representation of the spots created using the example image the resolution of the individual spots, and N for the total number of spots created using the example image.

Additionally, each patch's matrix of dimensions (P, P) is flattened into a matrix of dimensions $(1, p^2)$. To create a linear patch projection (p^2, D) , these compressed patches (E $(1, p^2)$ are transmitted via a single feed-forward layer that contains the embedding matrix (F). A conceptual overview of the ViT model is provided by patch embeddings (E), which are fixed latent vectors of modified size D (projection dimension) (1, D). They describe how patch embeddings are generated. Linear patch projections are thought of as learnable [class] embeddings. Since data is fed instantly, the sequence of inheritance must be done organically.

$$z_0 = \left[I_{class}; x_1 E; x_2 E; \dots x_n E + E_{pos} \right]$$
(12)

$$E \epsilon R^{(P^2C) \times D}, E_{pos} \epsilon R^{(N+1) \times D}$$
(13)

In order to solve this, extend the connection matrix (I_{class}) by including a concatenated matrix made up of patch embeddings (E) and learnable class embeddings (E_{pos}) . The below equation provides the results of the extracted patch and as a result embedded sequentially with the token z_0 .

$$z'_{l} = MSA(LN(Z_{l-1})) + Z_{l-1}$$
(14)

$$Z_l = MLP(LN(z_l')) + z_l'$$
⁽¹⁵⁾

As illustrated, a combination of these implanting groupings serves as the input to a sensor encoder composed of L indistinguishable layers of multi-head self-attention (MSA) and multilayer perceptron (MLP) squares. The two transformer encoder components serve as the ultimate two skip associations taking after the normalization layer (LN), concurring to the scientific detail.

$$SA = Softmax \left(QK^T \sqrt{d_k} \right) \times V = W_{attention} \times V \tag{16}$$

$$MSA = concat(SA_1, SA_2, SA_3, \dots, SA_h) \times W^o$$
(17)

$$w^o \in R^{hd_k \times D} \tag{18}$$

The attention block's beginning weight sets W_{q} , W_{k} , and W_{v} are multiplied by the MSA block's associated action and input vectors to create three unique matrices with queries Q, keys K, and values V. Make a matrix of attention. The inner product of each query in matrix Q and each key in matrix K is the transpose of matrix Q multiplied by matrix K. Use the same scale dot product as the typical dot product for self-attention (SA) blocks while taking the dimensionality of the key dk into account as a scale factor parameter. Softmax has an internal result that it uses to calculate the attention weights. The MSA block of the transencoder is used to compute the scaled internal attention for each h-head. Using learnable weights W0 as given in Eq, the connection result of

each attention head is sent to the feed-forward layer. The MLP block is composed of the highly nonlinear feed-forward dense layers of GeLU. In the last layer of encoding, the top components of the sequence ZL are passed to the exogenous classifier to estimate the class label. Figure 2 depicts the structure of the proposed ViT model with trans encoder and multi-attention blocks.

Datasets

During experiments such as HAM10000 and ISIC2018, two datasets are used to accomplish this task. Here are the details of the two datasets:

HAM1000 dataset

The HAM10000 "People Against Machines with 10,000 Preparing Pictures" dataset contains an addition of up to 10,015 dermoscopic pictures for the discovery of pigmented skin injuries freely open through the ISIC store, which is one of. The dataset incorporates an assortment of injuries such as melanocytic nevus (nv = 6705), actinic keratosis (akiec = 327), dermatofibroma (df = 115), basal cell carcinoma (bcc = 514), and vascular injury. Seven distinctive categories of pictures were classified as (vacs = 115), positive keratosis (bkl = 1099), and melanoma (mel = 1113). The dataset contains pictures of 54% male and 45 male skin injuries. Classification of these skin classes is not simple, and misclassification rates can be tall due to the issue of moo inter-class changeability and tall intra-class inconstancy in composite datasets with numerous pictures of skin injuries. A few illustration pictures are appeared in Fig. 3.

ISIC-2018 dataset

The Universal Skin Imaging Collaboration (ISIC) has discharged the ISIC-2018 dataset, a sizable collection of dermoscopic pictures that incorporates more than 12,500 pictures. Injury division, quality recognizable proof, and sickness classification are the three errands that each dataset completes. The dataset contains more than 10,000 pictures partitioned into seven course categories for classification errands. Figure 4 shows an outline of a pictures. The ISIC-2018 challenge has two essential issues. The primary is that certain categories as it was have a little number of pictures. At the moment, it is challenging for the classifier to adjust the categorization because of the unequal sum of pictures in each lesson.

Results and discussion

The experimental findings of the suggested framework are presented in this section. Analyze the categorization and segmentation of lesion outcomes. Lesion segmentation takes into account two factors: accuracy rate and error rate. The execution time of each test image is also recorded after the final segmentation. We experimented with several classifiers during the classification process to evaluate how well the CNN performed. The execution time of each test image is also recorded after the final segmentation. We experimented with several classifiers during the classification process to evaluate how well the CNN performed.







Fig. 3 Test pictures of HAM1000 dataset. Melanoma (mel), melanocytic nevus (nv), basal cell carcinoma (bcc), actinic keratosis intraepithelial carcinoma (akiec), generous keratosis (bkl), dermatofibroma (df), and vascular injury (vasc)



Performance metrics

Six criteria—accuracy, specificity, accuracy, F1 score, sensitivity, and Matthew correlation coefficient—were utilized to assess performance analysis (MCC). The following is the measuring formula's mathematical model:

$$Specifity = \frac{TN}{TN + FP} \times 100 \tag{19}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FN} \times 100$$
(20)

$$Precision = \frac{TN}{TN + FP} \times 100 \tag{21}$$

$$Sensitivity = \frac{TN}{TN + FP} \times 100$$
⁽²²⁾

$$F1score = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \times 100$$
(23)

$$MCC = \frac{TP \times TN - TP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \times 100$$
(24)

where TP and TN stand for the quantity of pixels that were properly identified as background and object, respectively. The number of pixels allotted to the background specifier and the background specifier object, respectively, is represented by the values FN and FP.

Preprocessing

Dermoscopic images of skin lesions undergo two rounds of preprocessing, with the first stage being the most crucial. The idea of hair removal is based on morphological modifications, which makes picking the right ROL simpler. To take advantage of intensity-based images enhancement, there is additional preprocessing step after the depilation procedure. It is going to receive better hairless shots after preprocessing, and you can utilize ROL to more precisely and quickly segment skin images. Figure 5 displays the outcomes of the preprocessing procedure.

Segmentation

Here, the lesion segmented numerical findings are shown alongside the visually segmented image. Additionally, accuracy ratings are contrasted with current best practises. The findings presented in this table were computed using the mean of the segmentation precision for each images chosen. The segmented images are then compared to the given real images using the newly developed ViT model. The database treats each images added in the same way. The average accuracy, FNR, and overall running time for each dataset are then calculated. The suggested segmentation



Fig. 5 a Original images and b preprocessed image

approach in ELM yields an average accuracy of 96.28%. A total of 52.3652 s passed during the lesion segmentation examination, and the error rate was 4.69%. Using KELM restores accuracy by 96.49% with a 4.32% error rate. This dataset's estimated test time is 60.5160 s. MSVM, a challenging split, attained accuracy of 93.70%. The error rate is 7.45%, and the execution time is 68.4202 s. Finally, we introduce the characteristics of EfficientNet. We achieved a 99.78% accuracy rate and a 1.5% ITSU error rate. The amount of time that has passed is 29.4356 s. Therefore, it follows that the execution time will be greater, the larger the dataset. For instance, ITSU-EfficientNet only needs 29.4356(s) to process 100 images. The simulation results of skin cancer segmentation and classification are shown in Figs. 6 and 7.

Classification

As a result of our calculations utilizing the recommended system, the numbers are shown in Table 1. Within the recommended system, we utilized the ViT classifier. For comparison, we too utilized the credulous Bayes, ELM, MSVM, and KELM classifiers. The table appears that the proposed ViT demonstrate and accomplished the normal exactness of 99.78% and 10.43% FNR in with computation time of 134.4632 s of ISIC-2018 dataset. The ordinary MSVM classifier execution measurements are FNR of 15.34%, a time of 122.5230 s, and the normal exactness is 96.45%. Exactness values for

INPUT	PRE	SEGMENTATIO	TYPE OF	ACCURA
IMAGE	PROCESSING	Ν	CANCER	СҮ
	IMAGE			
	Executing trades to hort-balance to their	Critor perior	Benign – Melanocytic Nevi	98.83
•			Benign – Melanocytic Nevi	99.87
	Encode Journey So Lori Lances const.	Case parts	Malignant – Squamous Cell Carcinoma	96.93
	Estandly bardy to be allowed as a set of the		Benign – Melanocytic Nevi	98.99
*		Georgepte	Benign - Acrochordon	98.99
0	Educatly using the for Lemma must	Castor porter	Benign – Melanocytic Nevi	99.57
-		Cor yes	Malignant - Melanoma	98.72
	Estandly starts to be includes in the	Georgentes	Benign - Acrochordon	99.45
-		Corporte Corporte	Benign – Melanocytic Nevi	98.64
•		Georgente	Malignant - Melanoma	99.03

Fig. 6 Proposed lesion location, with findings identified



Fig. 7 Average accuracy of skin cancer prediction

Table 1 Performance analysis of classification of ISIC-2018 dataset

Classifier	Performance measures			
	Accuracy (%)	Sensitivity (%)	FNR (%)	Prediction time (s)
Naive Bayes	93.34	92.24	19.23	158.2304
ELM	94.23	94.15	16.12	139.3352
MSVM	96.45	96.23	15.34	122.5230
Proposed ViT	99.78	97.27	10.43	134.4632

Table 2 Perform	mance analysis of	classification	of HAM1000 dataset
-----------------	-------------------	----------------	--------------------

Classifier	Performance mea	Performance measures			
	Accuracy (%)	Sensitivity (%)	FNR (%)	Prediction time (s)	
Naive Bayes	92.36	91.28	18.21	155.2104	
ELM	93.22	93.17	15.13	138.4731	
MSVM	95.44	95.23	14.35	121.8904	
Proposed ViT	99.16	99.28	9.38	133.4782	

Credulous Bayes, ELM, and KELM are 93.34%, 94.23%, and 93.45%, individually. The proposed ViT show accomplished the normal precision of 99.16% and 9.38% FNR with computation time of 133.4782s of HAM1000 dataset as in Table 2.

Confusion matrix

The disarray network of the proposed demonstration of ViT of the ISIC-2018 and HAM10000 appeared in figures. It makes categorization comes about simpler to imagine. The number of distinctive skin injuries classified by the model is appeared within the disarray lattice beneath each category. It illustrates the model's execution in classifying the particular skin injury well. It is a relationship framework between the genuine name and the model's classification (anticipated name). The perplexity framework for the proposed ViT demonstration of ISIC-2018 and HAM10000 datasets appeared in Figs. 8 and 9.

Receiver operating characteristic (ROC)

Demonstrative viability was surveyed utilizing the AUC collector working characteristic bend (ROC), portraying the model's categorization viability as a work of two parameters: genuine positives and wrong positives. The AUC is calculated as the zone



Fig. 8 Confusion matrix for the proposed ViT model of ISIC-2018



Fig. 9 Confusion matrix for the proposed ViT model of HAM10000



Fig. 10 ROC curve for ViT model of ISIC-2018



Fig. 11 ROC curve for ViT model of HAM10000

beneath the ROC bend secured by little trapezoidal portions. As appeared in figure, we performed ROC investigations employing a CNN show with a range of 0.83. The best-case ROC result for the recommended show after fine-tuning utilizing ViT show of ISIC-2018 is appeared in Fig. 10.

After a few times of fine-tuning, we got the finest which comes about with 100 epochs and gotten the zone beneath bend (AUC) esteem of 0.847 that appeared in Fig. 11. The proposed ROC bend for ViT showing HAM10000 can be sent in a CAD framework that will offer assistance viably to classify the skin injuries at the early arrangement. In expansion, the early location of harmful growth in the skin

particularly for those without induction to specialists can altogether empower them to induce the treatment and enhance the survival plausibility.

Results and discussion

A few strategies have been proposed within the literature to move forward decisionmaking within the conclusion of skin injuries, especially melanoma. The creators proposed a careful comparison of U-Net and attention-based skin injury picture division strategies. At long last, this paper recommends that a pre-trained Vision Transformer be fine-tuned for skin cancer classification. In comparison to other strategies, the proposed strategy beats existing state-of-the-art CNN models. Early ceasing and versatile learning rate are utilized to anticipate overfitting amid the preparing handle, which contributes to the model's execution enhancement. On two benchmark datasets, the test comes about appear that the proposed strategy beats existing approaches. The proposed building arrangement equalizations computation taken a toll and classification execution. The ponder permitted for the improvement of a prescient demonstration with greatly tall affectability and specificity, compared to greatly low amounts of falses (both positive and negative). This is often due to the self-attention instrument which, by taking under consideration the correspondences between patches, can superiorly get it the image's substance.

Conclusions

Skin cancer pros can physically distinguish harmful spots utilizing dermoscopy pictures but typically still a challenging effort; thus, computerized approaches were made to form the method simpler. In this work, we categorize skin cancer images utilizing the HAM1000 and ISIC-2018 databases. The preprocessing strategies required to form a robotized skin cancer discovery framework are portrayed in this article. The forms incorporate nitty-gritty depictions of all the stages included in viable strategies for improving skin cancer pictures as well as down-to-earth channels for picture clamor evacuation and picture smoothing. The recently formulated TSU-EfficientNet combines the ITSO calculation, UNet, and EfficientNet, wherein the learnable parameters of the profound learning demonstration are tuned utilizing the ITSO calculation. Here, the adjusted heightened and enhancement stage of the ITSO calculation kills data misfortune amid the information learning stage. It is the comparison of the viability of EfficientNet with four particular Naive narrows, ELM, and MSVM-based hyperparameter-optimized parts. In differentiating customary approaches, it comes about illustrating improved execution. For exact division, it differentiates development strategy particularly increments division precision. Computation time is one of our work's downsides; in any case, we arrange to illuminate it in consequent endeavors. In our up and coming work, we will extend our division strategy to avoid preparing profound models on futile pictures information.

Abbreviations

ITSOImproved tuna swarm optimizationMFS-NetMulti-focus segmentation networkRDCNNResidual deep convolutional neural networkDCNNDeep convolutional neural networkTSOTuna school optimization

MSA	Multi-head self-attention
MLP	Multilaver perceptron

ISIC International skin imaging collaboration

Acknowledgements

Not applicable

Authors' contributions

MR contributed to technical and SNG conceptual content and architectural design. PR contributed to guidance and ENG counseling on the writing of the paper. All authors read and approved the final manuscript.

Funding

No funding received by any government or private concern

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 2 September 2023 Accepted: 24 February 2024 Published online: 14 March 2024

References

- 1. Ali R, Manikandan A, Xu J (2023) A novel framework of adaptive fuzzy-GLCM segmentation and fuzzy with capsules network (F-CapsNet) classification. Neural Comput Applic. https://doi.org/10.1007/s00521-023-08666-y
- Annamalai M, Muthiah P (2022) An early prediction of tumor in heart by cardiac masses classification in echocardiogram images using robust back propagation neural network classifier. Braz Arch Biol Technol 65. https://doi.org/10. 1590/1678-4324-2022210316
- Manikandan A, PonniBala M (2023) Intracardiac mass detection and classification using double convolutional neural network classifier. J Eng Res 11(2A):272–280. https://doi.org/10.36909/jer.12237
- Sheikdavood K, Surendar P, Manikandan A (2016) Certain investigation on latent fingerprint improvement through multi-scale patch based sparse representation. Indian J Eng 13(31):59–64
- Wang J, Zhu L, Wu B, Ryspayev A (2022) Forestry canopy image segmentation based on improved tuna swarm optimization. Forests 13:1746. https://doi.org/10.3390/f13111746
- Yang G, Luo S, Greer P (2023) A novel vision transformer model for skin cancer classification. Neural Process Lett 55:9335–9351. https://doi.org/10.1007/s11063-023-11204-5
- Venmathi AR, David S, Govinda E, Ganapriya K, Dhanapal R, Manikandan A (2023) An automatic brain tumors detection and classification using deep convolutional neural network with VGG-19. In: 2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, pp 1–5. https://doi.org/10.1109/ICAECA56562.2023.10200949
- Balamurugan D, Aravinth SS, Reddy PC, Rupani A, Manikandan A (2022) Multiview objects recognition using deep learning-based Wrap-CNN with voting scheme. Neural Process Lett 54:1–27. https://doi.org/10.1007/ s11063-021-10679-4
- Manikandan A, Jamuna V (2017) Single image super resolution via FRI reconstruction method. J Adv Res Dyn Control Syst 9(2):23–28
- 10. Sharif M, Akram T, Kadry S, Hsu C-H (2021) A two-stream deep neural network-based intelligent system for complex skin cancer types classification. Int J Intell Syst 37 [CrossRef]
- Alhaisoni M, Tariq U, Hussain N, Majid A, Damaševičius R, Maskeliunas R (2021) COVID-19 case recognition from chest CT⁻ images by deep learning, entropy-controlled firefly optimization, and parallel feature fusion. Sensors 21:7286
- 12. Manikandan A, Suganya K, Saranya N, Sudha V, Sweetha S (2017) Assessment of intracardiac masses classification. J Chem Pharm Sci 5:101–103
- 13. Swamy KCT, Kishore VV, Ahmed ST, Farida MA (2021) Investigation of GPS-TEC inconsistency and correlation with SSN, solar flux (F10.7 cm) and Ap-index during low and high solar activity periods (2008 and 2014) over Indian equatorial low latitude region. In: 2021 International Conference on Intelligent Technologies (CONIT), Hubli, India, pp 1–9. https://doi.org/10.1109/CONIT51480.2021.9498292
- Kalpana V, Vijaya Kishore V, Praveena K (2020) A common framework for the extraction of ILD patterns from CT image. In: Hitendra Sarma T, Sankar V, Shaik R (eds) Emerging Trends in Electrical, Communications, and Information Technologies. Lecture Notes in Electrical Engineering, vol 569. Springer, Singapore. https://doi.org/10.1007/978-981-13-8942-9_42
- Vijaya Kishore V, Kalpana V (2020) ROI segmentation and detection of neoplasm based on morphology using segmentation operators. In: Hitendra Sarma T, Sankar V, Shaik R (eds) Emerging Trends in Electrical, Communications, and Information Technologies. Lecture Notes in Electrical Engineering, vol 569. Springer, Singapore. https://doi.org/ 10.1007/978-981-13-8942-9_41

- Vijaya Kishore V, Kalpana V (2020) Effect of noise on segmentation evaluation parameters. In: Pant M, Kumar Sharma T, Arya R, Sahana B, Zolfagharinia H (eds) Soft Computing: Theories and Applications. Advances in Intelligent Systems and Computing, vol 1154. Springer, Singapore. https://doi.org/10.1007/978-981-15-4032-5_41
- Kalpana V, Vijaya Kishore V, Satyanarayana RVS (2023) MRI and SPECT brain image analysis using image fusion. In: Marriwala N, Tripathi C, Jain S, Kumar D (eds) Mobile Radio Communications and 5G Networks. Lecture Notes in Networks and Systems, vol 588. Springer, Singapore. https://doi.org/10.1007/978-981-19-7982-8_48
- Xie Y, Zhang J, Xia Y, Shen C (2020) A mutual bootstrapping model for automated skin lesion segmentation and classification. IEEE Trans Med Imaging 39:2482–2493 ([CrossRef])
- Khan MA, Zhang Y-D, Sharif M, Akram T (2021) Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification. Comput Electr Eng 90:106956 ([CrossRef])
- Jin Q, Cui H, Sun C, Meng Z, Su R (2020) Cascade knowledge diffusion network for skin lesion diagnosis and segmentation. Appl Soft Comput 99:106881 ([CrossRef])
- Alsahafi Y, Kassem M, Hosny K (2023) Skin-Net: a novel deep residual network for skin lesions classification using multilevel feature extraction and cross-channel correlation with detection of outlier. J Big Data 10:105. https://doi. org/10.1186/s40537-023-00769-6
- Hosny K, Kassem M (2022) Refined residual deep convolutional network for skin lesion classification. J Digit Imaging. 35. https://doi.org/10.1007/s10278-021-00552-0
- 23. Hosny K, Kassem M, Fouad M (2020) Skin melanoma classification using ROI and data augmentation with deep convolutional neural networks. Multimed Tools Appl 79. https://doi.org/10.1007/s11042-020-09067-2

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.