ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Designing a Two-Stream Network Based Unsupervised Learning for Skin Cancer Recognition

Aekkarat Suksukont[1], Anuruk Prommakhot[2], Wichian Ooppakaew[3] and Jakkree Srinonchat[4]

ABSTRACT

Computer vision is crucial in identifying and diagnosing diseases like skin cancer, which can become life-threatening if not detected early. Although numerous methods have been developed, these techniques often face challenges due to the varied nature of skin cancer, which frequently presents irregular shapes and ambiguous structures. In this study, we introduce the design of an unsupervised two-stream network capable of simultaneously learning from datasets of various sizes. The network parameters are organized from smallest to largest to improve the efficiency of feature extraction. Additionally, the network incorporates residual blocks, bidirectional long short-term memory, and an attention layer to help reduce training loss. The proposed method was tested using the PAD-UFES-20 dataset, using mean square error to measure training loss and accuracy to check how well it recognizes skin cancer. The results showed a loss of 0.0079 and a training time of only 0.53 minutes, performing better than other advanced methods in both loss and speed. Our approach showed better results than previous methods, accurately recognizing skin cancer and showing potential for use in a mobile app to help with early detection and diagnosis.

Article information:

Keywords: Skin Cancer Recognition, Two-stream Network, Convolutional Neural Network, Deep Learning

Article history:
Received: November 7, 2024
Revised: June 8, 2025
Accepted: August 7, 2025
Published: August 9, 2025 (Online)

DOI: 10.37936/ecti-cit.2025193.259691

1. INTRODUCTION

Skin cancer begins in the layers of the skin and often appears as bumps or sores, especially on the face [1-3]. To diagnose it, doctors typically take a small tissue sample from the area for closer examination. It helps them identify the type of cancer cells and better understand their characteristics. Although there are image processing and computer vision tools that help detect features of skin cancer, these tools still have limitations. The complex nature of lesion features and various factors affecting results mean these methods cannot fully replace a histological diagnosis [4, 5].

In recent years, deep learning (DL) has shown promise in speeding up diagnosis and improving accuracy [6-11] to work in many areas. It has also become necessary to develop histological methods for detecting skin cancer. DL has attracted much attention and has been widely researched by analyzing the texture, size, and shape of nodules in images. J. Chaki *et al.* [12] introduced the DarkNet19, InceptionV3, and ResNet101 models, combining features to create deep feature vectors and selecting the most important ones using the iterative ReliefF feature. These vectors were then input into a k-nearest neighbor classifier, tested with a Bayesian optimizer, and trained on three different datasets. S. M. M. Elkhol *et al.* [13] developed an intelligent system for predicting and classifying diseases using an improved deep belief Network (DBN) as the main algorithm. They used softmax and categorical cross-entropy functions to measure accuracy and check the training results. S. D. Pande *et al.* [14] created a DL

[1]The author is with the Department of Computer Engineering, Faculty of Industrial Education, Rajamangala University of Technology Suvarnabhumi, Sam Chuk, Suphanburi, 72130, Thailand., Email: aekkarat.s@rmutsb.ac.th
[2]The author is with the Department of Computer Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand., Email: anuruk.p@en.rmutt.ac.th
[3,4]The authors are with the Department of Electronics and Telecommunication Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi, Pathum Thani, Thailand., Email: wichian.o@en.rmutt.ac.th and jakkree.s@en.rmutt.ac.th
[4]Corresponding author: jakkree.s@en.rmutt.ac.th

model, YOLOv8n-cls, to recognize different types of diseases. This method uses a convolutional neural network (CNN) as its base to analyze images in one step, producing predictions on the labels of objects in the images. F. L. Tiro *et al.* [15] created an in-vivo endoscopic image dataset and tested models based on both shallow and DL, such as XGBoost and InceptionV3. InceptionV3 showed higher accuracy than traditional methods that use ex-vivo data, though it also had some limits in classification. R. Magherini *et al.* [16] introduced a method that uses deep features with nnU-net to segment tumors and gather important features. The models were trained and tested on a mix of KiTS2019 data from Careggi University Hospital, using radio mic features from medical images. R. Gaikar *et al.* [17] suggested a method for segmenting images using a 2D attention U-Net model and tested five advanced methods to detect RMs in mpMRI images. They compared the results based on precision, recall, specificity, and the dice similarity score. A. Prommakhot *et al.* [18] created a combined model using VGG19 and VGG16 in a dual design to improve classification efficiency and reduce computing costs. While these studies used DL to boost recognition and diagnosis, some issues remain, such as changes in disease features that make it hard for the networks to process details effectively. The design of deep networks also affects training time, slowing progress in achieving better results.

Developing deep learning for skin cancer detection is another widely studied research approach. M. I. Faizi *et al.* [19] proposed an automated system that begins with template matching and uses k-means clustering to highlight regions of interest and cluster the data. The system enhances image sharpness with hu moments, analyzes shape, and uses GLCM to extract texture features before feeding them to KNN, SVM, and Random Forest for skin cancer classification. Z. Lan *et al.* [20] improved a capsule network (FixCaps) for skin cancer classification. The design employs a 31×31 kernel to enhance image recognition performance and reduce spatial information loss through a convolutional block attention module. It also utilizes group convolution to prevent incomplete training. L. We *et al.* [21] proposed a feature extraction and classification module for lesion classification, using a lightweight CNN to extract features from image samples. They developed a lightweight segmentation model based on the U-Net structure and conducted experiments with the ISBI 2016 dataset. N. Q. Nguyen *et al.* [22] proposed a new polyp segmentation method using the CDED-net architecture, which combines multiple deep encoder-decoder networks to gather multi-scale information, emphasize object boundaries with multi-scale decoders, and improve accuracy through a new dice loss function, making CDED-net more precise in cancer segmentation. P. A. Lyakhov *et al.* [23] proposed a multimodal neural network system with an improved cross-entropy loss function for imbalanced skin data to enhance the accuracy of skin cancer detection. S. S. Zareen *et al.* [24] proposed a ResNet-50-based CNN-RNN model for feature extraction to enhance performance in skin cancer classification, designed to LSTM for sequence detection. C. Kavitha *et al.* [25] proposed CNN and R-CNN techniques to enhance the efficiency of skin cancer detection and classification. The process begins with image pre-processing to remove noise and clarify information, followed by CNN for skin cancer classification and R-CNN for data analysis, aiming to develop an accurate automated system for early skin cancer diagnosis. J. S. T. Purni *et al.* [26] proposed a coronavirus optimization algorithm, an Ebola optimization search algorithm, and an improved canny edge detector for target edge detection designed to work with CNN for skin cancer classification. The experiments were conducted using the ISIC 2018 and 2019 datasets. M. S. Ali *et al.* [27] proposed a DCNN model to classify benign and skin cancer by denoising and normalizing images and using data augmentation to increase the number of images for training. This experiment uses AlexNet, ResNet, and VGG16 to train on the HAM10000 dataset. The study aims to optimize the DL for improved results. However, its implementation still requires consideration of factors such as time and data diversity, which contribute to the study suboptimal performance.

This study suggests an unsupervised network (UL) design to create a more unified feature representation. It includes designing a network that can learn from different data sizes by arranging network parameters hierarchically to capture features efficiently. The network also includes a residual block (RB), bidirectional long short-term memory (Bi-LSTM), and an attention layer (AL) to boost efficiency. The design aims to understand how two-stream networks learn to recognize skin cancer. The proposed method is tested using mean square error to measure UL's learning performance and compare the results with other advanced methods. Accuracy is used to show how well the network recognizes skin cancer.

2. METHOD

2.1 Convolutional neural network

Convolutional neural networks (CNN) [18, 28] have a different structure from neural networks (NN), with layers specifically designed for images, as presented in many research works [29-32]. The convolutional layer extracts image features by sliding a kernel across different image parts to capture spatial details. This process downsizes the feature map, reducing parameters and simplifying the model, forming the basis of the network's feature extraction, as shown in equation (1).

$$y_{i;j} \sum_{m} \sum_{n} x_{(i+m),(j+n)} \cdot k_{m,n} + b \quad (1)$$

When x represents the input feature map, k is the kernel or matrix for extracting features, b is the bias value, and y is the output feature map produced by the convolutional operation. This operation shifts the k filters across the input image x to calculate values at each point, resulting in a smaller feature map that captures spatial details. This feature map is then sent to the fully connected layer, the final layer that combines all the features from previous layers to make predictions or recognize the target, as shown in equation (2).

$$y = f(W \cdot x) + b \quad (2)$$

When W is the weight vector, x is the feature vector from the previous layer, b is the bias value, and f is the activation function for recognizing the target.

2.2 Residual layer

A Residual layer (RL) [33] uses skip connections, or shortcuts, that let data bypass particular layers directly. It gives the data an extra path through the network, helping the model learn more effectively and making it easier to train deeper models, as shown in equation (3). In this setup, $H(x)$ is the function the model aims to learn, $F(x)$ is the actual function processed by the residual function, which might be calculated with a convolutional layer and an activation function, and x is the original input that passes through the skip connection.

$$H(x) = F(x) + x \quad (3)$$

In the equation, the input x is added to the function $F(x)$ output to produce the result of the Residual Layer, which is then passed to the next layer in the network.

2.3 Attention Layer

The Attention layer (AL) [34] assigns a weight to each data element to show how important each one is. It has three main parts: Query (Q), which represents the data used to identify essential parts, Key (K), which helps signal the importance and Value (V), which is the actual data. This study uses Self-Attention, where each position, including itself, is in the same order. This method works well for data that follows a continuous sequence, as shown in equation (4).

$$Attention(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (4)$$

$$MH(Q, K, V) = \text{Concat}(head_1, \ldots, head_h).W^{\circ} \quad (5)$$

Where head $= attention(QWiq, KWik, VWiw)$ and W° is the weight used to combine information from multiple heads, multi-head attention enables the model to learn complex features from different sources, as in equation (5) , leading to more varied and accurate processing.

2.4 Bidirectional Long Short-Term Memory

Bidirectional long short-term memory (Bi-LSTM) [35, 36] is a neural network that improves on LSTM by processing data in two directions, forward and backward, allowing it to understand the context from previous and following data. It makes it useful for tasks that need sequential context. Each LSTM has three main gates: the Forget Gate, which decides whether to discard information from the previous state, as shown in equation (6); the Input Gate, which decides which new information to store and calculates a new cell state candidate (Ct), as shown in equation (7), (8); and the Output Gate, which determines whether to send a value to the next state and calculates the hidden state, as shown in equation (9), (10).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f \quad (6)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i \quad (7)$$

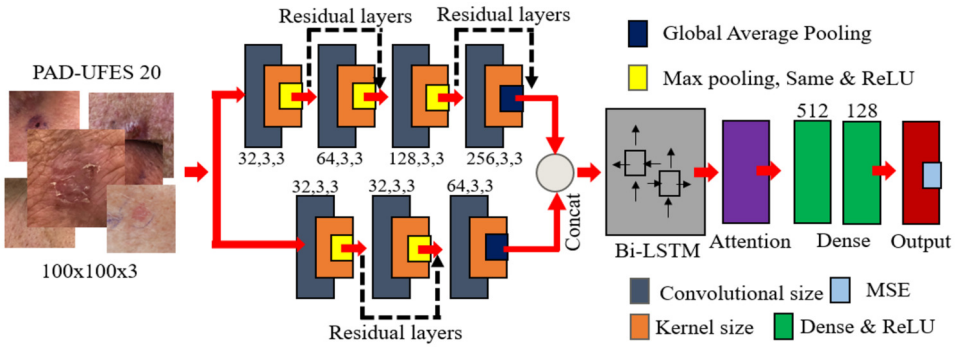


Fig.1: *Proposed method.*

$$\bar{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C \quad (8)$$

$$O_t = \sigma(W_C \cdot [h_{t-1}, x_t] + b_o \quad (9)$$

$$h_t = o_t * \tanh(C_t) \quad (10)$$

Here, σ is a sigmoid function, tanh is a hyperbolic tangent function that outputs values between -1 and 1, and W and b are the weights and biases the model learns. In Bi-LSTM, data moves through the LSTM in two directions—forward and backward—and the hidden states from both directions are combined to produce the Bi-LSTM output, as shown in equation (11).

$$h_t = [h_t^{forward}; h_t^{backward}] \quad (11)$$

Here, *[;]* represents the joining of two vectors, allowing the model to have information from both before and after each point.

3. PROPOSED

his study introduces an unsupervised dual-stream network designed for end-to-end learning, meaning the task runs automatically from start to finish without intervention. As shown in Fig. 1, the convolutional layers, as shown in equations (1) and (2), are arranged with sizes from 32 to 256 and 32 to 64, using a 3×3 kernel to capture different features. 2×2 Max-pooling reduces the feature map size, reducing the number of parameters and model complexity. As shown in equation (3), RL is added between networks, allowing gradients to pass smoothly through shortcuts for stable learning. Global average pooling (GAP) filters patches in the final layer to further reduce the parameters. The patches are combined using a concatenation layer [18, 31] to integrate features extracted from the convolutional layer and Bi-LSTM (64 and 128 units), as shown in equation (11), and the Attention layer, as in equation (4) and (5), to enhance feature quality, capture contextual information from previous image processing steps, and highlight key features before reaching the dense layer (512 and 128 nodes) as shown in equation (2), with a linear function. This network takes advantage of unsupervised learning, cutting down on data preparation time and resources, making it suited for recognizing complex.

4. EXPERIMENT SETUP

The network shown in Fig. 1 was tested on images sized 100×100×3 pixels and trained 50 epochs. During training, the learning rate (LR) was set to 10^{-3}, 10^{-4}, 10^{-5} using the Adam optimizer, with $\beta 1$ set to 0.9 and $\beta 2$ set to 0.0009. The mean squared error (MSE) function was used as the learning function, with the ReLU function for feature extraction and the Sigmoid function in the final layer for recognizing, the batch size was set to 64. The hyper-parameter details are shown in Table 1.

Table 1: *Hyper-parameter for training model.*

Parameter	Value
Input size	100x100x3
Max epochs	50
$\beta 1$, $\beta 2$ Adam optimizer	0.9, 0.0009
Learning rate	10^{-3}, 10^{-4}, 10^{-5}
Function	Mean squared error
Batch size	64

4.1 Dataset

This study uses the PAD-UFES-20 dataset [37] for researching medical skin cancer recognition. The dataset includes 2,298 images collected from 1,373 patients, covering 1,641 skin lesions. The samples represent six types of skin lesions: 3 cancer types (basal cell carcinoma, squamous cell carcinoma, and melanoma) and 3 non-cancerous types (actinic keratosis, seborrheic keratosis, and nevus). Around 58% of the samples are biopsy-proven, especially for cancerous cases. All images vary in size, as they were collected using different smartphone devices. This study's dataset is split into training and testing sets, where 80% is used for training (60% for the training set and 20% for the validation set) and 20% for testing. Examples of the skin cancer images can be seen in Fig. 2.

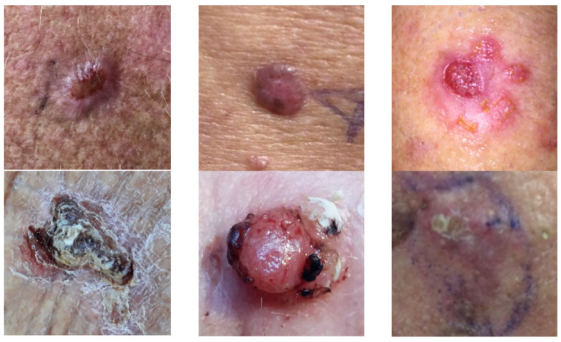


Fig.2: *PAD-UFES-20 dataset.*

4.2 Evaluation

In UL tasks, performance is measured by creating new features, reducing the data's dimensionality, and comparing the difference between the original data and the new data generated after training. MSE is used to measure learning loss, as shown in equation (12), where y_i is the actual value, and $\hat{y}_i$ is the value generated by the model.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (12)$$

Accuracy is a common metric for measuring model performance. In this study, it is used to show how well the model recognizes skin cancer. Accuracy is calculated as the number of correct predictions (NCP) divided by the total number of samples (TNS), as shown in equation (13).

$$\text{Accuracy} = \frac{NCP}{TNS} \tag{13}$$

5. RESULT

5.1 Training Loss

The losses of the proposed model are shown in Fig. 3, displaying the model training loss with different LR: 10^{-3}, 10^{-4}, 10^{-5}. The LR of 10^{-3} causes the loss to drop quickly at the start, reaching a low of about 0.0079, which indicates fast and efficient learning. The LR of 10^{-4} results in the lowest loss of around 0.0106, decreasing slightly slower. The LR of 10^{-5} causes the loss to decrease very slowly, leveling off around 0.0178, showing that an LR of 10^{-5} can make the model learn too slowly and adapt less effectively.

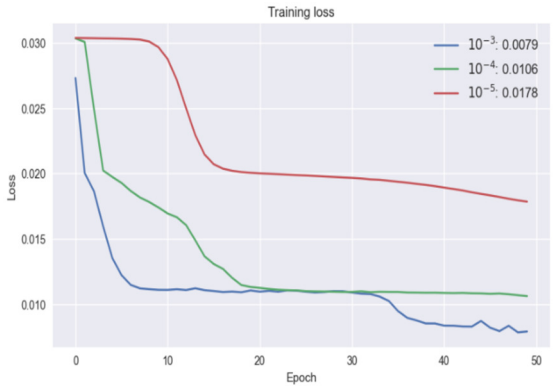


Fig.3: *The training loss results based on different LR.*

Fig. 4 shows the validation loss of the model trained with different LR, with each line representing the loss decrease in each epoch. The LR of 10^{-3} causes the loss to drop quickly, reaching as low as 0.0079, indicating that this rate allows the model to achieve the lowest loss. The LR of 10^{-4} results in a minimum loss of 0.0106, which is still effective but decreases more slowly. The LR of 10^{-5} causes the loss to drop very slowly, stabilizing around 0.0185, suggesting that a learning rate that is too high or too low may not yield optimal training results.

In Fig. 3 and 4, the learning curve of 10^{-5} shows slight fluctuations after the 10 epochs compared to the training loss in the second graph, which is smoother and decreases steadily. It may indicate slight overfitting when the LR of 10^{-3} is too high. In contrast, a more suitable LR of 10^{-4} provides more stable results in both graphs.

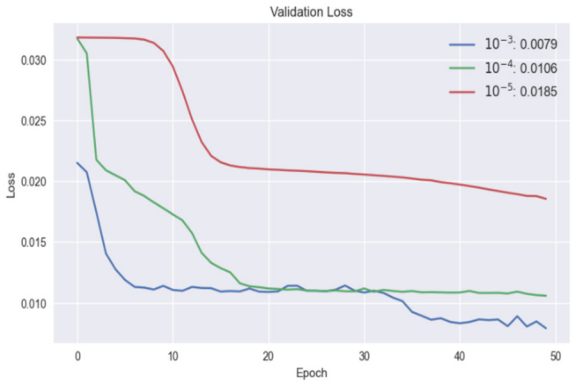


Fig.4: *The validation loss results based on different LR.*

Fig. 5 shows the model's training time in minutes for different LR. The results indicate that higher LR of 10^{-3} and 10^{-4} take about the same time, approximately 0.53 minutes, while the lower 10^{-5} takes slightly longer, around 0.59 minutes. This may be because the low LR requires the model to undergo extra training steps to improve, leading to a slight increase in total training time.

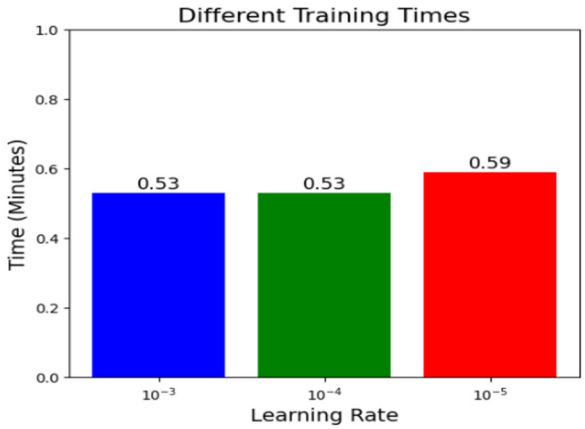


Fig.5: *The training time results based on different LR.*

5.2 Recognition Performance

In Fig. 6(a), the accuracy results of the model trained with an LR of 10^{-3} for skin cancer recognition. The model processes the input lesion image, identifies it, and produces the recognized image as output, with the recognition accuracy percentage shown below. The results show high accuracy (between 97.77% and 99.85%), demonstrating the model's effectiveness in accurately identifying skin cancer in the images. Although the LR is high, it enables the model to train and test quickly with high accuracy.

Fig. 6(b), the accuracy results of the model trained with an LR of 10^{-4} for skin cancer recognition. The

model outputs the predicted image and the recognition accuracy, ranging from 72.32% to 100%. This shows that using an LR of 10-4 causes some variation in recognition accuracy. While some images reach 100% accuracy, others fall below 80%, suggesting that this LR helps the model learn well but may still struggle with recognizing certain images.

Fig. 6(c) shows the test results of the model trained with a very low LR of 10^{-5} for skin cancer recognition. The model predicts the output image and shows the recognition accuracy, which ranges from 43.30% to 92.04%. This indicates that with a low LR of 10-5, the model has trouble learning effectively, leading to low accuracy, especially for some images where accuracy is below 50%. This may suggest that the LR is too low, preventing the model from learning enough important features and reducing its ability to recognize skin cancer accurately.

Comparing the results in Fig. 6 (a), (b), and (c), the test results with different LR show that an LR of 10^{-3} gives the best results, with high and stable accuracy between 97.77% to 99.85%, showing strong performance for skin cancer recognition. While an LR of 10^{-4} reaches high accuracy (up to 100%), it shows more fluctuation. The LR of 10-5 has the lowest accuracy, with some images below 50%, indicating issues with learning. Therefore, a LR of 10-3 is the most suitable choice for skin cancer recognition in this dataset.

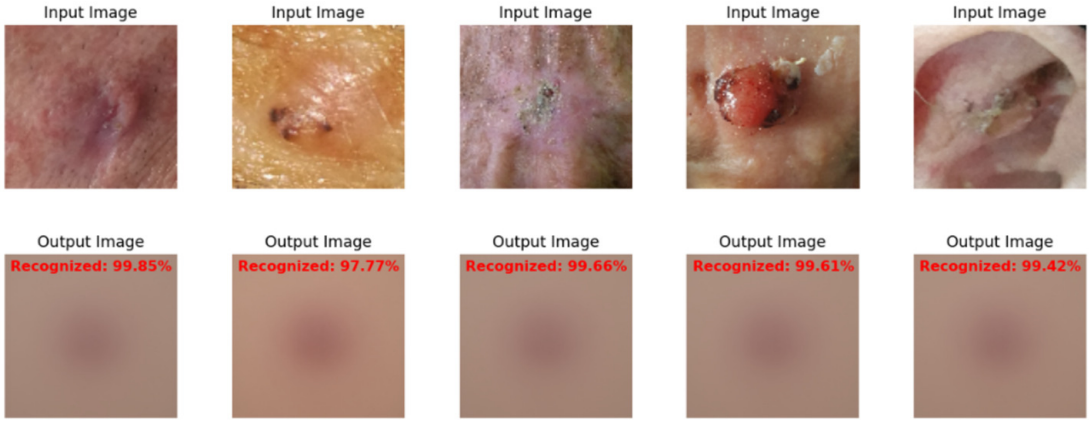
5.3 Comparison

We compare our experimental results with recent deep learning [38-41], including VGG19, ResNet50, DenseNet121, MobileNet, and the CNN model [42]. The authors modified these networks to function in an unsupervised setting for standard applications. Tuning and training were done on the same dataset, using the same training parameters and environment as the proposed method. Summarizes the training loss and training time of each model in Table 2.

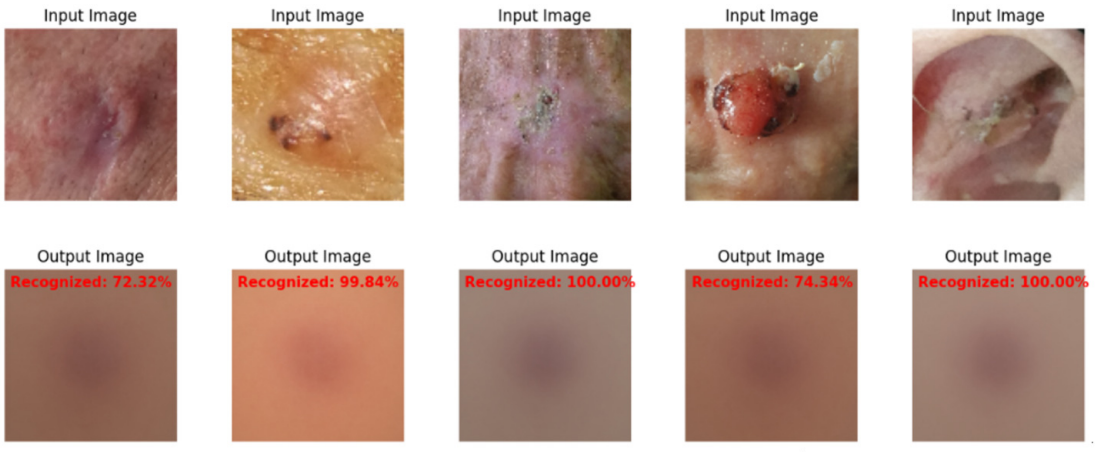
Table 2: *Comparision of loss and time.*

Year/Paper	Method	Loss	Time (min)
2020/ [38]	VGG19	0.0104	1.32
2020/ [39]	ResNet50	0.0162	1.33
2020/ [40]	DenseNet121	0.0036	1.36
2020/ [41]	MobileNet	0.0032	1.19
2020/ [42]	CNN	0.0043	0.49
Proposed	Two-Stream Network	0.0079	0.53

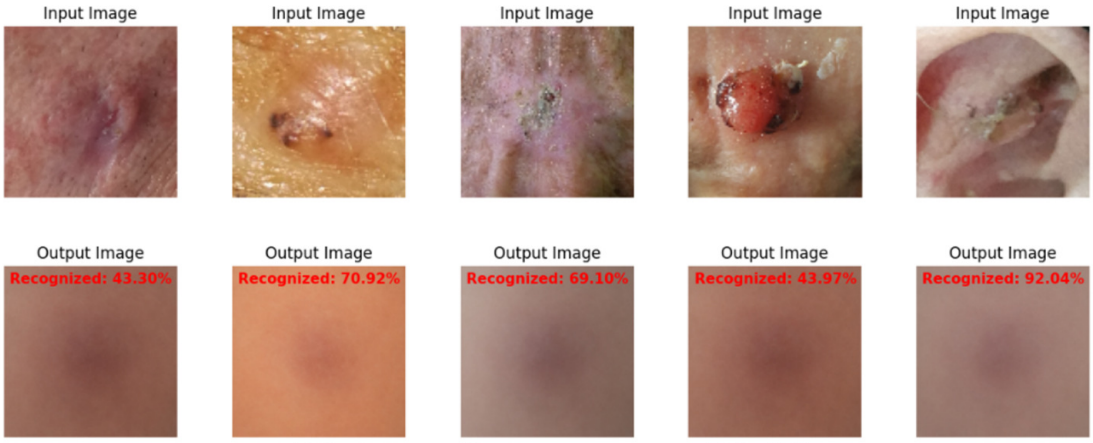
Table 2 compares the training loss and training time of different models. The proposed has a



(a) Performance of skin cancer recognition with LR of 10^{-3}



(b) Performance of skin cancer recognition with LR of 10^{-4}



(c) Performance of skin cancer recognition with LR of 10^{-5}

Fig.6: *Result of skin cancer recognition.*

loss of 0.0079. While this is slightly higher than DenseNet121 (0.0036), MobileNet (0.0032), and CNN (0.0043), it performs better than VGG19 (0.0104) and ResNet50 (0.0162). Regarding training time, the Two-stream network takes 0.53 minutes, which is faster than VGG19, ResNet50, and DenseNet121, each taking over 1 minute. Although MobileNet has the lowest losses, the two-stream network still offers a balance between accuracy and processing speed.

6. DISCUSSION

This study suggests an unsupervised DL model for skin cancer recognition. Testing different LR shows that a rate of 10^{-3} gives the most accurate and consistent results, with accuracy ranging from 97.77% to 99.85%, indicating strong learning and skin cancer recognition. While accurate in some cases, the LR of 10^{-4} has more variation. The lowest rate, 10^{-5}, has the lowest accuracy due to slower learning. In conclusion, the LR of 10^{-3} is the most suitable choice.

Analyzing the proposed Two-stream network design shows that using RL helps prevent the vanishing gradient problem, making DL training more stable. Combined with Bi-LSTM and AL, the models learn the context of deep features well, which boosts training efficiency and improves recognition results [43]. Additionally, using layers like global average pooling and max pooling enables the network to learn features from data of different sizes, as shown in [44]. However, this design is complex due to the mix of different layers, the need for high computational resources, and the fact that it is prone to overfitting. In this study, optimizing the training parameters helped address some of these issues and improve training results, which aligns with findings in [45, 46].

In Fig. 7, PCA shows the skin cancer samples from the PAD-UFES-20 dataset by turning their 26 features into just two main parts, PCA1 and PCA2. This helps show how each sample is placed and how close or different they are. The chart helps compare how similar or different the samples are. However, most points are grouped in the center, meaning many skin cancer samples share similar information. This makes it harder to tell them apart using only the given data.

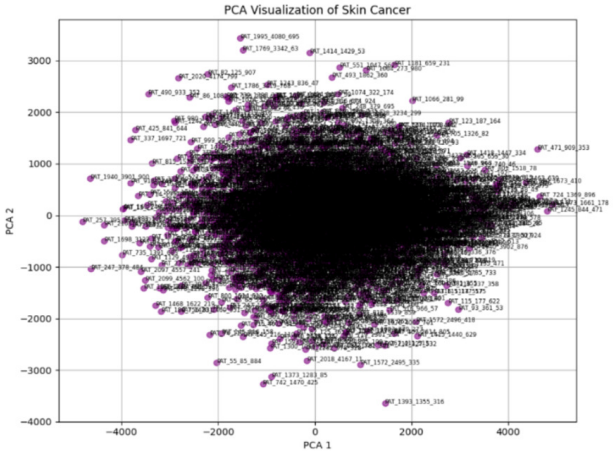


Fig.7: *Visualization of skin cancer features.*

In this study, the network was built in a way that helps reduce overfitting. It uses residual blocks, max pooling, and global average pooling to keep only the important parts and avoid unnecessary parameters. This helps the model focus better on features related to skin cancer. The input images are $100\times100\times3$ size, and the overall structure is simple. It includes convolutional layers, Bi-LSTM, and an attention mechanism, making the training faster and less demanding on hardware. One limitation is that the model is based on unsupervised learning. This makes it hard to compare results using typical disease-related labels directly. To improve this, switching to a self-supervised learning method in the future might help make the model more useful in real clinical settings.

7. CONCLUSION

We propose an unsupervised DL network for skin cancer recognition. The network is structured with a two-stream parallel feature extraction layer that increases data flow paths, allowing it to handle diverse data more effectively. It also includes a residual layer to prevent learning issues as the network gets deeper. Additionally, Bi-LSTM and AL are combined to learn features in both directions, reducing reliance on external data and improving the detection of complex features, leading to higher recognition accuracy.

We compared the performance of the proposed model with ML and CNN methods by measuring loss and computation time. The results show that our model has lower losses, indicating good accuracy, and uses less computation time than other networks. This balance between performance and time is precious for tasks needing high accuracy without adding extra time or computational demands.

In the future, researchers plan to improve the network to increase recognition efficiency and make it more flexible and comprehensive for mobile applications. This includes developing a system that can assist doctors in quickly detecting and diagnosing patients.

AUTHOR CONTRIBUTIONS

Conceptualization, A.S., A.P., W.O., and J.S.; methodology, A.S., and W.O; software, A.P and J.S.; validation, A.S., A.P., W.O., and J.S.; formal analysis, A.S.; investigation, A.S.; data curation, A.S.; writing—original draft preparation, A.S.; writing—review and editing, A.S., A.P., W.O., and J.S.; visualization, A.S. and J.S.; supervision, A.S., and J.S.; funding acquisition, A.P., and W.O. All authors have read and agreed to the published version of the manuscript.

References

[1] O. Y. Carrasquillo, J. Lambert and B. G. Merritt, "Comment on "Disparities in nonmelanoma skin cancer in Hispanic/Latino patients based on Mohs micrographic surgery defect size: A multicenter retrospective study," *Journal of the American Academy of Dermatology*, vol. 87, no. 4, pp. e129-e130, 2022.

[2] E. Rezk, M. Eltorki and W. El-Dakhakhni, "Human knowledge-based artificial intelligence methods for skin cancer management: Accuracy and interpretability study," *Smart Health*, vol. 36, Art. no. 100540, 2025.

[3] L. L. Locurcio, M. Breen and J. Haq, "Can a Single-Stage Approach Using a Dermal Regeneration Template Lead to Satisfactory Scalp Defect Reconstruction After Skin Cancer Excision?," *Journal of Oral and Maxillofacial Surger*, vol. 82, no. 3, pp. 341-346, 2024.

[4] A. Imran, A. Nasir, M. Bilal, G. Sun, A. Alzahrani, A. Almuhaimeed, "Skin Cancer Detection Using Combined Decision of Deep Learners," in *IEEE Access*, vol.10, pp. 118198 – 118212, 2022.

[5] W. K. Mridha, M. M. Uddin, J. Shin, S. Khadka and M. F. Mridha, "An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System," in *IEEE Access*, vol. 11, pp. 41003 – 41018, 2023.

[6] O. Akinrinade and C. Du, "Skin cancer detection using deep machine learning techniques," *Intelligence-Based Medicine*, vol. 11, Art. no. 100191, 2025.

[7] M. Subramanian, J. Cho, V. E. Sathishkumar and O. S. Naren, "Multiple Types of Cancer Classification Using CT/MRI Images Based on Learning Without Forgetting Powered Deep Learning Models," in *IEEE Access*, vol. 31, pp. 10336 - 10354, 2023.

[8] A. Strittmatter and F. G. Zöllner, "Multistep Networks for Deformable Multimodal Medical Image Registration," in *IEEE Access*, vol. 12, pp. 82676 - 82692, 2024.

[9] S. Pohtongkam and J. Srinonchat, "Object Recognition for Humanoid Robots Using Full Hand Tactile Sensor," in *IEEE Access*, vol. 11, pp. 20284-20297, 2023.

[10] T. Kaewrakmuk and J. Srinonchat, "Multi-Sensor Data Fusion and Time Series to Image Encoding for Hardness Recognition," *IEEE Sensors Journal*, vol. 24, Issue: 16, pp. 26463 – 26471, 2024.

[11] C. Chansri and J. Srinonchat, "Enhance Egocentric Grasp Recognition Based Flex Sensor Under Low Illumination," *Computers, Materials and Continua*, vol. 71, No.3, pp. 4377-4389, 2022, DOI:10.32604/cmc.2022.024026

[12] J. Chaki and A. Uçar, "An Efficient and Robust Approach Using Inductive Transfer-Based Ensemble Deep Neural Networks for Kidney Stone Detection," IEEE Access, pp. 32894 - 32910, 2024.

[13] S. M. M. Elkholy, A. Rezk and A. A. E. F. Saleh, "Early Prediction of Chronic Kidney Disease Using Deep Belief Network," in *IEEE Access*, vol. 9, pp. 135542 – 135549, 2021.

[14] S. D. Pande and R. Agarwal, "Multi-Class Kidney Abnormalities Detecting Novel System Through Computed Tomography," in *IEEE Access*, vol. 12, pp. 21147 – 21155, 2024.

[15] F. L. Tiro, V. Estrade, J. Hubert, D. F. Araiza, M. G. Mendoza and G. Ochoa, "On the In Vivo Recognition of Kidney Stones Using Machine Learning," in *IEEE Access*, vol. 12, pp. 10736 – 10759, 2024.

[16] R. Magherini, M. Servi, Y. Volpe, R. Campi

and F. Buonamici, "Distinguishing Kidney Tumor Types Using Radiomics Features and Deep Features," in *IEEE Access*, vol. 12, pp. 84241 – 84252, 2024.

[17] R. Gaikar, A. Azad, N. Schieda and E. Ukwatta, "Fully Automated Deep Learning-Based Renal Mass Detection on Multi-Parametric MRI," in *IEEE Access*, vol. 12, pp. 112714 – 112728, 2024.

[18] A. Prommakhot and J. Srinonchat, "VGGNet Integration for Kidney Tumor Classification," *International Electrical Engineering Congress (iEECON)*, Pattaya, Thailand, Mar. 06-08, 2024.

[19] M. I. Faizi and S. M. Adnan, "Improved Segmentation Model for Melanoma Lesion Detection Using Normalized Cross-Correlation-Based k-Means Clustering," in *IEEE Access*, vol. 12, pp. 20753 – 20766, 2024.

[20] Z. Lan, S. Cai, X. He and X. Wen, "FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer," *IEEE Access*, vol. 10, pp. 76261 – 76267, 2022.

[21] L. Wei, K. Ding and H. Hu, "Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network," in *IEEE Access*, vol. 8, pp. 99633 – 99647, 2020.

[22] N. Q. Nguyen and S. W. Lee, "Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-Decoder Network," in *IEEE Access*, vol. 7, pp. 33795 – 33808, 2019.

[23] P. A. Lyakhov, U. A. Lyakhova and D. I. Kalita, "Multimodal Analysis of Unbalanced Dermatological Data for Skin Cancer Recognition," in *IEEE Access*, vol. 11, pp. 131487 – 131507, 2023.

[24] S. S. Zareen, G. Sun, M. Kundi, S. F. Qadri and S. Qadri, "Enhancing Skin Cancer Diagnosis with Deep Learning: A Hybrid CNN-RNN Approach," *Computers, Materials and Continua*, vol. 79, no. 1, pp. 1497-1519, 2024.

[25] C. Kavitha, S. Priyanka, M. Praveen Kumar and V. Kusuma, "Skin Cancer Detection and Classification using Deep Learning Techniques," *Procedia Computer Science*, vol. 235, pp. 2793-2802, 2024.

[26] J. S. T. Purni and R. Vedhapriyavadhana, "EOSA-Net: A deep learning framework for enhanced multi-class skin cancer classification using optimized convolutional neural networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 3, Art. No. 102007, 2024.

[27] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, Art. No. Art. No. 100036, 2021.

[28] E. S. Rahayu, E. M. Yuniarno, I. K. E. Purnama and M. H. Purnomo, "A combination model of shifting joint angle changes with 3D-deep convolutional neural network to recognize human activity," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 1078–1089, 2024.

[29] A. Prommakhot and J. Srinonchat, "Scaled Dilation of DropBlock Optimization in Convolutional Neural Network for Fungus Classification," *Computers, Materials and Continua*, vol.72, No.2, pp. 3313-3329, 2022.

[30] S. Yaemprayoon and J. Srinonchat, "Exploring CNN Model with Inrush Current Pattern for Non-Intrusive Load Monitoring," *Computers, Materials and Continua*, vol.73, No.2, pp. 3667-3684, 2022.

[31] A. Prommakhot and J. Srinonchat, "Combining Convolutional Neural Networks for Fungi Classification," in *IEEE Access*, Vol. 12, pp. 58021-58030, 2024.

[32] C. Chansri and J. Srinonchat, "Utilizing Gramian Angular Fields and Convolution Neural Networks in Flex Sensors Glove for Human–Computer Interaction," *IEEE Transactions on Human-Machine Systems*, Vol. 54, Issue: 4, pp. 475 – 483, 2024.

[33] G. Xu, X. Wang, X. Wu, X. Leng and Y. Xu, "Development of residual learning in deep neural networks for computer vision: A survey," *Engineering Applications of Artificial Intelligence*, vol. 142, Art. no. 109890, 2025.

[34] J. Zhu, M. A. A. Al-qaness, D. AL-Alimi, H. Z. Tao and S. H. Alsamhi, "Multiple lung diseases detection using advanced deep learning model with attention mechanisms and upsampling features," *Engineering Applications of Artificial Intelligence*, vol. 156, Art. no. 111038, 2025.

[35] D. Haritha, N. Satyavani and A. Ramesh, "Generation of missing well log data with deep learning: CNN-Bi-LSTM approach," *Journal of Applied Geophysics*, vol. 233, Art. no. 105628, 2025.

[36] H. Ye, B. Cao, Z. Peng, T. Chen, Y. Wen and J. Liu, "Web Services Classification Based on Wide & Bi-LSTM Model," in *IEEE Access*, vol. 7, pp. 43697-43706, 2019.

[37] A. G. Pacheco *et al.*, "PAD-UFES-20: A skin lesion dataset composed of patient data and clinical images collected from smartphones," *Data in Brief*, vol. 32, Art. No. 106221, 2020. [online] `https://www.kaggle.com/datasets/mahdavi1202/skin-cancer/data`.

[38] A. Barbadekar, V. Ashtekar and A. Chaudhari, "Skin Cancer Classification and Detection Using VGG-19 and DesNet," *International Conference on Computational Intelligence, Networks and Security (ICCINS)*, Mylavaram, India, 2023.

[39] E. Firasari, N. Khasanah, F. L. D. Cahyanti,

D. N. Kholifah, U. Khultsum and F. Sarasati, "Performance Evaluation of ResNet50 and MobileNetV2 in Skin Cancer Image Classification with Various Optimizers," *International Conference on Information Technology Research and Innovation (ICITRI)*, Jakarta, Indonesia, 2024.

[40] S. S. Reka, H. L. Karthikeyan, A. J. Shakil, P. Venugopal and M. Muniraj, "Exploring Quantum Machine Learning for Enhanced Skin Lesion Classification: A Comparative Study of Implementation Methods," in *IEEE Access*, vol. 12, pp. 104568 – 104584, 2024.

[41] V. D. Hoang, X. T. Vo and K. H. Jo, "Categorical Weighting Domination for Imbalanced Classification With Skin Cancer in Intelligent Healthcare Systems," in *IEEE Access*, vol. 12, pp. 105170 – 105181, 2023.

[42] M. A. Rasel, U. H. Obaidellah and S. A. Kareem, "Convolutional Neural Network-Based Skin Lesion Classification With Variable Nonlinear Activation Functions," in *IEEE Access*, vol. 10, pp. 83398-83414, 2022.

[43] Q. -J. Lv, H. Y. Chen, W. B. Zhong, Y. Y. Wang, J. Y. Song and S. D. Guo, "A Multi-Task Group Bi-LSTM Networks Application on Electrocardiogram Classification," in *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 8, Art. No. 1900111, 2019.

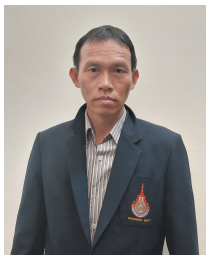
[44] H. Guo, H. Bai and W. Qin, "ClouDet: A Dilated Separable CNN-Based Cloud Detection Framework for Remote Sensing Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9743 – 9755, 2021.

[45] C. Aliferis and G. Simon, "Overfitting, Underfitting and General Model Overconfidence and Under-Performance: Pitfalls and Best Practices in Machine Learning and AI," *Artificial Intelligence and Machine Learning in Health Care and Medical Research*, Springer, vol. 10, pp. 123-145, 2023.

[46] J. Yeomans, S. Thwaites, W. S. P. Robertson, D. Booth, B. Ng and D. Thewlis, "Simulating Time-Series Data for Improved Deep Neural Network Performance," in *IEEE Access*, vol. 7, pp. 131248-131255, 2019

Aekkarat Suksukont received a B.Eng. degree in Electronics and Telecommunication Engineering from Rajamangala University of Technology Thanyaburi (RMUTT), in 2011, and the M.Eng. degrees in Electronics and Telecommunication Engineering from Rajamangala University of Technology Thanyaburi, in 2017. His research interests include speech and image processing, computer vision, and applied AI. He can be contacted at email: aekkarat.s@rmutsb.ac.th.

Anuruk Prommakhot received a B.Eng. degree in Computer Engineering from Rajamangala University of Technology Isan, Sakonnakhon Campus, in 2015, and an M.Eng. degree in Electronic and Communication Engineering from Rajamangala University of Technology Thanyaburi (RMUTT), in 2018. He completed his Ph.D. degree in Electrical Engineering from RMUTT in 2024. His research interests include deep learning, optimization techniques in deep learning, computer vision, AI Vision, Robot Vision, and applied AI for medical imaging. He can be contacted at anuruk.p@en.rmutt.ac.th.



Wichian Ooppakaew earned his Bachelor of Engineering in Electrical Engineering from Rajamangala University of Technology Thanyaburi (RMUTT), Thailand, in 1999. He obtained a Master of Engineering in Electrical Engineering from Chulalongkorn University, Thailand. He obtained a PhD from the University of Northumbria, Newcastle, United Kingdom, in 2012. He is currently employed as a lecturer and researcher at RMUTT. His research focuses on image processing and the use of artificial intelligence

Jakkree Srinonchat received a B.Eng. degree in electrical engineering from Rajamangala University of Technology Thanyaburi (RMUTT), in 1995 and a Ph.D. degree in electrical engineering from the University of Northumbria at Newcastle, UK, in 2005. Since then, he has been a researcher and lecturer Signal Processing Research and Laboratory at RMUTT. His research interests include speech and image processing, robot controller, machine learning and applied AI for signal processing. He can be contacted at email: jakkree.s@en.rmutt.ac.th