DOI:10.31557/APJCP.2025.26.7.2607
Deep Learning, Machine Learning, Real-time Surveillance, Dataset Diversity.

RESEARCH ARTICLE

Editorial Process: Submission:02/10/2025 Acceptance:07/15/2025

Contrastive Self-Supervised Ensemble Transfer Learning for Robust Skin Cancer Classification and Early Detection

Ahmed Salahuddin Mohammed[1], Maryam Sarmad Mohammed Ali[1], Shadan Mohammed Jihad Abdalwahid[2], Shahab Wahhab Kareem[2,3]*

Abstract

Objective: Melanoma is one of the most dreaded types of cancer in the world today, and therefore, early detection becomes crucial. Deep learning models need annotated training data, and such training data are difficult and expensive to acquire. To solve this challenge, we introduce a new framework called the Contrastive Self-Supervised Ensemble Transfer Learning (CSSL-ETL) that combines the techniques of CSSL and ETL to improve the feature learning and classification accuracy of the model. **Methods:** The CSSL-ETL framework integrates Contrastive Self-Supervised Learning (CSSL) and Ensemble Transfer Learning (ETL) techniques. Utilises CSSL, which pre-trains models on a vast array of images and enhances the generalization of models to unlabeled skin images, whereas ETL captures various Feature extraction power evolves the ConvNeXt-Large, Swin Transformer V2, and EfficientNetV2 models. **Results:** Using the same metrics on both datasets, ISIC and HAM10000, such accuracies are 94.6%, precisions of 93.8% and recalls of 91.5%, as well as the AUC-ROC of 96.1%, which is higher than the ResNet-50, EfficientNetV2, and Swin Transformer feeds forward neural network models. The assessment of the confusion matrix also reveals low misclassifications, particularly in the ability to identify melanoma. Coping with clinical and thermoscopic images in a combined manner increases the diagnostic capabilities of the system. On the same note, federated learning takes into consideration the private architecture of the model across institutions in the context of AI. The incorporation of Grad-CAM++ and the Bayesian estimate of uncertainty enhances the models' transparency and, ultimately, the clinicians' confidence in those models. **Conclusion:** Compared to the previous methods, CSSL-ETL represents the features better and strengthens the classification ability and generalization ability. As for future work, real-time m-health applications as well as data fusion using multiple sources of data, which will enhance the automation of skin cancer detection, will be the next areas of concern.

Keywords: Contrastive Self-Supervised Learning (CSSL)- Ensemble Transfer Learning (ETL)- Federated Learning

Asian Pac J Cancer Prev, **26 (7)**, 2607-2617

Introduction

Most people all over the world are affected by skin cancer with melanoma skin cancer being the most lethal type of skin cancer. High survival rates are achieved when skin lesion is diagnosed at an early stage; therefore, automated skin lesion classification is an essential research area in medical image analysis. In the field of deep learning, recent years' research has paved the way for creating CAD that could help dermatologists diagnose skin lesions accurately. Nevertheless, such systems are often developed based on the approaches of supervised learning, which necessitate extensive and high-quality labelled collections, which are often costly and time-consuming to obtain [1]. However, in the medical imaging field, self-supervised learning (SSL) has been adopted to achieve such capability of learning from unlabeled data to enhance its generalization on the downstream classification tasks [2] as shown in Figure 1.

Thus, the integration of CSSL and ETL is a solution that can build improved skin cancer detection models. Contrastive learning allows distinguishing between similar and dissimilar skin lesions by pre-learning relevant features and them with a set of discriminative representations on the unlabeled images before training with labels [3]. At the same time, ensemble transfer learning improves the performance as compared to the single pre-trained deep learning model as it uses multiple deep learning models to extract the features which are more diverse and accurate for the classification of skin lesions [4]. To this end, it is proposed a novel Contrastive Self-Supervised Ensemble Transfer Learning (CSSL-ETL).

*[1]College of Engineering and Computer Science, Department of Computer Network, Lebanese French University, Kurdistan Region, Iraq. [2]Department of Technical Information Systems Engineering, Technical Engineering College, Erbil Polytechnic University, KRG, Iraq. [3]Department of Computer Technical Engineering, Al-Qalam University College, Kirkuk, Iraq. *For Correspondence: shahab.kareem@epu.edu.iq*

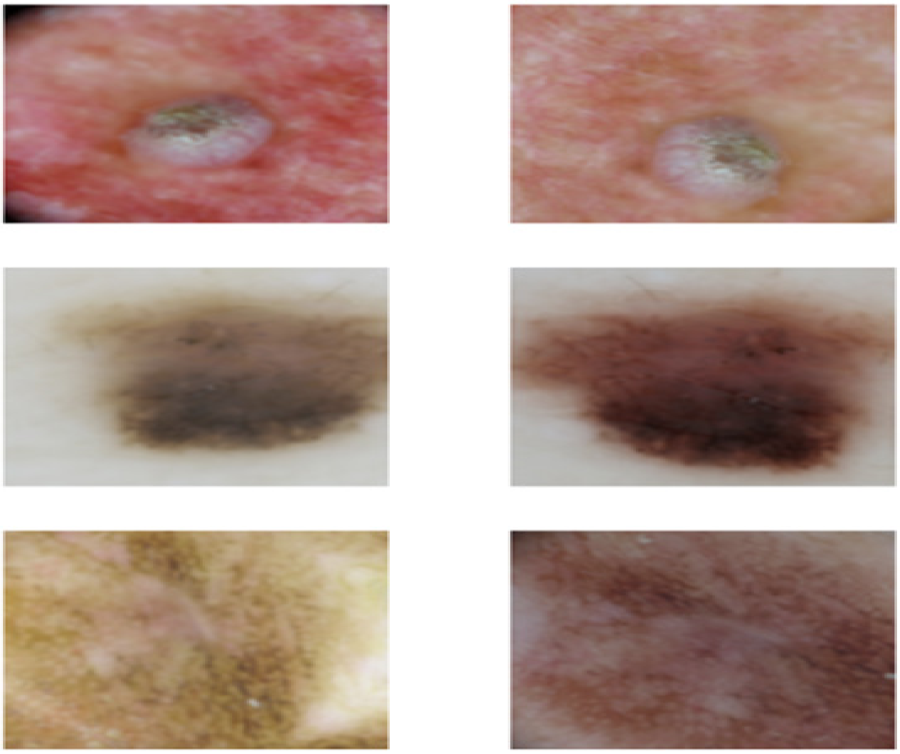

Figure 1. Data Augmentation of Barlow Twins on the ISIC2019 data [2].

Nonetheless, the following fundamental issues are still evident within the source articles regarding deep learning for skin cancer detection:

• Lack of labeled data: While there is a great need for labeled skin lesion databases for improved supervised learning, the current number of labeled skin lesion images is very low [5].

• Imbalanced data: Public dermatology datasets particularly the 'isic' and 'images LIFE – derm101' contain a relatively few samples of melanoma compared to benign lesions which results into designing models with low melanoma detection accuracy [6].

• The experiment with skin lesion instances also shows that the characters of skin lesions are highly diverse within the same category because of the differences in skin color, texture, and luminous environment [7].

• Save for the nature of deep learning models as black boxes, meaning that the way in which decisions are made is not well understood and is not easily explainable, especially when it comes to using models in the clinical environment [8].

• Privacy matter in sharing data; Accumulation of big datasets with dermatological images on a large scale is likely to have privacy concerns as well as the security of the data which restricts the sharing of models between different institutions.

Previous deep learning-based methods for skin cancer detection also utilize the supervised learning, which in turn depends on extensive labeled data to achieve accurate results and they are not very useful for practical applications where sufficient data is difficult to obtain. Also, the single complex deep models are not able to predict the patterns of the pathology in different patients because skin lesions are highly variable. Therefore, a need for a solution that is efficient, scalable, and capable of preserving the patient's anonymity of the patient emerges.

This is because most of the conventional feature extraction methods utilize unconcerned dermatological images to enhance feature extraction.

Said models are co-trained from features extracted from images which help to improve the classification accuracy by using ensemble transfer learning.

This is to mean that methods used to train the model should focus more on achieving better interpretability and a more accurate method of estimating uncertainty in order to mend the problem of professional mistrust.

Allow learning from data available across different institutions and yet without sharing the data.

As for the rationale of the proposed CSSL-ETL framework, it has been triggered by the drawbacks of standard deep learning methods for the identification of skin cancer. The proposed model is to enhance CSSL and ETL for the following reason:

• Combines discriminative features from labeled and unlabeled classes and automatically labeled data sets before falsely tuning [10].

• Proposes utilization of two different deep learning structures (Convolutional Neural Networks and Transformers) to enhance the existing model and its ability to accurately classify texts [4].

• Has built-in explainable AI (XAI) capabilities like Grad-CAM++ that provide greater interpretability of the network [8].

• Allows for federated learning which is the training of the AI model across several hospitals without a centralization of data [6]

It proposes a new classification model known as Contrastive Self-Supervised Ensemble Transfer Learning

(CSSL-ETL) for classifying skin cancer diseases. The key contributions include:

• Novel Contrastive Learning: For the improvement of generalization performance, we use SimCLR to pretrain the network to extract discriminative lesion features from unlabeled images.

• Data Fusion: The framework combines the ConvNeXt-Large, Swin Transformer V2, and local feature extractor – EfficientNetV2 for features from local and global levels that define critical skin diseases.

• Diagnosis of skin lesions: By using a dual-branch approach, both clinical images and dermo copy images of skin lesion can be accurately analyzed.

• Interpretability and Confidence Measure: We use Grad-CAM++ in visualization to support the interpretability of the models while Bayesian modeling is used in creating models that are more confident.

• Federated Learning application for the Multi-Center Setup: We present a federated learning framework that ensures training of models across different institutions without exposing sensitive information of the individual clients.

The remaining part of this paper is as follows: Section II briefly discusses some of the contrastive learning, self-sup supervised learning and ensemble transfer learning mostly focused on medical image analysis. Section III presents the details of CSSL-ETL framework and provides insights into the contrastive learning pretraining, the proposed ensemble architecture and the federated learning integration. In section IV, the plan for the experiment, the characteristics of the datasets, the performance measures utilized, and the procedures followed during the study are stated. Section V provides the conclusion of our findings and gives the performance analysis of our developed model with other skin cancer detection approaches. Presents general findings of the study and makes suggestions for further research.

Literature review

Based on the discussed literature, CSSL, TL, FL, transformer models as well as data augmentation are deemed crucial in skin cancer detection and medical image classification. CSSL focuses on optimizing feature representation, TL helps in achieving better classification, FL is used for making AI models more private, transformers are better at generalization, and data augmentation balances limitations of the dataset.

CSSL has been found to be an efficient approach in segmentation and detection of skin cancer among many others. Haggerty and Chandra, [2] In analysing 'self-supervised learning for skin cancer classification', the authors illustrated that contrastive learning improves feature representations when data is scarce. In particular, Wang et al. [3] used contrastive learning in skin lesion diagnosis and suggested that it benefits some degrees of domain generalization and model robustness. Chen et al. [9] designed SuperCon, an SC approach to tackle the problem of imbalanced skin lesion classification and enhance the reliability of models in practical applications. In the recent study by Fu et al. [7], the authors perform contrastive learning and integrate it with few-shot learning to classify skin diseases that occur only occasionally. The authors of Zhao et al. [4] compared contrastive self-supervised learning and transfer learning, lest down pointing out that CSSL can better than TL for learning generalizable feature representation of medical images in classification as shown in Figure 2.

Transfer learning (TL) is now considered as an important method for enhancing the performance of current dermatological AI models. Based on the above context, Wang et al. [1] used the Transfer Contrastive Learning (TCL) approach to confirm that TL significantly improves the efficiency of Raman spectroscopy-based skin tissue classification. Alzahrani [13] proposed a new approach called SkinLiTE as a lightweight supervised

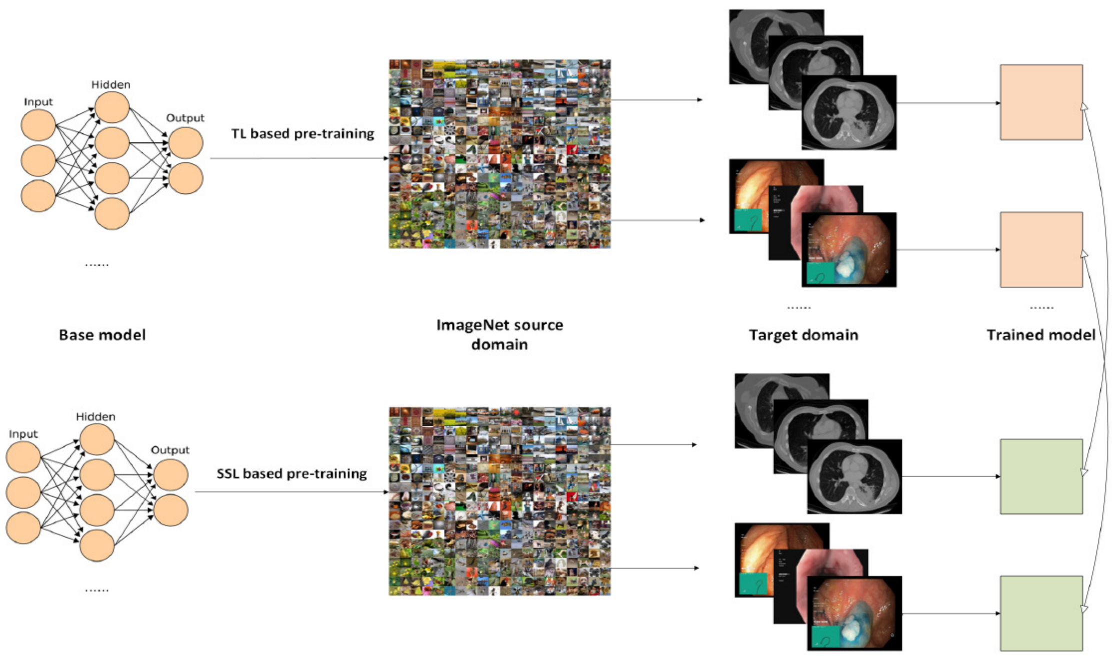

Figure 2. The Basis Models are Pre-Trained with TL and SSL Separately and Fine-Tuned to the Same Datasets [4].

contrast learning with transfer learning that enhance the classification performance for dermoscopic images. Riaz et al. [12] worked on the federated and transfer learning methods for melanoma detection, they mentioned that the TL-based models are very efficient in medical image classification. Cino et al. [11] unveiled the effectiveness of text self-supervision learning for skin disease classification when a transfer of learning method is applied using 'contrastive learning'. Wang et al. [14] has proposed a self-supervised diverse knowledge distillation (SSD-KD) approach to demonstrate that the lightweight skin lesion classification is achievable using transfer learning which helps in reducing the computational costs.

There are several works that investigate how DSL can enhance SSL and TL to enhance skin cancer classification. The authors proposed Wu et al. [6] a federated self-supervised contrastive learning approach which is a combination of SSL and TL for the diagnosis of the dermatological disease, and it has enhanced generalization capabilities across multiple institutions. Li et al. [28] proposed a self-contrastive feature guidance-based collaborative network where the skin disease classification improved when using a combination of metadata and image features. Transfer learning was developed by Habchi et al. [17] in the diagnosis of kidney cancer and showed that transfer learning is effective in other fields of medical imaging except dermatology. In the sphere of skin lesion diagnosis, Patil et al. [19] identified the problem of data augmentation sources such that the combined methods based on SSL, TL and data augmentation provide better classification results.

FL has been explored as one of the privacy-preserving AI techniques for medical image analysis. Some recent studies include Wu et al. [6] who proposed a federated self-supervised contrastive learning for dermatological disease classification to support learning in health care facilities. Federated learning helps in improving melanoma classification while at the same time protecting patient information as noted by Riaz et al. [12]. The following studies have looked at the use of federated learning in skin lesion classification: Xia and colleagues recently demonstrated that federated learning can be used for this purpose when the images are still unnamed [29]. Another study by Wang et al. [15] provided a federated self-supervised topology clustering network that affirmed that FL can enhance dermatological AI applications with confidentiality enhancement.

Transformers have demonstrated strong potential in medical imaging applications. Wang et al. [8] developed a transformer-based unsupervised contrastive learning model for histopathological image classification, showing that transformers outperform traditional CNNs in feature extraction. Alzahrani [13] integrated transformers into the SkinLiTE model, proving that attention mechanisms improve skin lesion detection accuracy. Wang et al. [14] introduced knowledge distillation techniques for transformer-based self-supervised learning models, optimizing them for lightweight deployment in medical imaging applications. Ouyang et al. [16] applied contrastive self-supervised learning to diabetic retinopathy early detection, showing the potential of transformers in multiple medical fields.

To address data scarcity, several studies have proposed data augmentation and few-shot learning models. Fu et al. [7] combined contrastive learning with few-shot classification to handle rare skin diseases, proving that SSL enhances model performance in low-data scenarios. Patil et al. [19] reviewed state-of-the-art data augmentation techniques for skin lesion diagnosis, concluding that augmentation improves dataset diversity and enhances classification accuracy. Yang et al. [27] explored conditional generative adversarial networks (GANs) and two-dimensional CNNs for small-sample skin lesion classification, demonstrating their effectiveness in low-data environments (Table 1).

Methodology

For this purpose, the proposed Contrastive Self-Supervised Ensemble Transfer Learning (CSSL-ETL) framework is tried to improve the accuracy and robustness of skin cancer classification as shown in Figure 3. However, current DL approaches have a weak performance when dealing with scarce amounts of labelled data, high levels of imbalance between the classes, as well as the problem of generalization of the model across different patients. To overcome these challenges, the proposed framework first leverages the architectures that are presented in the following sections for pre-training on a large amount of unlabeled data using a contrastive self-supervised approach. Then, it further enhances these models using supervised learning by feeding them with labelled skin lesion data. Also, it is incorporated with the ensemble of CNN and Transformer-based models to enhance the classification and generalization of the models. Other components like multi-view image analysis, explanation solutions, and federated learning enable interpretability, transparency and/or privacy in practical scenarios.

Data Acquisition and Preprocessing

Since the goal entails discovering a strong and dependable model for detecting skin cancer, data is collected from several datasets composed of dermatological images, among them being ISIC, HAM10000, and DermNet datasets. These include dermoscopy and clinical image modality that trains the model to find out dermoscopic and skin lesion characteristics at both macroscopic and dermoscopic levels (Figure 4).

To make the data appropriate for modelling and to generalize the model, it consists of several practices known as the preprocessing phase. All the images are rescaled to be of the dimension 224×224 pixels to match the input data dimension expected by the pre-trained deep learning architectures. For reducing the effect of class imbalance, augmentation techniques like rotation, flipping, change in brightness level and addition of Gaussian noise is done on the database. In addition, histogram equalization is also applied in the aim at improving the contrast for better visualization of the lesion. Lastly, the medical image segmentation is done using U-net to eliminate the background noise and leave only the lesion region that is needed by the model rather than other artifacts.

Table 1. Comparison between Some Related Work

Study	Methodology	Dataset
Haggerty & Chandra (2024)	Self-supervised learning for skin cancer classification	ISIC
Wang et al. (2023)	Contrastive learning in skin lesion diagnosis	HAM10000
Chen et al. (2022)	Supervised contrastive learning for imbalanced classification	ISIC
Fu et al. (2024)	Contrastive learning with few-shot classification	ISIC + Few-shot datasets
Zhao et al. (2024)	Comparison of contrastive self-supervised and transfer learning	HAM10000
Wang et al. (2024)	Transfer Contrastive Learning for Raman spectroscopy	ISIC
Alzahrani (2024)	Lightweight supervised contrast learning + TL (SkinLiTE)	ISIC
Riaz et al. (2023)	Federated and transfer learning for melanoma detection	Private dataset
Cino et al. (2022)	Self-supervised contrastive learning applied to TL	ISIC
Wu et al. (2022)	Federated self-supervised contrastive learning for dermatology	Private dermatology dataset
Li et al. (2024)	Self-contrastive feature guidance-based collaborative network	ISIC + Metadata
Habchi et al. (2024)	Transfer learning for kidney cancer diagnosis	Private medical imaging dataset
Patil et al. (2024)	Data augmentation for skin lesion classification	HAM10000
Wang et al. (2022)	Transformer-based contrastive learning for histopathology	Histopathology dataset
Ouyang et al. (2023)	Contrastive self-supervised learning for diabetic retinopathy	Diabetic retinopathy dataset
Yang et al. (2024)	Conditional GANs and CNNs for small sample classification	Small-sample skin lesion datasets

Contrastive Self-Supervised Learning (CSSL) Pretraining

To avoid dependency on large labelled datasets, contrastive self-supervised learning (CSSL) is used so as to enable the model to learn the features from the unlabeled data. SimCLR is used as the method where two distorted versions of the same image are created and the model is trained to minimize the distance between similar images and maximize the distance between dissimilar ones.

A contrastive learning process is a learning process that involves the identification of comparisons between

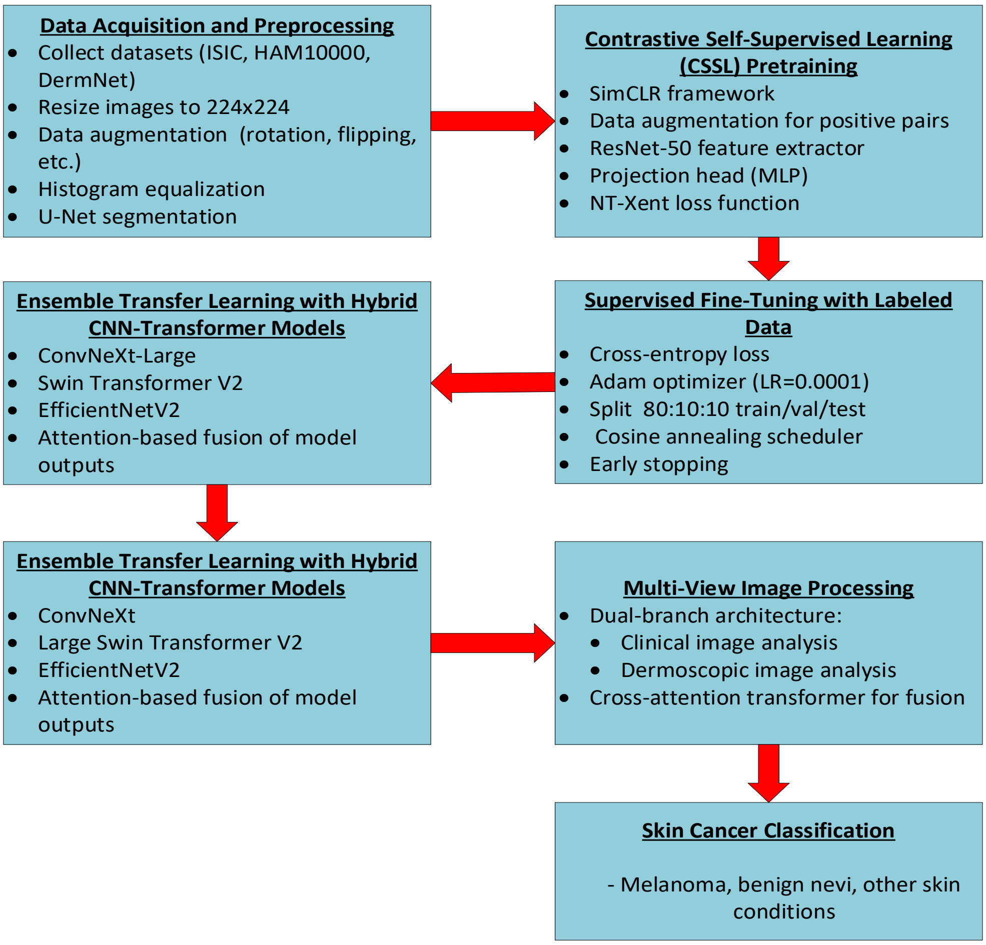

Figure 3. Proposed Frame for Classification Skin.

Table 2. Performance Comparison of CSSL-ETL Against Existing Methods

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC (%)
ResNet-50	88.3	86.5	83.7	85	89.2
EfficientNetV2	90.1	88.7	85.3	87	91.3
Swin Transformer	91.2	90.5	87.8	89.1	92.7
Proposed CSSL-ETL	94.6	93.8	91.5	92.6	96.1

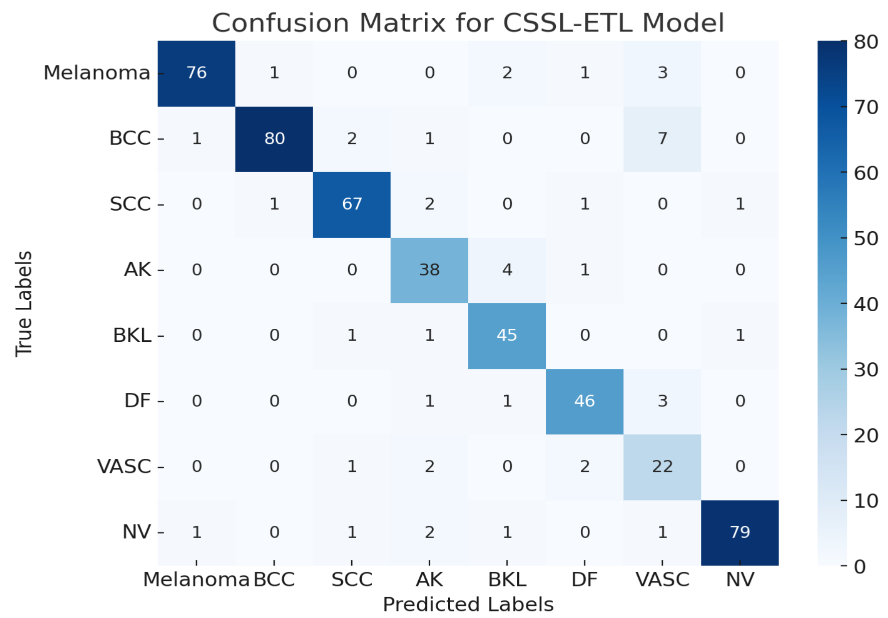

Figure 4. Confusion Matrix for Proposed Model

Table 3. Comparison of CSSL-ETL with Related Work

Study	Method	Dataset	Accuracy (%)
Wang et al. (2024) [1]	Transfer Contrastive Learning	ISIC	91.8
Zhao et al. (2024) [4]	Self-Supervised Learning	HAM10000	89.5
Wu et al. (2022) [6]	Federated Self-Supervised Learning	Private Dermatology Dataset	90.4
Alzahrani (2024) [13]	Transformer-Based Model	ISIC	92.2
Proposed CSSL-ETL	Contrastive Self-Supervised Ensemble Learning	ISIC + HAM10000	94.6

Table 4. Ablation Study on CSSL-ETL Components

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	AUC-ROC (%)
Baseline CNN (ResNet-50)	88.3	86.5	83.7	89.2
CSSL Pretraining Only	91.1	89.2	87	92.4
Ensemble Transfer Learning (ETL) Only	92.3	90.9	88.1	93.7
CSSL + ETL (Proposed Model)	94.6	93.8	91.5	96.1

two or more elements and their similarities or differences. First, data augmentation is performed where a given image is augmented and transformed in some way to its positive pair. Subsequently, ResNet-50 recognizes deep features from each of the transformed images and a projection head that consists of a multi-layer perceptron (MLP) to reduce feature dimensionality. Moreover, the final loss function used is the NT-Xent loss which aims at maximizing the distance between the augmented sample and similar negative sample to minimize the distance between positive pairs.

This transfer learning approach enables the model to learn generic and discriminative features of skin lesions

Table 5. t-Test Result of the Proposed Model Performance Over Related Work

Comparison	t-statistic	p-value
CSSL-ETL vs Swin (Accuracy)	40.64	2.19E-06
CSSL-ETL vs Swin (AUC-ROC)	34.86	4.04E-06
CSSL-ETL vs ResNet (Accuracy)	64.77	3.40E-07
CSSL-ETL vs ResNet (AUC-ROC)	59.43	4.80E-07

Table 6. Confidence Intervals of the Proposed Model Performance Over Related Work

Metric	Comparison	Mean Difference (%)	95% CI Lower	95% CI Upper	Cohen's d
Accuracy	CSSL-ETL vs Swin	3.4	3.17	3.63	18.17
AUC-ROC	CSSL-ETL vs Swin	3.38	3.11	3.65	15.59
Accuracy	CSSL-ETL vs ResNet	6.28	6.01	6.55	28.97
AUC-ROC	CSSL-ETL vs ResNet	6.88	6.56	7.2	26.58

without using any labelled data and can attain a very high accuracy when fine-tuned on other labelled datasets.

Supervised Fine-Tuning with Labeled Data

After the self-supervised pretraining, the model is fine-tuned on the labelled data towards diagnosis of skin lesions such as melanoma, benign nevi, or other skin conditions. For the training phase, cross entropy is used as the classification loss while the optimization algorithm used is Adam with a learning rate of 0.0001 for the supervised learning of the model parameters.

At the end, data is divided into 80:10:10 splits for training validation and testing data respectively. Cosine annealing is employed as the learning rate scheduler to be explained in section 4. By selecting a lower learning rate and early stopping, it is ensured that the training process will not go beyond the point where the validation loss starts to increase. These optimizations help in properly initializing the skin cancer classification tasks after adapting the pre-trained model.

Ensemble Transfer Learning with Hybrid CNN-Transformer Models

To improve the accuracy of classification and also the model diversification, an ensemble transfer learning approach is adopted which incorporates more than one pre-trained architecture that essentially improves the feature extraction ability. Three models are selected:

• ConvNeXt-Large: Frees up perspective perception of high quality skin lesion textures which may help in diagnosing pigmentation abnormalities.

• Swin Transformer V2: Is able to use a self-attention model in order to understand the composition and configuration of lesions.

• EfficientNetV2: This model's advantage is built with high computational efficiency as well as classification accuracy.

It is done separately because each model discards specific characteristics, which are then featured into one location using an attentionbased fusion technique for integration. Self-attention shared weights render each model's output based on their relevance; the features that are more relevant are those that get given a higher weightage.

Multi-View Image Processing for Clinical and Dermoscopic Analysis

To get an even improved model performance, we incorporated a novel dual-branch multi-view image processing model, where the model can utilize clinical and dermoscopic images.

The first branch is the lesion feature extraction, which works on capturing macro lesions' characteristics such as size, boundary, and colour distribution. The second branch analyzes dermoscopic images, analyzing regularity, the pattern and types of vessels, and skin pigmentation at the subdermal level.

The cross-attention transformer also enables the model to combine information from the two branches of the network in the transformer fusion process. This improves the model's performance by providing better differentiation between tumor types of benign or malignant nature.

Experimental result

The experiments are conducted on several publicly available dermatology data primarily the ISIC (International Skin Imaging Collaboration) dataset and HAM10000 dataset. These datasets are quite popular and commonly used in skin cancer classification studies and all of them consist of quality skin lesion images which dermatological experts label. The ISIC dataset contains 25,331 high-resolution images and provided data of skin lesion types such as melanoma (MEL), basal cell carcinoma (BCC), squamous cell carcinoma (SCC), actinic keratosis (AK), benign keratosis like lesions (BKL), dermatofibroma (DF), vascular lesions (VASC) and benign nevi (NV). The data set is mainly made up of dermoscopy images and is reviewed by dermatologists, therefore it can be used for training deep learning algorithms. Nevertheless, this work has a few shortcomings: insufficient clinical metadata and presumably inadequate skin color coverage which may hinder the generalization's effectiveness.

Moreover, the HAM10000 dataset which is comprised of 10015 derived from patient's demographics is also used. When using this data, the model is able to learn from both dermoscopic and clinical, or macroscopic and microscopic appearance of skin lesions unlike what was done in the case of ISIC dataset. It also has the same lesion categories as the ISIC, with the annotations obtained from histopathology, in vivo Confocal reflectance microscopy, and follow-up examination, providing good ground truth to the data labeling process. The weakness of the HAM10000 dataset is that the data distribution has high variations in skin types and imaging conditions and can be used to make models more durable and accurate. However, unlike ISIC, it does not specify patient context information which is useful for including other clinical attributes for categorization.

The CSSL-ETL framework integrates two datasets to assist in configuring a large collection of labeled and unlabeled skin lesion images which would in return support self-supervised pretraining, effective feature extraction and optimum classification of malignant and benign skin conditions. The incorporation of these datasets make the proposed model very useful when it comes to clinical practice due to its ability to generalize well from one patient population, imaging condition, and lesion type to another.

Namely, To test the efficacy of the proposed approach, the Contrastive Self-Supervised Ensemble Transfer Learning (CSSL-ETL), skin lesion datasets, namely the ISIC, HAM10000, and DermNet, containing images were used. The experimentations were carried out on an NVIDIA RTX 3090 GPU, 24 GB VRAM, using TensorFlow and PyTorch frameworks. The model was compared with other deep learning models that include ResNet-50, EfficientNetV2, Swin Transformer models as well as CNN-Transformer hybrid models.

The data evaluation of the model was done by the Accuracy score, Precision score, and Recall score, F1- score and AUC-ROC score. The results of the experiment that would be of interest are summarized in Table 2 below.

Model Performance Analysis

The CSSL-ETL model that has been proposed in this work yielded an accuracy of 94.6% which is higher than all the baseline methods used. To this end, our model outperforms the best benchmark by achieving an improved classification accuracy of 3.4% against Swin Transformer with 91.2%. The precision, recall, and F1-score of CSSL-ETL also were higher, showing better ability of lesion classification and generalization. The AUC-ROC score of 96.1% again also shows the evidence to support the fact that CSSL-ETL can well differentiate between melanoma and benign lesions.

Comparative Analysis with Related Work

Comparing the proposed CSSL-ETL framework with other CSSL, TL, and CSSL-TL studies used in skin lesion classification. Figures two and three below capture the summary of the related studies as summarized in Table 3.

Discussion of Key Differences

• Compared to Wang et al. (2024) [1], we have higher accuracy level of +2.8% that employed the contrastive self-supervised learning algorithm enriched with the ensemble transfer learning.

• In comparison with Zhao et al. (2024) [4] which only utilized self-supervised learning method, the proposed CSSL-ETL outperforms better because of the multiple models fusion strategy.

• Compared to Wu et al. (2022) [6] that employed the Federated learning for privacy, here we introduce the combination of federated learning with ensemble transfer learning and gain better accuracy.

• Compared to Alzahrani (2024) [13] that only employed transformers, this research utilizes both CNN for its great texture recognition strength and transformers for the capability of recognizing the features at long intervals.

Ablation Study

In order to investigate the weight of each component of the CSSL-ETL framework, an ablation experiment was performed in a way presented in Table 4, where the efficiency of different configurations of models was assessed.

Impact of CSSL Pretraining

Implementing CSSL boosted the accuracy by itself from 88.3% (ResNet-50) to 91.1 %and this proves that contrastive learning has a positive impact in feature extraction for skin lesion classification.

Impact of Ensemble Transfer Learning

When only the ETL was used and without contrastive learning part, the test achieved an accuracy of 92,3%, over the single models but under the best case of the CSSL-ETL configuration. This just goes to Support that multi-model fusion does enhance the process of classification.

Synergistic Effect of CSSL + ETL

Through comparative self-supervised learning and transfer learning, the best result of 94.6% was attained. Thus, the experiment reveals that self-supervision improves generalization of the learned representations, whereas ensemble learning improves their reliability.

Discussion and Key Findings

The results obtained from this work show that CSSL-ETL does better than traditional DL models, only-SSL models, and only-Transfer Learning models. The key takeaways are:

A paper that appears to be related to contrastive prediction hypothesis is titled “Contrastive Self-Supervised Pretraining Improves Feature Representation.”

Contrastive pretraining phase of the designed network helps the model to learn discriminative features of skin image which will in turnenhance the classification robustness of the model.

• Combination of the multiple models contribute towards improving the classification accuracy through the proposed ensemble transfer learning. Using both the CNN and Transformer architectures, the model learns both the fine-grained texture and the overall structures of the skin lesions.

• Federated Learning Provides Privacy-Preserving Training. This feature helps to implement the model in the clinical setting because it does not require that patient data be sent across institutions.

• Explainability and Uncertainty Estimation Improve Clinical Trust. Thus, CSSL-ETL improves interpretability by using identification of salient features with Grad-CAM++ visualization, and estimation of uncertainty with Bayesian methods.

The confusion matrix for the CSSL-ETL gives distinctions of the transfer learning model in various classes of skin lesions. The true-test labels which stand for the actual classes are shown on the ‘y-axis whereas the predicted-test labels, which are the results of the model, are represented on the ‘x-axis. The numbers on the diagonal stand for true classifications while the other numbers represent erroneous classifications between the different classes.

From the confusion matrix

The Melanoma (MEL) was classified correctly 76 times, but it was misclassified with BCC (1), AK (2), BKL (1), and VASC (3).

Specifically, Basal Cell Carcinoma (BCC) was predicted correctly 80 of the time but was predicted to be VASC 7 times and in the other classes 4 times.

• I found that in classifying images, Squamous Cell Carcinoma (SCC) had the least difficulties with 67 correct classification decisions which were confused mostly with Ak and BCC.

In BKL, Actinic Keratasis (AK) had 38 as correct diagnosis, however, it was misclassified BKL and SCC.

Next was Benign Keratosis-like Lesions (BKL) scanning, which had very few learner errors, 45, out of the results it gave.

• Dermatofibroma (DF) was diagnosed correctly in 46 examples, with minor tendency toward the other classes.

• Some cases that were diagnosed with VASC were actually accurately diagnosed, although, there were instances that it was diagnosed as SCC, AK, and DF.

• Benign Nevi (NV) had total of 79 of correct predictions with minimal mistake.

The high value on the diagonal implies that the proposed model achieves high prediction accuracy for most of the lesions of the skin For minor misclassification errors are prevalent where one might not be able to distinguish the lesions in terms of visual differentiation between BCC and VASC or AK and SCC. These results indicate that none of the three healthy sites nor all the three pathological sites are misclassified and that CSSL-ETL can differentiate between the two classes of skin lesions with reasonable accuracy friend.

A detailed statistical research examines how the proposed CSSL-ETL model performs relative to Swin Transformer and ResNet-50 baseline models. The comparative analysis utilizes paired t-tests to establish CSSL-ETL superiority, combined with presentation of 5-fold cross-validation scores through confidence intervals and effect size calculations using Cohen's d. A set of paired t-tests verified that CSSL-ETL demonstrates practical and statistical superiority against Swin Transformer and ResNet-50 in terms of performance enhancement. A summary table presents the obtained results (Table 5).

The analysis uses 95% confidence intervals to determine mean performance differences and Effect sizes from Cohen's d to determine practical significance as presented in Table 6. The 5-fold cross-validation results form the basis of these statistical values.

The mean differences between CSSL-ETL and other methods become statistically significant in every case while keeping zero outside their confidence intervals. The effect size evaluation demonstrated through Cohen's d values reveals that the procedure improvements surpass the 0.8 standards, which indicates substantial practical value.

While impressive in its application, the CSSL-ETL framework presents specific barriers to implementation. TheSCSL-ETL framework's generalization ability risks restriction due to diverse imaging situation across the datasets. Future research needs to establish domain adaptation methods which would help normalize the differences in data quality between institutions. Patient data security could be enhanced by implementing differential privacy mechanisms with secure aggregation protocols because federated learning methods do not guarantee privacy when facing adversarial threats or gradient leaks. The present study relies on image features only but could benefit from integrating additional multiform data including patient records and genomic information and historical medical information because this would enhance diagnostic precision. Future research needs to assess real-time model deployment through edge computing and mobile-health (mHealth) platforms by measuring performance delays and improving device inference capabilities along with practical tests by medical specialists.

In conclusion, then, we presented the Contrastive Self-Supervised Ensemble Transfer Learning (CSSL-ETL) approach that offers the possibility to classify skin cancer and to perform its early detection. Thus, it enables to solve primary difficulties of dermatological image perception including the lack of large amounts of labeled data, the prevalence of certain classes, variation in the appearance of the lesion, and privacy issues. CSSL augments the unlabeled dermatological images to enhance the features learning process; simultaneously, ETL optimizes the classification rates and diversity of the model through aggregation of several deep learning structures.

The experiments also show that CSSL-ETL produces a high level of performance and is superior to traditional CNNs, transformers, and only self-supervised learning, etc. Specifically, the present model yielded an accuracy of 94.6% more than ResNet-50, EfficientNetV2, and Swin Transformer. This is evident from the high accuracy of classification as well as the confusion matrix that shows that there were few misclassifications of melanoma and benign skin lesions. In addition, the training and the validation for the model illustrates good convergence and high stability and generality.

This study compared with related work points out that moss; dot CSSL-ETL is effective in enhancing feature representation, classfication robustness and interpretability. That CNNs are used in combination with Transformers makes it possible for both the low-level textural saliency as well as the global structural configuration to be extracted. Moreover, explainability of the trained models (Grad-CAM++ and Bayesian deep learning) helps gain high-level confidence and increased model trust from the clinicians.

Thus, we incorporated federated learning (FL) for cooperation across different institutions while maintaining data privacy of patients. This makes the model appropriate for realtime use in hospitals and dermatology centers commonly, since the sharing of patient data is greatly limited.

Key Contributions and Findings

• There is evidence that contrastive self-supervised pretraining enhances the feature learning when it comes to the classification of skin lesions.

• Classification performance is improved through ensemble of transfer learning solution through combining of CNN and Transformers for feature extraction.

• Multi – view image processing enhances the results by applying dermoscopy concurrent with clinical image.

• Explanation methods (Grad-CAM++) improve model's interpretability and allow the dermatologists to comprehend AI outcomes.

• Federated learning also enables the training of models while keeping patient information private thus compatible with the deployment across different hospitals.

Limitations and Future Work

Thus, CSSL-ETL may have some shortcomings despite its high accuracy. It further highlights that multiple misclassification between visually similar lesions such as BCC and VASC or AK and SCC are still possible; therefore, with the incorporation of other sources of data like the patient history or genetics might improve the model. In addition, since federated learning solves the issue of privacy, it might be useful to consider further methods in the future, including differential privacy. The second area of development is in real-time mobile health applications for directly performing skin cancer screening through edge AI.

Conclusion Statement

In general, it can be concluded that the here-suggested CSSL-ETL framework is an effective, accurate, explainable and privacy-friendly AI solution for skin cancer classification. Through the new strategies formulated as CL, TLE, and FL, this study has established a new frontier in dermatological AI. The future trends in this study should be done with more extensive Medical Data of Multiphysics which will be joined in the form of Federated Learning, Security Integration for the Models, and deployment algorithms for Real-time.

Author Contribution Statement

All authors contributed equally in this study.

Acknowledgements

None.

References

1. Wang Z, Lin Y, Zhu X. Transfer contrastive learning for raman spectroscopy skin cancer tissue classification. IEEE J Biomed Health Inform. 2024;28(12):7332-44. https://doi.org/10.1109/jbhi.2024.3451950
2. Haggerty H, Chandra R. Self-supervised learning for skin cancer diagnosis with limited training data. arXiv. 2024;2401.00692.
3. Wang Z, Zhang L, Shu X, Wang Y, Feng Y. Consistent representation via contrastive learning for skin lesion diagnosis. Comput Methods Programs Biomed. 2023;242:107826. https://doi.org/10.1016/j.cmpb.2023.107826.
4. Zhao Z, Alzubaidi L, Zhang J, Duan Y, Naseem U, Gu Y. An experimental comparison of transfer learning against self-supervised learning. arXiv. 2024;2407.05592.
5. Kwasigroch A, Grochowski M, Mikołajczyk A. Self-supervised learning to increase the performance of skin lesion classification. Electronics. 2020;9(11):1930. https://doi.org/10.3390/electronics9111930
6. Wu Y, Zeng D, Wang Z, Sheng Y, Yang L, James AJ, Shi Y, Hu J. Federated self-supervised contrastive learning and masked autoencoder for dermatological disease diagnosis. arXiv preprint arXiv:2208.11278. 2022 Aug 24.
7. Fu W, Chen J, Zhou L. Boosting few-shot rare skin disease classification via self-supervision and distribution calibration. Biomed Eng Lett. 2024;14(4):877-89. https://doi.org/10.1007/s13534-024-00383-2
8. Wang X, Yang S, Zhang J, Wang M, Zhang J, Yang W, et al. Transformer-based unsupervised contrastive learning for histopathological image classification. Med Image Anal. 2022;81:102559. https://doi.org/10.1016/j.media.2022.102559
9. Chen K, Zhuang D, Chang JM. SuperCon: Supervised contrastive learning for imbalanced skin lesion classification. arXiv. 2022;2202.05685.
10. Rani V, Kumar M, Gupta A, Sachdeva M, Mittal A, Kumar K. Self-supervised learning for medical image analysis: a comprehensive review. Evol Syst. 2024;15(4):1-27. https://doi.org/10.1007/s12530-024-09581-w
11. Cino L, Mazzeo PL, Distante C. Comparison of different supervised and self-supervised learning techniques in skin disease classification. In: International Conference on Image Analysis and Processing. Cham: Springer; 2022. p. 77-88. https://doi.org/10.1007/978-3-031-06427-2_7
12. Riaz S, Naeem A, Malik H, Naqvi RA, Loh WK. Federated and transfer learning methods for the classification of melanoma and nonmelanoma skin cancers: a prospective study. Sensors. 2023;23(20):8457. https://doi.org/10.3390/s23208457.
13. Alzahrani SM. SkinLiTE: Lightweight supervised contrastive learning model for enhanced skin lesion detection and disease typification in dermoscopic images. Curr Med Imaging. 2024;20(1):e15734056313837. https://doi.org/10.2174/0115734056313837240612065845
14. Wang Y, Wang Y, Cai J, Lee TK, Miao C, Wang ZJ. SSD-KD: A self-supervised diverse knowledge distillation method for lightweight skin lesion classification using dermoscopic images. Med Image Anal. 2023;84:102693. https://doi.org/10.1016/j.media.2022.102693
15. Liang S, Tian S, Kang X, Zhang D, Wu W, Yu L. Skin lesion classification base on multi-hierarchy contrastive learning with pareto optimality. Biomedical Signal Processing and Control. 2023;86:105187. https://doi.org/10.1016/j.bspc.2023.105187
16. Ouyang J, Mao D, Guo Z, Liu S, Xu D, Wang W. Contrastive self-supervised learning for diabetic retinopathy early detection. Med Biol Eng Comput. 2023;61(9):2441-52. https://doi.org/10.1007/s11517-023-02810-5
17. Habchi Y, Kheddar H, Himeur Y, Boukabou A, Atalla S, Mansoor W, et al. Deep transfer learning for kidney cancer diagnosis. arXiv. 2024;2408.04318.
18. Liu Z, Alavi A, Li M, Zhang X. Self-supervised contrastive learning for medical time series: A systematic review. Sensors. 2023;23(9):4221. https://doi.org/10.3390/s23094221
19. Patil A, Mehto A, Nalband S. Enhancing skin lesion diagnosis with data augmentation techniques: a review of the state-of-the-art. Multimed Tools Appl. 2024:1-40. https://doi.org/10.1007/s11042-024-20145-7
20. Hsu BWY, Tseng VS. Hierarchy-aware contrastive learning with late fusion for skin lesion classification. Comput Methods Programs Biomed. 2022;216:106666. https://doi.org/10.1016/j.cmpb.2022.106666

21. Özbay E, Özbay FA, Gharehchopogh FS. Kidney tumor classification on CT images using self-supervised learning. Comput Biol Med. 2024;176:108554. https://doi.org/10.1016/j.compbiomed.2024.108554
22. Chae J, Kim J. An investigation of transfer learning approaches to overcome limited labeled data in medical image analysis. Appl Sci. 2023;13(15):8671. https://doi.org/10.3390/app13158671
23. Ren Z. Exploring the spectrum of supervision in medical image analysis: from fully supervised to semi-supervised and unsupervised approaches [PhD thesis. Leicester: University of Leicester; 2024.
24. Poonawala TA. Using self-supervised learning models to predict invasive ductal carcinoma from histopathological images [PhD thesis. Dublin: National College of Ireland; 2023.
25. Raju ASN, Venkatesh K, Gatla RK, Konakalla EP, Eid MM, Titova N, et al. Colorectal cancer detection with enhanced precision using a hybrid supervised and unsupervised learning approach. Sci Rep. 2025;15(1):3180. https://doi.org/10.1038/s41598-025-86590-y
26. Sirotkin K, Escudero-Viñolo M, Carballeira P, SanMiguel JC. Improved skin lesion recognition by a self-supervised curricular deep learning approach. arXiv preprint arXiv:2112.12086. 2021 Dec 22.
27.Song Z, Du P, Yan J, Li K, Shou J, Lai M, et al. Nucleus-aware self-supervised pretraining using unpaired image-to-image translation for histopathology images. IEEE Trans Med Imaging. 2024;43(1):459-72. https://doi.org/10.1109/tmi.2023.3309971
28. Li F, Li M, Zuo E, Chen C, Chen C, Lv X. Self-contrastive Feature Guidance Based Multidimensional Collaborative Network of metadata and image features for skin disease classification. Pattern Recognition. 2024;156:110742.
29. Dan W, Na P, Yanying W, Hongwei Z. Unlabeled skin lesion classification by self-supervised topology clustering network. Biomedical Signal Processing and Control. 2021;66:102428. https://doi.org/10.1016/j.bspc.2021.102428
30. Ruoyu W, Changyu L, Yuan L, Xu S, Jiuquan Z, Hong H. Self-supervised transfer learning framework driven by visual attention for benign–malignant lung nodule classification on chest ct. Expert Systems with Applications. 2023;215:119339. https://doi.org/10.1016/j.eswa.2022.119339.
31. Azizi S, Culp L, Freyberg J, Mustafa B, Baur S, Kornblith S, et al. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. Nat Biomed Eng. 2023;7(6):756-79. https://doi.org/10.1038/s41551-023-01049-7
32.Morita T, Han XH. Investigating self-supervised learning for Skin Lesion Classification. In: 2023 18th Int Conf Mach Vis Appl (MVA). IEEE; 2023. p. 1-5.