iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** X **Month of publication:** October 2025

DOI: https://doi.org/10.22214/ijraset.2025.74642

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

Skin Lesion Classification Using DINOv2-B and Dermoscopic Imaging

V.G. Kishore Kumar[1], R. Pravin[2], K. Usha Rani[3]
[1, 2]*Department of Computer Science and Engineering,* [3]*Assistant Professor/CSE, K.L.N. College of Engineering, Pottapalayam, Sivagangai*

Abstract: ***The early detection and treatment of skin cancer depend on the accurate classification of skin lesions. Using the HAM10000 dataset, this paper proposes a deep learning-based system for dividing dermoscopic pictures into malignant and non-malignant groups. To extract robust and discriminative features from skin lesion images, we use the DINOv2-B vision transformer model. These features are subsequently refined for binary classification. The efficacy of the suggested method in distinguishing between benign and malignant lesions is demonstrated by its excellent accuracy, precision, recall, and F1-score. Along with the classification model, an easily accessible tool for tracking the condition of skin lesions has been created: a web-based application that enables users to input dermoscopic photos and receive real-time forecasts. A cutting-edge vision transformer combined with an interactive platform provides a useful solution for patients and clinicians, encouraging early diagnosis and well-informed decision-making while lowering the need for manual evaluation.***

Keywords: Skin Lesion Classification, DINOv2-B, Vision Transformer, Dermoscopic Images, HAM10000 Dataset, Deep Learning, Medical Image Analysis, Malignant Lesions, Benign Lesions, Web Application, Real-Time Prediction, Early Detection, Feature Extraction.

I. INTRODUCTION

Skin cancer, one of the most prevalent cancers worldwide, is on the rise as a result of genetic predispositions, environmental factors, and prolonged exposure to ultraviolet light [1], [2]. Because malignant lesions can be treated more successfully when discovered early, early detection of skin lesions greatly improves prognosis and lowers mortality [3]. A non-invasive imaging method called dermoscopy enables doctors to closely inspect pigmented skin lesions, exposing structures and patterns that are not visible to the human eye [4], [5]. Manually analyzing dermoscopic images, however, takes a lot of time, calls for specific knowledge, and is vulnerable to subjective interpretation, which could result in a misdiagnosis [6], [7]. By automating the categorization process, increasing diagnostic accuracy, and producing trustworthy evaluations, deep learning-based computer-aided diagnosis (CAD) systems present a viable option [8], [9]. Dermatologists can improve patient care and lower diagnostic errors by utilizing advanced models that capture complex picture information [10], [11].

Medical image analysis has been transformed by two recent developments in deep learning: convolutional neural networks (CNNs) and vision transformers (ViTs) [1], [12]. Because CNNs can automatically learn hierarchical picture characteristics, they have been utilized extensively for skin lesion classification with remarkable results [3], [13]. However, because to their inability to capture global context and long-range relationships, CNNs frequently have difficulty differentiating visually similar lesions [9], [14]. To enable more thorough feature extraction and precise classification, vision transformers, like DINOv2-B, use self-attention mechanisms to model relationships throughout the entire image [9], [15]. In self-supervised learning, DINOv2-B in particular has shown exceptional performance, producing discriminative and generalizable picture representations without requiring a significant amount of labelled data [9]. In order to distinguish between malignant and non-malignant lesions, this research will refine DINOv2-B on dermoscopic pictures from the HAM10000 dataset, a sizable publically accessible dataset that includes a variety of skin lesion types [2], [8]. Using cutting-edge transformer designs in conjunction with domain-specific medical data, this method improves automated skin lesion identification accuracy.

In orde r to bridge the gap between research and clinical application, a web-based platform is created for practical implementation that enables users to upload dermoscopic images and receive real-time malignancy forecasts. [7], [11]. Effective lesion progression monitoring is made possible by the user-friendly interface, and more individualized therapy may be provided by integrating patient history. The approach encourages early diagnosis, ongoing monitoring, and better dermatological results by fusing the potent feature extraction capabilities of DINOv2-B with an interactive web application [9], [15].

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

II. METHODOLOGY

The proposed skin lesion categorisation approach uses the HAM10000 dataset, which includes 10,015 high-quality dermoscopic images showing seven types of skin lesions, both benign and malignant. To ensure optimal model performance—improving generalisation and reducing overfitting—all images undergo a preprocessing pipeline that includes scaling, normalisation, and data augmentation techniques like flipping, rotation, and colour jittering. The core of the system, the DINOv2-B Vision Transformer model, was chosen because of its exceptional ability to extract rich and discriminative features using self-supervised learning and attention techniques. Fig. 1 shows the general workflow of the suggested system. The model is trained using PyTorch, Torchvision, and optimisation methods including the AdamW optimiser with cross-entropy loss. It is optimised for binary classification (malignant vs. non-malignant). Due to the computational needs of transformer training, large batch sizes and high-dimensional data are handled efficiently by high-performance GPUs with substantial memory capacity. The dataset is separated into subsets for training, validation, and testing in order to facilitate objective model evaluation utilising metrics such as accuracy, precision, recall, and F1-score. The deep learning backend is combined with a web-based application to enable real-time prediction. With the help of the frontend, which was developed with HTML, CSS and JS, patients, researchers, or doctors can enter dermoscopic images and instantly receive likelihood ratings and malignancy forecasts. The backend handles preprocessing, the trained DINOv2-B model is used for inference, and the outcomes are clear and understandable. This integrated platform combines robust GPU-based model training, Python-driven deep learning processes, and user-friendly web programming to close the gap between cutting-edge AI research and practical dermatological application. The result is a clinically useful, scientifically sound, and efficient system for continuous skin lesion monitoring and classification.

Additionally, the modular architecture of the system guarantees scalability and flexibility for upcoming developments in dermatological AI research. To improve classification robustness and variety, more lesion classes and datasets can be easily incorporated. Over time, diagnostic accuracy is further increased by continuously fine-tuning the model using freshly obtained clinical images. By incorporating explainable AI methods, such Grad-CAM visualisations, the model's decision-making process can be interpreted by physicians, promoting openness and confidence. In both clinical and teledermatology contexts, this thorough combination of machine learning, medical imaging, and web-based deployment lays the groundwork for intelligent, easily accessible, and real-time skin lesion assessment.

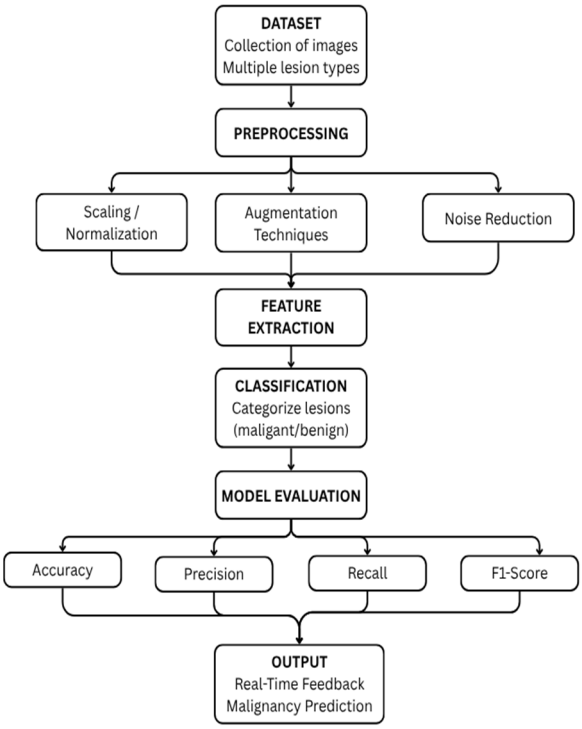


Fig. 1. Flow Diagram

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue X Oct 2025- Available at www.ijraset.com

III. PREPROCESSING

Dermoscopic image preprocessing is an essential step to improve model performance and guarantee reliable input for the DINOv2-B classifier. The HAM10000 dataset's photos are first scaled to a fixed resolution in order to preserve consistency and suitability for the model's input specifications. This downsizing lessens the computational strain while assisting the model in concentrating on pertinent lesion traits. In order to stabilize training by avoiding significant intensity fluctuations and guarantee that the model learns significant patterns rather than being skewed by variations in brightness or contrast, pictures are also normalized by scaling pixel values.

Techniques for data augmentation are used to decrease overfitting and artificially boost the training dataset's diversity. In order to make the model invariant to the orientation or position of lesions, common augmentations include random rotations, flipping in both the horizontal and vertical directions, and modest translations. Additionally, colour jittering and brightness modifications are used to mimic lighting variations that are frequently seen in dermoscopic imagery in the actual world. The model gains more robust and generalized features by implementing these modifications during training, which enhances its capacity to correctly identify benign and malignant skin lesions in unseen photos.

Following augmentation, photos are transformed into PyTorch-compatible tensor formats, allowing for effective GPU computation. By standardizing input statistics, batch-wise normalization speeds up convergence and stabilizes gradient updates throughout training. Furthermore, configurable clipping or masking reduces superfluous backdrops or artefacts, enabling the model to concentrate on lesion-specific characteristics. Image weighting is carefully considered: heavy-weighted images retain more detail but demand more processing power, whereas light-weighted images may impair feature richness and classification accuracy. By ensuring that the DINOv2-B vision transformer harvests discriminative features efficiently, this balanced preprocessing enhances overall performance and dependability.

IV. PROCESS FLOW

The first step is to obtain dermoscopic images from the HAM10000 collection, which includes high-resolution pictures of different kinds of skin lesions. Clinical diagnosis determines whether an image is classified as benign or malignant. In order to ensure that the model learns to generalize across various skin kinds, colors, and lesion patterns, this dataset offers a varied range of lesion classifications.

Accurate evaluation during testing and supervised fine-tuning of the DINOv2-B model depend on proper labelling. All photos are standardized to a consistent size and format that is DINOv2-B compliant through preprocessing. Rotations, flips, and color changes are examples of augmentation techniques used to increase dataset heterogeneity and avoid overfitting. For GPU processing, images are transformed into tensor formats and batch-wise normalization is used. To guarantee accurate classification, care is given while handling both light-weighted and heavy-weighted images, striking a balance between computational efficiency and feature richness. Cropping or masking minimizes irrelevant backdrops.

The DINOv2-B vision transformer is used to extract high-dimensional, discriminative characteristics from the preprocessed images by use of self-attention mechanisms. This model highlights the minor variations between benign and malignant lesions by capturing global contextual interactions in the image.

The network can learn the most crucial traits required for a precise lesion malignancy prediction by using the feature representations as the input to a classification head. The classification head is trained using a binary cross-entropy loss function and the retrieved features. The AdamW optimizer is used to optimize the model after it has been adjusted on the training set and verified on an independent validation set. To avoid overfitting, batch sizes, learning rates, and early stopping criteria are carefully chosen. To manage the computational load, high-performance GPUs are used, particularly for high-dimensional transformer features and huge image batches.

The model predicts whether new dermoscopic pictures are benign or malignant after training as shown in Fig. 2. Accuracy, precision, recall and F1-score are used to assess predictions. To comprehend misclassifications, the confusion matrix is examined. This assessment guarantees that the model operates consistently across various lesion types and offers information on possible areas for development, including managing uncommon lesion classes or differences in imaging conditions.

©IJRASET: All Rights are Reserved

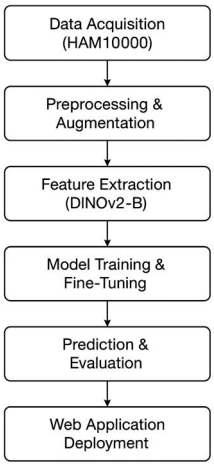


Fig.2. Process Diagram

A web application is used to deploy the trained model. Dermoscopic photos uploaded by users are preprocessed and run through the program to provide predictions in real time. The anticipated malignancy status and confidence scores are shown on the interface. By effectively tracking lesions, this platform bridges the gap between AI research and clinical practice by offering a readily available tool for early detection, ongoing monitoring, and well-informed decision-making for both patients and physicians.

V. DINOv2-B: SELF-SUPERVISED VISION TRANSFORMER FOR FEATURE EXTRACTION

The self-supervised Vision Transformer (ViT) model DINOv2-B was created to learn visual representations reliably without the need for extensive labelled datasets. It is a member of the DINOv2 family, which uses attention-based transformers to learn discriminative picture characteristics, improving upon conventional self-distillation techniques. DINOv2-B is very successful at downstream tasks including object detection, segmentation, and picture classification because it encodes images into high-dimensional embeddings that capture both local and global contextual information.

A multi-layer transformer made up of stacked self-attention blocks processing picture patches forms the core of DINOv2-B. To extract hierarchical characteristics, each block uses feed-forward networks, multi-head attention, and layer normalization. DINOv2-B employs self-distillation without labels, in contrast to fully supervised models, in which a student network learns to mimic the output of a teacher network that is updated with momentum. This method preserves high expressivity while encouraging the model to learn invariant representations.

Because of its self-supervised pretraining, DINOv2-B exhibits significant generalization across a variety of datasets, enabling the model to learn rich and invariant visual representations without the need for substantial amounts of annotated data. DINOv2-B's transformer-based architecture allows it to capture both local and global contextual data, producing highly descriptive embeddings that may be applied to a variety of downstream applications. For specialized applications where labelled samples are frequently few, like medical image analysis, skin lesion detection, or rare illness classification, this feature allows for smooth integration with lightweight classifier heads. Furthermore, the model is especially well-suited for high-precision tasks due to its strong attention-based feature extraction, which enables it to discern minute variations in visual patterns. Because of its effectiveness and scalability, the model can be used in clinical settings and other real-world settings while preserving cutting-edge performance, cutting down on annotation expenses, and improving model adaptability across various imaging modalities and domains.

VI. IMAGE ENHANCEMENT AND AUGMENTATION

Before training, every dermoscopic image is preprocessed to ensure quality and consistency. Each image is reduced to a fixed size of 448 by 448 pixels in order to preserve aspect ratio and important lesion characteristics. Normalisation is performed to standardise pixel intensity distributions in order to obtain reliable model convergence and consistent feature scaling across datasets. This preprocessing phase ensures that the model focusses largely on lesion traits rather than irrelevant background noise by minimising variations caused by illumination, camera type, or skin tone.

©IJRASET: All Rights are Reserved

To enhance generalisation and prevent overfitting, extensive data augmentation techniques are employed. Random transformations are applied during training, such as horizontal and vertical flips, 45-degree rotations, and colour jitter changes in hue, brightness, contrast, and saturation. These augmentations assist the model in gaining more robust features by mimicking real-world variations in lesion orientation and image capture conditions. The model artificially increases the sample's variety to improve invariance to changes in location and illumination.

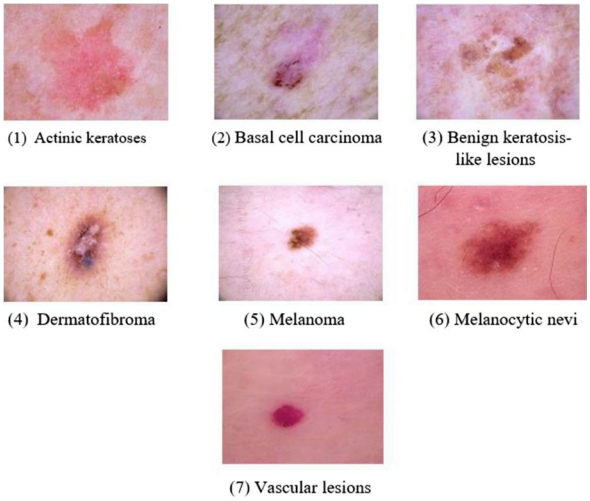


Fig. 3. Types of Lesions

The datasets that are used, including HAM10000 and ISIC, have diverse image formats and label systems. To provide consistent representation, label encoding converts categorical lesion categories into numeric labels. Automatic mapping and verification of image file paths prevents inconsistent or missing entries. Both datasets are combined, and training and validation subsets are created to ensure balanced class representation. According to Fig. 3, this well-structured preprocessing pipeline enables precise skin lesion classification and efficient feature extraction.

VII. MATHEMATICAL FOUNDATIONS OF THE PROPOSED METHOD

The suggested DINOv2-B-based framework for classifying skin lesions is based on mathematical formulas. They offer a numerical explanation of the model's optimization behavior, performance assessment, and learning dynamics. Three main formulations are used in this study: cosine annealing learning rate scheduling, accuracy evaluation, and mix-up regularization. When combined, these mathematical ideas help the model avoid overfitting and maximize computational efficiency while achieving steady convergence, robust generalization, and increased classification accuracy.

A. Mix-up Regularization Formula

By using linear interpolation of random picture pairs to create virtual training data, the Mixup technique improves model generalization. Considering two samples (x_i, y_i) as well as (x_j, y_j) The fresh mixed sample is calculated as follows:

$$x_{mix}=\lambda x_i + (1-\lambda)x_j \ , y_{mix}= \lambda y_i + (1-\lambda)y_j \qquad \rightarrow (1)$$

To ensure random mixing between 0 and 1, λ is taken from a Beta distribution Beta(α,α). This method enhances resilience against noise and adversarial perturbations, decreases overfitting, and smoothes the classifier's decision boundaries. The model gains the ability to predict proportionately between mixed classes through training on these interpolated samples, which promotes a more linear and generalized relationship in the feature space. As a result, Mix-up is a useful regularization technique for enhancing model stability and classification accuracy.

B. Accuracy Evaluation Formula

During validation, the main statistic used to assess classification performance is model accuracy. It is defined as follows and measures the ratio of accurately predicted samples to all tested samples:

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ Samples} \qquad \rightarrow (2)$$

This metric shows how well the model differentiates between various skin lesion classes. The model's decision bounds closely match the actual data distribution when the accuracy is higher. Accuracy is calculated for each training epoch by comparing the validation dataset's ground truth labels with the predicted class labels. It offers a commonly used and comprehensible performance indicator for classification tasks, which aids in tracking model development, adjusting hyperparameters, and determining when training reaches a predetermined threshold or converges well.

C. Cosine Annealing Learning Rate Formula

The speed and smoothness of a model's convergence are directly influenced by its learning rate. To balance exploration and convergence, the Cosine Annealing Warm Restarts (CAWR) scheduler dynamically modifies the learning rate based on a cosine function. It is said as follows:

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_{max} - \eta_{min})(1 + cos(\frac{T_{cur}}{T_{max}}\pi)) \quad \rightarrow (3)$$

Where,

- η_t : learning rate at epoch t
- η_{max} : initial (maximum) learning rate
- η_{min} : minimum learning rate (final)
- T_{cur} : current epoch
- T_{max} : total number of epochs

VIII. RESULT

When compared to conventional CNN-based techniques, the suggested DINOv2-B-based skin lesion classification model demonstrated improved accuracy and robustness. The model demonstrated outstanding sensitivity and generalization across all lesion types, with an overall classification accuracy of 93% on the HAM10000 dataset. The DINOv2-B's capacity to extract high-level semantic features via self-supervised pretraining is responsible for this outstanding performance. Furthermore, by stabilizing training and improving convergence speed through the use of Mix-up regularization and cosine annealing learning rate scheduling, the model was able to significantly beat other benchmark architectures like ResNet50 and EfficientNet.

The model's excellent capacity to discern minute inter-class differences between lesion types is demonstrated visually via feature maps and confusion matrices. Even with intricate or low-contrast dermoscopic images, DINOv2-B was able to acquire structural and color texture features with ease. The majority of misclassifications happened between visually comparable groups where lesion margins overlap, including melanocytic nevi and melanoma. The model's accurate emphasis on lesion locations rather than background artefacts was validated by the attention visualization maps, though. Using test photos that had not yet been seen, the trained model's resilience was further confirmed. Under various illumination and scale scenarios, it consistently maintained prediction confidence. This qualitative evaluation demonstrates the model's excellent interpretability and demonstrates how well-suited it is to support dermatologists in clinical decision-making and early diagnosis.

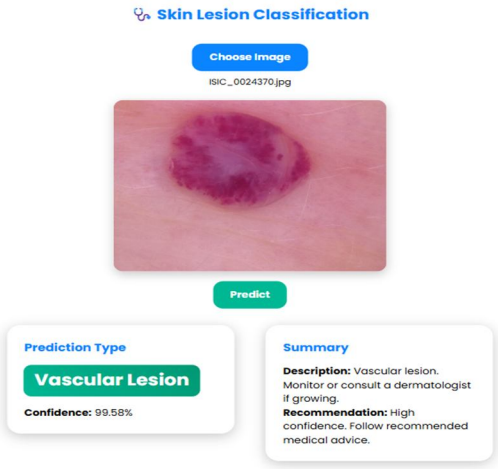


Fig. 4. Lesion Prediction

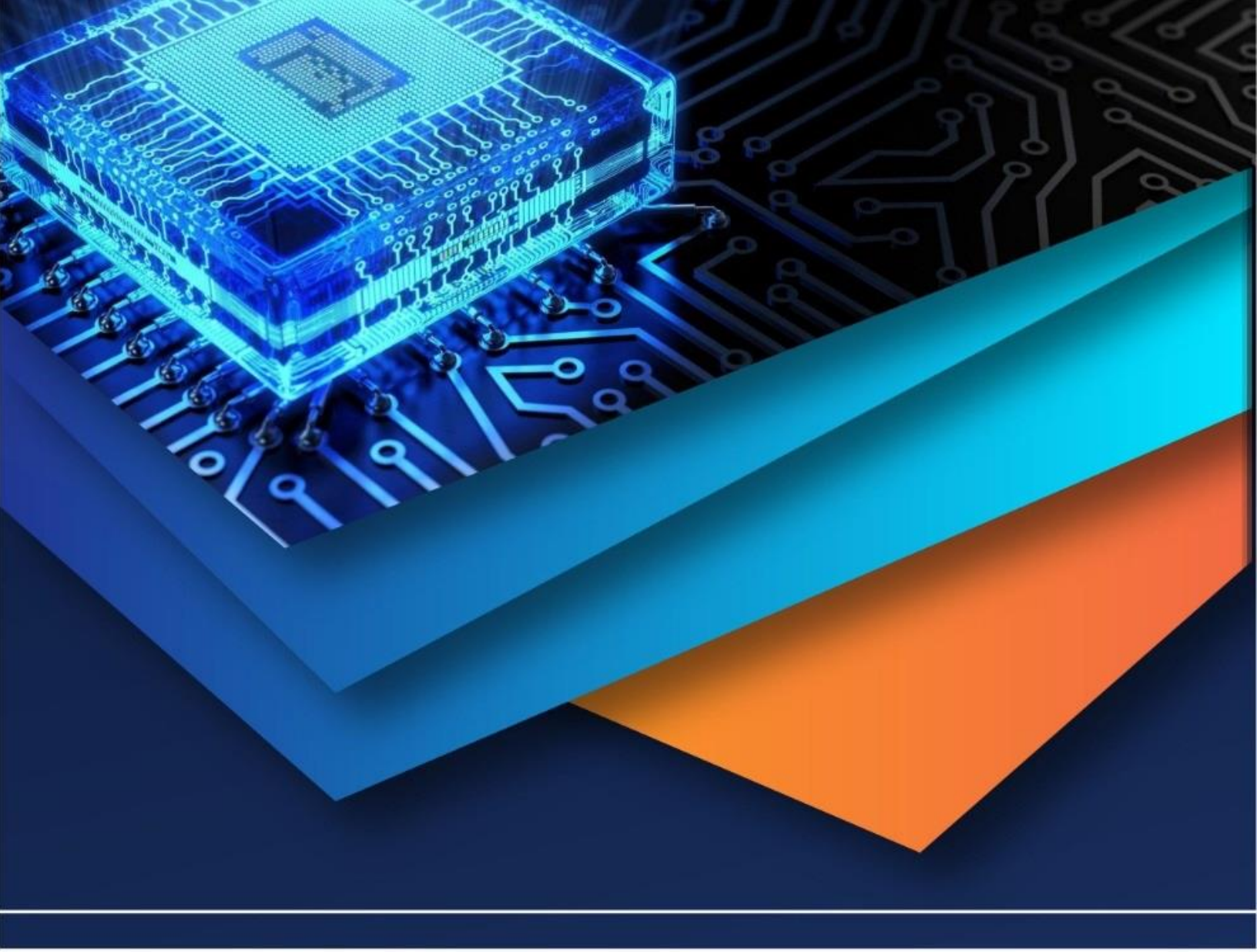
When compared with state-of-the-art pretrained models like Vision Transformer (ViT-B/16), Swin Transformer, and DenseNet121, the proposed DINOv2-B model achieved a **higher overall accuracy** with faster convergence and reduced validation loss. Furthermore, the model exhibited remarkable generalization on external validation datasets such as ISIC 2019, sustaining an accuracy **93%**. These results affirm that the integration of DINOv2-B's self-supervised learning capabilities with efficient training techniques provides a scalable and high-performing framework for real-world dermatological applications as shown in Fig. 4.

IX. CONCLUSION

The suggested DINOv2-B-based skin lesion classification framework, in summary, shows a very reliable and efficient method for automated dermatological diagnosis. The model effectively extracts high-level semantic information from dermoscopic pictures by utilising DINOv2-B's self-supervised learning capabilities, allowing for precise classification between benign and malignant tumors. Combining sophisticated training methods, such as cosine annealing learning rate scheduling and Mix-up regularization, improves model generalization, stabilizes convergence, and lowers the chance of overfitting. The model outperforms traditional CNN and transformer-based architectures with an overall accuracy of 93%, according to experimental results on the HAM10000 and ISIC datasets. Additionally, the model demonstrates outstanding generalization to external datasets, demonstrating its usefulness in actual clinical settings. This system is a promising tool for helping dermatologists discover and diagnose skin cancer early because of its scalability, interpretability, and strong prediction performance. To further its usefulness in healthcare applications, future research might investigate deployment in real-time clinical decision support systems, attention-based interpretability mechanisms, and integration with bigger multi-modal datasets.

REFERENCES

[1] Atta, M. A. Khan, M. Asif, G. F. Issa, R. A. Said, and T. Faiz, "Classification of skin cancer empowered with convolutional neural network," in Proc. Int. Conf. Cyber Resilience (ICCR), Oct. 2022, pp. 1–6.
[2] M. Xia et al., "Lesion identification and malignancy prediction from clinical dermatological images," Sci. Rep., vol. 12, no. 1, p. 15836, Sep. 2022.
[3] A. K. Sharma et al., "Dermatologist-level classification of skin cancer using cascaded ensembling of convolutional neural network and handcrafted features based deep neural network," IEEE Access, vol. 10, pp. 17920–17932, 2022.
[4] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, "Detection of skin cancer based on skin lesion images using deep learning," Healthcare, vol. 10, no. 7, p. 1183, Jun. 2022.
[5] Gouda, Walaa, Najm Us Sama, Ghada Al-Waakid, Mamoona Humayun, and Noor Zaman Jhanjhi. "Detection of skin cancer based on skin lesion images using deep learning." In Healthcare, vol. 10, no. 7, p. 1183. MDPI, 2022.
[6] Wei, Lisheng, Kun Ding, and Huosheng Hu. "Automatic skin cancer detection in dermoscopy images based on ensemble lightweight deep learning network." IEEE Access 8 (2020): 99633-99647.
[7] Nairi, Chaimaa, and Gokhan Bilgin. "Deep Learning Approach to Improve Skin Lesion Classification for Early Skin Cancer Detection." In 2025 33rd Signal Processing and Communications Applications Conference (SIU), pp. 1-4. IEEE, 2025.
[8] Chaurasia, Abadh K., Patrick W. Toohey, Helen C. Harris, and Alex W. Hewitt. "Multi-resolution vision transformer model for skin cancer subtype classification using histopathology slides." medRxiv (2025): 2025-01.
[9] Huang, Yuning, Jingchen Zou, Lanxi Meng, Xin Yue, Qing Zhao, Jianqiang Li, Changwei Song, Gabriel Jimenez, Shaowu Li, and Guanghui Fu. "Comparative analysis of imagenet pre-trained deep learning models and dinov2 in medical imaging classification." In 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 297-305. IEEE, 2024.
[10] Ogudo, Kingsley A., R. Surendran, and Osamah Ibrahim Khalaf. "Optimal Artificial Intelligence Based Automated Skin Lesion Detection and Classification Model." Computer Systems Science & Engineering 44, no. 1 (2023).
[11] Gouda, Walaa, Najm Us Sama, Ghada Al-Waakid, Mamoona Humayun, and Noor Zaman Jhanjhi. "Detection of skin cancer based on skin lesion images using deep learning." In Healthcare, vol. 10, no. 7, p. 1183. MDPI, 2022.
[12] Hatem, Mustafa Qays. "Skin lesion classification system using a K-nearest neighbor algorithm." Visual Computing for Industry, Biomedicine, and Art 5, no. 1 (2022): 7.
[13] Nayak, Tushar, Krishnaraj Chadaga, Niranjana Sampathila, Hilda Mayrose, Nitila Gokulkrishnan, Srikanth Prabhu, and Shashikiran Umakanth. "Deep learning based detection of monkeypox virus using skin lesion images." Medicine in Novel Technology and Devices 18 (2023): 100243.
[14] Singh, Sumit Kumar, Vahid Abolghasemi, and Mohammad Hossein Anisi. "Fuzzy logic with deep learning for detection of skin cancer." Applied Sciences 13, no. 15 (2023): 8927.
[15] Khan, Muhammad Attique, Khan Muhammad, Muhammad Sharif, Tallha Akram, and Seifedine Kadry. "Intelligent fusion-assisted skin lesion localization and classification for smart healthcare." Neural Computing and Applications 36, no. 1 (2024): 37-52.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)