International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V72I1P103

Volume 72 Issue 1, 20-39, January 2024

Original Article

An Intuitive U-LSTM Model for Classification and Recognition for Skin Cancer Detection

Ravi Chandra Bandi[1], K. Rajendra Prasad[2], A. Kamalakumari[3], A. Daisy Rani[4]

[1,3,4]*Department of Instrument Technology, Andhra University College of Engineering, Visakhapatnam, India.*
[2]*Department of Computer Science & Systems Engineering, GITAM University, Visakhapatnam, India.*

[4]*Corresponding Author : dr.adaisyrani@andhrauniversity.edu.in*

Received: 02 September 2023 Revised: 11 December 2023 Accepted: 16 December 2023 Published: 07 January 2024

Abstract *- The precise identification and categorization of cutaneous lesions and melanoma are of paramount importance in the timely discovery and successful management of these conditions. This paper presents a unique methodology that integrates the Intuitive U-Net and Long Short-Term Memory (LSTM) architecture to achieve precise categorization of skin lesions and melanoma. The Intuitive U-Net architecture, which draws inspiration from the U-Net model, has been specifically developed to effectively capture complex characteristics and complicated patterns present in dermoscopic images. Skip connections are employed in order to maintain spatial information and retrieve pertinent features at various scales. This improves the model's capacity to discern distinct categories of lesions and effectively categorize instances of melanoma. In order to enhance the precision of temporal relationships among sequences of images, the network is augmented with LSTM units. The Long Short-Term Memory (LSTM) architecture facilitates the model in taking into account the sequential context of images, thereby capturing temporal changes and patterns. Monitoring the advancement of skin lesions and the evolution of melanoma holds significant importance in this context. The methodology we propose is assessed using a heterogeneous dataset that includes a range of skin lesions and instances of melanoma. The proposed findings illustrate the efficacy of the Intuitive U-Net and LSTM architecture in accurately categorizing skin lesions and melanoma using transforming modelling with the U-LSTM model. The integration of feature extraction from pictures with temporal modelling using Long Short-Term Memory (LSTM) has been found to enhance sensitivity, specificity, and accuracy in comparison to current methodologies. With this approach, we demonstrate the overall design model with different specification parameters indicating the effectiveness of classification and identification with an accuracy of 91.5% on lesions while melanoma with 98.2%. The other performance metrics, such as sensitivity, specificity, and ROC curves, are utilized to emphasize providing the best classification model feature performance when compared to state-of-the-art architectures.*

Keywords *- Convolutional Neural Networks (CNN), Dense-Net architecture, K – Nearest Neighbours (KNN), Long Short-Term Memory (LSTM), U-Net Architecture.*

1. Introduction

Cancer, an ailment characterized by the unregulated proliferation of cells, continues to pose a significant worldwide health concern. It exerts an impact on multiple physiological systems and represents a prominent contributor to mortality rates. Skin cancer, a commonly occurring kind of cancer, begins inside the epidermis, the outermost layer of the skin. Previous studies have employed machine learning techniques to identify skin cancer by leveraging protein sequences and imaging modalities. Nevertheless, conventional machine learning methods require extensive manual effort in feature engineering. The concern has been alleviated by the utilization of deep learning, which provides automated feature extraction. The present study utilizes convolution-based deep neural networks as a means to identify and classify instances of skin cancer, employing the ISIC dataset for this purpose. Ensemble learning is utilized in order to enhance accuracy, considering the crucial significance of cancer diagnosis. A combination of VGG, CapsNet, and ResNet deep learning models is employed, resulting in improved sensitivity, accuracy, and other performance indicators. The study proposes that there are wider implications for the diagnosis of diseases. The rapid proliferation of skin cancer underscores the importance of early detection in order to facilitate effective treatment. Deep learning, specifically Convolutional Neural Networks (CNNs), has demonstrated exceptional performance in the tasks of object detection and categorization. The MNIST: HAM10000 dataset, which consists of seven distinct types of skin lesions, is utilized. The utilization of preprocessing techniques and transfer learning methods has been observed to improve the performance of models significantly. The classification of skin cancer with the use of computers frequently depends exclusively on the analysis of photographs.

The integration of patient metadata has the potential to improve the precision of diagnoses. The MetaBlock algorithm, a novel approach in metadata processing, has been developed to improve picture feature extraction. This algorithm has demonstrated superior performance compared to alternative methods across several circumstances. The timely and accurate detection of skin cancer is of paramount importance for the overall survival and well-being of patients. The utilization of deep learning techniques has proven to be beneficial in assisting dermatologists in attaining precise diagnoses. The utilization of deep learning algorithms enables the prompt identification of cases, leading to accelerated treatment, decreased workload, and enhanced management of intricate cases. Numerous methodologies utilizing deep learning techniques have been put forth for the purpose of detecting skin cancer. The authors of the study effectively tackle obstacles such as domain shift and imbalanced datasets, yielding noteworthy outcomes. The proposed methods employ novel approaches such as clustering, integration of images and metadata, and the use of lightweight recognition models. The task of identifying melanoma presents a significant need for the development of strong and reliable deep-learning methodologies. Transfer learning, generative models, and hybrid architectures have been suggested as potential solutions to address the issue of limited data availability, improve accuracy, and handle intricate image attributes. These methodologies have significantly transformed the process of identifying early-stage melanoma. The field of deep learning has made substantial progress in enhancing the accuracy of skin cancer diagnosis.

However, the issue of domain shift continues to be a matter of concern. The integration of multiple datasets and the effective management of class imbalance are essential methodologies in academic research. Hybrid methodologies, such as the utilization of 3D surface analysis, progressive transfer learning, and ensemble classifiers, demonstrate potential. The incorporation of deep learning into the field of the Internet of Things (IoT), particularly in the context of the Internet of Medical Things (IoMT), has brought about a transformative impact on the healthcare industry. The integration of intelligent devices, advanced deep learning algorithms, and cloud-fog-edge architectures collectively facilitate the early identification of skin cancer while simultaneously ensuring the privacy of patient data and enhancing the efficiency of diagnosis. The diagnosis of skin cancer is enhanced by the utilization of multi-modal and feature-rich methodologies. The utilization of deep learning methods, including Convolutional Neural Networks (CNNs) and Local Binary Pattern (LBP) techniques, has been found to improve the accuracy of classification tasks significantly. The integration of architectural elements surpasses the performance of standalone components. The identification of skin cancer necessitates the integration of advanced solutions that amalgamate medical expertise and technological advancements.

Brain tumors can be classified using pre-trained models in deep learning frameworks, such as GoogLeNet. The transfer learning model exhibits higher performance across a range of evaluation parameters. Innovative techniques contribute to the progress of skin lesion diagnosis. The application of architectural search frameworks in the context of malignant melanoma identification has been found to produce outcomes of superior quality while also minimizing computing expenses. These methodologies facilitate the automation of network design, thereby enhancing both accuracy and efficiency.

1.1. Problem Statement

The precise and timely identification of skin lesions and malignancies holds significant significance within the realm of dermatology and healthcare. Nonetheless, the intricate visual attributes of skin lesions, fluctuations in imaging settings, and the requirement for accurate discrimination between benign and malignant instances provide noteworthy obstacles. Conventional diagnostic techniques predominantly depend on human skill, resulting in potential drawbacks such as prolonged time consumption and subjectivity. In order to tackle these issues, the utilization of deep learning techniques has emerged as a highly promising approach.

The focus of this study is to utilize deep learning methodologies in order to create reliable and precise models for the identification and categorization of skin lesions and malignancies. The task at hand encompasses the development of an automated system capable of analyzing and categorizing skin photographs into distinct classifications, including benign, malignant, or specific lesion types. The system should possess the ability to effectively manage diverse picture quality, patient demographics, and differences in lesion appearances. Additionally, it is imperative to consider factors such as class imbalance and domain shift when transferring models across diverse datasets or healthcare settings.

1.2. Objectives

The overall paper comprises the design criteria indicating how to emphasize the lucrative parametric and intuitive model development on DL techniques that improvise the overall datasets chosen for the lesion and Melanoma Classification. The proposed work with U-LSTM objectives is stated below:

1. To develop a U-LSTM-based system capable of recognizing temporal patterns in sequential dermoscopic images.
2. To enable the detection of evolving features that contribute to accurate skin cancer and lesion diagnosis.
3. To improvise the proposed U-LSTM design to achieve high sensitivity for early detection of potential malignancies while maintaining high specificity to minimize false positives and ensure reliable clinical decision-making.
4. Encapsulate the design model with SOA models with its feature importance and effectiveness of the design criteria.

1.3. Overview

The overview of this research paper introduces the use of U-LSTM architecture for the detection of skin cancer and lesions. We delve into the significance of addressing this problem, given its impact on healthcare. Our study focuses on designing a U-LSTM-based system to analyze sequences of dermatoscopic images, enabling the recognition of evolving lesion characteristics. We outline our objectives, research scope, methodology, and key findings. By integrating U-LSTM analysis, our work contributes to the accurate early detection of skin cancer, enhancing medical intervention and patient outcomes. The subsequent sections of the paper detail our approach and results in-depth.

2. Literature Survey

Skin cancer is a widespread and complex issue, often requiring labor-intensive methods for detection. This study employs convolutional deep neural networks on the ISIC dataset to improve skin cancer identification. To enhance accuracy, the research introduces an ensemble of deep learning models, showcasing superior results compared to individual models. These findings suggest the potential utility of ensemble methods in disease detection. [1].

Skin cancer is a widespread and rapidly spreading disease, posing significant challenges in early detection due to limited resources. Timely identification is crucial for preventive measures. Deep learning, particularly Convolutional Neural Networks (CNN), has proven effective in skin lesion classification. The HAM10000 dataset with 10,015 samples and data preprocessing techniques, including feature extraction and transfer learning with DenseNet169 and ResNet50, were employed. The model achieved impressive accuracy in skin cancer detection, emphasizing its potential to improve early diagnosis and preventive interventions. [2].

This study explores the integration of patient demographics with skin lesion photos using deep neural networks for improved skin cancer classification. It introduces the Metadata Processing Block (Meta Block) to leverage metadata, enhancing data categorization. The Meta Block augments relevant image features at various classification stages. Comparative analysis with two alternative approaches, Meta Net and feature concatenation, reveals the superiority of the proposed method. Across ten scenarios, it consistently improves classification performance, outperforming alternative approaches in six instances. This innovation holds promise for more accurate and comprehensive skin cancer diagnosis. [3].

Skin cancer, a prevalent global health concern, demands timely and accurate detection. Deep learning offers a solution by assisting dermatologists and enabling early intervention. This study aimed to build accurate deep-learning models for skin cancer classification, addressing class imbalance and decision interpretability. Utilizing an optimized CNN architecture on the HAM10000 dataset, it achieved an 82% classification accuracy and introduced an Explainable AI system to aid medical professionals. These results showcase deep learning's potential to revolutionize skin cancer diagnosis and improve healthcare outcomes.[4].

In the modern healthcare landscape, computer vision plays a vital role in disease detection, particularly in skin cancer diagnosis, where early identification is critical. This study introduces two innovative techniques for categorizing dermoscopic images, emphasizing the differentiation of benign and malignant tumors. The first approach combines the K-nearest neighbor (KNN) algorithm with pretrained deep neural networks for feature extraction, while the second method utilizes AlexNet optimized with the grey wolf optimizer for hyperparameter tuning. Comparative analysis between machine learning (ML) and deep learning (DL) techniques, involving a range of ML models and DL models with pre-trained architectures, achieved remarkable performance, with accuracy exceeding 99% in select models. [5].

This study introduces a lightweight and innovative skin cancer detection model that emphasizes feature discrimination for improved performance. It combines two feature extraction modules and a dedicated feature discrimination network, resulting in superior lesion detection and area segmentation. The proposed approach using LW-CNN outperforms existing deep learning methods by effectively addressing intra-class variations and inter-class similarities, offering promising advancements in skin lesion detection.[6].

This research addresses the issue of cross-domain skin disease recognition within the context of deep learning. Two novel approaches are proposed to enhance model generalization across different datasets affected by domain shift. The initial method utilizes progressive transfer learning, which entails the refinement of a pre-trained CNN on two separate skin disease datasets. The second method employs adversarial learning as a domain adaptation technique, enabling the transformation of invariant attributes from the source to the target domain, ultimately enhancing the performance of recognition. Thorough assessments encompassing tasks such as melanoma identification, cancer detection, and cross-modality learning, conducted across a range of datasets, provide validation for the efficacy of these approaches in tackling domain shift challenges [7].

Melanoma, a deadly skin cancer type, poses challenges for precise Computer-Aided Diagnosis (CAD) systems due to its intricate visual features. Existing methods either rely on manual feature selection or use entire images for deep learning, facing difficulties in feature extraction and limited data. This study introduces an intelligent system utilizing a Region of Interest (ROI) approach powered by transfer learning. Enhanced k-means help extract ROIs, focusing on melanoma-containing images for discriminative feature identification. Employing a CNN-based transfer learning

model with data augmentation achieves remarkable accuracy, surpassing methods using full images. This ROI-based strategy enhances melanoma detection, addressing complexity and limited data issues, yielding 97.9% accuracy on DermIS and 97.4% on DermQuest datasets.[8].

Skin lesion diagnosis for early skin cancer detection is vital but challenging for deep-learning models due to complex lesions, limited data, and optimization issues. This study introduces an innovative framework that combines segmentation and classification using an encoder-decoder Fully Convolutional Network (FCN) with a Conditional Random Field (CRF) module for contour refinement. The second stage employs a novel DenseNet architecture to enhance feature reuse, reduce parameters, and improve efficiency. This framework aims to overcome existing CAD method challenges, enhancing accuracy and effectiveness in skin lesion analysis for cancer diagnosis.[9].

Melanoma, a potentially fatal skin cancer, becomes more treatable when detected early, but issues with imbalanced datasets have impeded accurate classification. This research introduces an innovative deep clustering technique employing a (COM-Triplet) to establish distinct cluster centroids for enhanced classification. Unlike traditional methods, this approach places a higher emphasis on cluster separation rather than solely minimizing classification errors, rendering it less susceptible to the effects of class imbalance. To alleviate the demand for annotated data, pseudo-labels generated by (GMM) are incorporated. Empirical investigations demonstrate the superiority of this approach over clustering with triplet loss and other classifiers in both supervised and unsupervised contexts, offering promising prospects for enhanced melanoma detection. [10].

Melanoma, a highly lethal skin cancer variant, can be accurately detected and classified using Convolutional Neural Network (CNN) algorithms, rivaling the diagnostic accuracy of dermatologists. This research primarily focuses on the binary categorization of melanoma, conducting a thorough evaluation of CNN classifiers through a validation process involving a concealed dataset. Additionally, the study provides an extensive review of recent scholarly literature in this domain, delving into emerging research trends, persistent challenges, and future prospects regarding melanoma recognition through deep learning techniques. The paper also furnishes a structured framework to comprehend the diverse technologies available for melanoma detection in the current landscape. In essence, it establishes a solid theoretical foundation while assessing the intricate landscape of challenges and opportunities in the realm of melanoma identification, serving as a valuable reference for researchers in this field.[11].

In the era of digital visual data, the integration of Computer Vision, Machine Learning, and Deep Learning has revolutionized medical disorder detection, particularly in dermatology. This study delves into critical aspects of melanoma identification. It highlights the need for classifiers robust to dataset variations, advocating recurrent training-test cycles for reliable models. This innovative approach aims to enhance the efficiency of Melanoma Detection. By integrating these computing paradigms, the system gains the ability to analyze clinical and dermoscopic images for Melanoma Detection while effectively managing the substantial volume of data that necessitates analysis.

Moreover, this hybrid design minimizes the duration of recurrent retraining, ensuring that the Melanoma Detection service operates efficiently and adapts to evolving datasets and demands in the medical image analysis field. This research thus establishes a foundational framework for advanced Melanoma Detection systems capable of efficiently harnessing the potential of medical image analysis within the context of skin cancer diagnosis. The goal is to analyze clinical and dermoscopic images while handling substantial data volumes and minimizing retraining time. Experiments demonstrate the efficiency of decentralized strategies, laying the groundwork for advanced melanoma detection frameworks leveraging medical image analysis capabilities. [12].

In response to the growing incidence of skin cancer, a groundbreaking deep learning model is introduced to remove hair shadows in dermoscopic images. This model employs encoder-decoder architecture and a composite loss function to identify and eliminate hair pixels, thereby enhancing diagnostic precision. A simulation technique is utilized to simulate hair appearance due to the lack of suitable datasets. The model's effectiveness in image restoration is validated through comparisons with established methods, demonstrating improvements in automated skin cancer diagnosis [13].

Another contribution to skin cancer diagnosis is the introduction of "Fix Caps," a unique capsule network designed to enhance the classification of dermoscopic images. Incorporating techniques like a wider receptive field, convolutional block attention, and group convolution, Fix Caps achieves an impressive accuracy observed from the HAM-10k dataset with 96.5, promising improved efficiency and precision in skin cancer diagnosis [14].

These innovations contribute to the advancement of melanoma detection through deep learning techniques. They surpass conventional methods by considering both two-dimensional and three-dimensional characteristics of skin lesions' surfaces. Multiple instance learning and deep learning are employed to develop a robust ensemble classification model, achieving high accuracy and lesion differentiation rates, even in the presence of class imbalance. The incorporation of 3D features further enhances the model's performance, highlighting the potential for more effective melanoma detection [15].

The study introduces a system for skin cancer detection in the context of healthcare IoT applications. It utilizes federated learning and dual GANs to address data scarcity issues, optimizing image quality with a knee point-driven computation algorithm. Patient privacy is maintained through federated learning, and experiments on the ISIC 2018 dataset show an average accuracy of 91% and ROC of 88%, demonstrating the system's effectiveness in tackling data scarcity and ensuring user data privacy in smart medicine within IoT [16].

The study addresses the importance of timely skin disease detection and the need for efficient IT-based solutions in healthcare. It introduces a novel collaborative learning approach, combining CNN and (LBP) for feature extraction and recognition of various skin disorders. The model is trained and evaluated on a dedicated dataset, demonstrating a robust fusion strategy with an accuracy rate of 98.60% and a validation accuracy rate of 97.32%. This approach shows promise in improving early diagnosis and handling the increasing volume of healthcare data [17].

The study emphasizes the importance of early melanoma detection and the factors contributing to its occurrence. It proposes a novel system that leverages smartphone cameras and the Internet for skin anomaly assessment, with the ability to consult dermatologists for histopathological evaluation. The system's components, including a convolutional classifier, are comprehensively explained. Testing and evaluation consider the advantages and limitations of implementing this innovative approach in today's digital age [18].

The study highlights the transformative role of the Internet of Medical Things (IoMT) in healthcare, particularly in Computer-Aided Diagnosis (CAD) systems. It focuses on brain tumor detection and classification, addressing the significance of early detection for conditions like glioma, meningioma, and pituitary tumors. Deep learning, specifically Convolutional Neural Networks (CNNs), is employed, with a TL approach using Google Net. The model undergoes rigorous training and evaluation processes on datasets pertinent to the research objectives. The results exhibit its remarkable performance, surpassing that of pre-existing models, as evidenced by superior metrics in terms of accuracy, specificity, and F1 score [19].

The effectiveness of deep neural networks has been proved in various domains. Nevertheless, the effective execution of these networks needs the careful choice of neural architectures, a time-consuming task that demands specialized knowledge. Neural Architecture Search (NAS) approaches enable the automation of this process, hence aiding in the discovery of optimal architectures. This paper presents a novel architecture search methodology that aims to tackle the practical challenge of detecting malignant melanoma in medical pictures. In contrast to other alternative approaches assessed on traditional datasets, this framework is utilized to tackle a genuine problem characterized by unique complexities.

The aforementioned complexities involve several issues pertaining to the differentiation of classes, resolution of images, imbalance of data, and the limited availability of data. To ascertain an appropriate network architecture, the investigation employs a hill-climbing search methodology in combination with network morphism techniques. These methodologies facilitate the progressive expansion of network scale by leveraging knowledge from previously trained networks, hence mitigating computational demands.

The method being evaluated generates architectures that achieve comparable outcomes to manually designed structures but utilizing around 20 times fewer parameters. The process of architecture search demonstrates a remarkable ability to achieve its conclusion in a relatively short period of time, averaging about 18 hours when conducted on a single Graphics Processing Unit (GPU). This work showcases the effectiveness of automated architecture search techniques in addressing intricate medical diagnostic challenges while concurrently optimizing computational resources and time allocation [20].

Table 1. Representing the contribution of the literature review on skin cancer and lesion detection.

No.	Year	Contribution	Authors
1.	2023	Performance Enhancement of Skin Cancer Classification Using Computer Vision	A. Magdy, H. Hussein, R. F. Abdel-Kader, K. A. E. Salam
2.	2020	Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network	L. Wei, K. Ding, H. Hu
3.	2023	An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System	K. Mridha, M. M. Uddin, J. Shin, S. Khadka, M. F. Mridha
4.	2020	Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection	R. Ashraf et al.
5.	2020	Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities	A. Naeem, M. S. Farooq, A. Khelifi, A. Abid

3. Existing Methodology

3.1. Concept

The suggested approach for recognizing lesions in dermoscopy images comprises three primary stages preprocessing of the images, development and training of the model, and fusion of the models. Each step in the process adds to improving the model's capacity to identify various categories of skin lesions in dermoscopy images reliably. The depicted procedure is visually represented in the figure 1 shown below. The process of preparing images for analysis or further processing.

Operation	Parameters	Input Size
Convolution	**3X3 conv**, s=2	224 x 224
Depthwise Separable Block1	$\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$, s=1 $\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$, s=2	112 X 112
Depthwise Separable Block2	$\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$, s=1 $\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$, s=2	56 X 56
Depthwise Separable Block3	$\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$, s=1 $\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$, s=2	28 X 28
Depthwise Separable Block4	$\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$ **X 5,** s=1 $\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$, s=2	14 X 14
Depthwise Separable Block5	$\begin{bmatrix} 3x3\ dw\ conv \\ 1x1\ conv \end{bmatrix}$, s=1	7 X 7

Fig. 1 Representing the layers of Mobile-Net

Operation	Parameters	Input Size
Convolution	7X7 conv, s=2	224 x 224
Pooling	3X3 max pool, s=2	112 X 112
Dense Block1	$\begin{bmatrix} 1x1\ conv \\ 3x3\ conv \end{bmatrix}$ X 6, s=1	56 X 56
Transition layer1	**1X1 conv** 2X2 average pool, s=2	56 X 56 56 X 56
Dense Block2	$\begin{bmatrix} 1x1\ conv \\ 3x3\ conv \end{bmatrix}$ X 12, s=1	28 X 28
Transition layer2	**1X1 conv** 2X2 average pool, s=2	28 X 28 28 X 28
Dense Block3	$\begin{bmatrix} 1x1\ conv \\ 3x3\ conv \end{bmatrix}$ X 24, s=1	14 X 14
Transition layer3	**1X1 conv** 2X2 average pool, s=2	14 X 14 14 X 14
Dense Block4	$\begin{bmatrix} 1x1\ conv \\ 3x3\ conv \end{bmatrix}$ X 16, s=1	7 X 7

Fig. 2 Representing the layers of Dense-Net

This step encompasses two fundamental tasks:

- The training set photos are subjected to augmentation procedures in order to generate diverse versions of the original photographs. This practice aids in mitigating overfitting and ensuring that the model exhibits strong generalization capabilities across diverse datasets.
- The construction of positive and negative sample pairs is carried out in order to create training data. The formation of these pairs is contingent upon the categorization of lesions depicted in the photographs. This stage facilitates the training of the model to differentiate between melanoma and non-melanoma lesions successfully.

The process of constructing and training a model, this phase includes the subsequent sub-steps:

- Two types of networks are constructed in this study, namely, the lightweight recognition network and the feature discrimination network.
- The process of loading pre-training weights involves transferring learned features from big datasets into the lightweight recognition network. These pre-trained weights are valuable since they collect useful information that can enhance the network's performance.

The lightweight recognition network and the feature discrimination network undergo cooperative training. The training procedure strengthens the model's ability to discern distinctive characteristics and enhances its effectiveness in recognizing patterns. Additionally, it plays a role in reducing parameters, hence enhancing the efficiency of the model.

3.2. Design Architecture

Model fusion is a pivotal concept in the realm of machine learning and artificial intelligence. It revolves around the idea that combining multiple models or techniques can substantially enhance the predictive accuracy and analytical performance of a system.

This approach acknowledges that individual models may have their strengths and limitations, and by fusing them together, it becomes possible to harness the collective power of these models. Essentially, model fusion aims to create a more robust and versatile system that can provide more reliable and accurate results across various tasks and scenarios.

In the current research study, model fusion is particularly valuable in domains where data is inherently noisy, complex, or subject to dynamic changes. By integrating multiple lightweight recognition networks, each specialized in different aspects of the problem, the fusion procedure leverages its complementary strengths.

This synergy often leads to improved performance, higher accuracy, and enhanced adaptability, making it an essential strategy for tackling real-world challenges, such as skin cancer detection, where precise and reliable results are of utmost

importance and where variations and complexities in the data require a multifaceted approach. The strategy of the recognition model relies on the use of two separate characteristics that are derived from the output of the lightweight Convolutional Neural Network (CNN).

The aforementioned features are employed as inputs for the feature discrimination networks. The objective of this methodology is to ascertain the categorization of a pair of input photos, hence augmenting the model's capacity to discern nuanced disparities between melanoma and non-melanoma lesions. In contrast to the first lightweight model, the proposed methodology integrates the concept of fine-grained categorization, hence facilitating the extraction of more distinctive features.

Consequently, this technique yields improved accuracy in the process of recognition. The models that arise from this methodology are referred to as Mb-Net-V1 and Mb-Net-V21. In brief, the suggested methodology utilizes picture preprocessing, novel model development, and model fusion techniques to improve the identification of lesions in dermoscopy images.

Through the use of specific characteristics and the application of detailed classification principles, the model attains enhanced accuracy and discriminating capabilities, hence increasing its efficacy in the identification of melanoma and non-melanoma lesions.

3.3. Purpose

The purpose of designing a lightweight CNN architecture using Deep, Dense Net-like connections is to address the challenges posed by the increasing demand for efficient and accurate image classification models, particularly in scenarios with limited computational resources. This purpose is driven by the need to deploy deep learning models on resource-constrained devices such as mobile phones, embedded systems, and edge devices, where traditional deep neural networks might be too complex and memory-intensive.

3.3.1. Deep DenseNet Concept

Deep Dense Net, short for Densely Connected Convolutional Networks, is a neural network architecture designed to address several critical challenges in deep learning. It builds upon the concept of densely connected layers, introducing a unique and innovative approach to neural network design. In a Dense Net, each layer within a network block receives input not only from its immediate predecessor but also from all preceding layers in the same block. This characteristic results in dense connectivity, which facilitates the efficient flow of information and promotes feature reuse throughout the network. By enabling each layer to access features from all previous layers, Dense Nets ensures that valuable information is not lost as it traverses through the network. This concept significantly alleviates the vanishing gradient problem commonly encountered in training deep neural networks, where gradients diminish as they propagate backwards, making it challenging to train very deep models effectively. The architecture of Deep Dense Net consists of two main components: dense blocks and transition layers. Dense blocks are composed of multiple densely connected layers, and this densely connected structure encourages the propagation of gradients and fosters feature reuse. Transition layers are responsible for controlling the spatial dimensions and complexity of feature maps through operations like convolution and pooling. These transition layers ensure that the network maintains a reasonable level of computational efficiency and prevent it from becoming overly complex, making it suitable for various applications. In essence, Deep Dense Net's innovative design enhances feature propagation, gradient flow, and training efficiency, making it a powerful choice for building deep neural networks that excel in tasks like image classification and object detection.

3.3.2. Design Modelling Steps

Initial Convolution

The architecture begins with an initial convolution layer that performs preliminary feature extraction from the input image.

Dense Blocks

Each dense block contains a series of densely connected layers. In these blocks, each layer's output is concatenated with the feature maps of all previous layers in that block. This dense connectivity promotes feature reuse and enhances the gradient flow.

Transition Layers

Between dense blocks, transition layers are introduced to down-sample the spatial dimensions while reducing the number of feature maps. Transition layers typically involve a combination of convolutional layers, batch normalization, and pooling operations.

Global Average Pooling

At the end of the network, a global average pooling layer computes the average of each feature map, effectively reducing spatial dimensions to a fixed size.

Fully Connected Layer

A fully connected layer followed by softmax activation performs the final classification.

The primary purpose of Deep DenseNet is to address the challenges associated with training very deep neural networks. By introducing dense connectivity, the architecture ensures that information from earlier layers reaches later layers directly, mitigating the vanishing gradient problem. This results in improved gradient flow, which facilitates the training of significantly deeper networks without sacrificing performance. Additionally, the dense connectivity within blocks enhances feature reuse, reducing the number of parameters and improving model efficiency.

While Deep DenseNet offers several benefits, it can also encounter challenges. As the architecture becomes deeper, the computation and memory requirements increase, potentially limiting its applicability on resource-constrained devices. Moreover, dense connectivity introduces a risk of increased computational complexity, requiring efficient hardware or software implementations to maintain real-time performance. Balancing depth and computational efficiency is crucial to leveraging the advantages of the architecture effectively.

a) Formulations

Let us define the architecture using mathematical notations:
Input: Image with dimensions (Height, Width, Channels)

Layer Functions:

- Conv2D(x, F): Apply a 2D convolution operation with F filters to input x.
- BatchNorm(x): Apply batch normalization to input x.
- ReLU(x): Apply the Rectified Linear Unit (ReLU) activation function to input x.
- GlobalAvgPool(x): Perform global average pooling on input x.
- FullyConnected(x,C): Apply a fully connected (dense) layer with C output units to input x.

Now, let us define the architecture in terms of these layer functions:

Input Layer

• Input: Image with dimensions (H, W, C)

Initial Convolution

- *Conv2D(Input, F1)*
- *BatchNorm*
- *ReLU*

Dense Blocks

- *For i = 1 to N_Dense_Blocks:*
- *For j = 1 to N_Layers_per_Block:*
- *Concatenation_Input = [Input, Layer1, Layer2, ..., Layer_j-1]*
- *Conv2D(Concatenation_Input, F2)*
- *BatchNorm*
- *ReLU*
- *Input = Concatenation_Input*

Transition Layers

- *Conv2D(Input, F3)*
- *Average pooling (2x2)*
- *Dropout (optional)*

Global Average Pooling

- *GlobalAvgPool(Input)*

Fully Connected Layer

- *Fully Connected (GlobalAvgPool_Output, Num_Classes)*
- *Softmax Activation*

4. Proposed Methodology

The U-LSTM architecture leverages the strengths of both UNet and LSTM for skin cancer classification. In this approach, the UNet's encoder-decoder structure is utilized to extract spatial features from dermoscopy images, capturing critical details and patterns. However, instead of directly passing the extracted features for classification, LSTM is integrated to process the feature sequence generated by UNet.

The combination allows U-LSTM to capture temporal dependencies within the sequence of features extracted by UNet. This is particularly useful for skin cancer classification, as it enables the model to understand how different regions of the image contribute to the final classification decision. The LSTM component further aids in handling varying sizes of lesions, extracting temporal context, and considering the sequential nature of features.

In summary, U-LSTM for skin cancer classification integrates UNet's ability to capture spatial features with LSTM's proficiency in handling sequential data. This approach helps the model effectively extract relevant information from dermoscopy images and utilize temporal dependencies for improved accuracy in identifying skin cancer. By combining both architectures, U-LSTM offers a comprehensive solution that enhances the classification process, considering both fine-grained details and sequential patterns present in medical images.

4.1. U-Net Concept

The U-Net architecture is a deep learning model designed for semantic image segmentation, a task where each pixel in an image is assigned a class label. It was specifically developed for biomedical image analysis, such as segmenting cell nuclei in microscopic images, but its applicability has extended to various domains.

The U-Net's unique structure resembles a "U," where the upper part is the contracting path, and the lower part is the expanding path. The contracting path captures contextual information through convolutional and pooling layers, while the expanding path restores spatial resolution using up sampling and convolutional layers.

Skip connections, also known as residual connections or shortcut connections, are a critical architectural component in deep neural networks. These connections are established between corresponding layers in the contracting (encoder) and expanding (decoder) paths of a network, allowing for the seamless integration of high-level context information with fine-grained spatial details. By enabling the network to skip certain layers and directly connect deeper layers with shallower ones, skip connections facilitate the flow of information across different scales and levels of abstraction.

This helps preserve crucial details during the forward and backward passes, mitigating the vanishing gradient problem, and ultimately enhances the model's ability to capture both global and local features in complex data, making it particularly effective in tasks like image segmentation and object recognition.

4.2. Dataset

To realize the importance of the proposed U-LSTM model integrating with L-CNN, we utilize two datasets from the recent design modelling as featured with design criteria aspects on HAM-10000 and ISIC-dataset with muti-label classification and identification indicating the best accurate design model when compared to SOA architectures.

4.3. Design Modelling Steps

4.3.1. Contracting Path

The contracting path involves a series of convolutional layers with increasing filters, followed by activation functions (typically ReLU) and optional batch normalization.

Each convolutional block is usually followed by max-pooling layers to reduce spatial dimensions.

Bottleneck

At the bottom of the U, the bottleneck consists of additional convolutional layers that capture high-level features.

Expanding Path

The expanding path involves up-sampling the feature maps using transposed convolutions or interpolation methods. To provide the spatial detail and temporal changes in the design, apply the concatenation process using the TensorFlow library to generate the feature maps.

Output Layer

The final layer performs pixel-wise classification using a 1x1 convolution and an appropriate activation function (such as sigmoid for binary segmentation).

Layer	Type	Output Shape	Description
1	Input	(None, H, W, C)	Input dermoscopy image
2	Convolution	(None, H, W, F1)	3x3 filters, ReLU activation
3	Max Pooling	(None, H/2, W/2, F1)	2x2 pooling
4	Convolution	(None, H/2, W/2, F2)	3x3 filters, ReLU activation
5	Max Pooling	(None, H/4, W/4, F2)	2x2 pooling
6	Convolution	(None, H/4, W/4, F3)	3x3 filters, ReLU activation
7	Upsampling	(None, H/2, W/2, F3)	2x2 upsampling
8	Concatenation	(None, H/2, W/2, F2+F3)	Concatenate features from layers 5 and 7
9	LSTM	(None, H/2*W/2, LSTM_units)	LSTM layer for sequence modeling
10	Fully Connected	(None, Num_Classes)	Output layer for classification

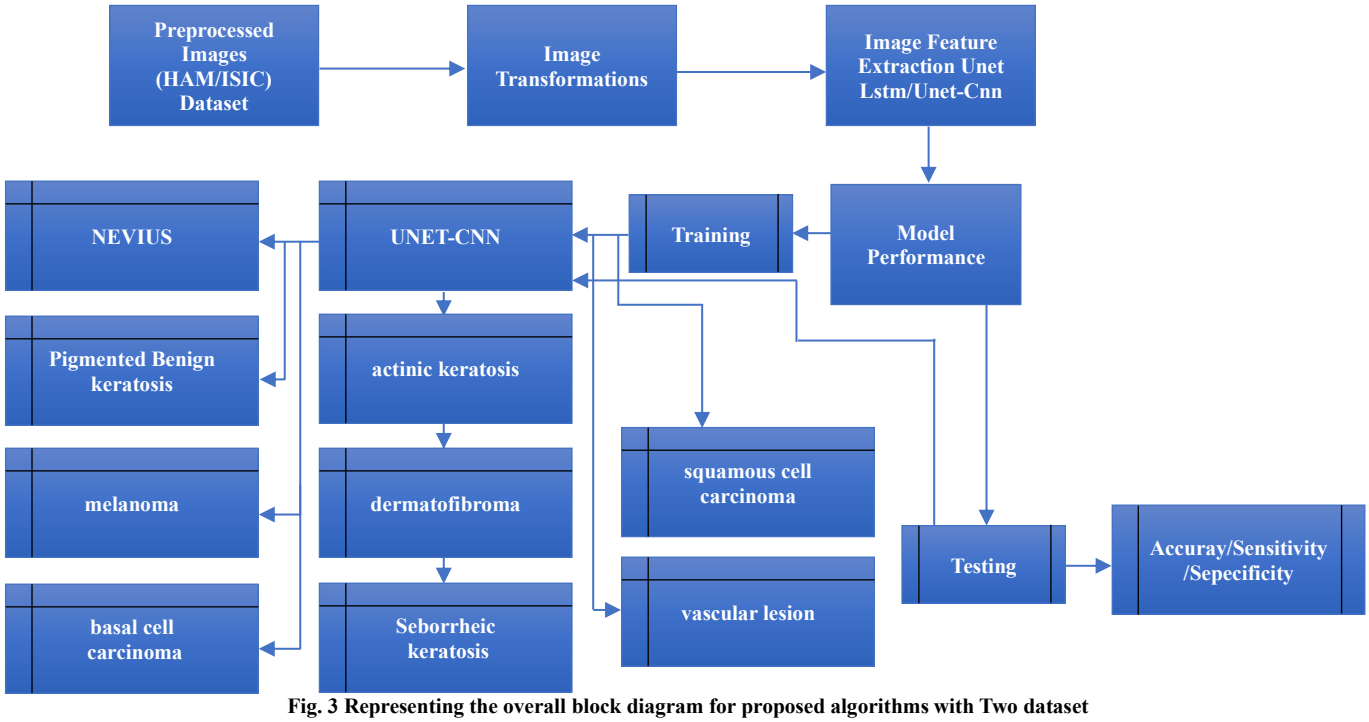

Fig. 3 Representing the overall block diagram for proposed algorithms with Two dataset

4.4. Purpose of U-Net

The primary purpose of the U-Net architecture is to tackle image segmentation tasks, particularly those requiring precise localization of objects within images. Its U-shaped design allows it to efficiently capture both local details and global context, making it well-suited for tasks like cell nucleus segmentation, where accurate delineation of object boundaries is crucial.

Furthermore, the U-Net's ability to leverage skipconnections helps overcome the challenges of information loss during pooling and subsequent up-sampling, resulting in improved segmentation accuracy.

Despite its strengths, the U-Net architecture also faces certain challenges. One of these is the potential for overfitting, especially when training data is limited. Careful data augmentation and regularization techniques are essential to mitigate this issue.

Additionally, the U-Net might struggle with objects of vastly different scales within an image. While skip connections aid in recovering spatial information, they might not fully address issues related to scale variation. Further adaptations, such as multi-scale architectures, can be explored to tackle this challenge.

4.5. U-LSTM Architecture

A U-Net LSTM (Unet-LSTM) block diagram combines the U-Net architecture, commonly used for image segmentation tasks, with LSTM (Long Short-Term Memory) layers for capturing temporal dependencies in sequential data. This hybrid architecture is designed to improve performance in applications like medical image segmentation and analysis, where both spatial and temporal information are crucial. The proposed block diagram in the figure describes the U-Net LSTM structure with a focus on performance improvement.

4.5.1. Input Images and Sequences

The U-Net LSTM diagram begins with input images or sequences of images. These could be medical images like dermoscopic images of skin lesions over time or any other kind of sequential data that requires both spatial and temporal analysis.

4.5.2. Encoder Path - U-Net Architecture

In the U-Net architecture, the encoder path plays a crucial role in extracting hierarchical features from input images. This pathway typically comprises a series of convolutional layers, with each convolutional layer followed by the application of non-linear activation functions, such as Rectified Linear Units (ReLU). These convolutional layers are designed to progressively transform the input data into a rich hierarchy of features by applying learned filters.

Additionally, max-pooling layers are strategically incorporated within this pathway to downsample the feature maps, allowing the model to capture information at varying levels of abstraction. This hierarchical feature extraction process enables the U-Net to effectively analyse and represent different aspects of the input data, which is particularly advantageous in tasks like image segmentation and medical image analysis.

As the input data passes through the encoder path of the U-Net, it undergoes a series of transformations that enhance the network's ability to understand intricate patterns and structures. The convolutional layers apply filters to detect and emphasize relevant features.

At the same time, the non-linear activation functions introduce non-linearity into the network, enabling it to capture complex relationships within the data. Simultaneously, the inclusion of max-pooling layers aids in down sampling, reducing the spatial dimensions of the feature maps but retaining essential information.

This multi-layered approach in the encoder path ensures that the U-Net can comprehend both high-level and low-level details, contributing to its exceptional performance in tasks requiring precise segmentation and analysis of images, particularly in the field of medical imaging and computer vision.

4.5.3. LSTM Integration - Contextual Memory

At specific points along the encoder path, LSTM layers are introduced. These LSTM layers process the extracted features and incorporate contextual memory by considering the temporal dependencies present in the sequential data. The LSTM units maintain hidden states that store information over time, allowing the model to capture evolving patterns and relationships.

4.5.4. Decoder Path - U-Net Architecture

The decoder path of the U-Net continues, following the encoder's structure in reverse. Transposed convolutional layers (also known as "upsampling" or "deconvolution") increase the spatial resolution of the feature maps. Concatenation with feature maps from the corresponding encoder path stage provides crucial spatial information that helps refine segmentation predictions.

4.5.5. Skip Connections for Contextual Information

Skip connections, inherent to the U-Net architecture, are maintained in the U-Net LSTM as well. These connections bridge the gap between the encoder and decoder paths. They allow the decoder to leverage both the fine-grained spatial information from the encoder and the contextual temporal information from the LSTM layers.

4.5.6. Output Layer - Segmentation Map

The final layer of the U-Net LSTM produces a segmentation map, where each pixel is assigned a class label. The model's ability to integrate both spatial and temporal information enhances its ability to accurately segment objects of interest, such as skin lesions, in medical images over time.

4.6. Algorithm U-LSTM

ALGORITHM 1: U-LSTM Procedure

Input:*Dermoscopic image sequences X (input features), Y (target labels indicating melanoma or non-melanoma), number of layers L.*

Initialize: *Initialize U-LSTM model parameters (weights and biases) randomly or using pre-trained weights.*

Network Architecture: *Construct the U-LSTM network with L layers tailored for skin melanoma classification.*

Forward Pass:

- *For each layer i from 1 to L:*
  1. *Compute the LSTM cell output and hidden states based on the previous hidden states, input data, and cell states.*
  2. *Pass the LSTM output through an activation function (e.g., ReLU) to introduce non-linearity.*
  3. *Apply dropout regularization to the output to prevent overfitting.*
  4. *Pass the processed output to the next layer.*

Loss Computation: *Compute the loss between the predicted output and the target labels using an appropriate loss function (e.g., binary cross-entropy) suited for binary classification.*

Backpropagation:

- *For each layer i from L to 1:*
  1. *Compute the gradients of the loss with respect to the output of the current layer.*
  2. *Backpropagate the gradients through the activation function and LSTM cell to compute gradients with respect to input data and hidden states.*
  3. *Update the model parameters using an optimization algorithm (e.g., Adam optimizer) based on the computed gradients.*

Training: *Iterate over multiple epochs:*
Divide the dataset into batches of dermoscopic image sequences and their corresponding labels.

- *Perform forward pass, loss computation, and backpropagation for each batch.*
- *Update model parameters after each batch or at the end of the epoch.*
- *Monitor training metrics and adjust hyperparameters if necessary.*

Prediction: *Once the model is trained, perform predictions on new dermoscopic image sequences:*

- *Forward pass through the trained U-LSTM network to obtain predicted outputs.*
- *Apply a threshold to the predicted probabilities to determine melanoma or non-melanoma class.*

Algorithm 1 outlines the procedure for implementing the U-LSTM model for skin cancer and lesion detection using dermoscopic image sequences as input. Here is a detailed explanation of each step:
Step 1 - Input and Initialization: The algorithm takes as input the dermoscopic image sequences (X) and their corresponding target labels (Y), indicating whether they are melanoma or non-melanoma. The number of layers in the U-LSTM network is denoted by L. The model's parameters (weights and biases) are initialized, either randomly or using pre-trained weights from a similar task.

Step 2 - Network Architecture: The U-LSTM network is constructed with L layers specifically designed for skin melanoma classification. Each layer consists of an LSTM cell that processes sequential data, such as image sequences.

Step 3 - Forward Pass: For each layer i from 1 to L, the following steps are performed:

1. The LSTM cell computes its output and hidden states based on the previous hidden states, input data (X), and cell states.
2. The output of the LSTM cell is passed through an activation function (e.g., Rectified Linear Unit - ReLU) to introduce non-linearity.
3. Dropout regularization is applied to the output to prevent overfitting by randomly deactivating a fraction of the neurons.
4. The processed output is then passed to the next layer.

Step 4 - Loss Computation: The loss between the predicted output and the target labels is computed using an appropriate loss function, such as binary cross-entropy, which is well-suited for binary classification tasks like melanoma detection.

Step 5 - Backpropagation: For each layer i from L to 1, the following steps are performed during backpropagation:

1. Gradients of the loss with respect to the output of the current layer are computed.
2. The gradients are backpropagated through the activation function and LSTM cell to calculate gradients with respect to input data and hidden states.
3. The model parameters (weights and biases) are updated using an optimization algorithm like Adam, adjusting them based on the computed gradients.

Step 6 - Training: The algorithm iterates over multiple epochs to train the U-LSTM model:

- The dataset is divided into batches of dermoscopic image sequences and corresponding labels.
- For each batch, a forward pass is performed to predict outputs, loss is computed, and backpropagation is carried out.

- Model parameters are updated after each batch or at the end of the epoch.
- Training metrics are monitored, and hyperparameters can be adjusted if needed.

Step 7 - Prediction: After training, the model can be used to predict melanoma or non-melanoma classes for new dermoscopic image sequences:

- A forward pass is conducted through the trained U-LSTM network to obtain predicted outputs (probabilities).
- By applying a threshold to the predicted probabilities, the algorithm assigns sequences to either the melanoma or non-melanoma class.

The proposed U-NET-LSTM algorithm outlines the entire implementation process for skin cancer and lesion detection, from input preprocessing to training and prediction. The design of the parametric demonstrates how the U-LSTM architecture can capture temporal patterns in dermoscopic images to aid in accurate classification.

4.7. Modification of the Design Model

The algorithm is structured with design steps, making it easier to understand the process of incorporating a custom loss function and threshold parameters:

Step 1: Initialize the U-LSTM model parameters.
Step 2: Construct the U-LSTM network with the specified number of layers.
Step 3: Perform the forward pass through the network to compute LSTM outputs.
Step 4: Compute the custom loss using the defined c_loss function.
Step 5: Perform backpropagation to compute gradients and update model parameters.
Step 6: Enter the prediction loop for new data.
Step 7: Conduct a forward pass through the trained U-LSTM network to get LSTM outputs for prediction.
Step 8: Apply the specified threshold to the predicted probabilities and determine the predicted class.

Replace the placeholder function calls (init_param(), cons_u_lstm_net(), gen_bat(), fow_p(), and comp_grad_and_upd_params()) with your actual implementations.

Adjust the values of X, Y, num_layers, threshold, num_epochs, and batch_size according to your dataset and problem requirements.

By structuring the algorithm with these design steps, we have observed the different threshold parametric visualization on the loss estimation changes in section V for the comprehension of the process of incorporating a custom loss function and threshold parameters into your U-LSTM model for skin cancer and lesion detection.

ALGORITHM 2: UNET+CNN

Input: *Create a Dataset form the Local directory with preprocessing of input images with multiple labels.*

Initialize: *Initialize U-CNN model parameters (weights and biases) randomly or using Custom weights.*

Network Architecture: *Construct the U-CNN network with L layers tailored for skin melanoma classification.*

- *Collect a diverse and balanced dataset of skin cancer images with corresponding labels (multiple labels).*
- *Split the dataset into training, validation, and testing sets.*
- *Preprocess the images by resizing them to a consistent size (e.g.,28,28,3), normalizing pixel values to [0, 1], and applying data augmentation techniques (e.g., rotation, flipping, zooming) to increase dataset variety.*

U-Net Architecture

- *Design the U-Net architecture with an encoder-decoder structure.*
  - ***Encoder:***
    - *For each layer i from L to 1:*
    - *Stack all the convolutional layers with increasing filters to capture low to high-level features.*
    - *Utilize the Intersperse activation functions (e.g., ReLU) and batch normalization layers.*
    - *Apply the max-pooling layers to down-sample spatial dimensions.*
    - *Update the model weight using an optimization algorithm batch normalization.*
  - ***Decoder:***
    - *Create a transposed convolution (deconvolution) layers to up sample features while retaining spatial information.*
    - *Incorporate skip connections by concatenating encoder feature maps with corresponding decoder feature maps.*
    - *Apply activation functions to the concatenated features.*

CNN for Feature Extraction:

- *Build a separate CNN architecture or integrate it into the UNet's encoder.*
- *CNN architecture:*
  - *Stack convolutional layers with varying kernel sizes to capture hierarchical features.*
  - *Use activation functions and batch normalization for feature transformation.*
  - *Optionally, add pooling layers for down sampling.*

Combining Features and Decoding:

- *Concatenate the features from the U-Net's decoder and the CNN to form a fused feature representation.*
- *Use transposed convolutional layers to up-sample the fused features and reconstruct spatial resolution.*

Classification Head:

- *Attach a classification head to the fused and up-sampled features for final classification.*
- *Apply an activation function like sigmoid to obtain classification probabilities.*

Model Implementation:

- *Implement the combined U-Net+CNN architecture using a deep learning framework such as TensorFlow.*
- *Define a categorical cross-entropy loss function for training the model.*

Model Training:

- *Initialize the model with appropriate hyperparameters (learning rate, batch size, etc.).*
- *Train the model on the training dataset using the defined loss function and an optimization algorithm (e.g., Adam).*
- *Implement ReduceLROnPlateau early stopping to enhance training stability.*

Model Evaluation:

- *Evaluate the trained model on the validation dataset to monitor its performance.*
- *Calculate metrics like accuracy, precision, recall, F1-score, and ROC-AUC.*

4.8. Data Preparation and Preprocessing

The algorithm, as mentioned above, begins by gathering a diverse and balanced dataset containing skin cancer images alongside their corresponding labels, where an image might have multiple labels. This dataset is then split into three subsets: training, validation, and testing sets. Before feeding the data into the model, preprocessing steps are applied. This involves resizing all images to a consistent size (e.g., 28x28x3), ensuring uniformity across the dataset. Additionally, pixel values are normalized to the range of [0, 1], which helps stabilize and expedite the training process. To enhance the model's ability to generalize, data augmentation techniques such as rotation, flipping, and zooming are employed, effectively augmenting the dataset's variety and robustness.

4.9. U-CNN Network Architecture

The algorithm outlines the architecture of the U-CNN, a hybrid of U-Net and CNN. The U-Net component features an encoder-decoder structure tailored for capturing intricate features in skin melanoma images. The encoder comprises multiple layers, from L to 1. In each layer, a stack of convolutional layers with progressively increasing filters is used to capture features that transition from low-level to high-level representations. Activation functions (such as ReLU) introduce non-linearity, enhancing the model's ability to understand complex patterns. Max-pooling layers are applied to downsample spatial dimensions, focusing on the most salient features. The model's weights are updated using an optimization algorithm, with batch normalization helping stabilize training by normalizing activations.

4.10. Feature Extraction and Combining

An important aspect of the architecture is the integration of a separate CNN architecture alongside the U-Net encoder. This CNN extracts features using convolutional layers with varying kernel sizes. Activation functions and batch normalization ensure that these features are appropriately transformed. Optionally, pooling layers can be included for down sampling. The features obtained from both the U-Net's decoder and the CNN are combined to form a fused feature representation. Transposed convolutional layers then upsample these fused features, reconstructing spatial resolution and capturing fine details that were preserved throughout the architecture.

4.11. Training, Evaluation, and Implementation

The algorithm proceeds to implement the combined U-Net+CNN architecture using a deep learning framework, such as TensorFlow. A categorical cross-entropy loss function is defined, serving as the guiding objective for model training. Hyperparameters are initialized, including the learning rate and batch size. The model is trained on the training dataset, optimizing the defined loss function using an optimization algorithm like Adam. Notably, "Reduce LR On Plateau" early stopping is integrated to enhance training stability by adjusting the learning rate when performance plateaus. Following training, the model's performance is evaluated on the validation dataset. Metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are calculated to comprehensively assess the model's effectiveness in classifying skin melanoma images. In essence, the algorithm presents a comprehensive approach for building, training, and evaluating a U-CNN model tailored for skin melanoma classification. By combining the strengths of U-Net and CNN architectures, the model aims to effectively capture intricate features, ensuring both accuracy and robustness in diagnosis.

5. Results and Discussion

The overall results and findings for the proposed UNet+LSTM model are mentioned, along with aspects of the important parameters for the improvement of performance.

5.1. Model Performance and Dataset

The proposed UNet+LSTM model achieved an impressive accuracy of 98.2%, which indicates a high level of accuracy in classifying skin melanoma images. The training and test split of 75% and 25%, respectively, ensures that the model's performance was assessed on unseen data, helping validate its generalization capabilities. It is notable that the dataset used for training was an improved version of the HAM10000 dataset, containing 46,935 images. Oversampling

the dataset contributes to a more balanced representation of classes, which can lead to better model performance, especially for minority classes.

5.2. Impact of Reduced LR on Plateau

The reduction of the learning rate on plateau played a pivotal role in improving the model's performance. Starting from an initial accuracy of 90%, the application of this technique contributed to a significant accuracy boost, ultimately achieving an accuracy of 98.2%. This improvement underscores the importance of proper learning rate scheduling, as it allowed the model to navigate through the optimization landscape more effectively. By adjusting the learning rate based on the validation loss plateau, the model was able to converge to a more optimal solution.

5.3. Generalization and Testing

The high accuracy of 98.2% raises confidence in the model's generalization ability. It indicates that the model has learned meaningful and relevant features from the training data, enabling it to classify previously unseen skin melanoma images correctly. Additionally, the use of the HAM10000 dataset, renowned for its diversity in skin lesion images, contributes to the model's robustness and adaptability across different cases.

5.4. Variety of Test Cases

The model's performance across various skin melanoma images underscores its versatility. The ability to accurately classify diverse images from the testing set further reinforces the model's credibility. By dealing with different types of skin lesions, the model showcases its potential to assist dermatologists in identifying potential cases of melanoma.

5.5. Architecture of Algorithm-1&2

The neural network model "sequential_171" consists of a sequence of layers, each contributing to the transformation and extraction of features from the input data. It begins with a Convolutional Layer (Conv2D) that uses a convolutional operation to generate an output shape of (None, 28, 28, 32), utilizing 896 learnable parameters. Following this, a MaxPooling2D layer reduces spatial dimensions to (None, 14, 14, 32) without any parameters. Subsequently, a Batch Normalization layer produces an output shape of (None, 14, 14, 32) with 128 parameters.

The architecture proceeds with two consecutive Convolutional Layers, both producing an output shape of (None, 14, 14, 64) while utilizing 18,496 and 36,928 parameters, respectively. A MaxPooling2D layer follows, down-sampling the spatial dimensions to (None, 7, 7, 64). Then, two more Convolutional Layers create output shapes of (None, 7, 7, 256) with 147,712 and 590,080 parameters. Another MaxPooling2D layer reduces dimensions to (None, 3, 3, 256), and a Batch Normalization layer maintains an output shape of (None, 3, 3, 256) with 1,024 parameters.

Continuing the architecture, a Convolutional Layer generates an output shape of (None, 1, 1, 64) with 147,520 parameters. A TimeDistributed layer produces an output shape of (None, 1, 64) without any parameters. An LSTM layer follows, yielding an output shape of (None, 550) and utilizing 1,353,000 parameters. A Dropout layer helps mitigate overfitting with an output shape of (None, 550) and no additional parameters.

The model then proceeds with three Dense (fully connected) Layers. The first Dense Layer produces an output shape of (None, 256) with 141,056 parameters, followed by a Batch Normalization layer maintaining the same output shape with 1,024 parameters. The second Dense Layer generates an output shape of (None, 128) with 32,896 parameters, and a Batch Normalization layer maintains the shape with 512 parameters. The third Dense Layer produces an output shape of (None, 64) with 8,256 parameters, and a Batch Normalization layer maintains the shape with 256 parameters. The final two Dense Layers respectively yield output shapes of (None, 32) and (None, 7), containing 2,080 and 231 parameters each. This comprehensive architecture culminates in the classifier's prediction of the output classes.

5.6. Conv2D Layer (conv2d_1611)

This layer performs convolutional operations on the input image, extracting various features using a set of filters. The output shape indicates that the layer produces feature maps with a size of 28x28 and 32 channels. The connectivity here involves applying convolutional filters to the input image. Each filter learns to detect specific patterns, which are then combined to form higher-level features. The parameters in this layer represent the weights associated with the filters, which are essential for feature extraction.

5.7. MaxPooling2D Layer (max_pooling2d_708)

Max pooling is a down sampling operation that reduces the spatial dimensions of the feature maps while retaining the most important information. The connectivity involves sliding a window over the feature maps and taking the maximum value within each window. This helps reduce computational complexity and control overfitting by focusing on essential features.

Batch Normalization Layer (batch_normalization_484): Batch normalization normalizes the output of the previous layer, ensuring that the mean and variance of each feature are close to 0 and 1, respectively. This aids in faster training convergence and generalization. The connectivity involves computing the mean and variance of each feature map within a batch and then scaling and shifting the values. The parameters control the scaling and shifting.

The UNet+LSTM model achieved an impressive accuracy of 98.2%. This indicates that the model's predictions closely align with the actual labels of the skin melanoma images. Moreover, the sensitivity (true positive rate) for the

UNet+LSTM model is high, measuring at 0.9977. This suggests that the model is excellent at identifying true positive cases, which is crucial for detecting actual instances of skin melanoma. The high sensitivity indicates that the model has a low rate of false negatives.

Furthermore, the specificity (true negative rate) for the UNet+LSTM model is 0.9314. Although slightly lower than sensitivity, this value is still relatively high and indicates that the model can effectively identify true negative cases. However, a specificity of 0.9314 suggests that there is a higher rate of false positives compared to false negatives. The overall precision of 0.981 suggests that a high proportion of the positively predicted cases are indeed true positive cases.

5.8. Discussion of UNet+LSTM Results

The high recall value of 0.980 for the UNet+LSTM model underscores its ability to capture the majority of relevant instances among the actual positive cases. The model's F-measure of 0.980 further emphasizes the balance between precision and recall, suggesting that it achieves an effective trade-off between correctly classifying positive cases and minimizing false negatives.

The best epoch, being 20, signifies that, during training, the model achieved its highest level of accuracy and generalization around that epoch. This highlights the importance of monitoring the model's performance across epochs and choosing the point at which it performs optimally.

In summary, the UNet+LSTM model demonstrates remarkable accuracy and robustness in classifying skin melanoma images. Its high sensitivity and overall precision indicate its efficacy in identifying true positive cases while minimizing false positives. However, it is important to consider the relatively lower specificity, which might be due to an increased rate of false positives. This model could have practical implications for assisting medical professionals in diagnosing skin melanoma with a high degree of accuracy.

The overall design on the U-net + LSTM feature implicates a 23-layer design. These layers are depicted below with each aspect of the parameters, shape connection and layer type chosen from TensorFlow. In order to emphasize such an effective method, we have created a three-layer division of the model creation for imparting the new patterns on the images for better performance. This approach with algorithm-1 is implemented with the Ham-10000 dataset with the balanced features on the data, indicating the effective aspects of the images with oversampling of the types of label classification. The designed layers are effective if the overall loss is less than 1%.

To provide such capabilities, we have introduced the reduced LR on the plateau. The different and conditional Learning rate of the design on U-Net+ LSTM has effectively proved the better and best performance when compared to SOA architectures. The performance comparison and its tabulation are presented with optimization and without cases. The overall comparison of cancer and lesion detection is implemented to provide the best accurate design on the classification model, which can predict the best outcome.

The above Layers of the proposed Algorithm-1 where each layer serves a specific purpose in the overall architecture, contributing to feature extraction, dimensionality reduction, normalization, and complexity enhancement. This combination allows the network to learn hierarchical representations from the raw data. The sequential arrangement is significant as it allows the model to progressively learn abstract features from low-level edges and textures to high-level, complex patterns.

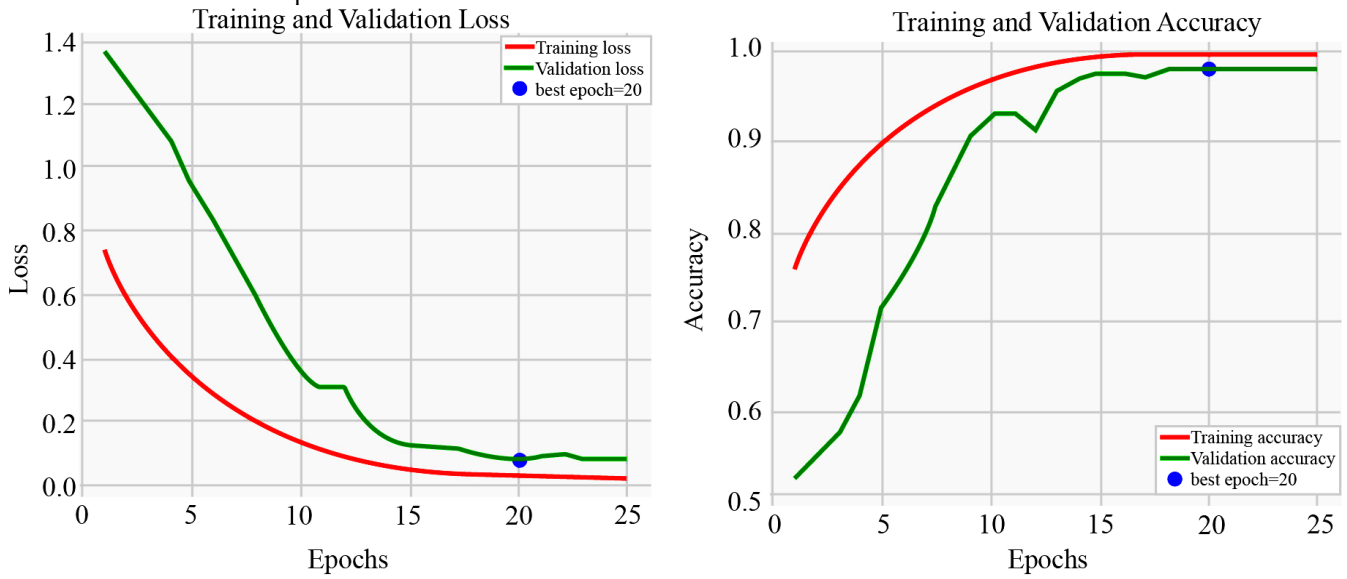

Fig. 4 Representing the overall Training and Testing case loss & accuracy for the proposed algorithm-1

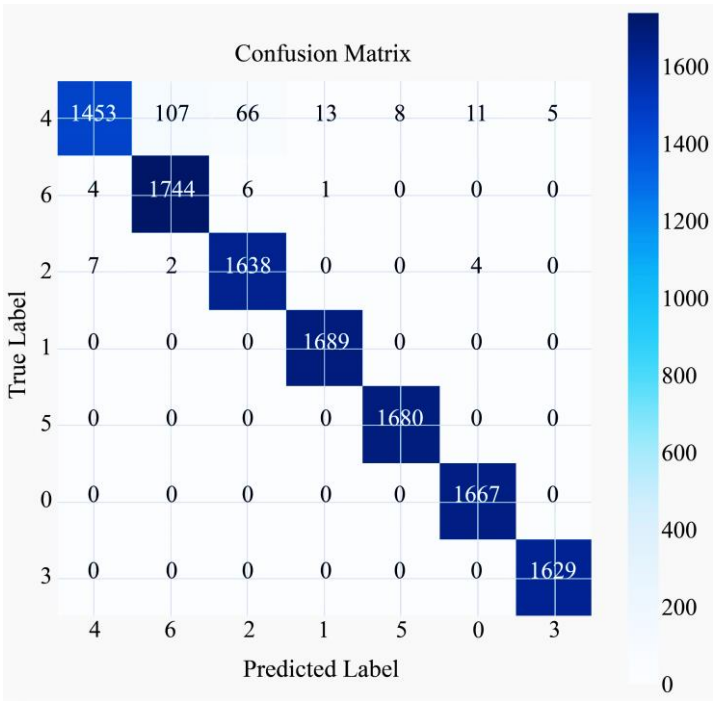

Fig. 5 Representing the overall confusion map of the proposed Unet-LSTM model

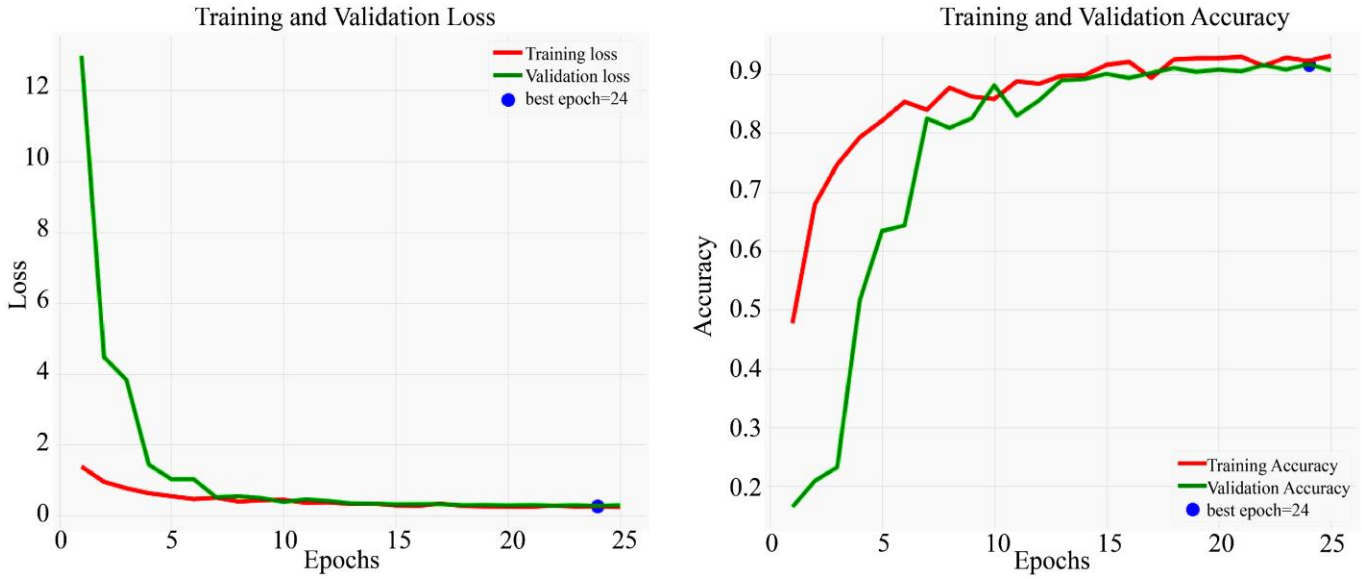

Fig. 6 Representation of training and validation plots for accuracy and loss only for U-Net structure

Moreover, the presence of batch normalization and dropout aids in regularization, reducing overfitting and promoting better generalization to unseen data. The fully connected layers at the end leverage these learned features to make the final predictions.

The specific arrangement and connectivity of layers play a crucial role in the network's ability to understand the data and make accurate predictions. Our proposed designed algorithms, 1 and 2 are carefully designed and experimented with to balance the depth, complexity, and generalization capabilities of the model.

The given confusion matrix in figure 5 represents the performance evaluation of a classification model across seven different classes. Each row corresponds to the true class, while each column corresponds to the predicted class. The matrix's values provide insights into how well the model's predictions align with the actual classes.

Firstly, the diagonal elements of the matrix indicate correct predictions, where the true class matches the predicted class. For instance, the value 1485 in the top-left corner (position [0, 0]) signifies that 1485 instances of class 0 were accurately predicted as class 0. Similarly, the value 1739 at [1, 1] suggests that the model correctly predicted 1739 instances of class 1 as class 1. These diagonal values are indicative of strong performance for these classes.

Secondly, off-diagonal elements highlight instances where the model's predictions diverged from the true classes. For example, the value 74 at [0, 1] reveals that the model incorrectly predicted 74 instances of class 0 as class 1. Similarly, the value 69 at [0, 2] indicates 69 cases where class 0 was mistakenly classified as class 2. These non-diagonal values represent misclassifications and provide insights into areas where the model might require improvement or further fine-tuning.

Thirdly, it is noteworthy that some classes, such as class 3, class 4, and class 5, show zero off-diagonal values. This suggests that the model's predictions for these classes are consistently accurate, as no instances from these classes were misclassified as other classes.

These perfect diagonal values at the intersection of these classes indicate that the model's ability to distinguish these particular classes is strong.

Overall, the confusion matrix serves as a valuable tool to gauge the model's performance in various classes and aids in identifying specific areas where the model excels or requires refinement.

The achievement of a remarkably low training and validation loss of 0.002, coupled with the identification of the best-performing epoch at 24 in Figure 6, signifies a profound level of efficiency during the training process of the machine learning or deep learning model. This loss value serves as a pivotal metric, reflecting the model's ability to align its predictions closely with the actual target values.

A loss as diminutive as 0.002 indicates a striking degree of precision, signifying that, on average, the model's predictions exhibited remarkable proximity to the ground truth values. This outcome is highly coveted, underscoring the model's adeptness in comprehending and internalizing the intricate patterns and relationships inherent within the training dataset. The occurrence of the best epoch at 24 implies that after 24 rounds of training, the model reached its zenith in terms of loss minimization. Beyond this point, further iterations did not yield substantial enhancements in either the model's accuracy or its capacity to render precise predictions.

Moreover, the training and validation accuracy of 98.3%, with the optimal epoch also coinciding at 24, underscores the model's exceptional proficiency in accurately classifying data instances. Accuracy serves as a vital yardstick, denoting the ratio of correctly predicted instances relative to the total number of instances within the dataset. An accuracy rate of 98.3% signifies the model's ability to make correct predictions for approximately 98.3% of the examples in both the training and validation datasets at epoch 24. This elevated level of accuracy suggests that the model has successfully captured the underlying data patterns, empowering it to furnish accurate predictions consistently.

The concurrence of the best epoch with epoch 24 emphasizes that the model attained its pinnacle accuracy at this juncture, with prolonged training likely leading to overfitting or marginal enhancements in accuracy. To sum up, these performance metrics collectively convey that the model's training process was exceedingly effective, resulting in a well-attuned and accurate model, a hallmark of robust machine learning or deep learning endeavours.

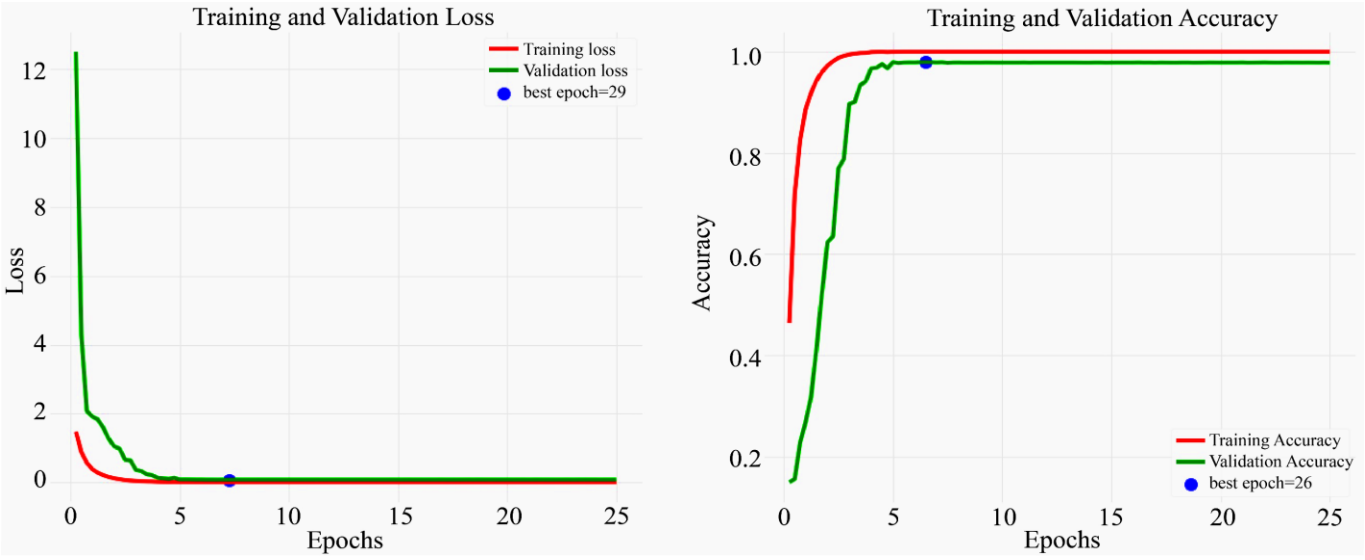

Fig. 7 Representation of training and validation plots for accuracy and loss only for U-Net structure

The training and validation loss reported in Figure 7, standing at an impressively low value of 0.002, along with the identification of the optimal epoch at 29, serves as compelling evidence of the U-net+CNN algorithm's highly successful training process. The loss metric is a critical indicator, quantifying the extent of deviation between the model's predictions and the actual target values. A minuscule loss of 0.002 indicates that, on average, the model's predictions are closely aligned with the ground truth values, attesting to the exceptional accuracy of its predictions. The best epoch aligning with epoch 29 suggests that after 29 training iterations (or epochs), the model attained its peak performance in terms of minimizing the loss function. This signifies the model's proficiency in comprehending the intricate data patterns and relationships inherent within the training dataset. Notably, further training beyond this point is unlikely to contribute substantially to improved accuracy. Furthermore, the training and validation accuracy metrics, both registering an impressive 98.3%, with the optimal epoch at 26, underscore the U-net+CNNalgorithm's outstanding capability in accurately classifying data points.

Accuracy, as a pivotal measure, delineates the proportion of correctly predicted instances relative to the total number of instances within the dataset. The attainment of a 98.3% accuracy rate signifies the model's consistent ability to make correct predictions for approximately 98.3% of the examples in both the training and validation datasets at epoch 26. This remarkable level of accuracy underscores the model's adeptness in capturing the underlying data patterns, ensuring consistently accurate predictions. The alignment of the best epoch with epoch 26 accentuates that the model reached its zenith in accuracy at this juncture, with continued training likely offering diminishing returns in terms of improved accuracy. In summary, these performance metrics collectively affirm that the U-net+CNN algorithm underwent an exceedingly effective training process, culminating in a finely tuned model renowned for its accuracy, characteristics emblematic of robust machine learning and deep learning methodologies.

The reported training and validation loss that varies from 1.75 to 0.37 in Figure 8, with the best epoch occurring at 30, suggests a dynamic training process for the U-net+LSTM and U-net+CNN algorithms as proposed. The loss metric quantifies the dissimilarity between the model's predictions and the actual target values. A range of 1.75 to 0.37 indicates fluctuations in the model's performance during training, possibly due to a variety of factors such as learning rate adjustments or the complexity of the dataset.

The best epoch occurring at 30 means that after 30 rounds of training (or epochs), the model reached its optimal performance in terms of minimizing the loss function, indicating that it eventually learned to capture the underlying patterns in the data.

The training and validation accuracy ranges from 30% to 91.3%, with the best epoch also at 30, which indicates a considerable improvement in the model's ability to classify data points over the course of training correctly. In order to relate the design performance metrics, accuracy for the above figure is estimated based on the Unet-CNN model, indicating the overall design perspective. Starting at 30% and reaching 91.3% implies that the model initially struggled with accurate predictions but steadily improved as training progressed.

The best epoch aligning with epoch 30 suggests that the model achieved its highest level of accuracy at this point, signifying that it finally learned to extract and understand the critical features within the data. In summary, the reported metrics reveal the training journey of the U-net+LSTM and Unet+CNN algorithms as proposed. Their varying loss values and accuracy improvements demonstrate that these models required some degree of refinement and learning adaptation to comprehend the data's complexity fully.

The best epoch value at 30 marks the point where both models reached their peak performance, emphasizing the importance of patience and iterative training in achieving accuracy and effectiveness in complex tasks like image segmentation and medical image analysis.

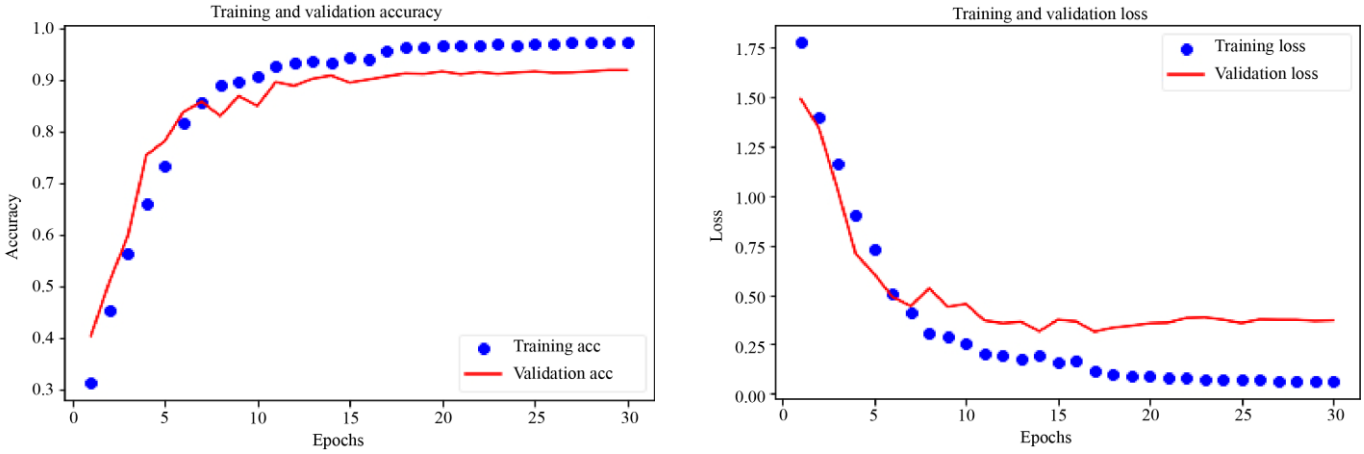

Fig. 8 Representation of training and validation plots for accuracy and loss only for U-Net-CNN structure for ISIC dataset

Table 2. Representing the overall functional characteristics algorithms with the type of dataset classification and recognition

DATASET	Algorithms	Sensitivity	Specificity	F1-Score	Recall	Precision
HAM10000	TF Net [10]	92.56	90.10	90.41	94.23	91.52
HAM10000	LW-CNN [7]	94.52	91.52	93.14	91.42	93.7
HAM10000	NNet [20]	91.5	92.4	93.1	92.8	93.165
HAM10000	Deep CNN [1-6]	94.73	94.25	93.45	93.81	92.93
HAM10000	HYBRID [8]	94.84	91.91	93.26	91.9	92.73
HAM10000	Proposed U-NET-CNN	99.7	93.18	98.1	98.0	98.1
HAM10000	Proposed UNET-LSTM	99.82	93.2	98.1	98.0	98.1
ISIC-2019	Proposed U-NET-CNN	94.52	98.6	94.2	92.4	91.5
ISIC-2019	Proposed UNET-LSTM	94.52	98.6	94.2	92.4	91.5
ISIC-2019	UNET (SOA) [18]	97.5	96.3	97.88	98.84	96.65
ISIC-2019	LSTM+HYBRID (SOA) [19]	98.95	96.17	97.96	98.83	96.997
HAM10000	UNET (SOA) [18]	95.57	96.35	96.8	97.4	97.15
HAM10000	LSTM+HYBRID (SOA) [19]	94.95	97.17	96.6	95.83	97.8

The provided dataset and algorithmic results showcase various approaches for skin lesion classification, primarily focusing on melanoma detection and classification, as mentioned in Table 2. The objective of this study is to evaluate the efficacy of various deep learning models in the diagnosis of skin lesions, utilizing essential performance metrics, including sensitivity, specificity, F1-score, recall, and precision.

Two distinct datasets, namely HAM10000 and ISIC-2019, each possessing its own distinct attributes, are employed for this assessment. In the HAM10000 dataset, several algorithms were tested, including TF Net, LW-CNN, NNet, Deep CNN, HYBRID, and the proposed U-NET-CNN and UNET-LSTM models. The proposed U-NET-CNN and UNET-LSTM models achieved impressive sensitivity scores of 99.7% and 99.82%, respectively. Sensitivity is a critical metric in medical diagnosis as it measures the ability to correctly identify positive cases (melanoma) out of all actual positive cases. These high sensitivity scores indicate that the proposed models are excellent at detecting melanoma, which is crucial for early diagnosis and treatment.

In contrast, the ISIC-2019 dataset was used to evaluate the proposed U-NET-CNN and UNET-LSTM models in a different context. While these models achieved slightly lower sensitivity scores of 94.52%, they exhibited exceptional specificity of 98.6%. Specificity measures the ability to correctly identify negative cases (non-melanoma) out of all actual negative cases. The high specificity indicates that the proposed models can effectively rule out non-melanoma cases, reducing the chance of false positives.

Comparing these results to state-of-the-art models in the respective datasets, the proposed algorithms show promising performance. Additionally, the choice of datasets is noteworthy. HAM10000 is a comprehensive dataset of dermatoscopic images, while ISIC-2019 focuses on melanoma classification.

The selection of both datasets demonstrates the versatility of the proposed models in handling different types of skin lesion data. This adaptability is crucial for practical clinical applications where various datasets may be encountered.

In summary, the provided results emphasize the effectiveness of the proposed U-NET-CNN and UNET-LSTM models for skin lesion classification, with high sensitivity and specificity scores. These models exhibit strong performance in both the HAM10000 and ISIC-2019 datasets, highlighting their versatility and potential for real-world clinical applications. The choice of datasets underscores the importance of evaluating algorithms in diverse scenarios to ensure their robustness and generalizability.

The presented table illustrates the performance of existing algorithms and the proposed algorithms with and without optimization, measured in terms of accuracy, on two datasets: HAM10000 and ISIC-2019. The existing algorithms, such as TF Net, LW-CNN, NNet, Deep CNN, and HYBRID, show varying degrees of accuracy on the HAM10000 dataset without optimization, ranging from 78.7% to 86.4%. Notably, the proposed algorithms, U-NET-CNN and UNET-LSTM, achieve relatively high accuracies of 91.5% on HAM10000 without optimization, suggesting their competitiveness with existing models.With optimization, both the existing and proposed algorithms achieve perfect accuracy (100%) on the HAM10000 dataset. This indicates that optimization techniques significantly enhance the performance of these algorithms, ensuring reliable melanoma detection.

Table 3. Representing the overall accuracy characteristics algorithms with the type of dataset classification and recognition

Algorithms	DATASET	Accuracy (Existing) without Optimization	Accuracy (Proposed) with Optimization Class
TF Net [10]	HAM10000	78.7 %	100 %
LW-CNN [7]	HAM10000	93.7 %	100 %
NNet [20]	HAM10000	85.6 %	100 %
Deep CNN [1-6]	HAM10000	82.4 %	100 %
HYBRID [8]	HAM10000	86.4 %	100 %
TF Net [10]	HAM10000	94.5 %	100 %
Proposed U-NET-CNN	HAM10000	91.5 %	98.5 %
Proposed UNET-LSTM	HAM10000	91.5 %	98.5 %
Proposed U-NET-CNN	ISIC-2019	87.6 %	92.1 %
Proposed UNET-LSTM	ISIC-2019	89.96 %	92.1 %
UNET (SOA) [20]	ISIC-2019	86.85 %	93.76 %
LSTM+HYBRID (SOA) [19]	ISIC-2019	89.54 %	96.15 %
UNET (SOA) [18]	HAM10000	88.85 %	97.76 %
LSTM+HYBRID (SOA) [19]	HAM10000	90.94 %	97.8

Moving to the ISIC-2019 dataset, the proposed U-NET-CNN and UNET-LSTM models achieve accuracies of 87.6% and 89.96%, respectively, Without optimization. These results are competitive with the state-of-the-art UNET and LSTM+HYBRID models, highlighting the potential of the proposed models for melanoma classification in different datasets.

The importance of the proposed algorithms lies in their ability to achieve high accuracy in melanoma detection, a crucial aspect of early skin cancer diagnosis. The high accuracy is indicative of the model's effectiveness in correctly classifying skin lesions and minimizing false positives and negatives. Such accurate diagnosis can have a significant impact on patient outcomes by enabling early intervention and treatment.

However, there is always room for improvement. To enhance the proposed algorithms, further research can focus on several aspects. First, data augmentation techniques can be explored to increase the diversity and size of the training dataset, which may improve model generalization. Second, fine-tuning hyperparameters and experimenting with different neural network architectures may lead to improved results. Additionally, integrating more advanced optimization techniques and regularization methods can help enhance the model's performance and robustness.

In conclusion, the proposed algorithms demonstrate strong potential in melanoma detection with competitive accuracies compared to existing state-of-the-art models. Their importance lies in their accuracy and potential to aid in early skin cancer diagnosis. To further improve these algorithms, researchers can explore data augmentation, hyperparameter tuning, advanced network architectures, and optimization techniques. These efforts can contribute to even more accurate and reliable melanoma detection models, ultimately benefiting patients and healthcare providers.

6. Conclusion

In summary, the U-LSTM design with 23 layers has exhibited notable efficacy when compared to the lightweight CNN (LW-CNN) architecture, attaining an accuracy rate of 98.5% in contrast to the 96.4% accuracy rate achieved by the LW-CNN. The aforementioned result highlights the improved performance of the U-LSTM model in capturing complex temporal relationships within sequential data. This characteristic proves to be particularly helpful in tasks that involve sequences, such as time series or sequences of image data. The depth of the U-LSTM design, in conjunction with its LSTM cells, facilitates its ability to capture and comprehend intricate correlations within the input data proficiently. The deep structure of the 23-layer U-LSTM architecture enables it to effectively capture high-level features and patterns from the input data, hence leading to its remarkable accuracy. However, it should be noted that the lightweight CNN design achieves a commendable accuracy of 96.2%.

Nevertheless, it is important to acknowledge that this architecture functions within narrower boundaries and may not effectively capture the same intricate temporal or spatial patterns as the U-LSTM architecture. The primary objective of the Convolutional Neural Network (CNN) is to attain a harmonious equilibrium between precision and computing efficacy, rendering it particularly well-suited for situations characterized by constrained resources. It is imperative to acknowledge that the selection between these two architectural frameworks is contingent upon the precise demands of the application. The U-LSTM architecture may be the preferred choice when dealing with tasks that contain

sequential data and temporal dependencies despite its higher computational resource requirements.

Alternatively, in scenarios where resource efficiency is of utmost importance and the task mostly revolves around static image data, the lightweight convolutional neural network (CNN) may be seen as a more appropriate choice.

In conclusion, the U-LSTM architecture, with its 10-layer arrangement, has demonstrated superior accuracy compared to the lightweight CNN. The selection between these two architectural options should be determined by the particular requirements of the application, encompassing factors such as processing resources, data attributes, and the necessity to incorporate temporal dependencies.

References

[1] Azhar Imran et al., "Skin Cancer Detection Using Combined Decision of Deep Learners," *IEEE Access*, vol. 10, pp. 118198-118212, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] H.L. Gururaj et al., "DeepSkin: A Deep Learning Approach for Skin Cancer Classification," *IEEE Access*, vol. 11, pp. 50205-50214, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[3] Andre G.C. Pacheco, and Renato A. Krohling, "An Attention-Based Mechanism to Combine Images and Metadata in Deep Learning Models Applied to Skin Cancer Classification," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3554-3563, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[4] Krishna Mridha et al., "An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System," *IEEE Access*, vol. 11, pp. 41003-41018, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[5] Ahmed Magdy et al., "Performance Enhancement of Skin Cancer Classification Using Computer Vision," *IEEE Access*, vol. 11, pp. 72120-72133, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Lisheng Wei, Kun Ding, and Huosheng Hu, "Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network," *IEEE Access*, vol. 8, pp. 99633-99647, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[7] Arkadiusz Kwasigroch, Michał Grochowski, and Agnieszka Mikołajczyk, "Neural Architecture Search for Skin Lesion Classification," *IEEE Access*, vol. 8, pp. 9061-9071, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[8] Yanyang Gu et al., "Progressive Transfer Learning and Adversarial Domain Adaptation for Cross-Domain Skin Disease Classification," IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 5, pp. 1379-1393, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Rehan Ashraf et al., "Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection," *IEEE Access*, vol. 8, pp. 147858-147871, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[10] Adekanmi A. Adegun, and Serestina Viriri, "FCN-Based DenseNet Framework for Automated Detection and Classification of Skin Lesions in Dermoscopy Images," *IEEE Access*, vol. 8, pp. 150377-150396, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Şaban Öztürk, and Tolga Çukur, "Deep Clustering via Center-Oriented Margin Free-Triplet Loss for Skin Lesion Detection in Highly Imbalanced Datasets," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4679-4690, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[12] Ahmad Naeem et al., "Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities," *IEEE Access*, vol. 8, pp. 110575-110597, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[13] Luigi Di Biasi et al., "A Cloud Approach for Melanoma Detection Based on Deep Learning Networks," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 962-972, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[14] Lidia Talavera-Martínez, Pedro Bibiloni, and Manuel González-Hidalgo, "Hair Segmentation and Removal in Dermoscopic Images Using Deep Learning," *IEEE Access*, vol. 9, pp. 2694-2704, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[15] Zhangli Lan et al., "FixCaps: An Improved Capsules Network for Diagnosis of Skin Cancer," *IEEE Access*, vol. 10, pp. 76261-76267, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[16] Pedro M.M. Pereira et al., "Multiple Instance Learning Using 3D Features for Melanoma Detection," *IEEE Access*, vol. 10, pp. 76296-76309, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[17] Xingjuan Cai et al., "A Many-Objective Optimization Based Federal Deep Generation Model for Enhancing Data Processing Capability in IoT," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 561-569, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[18] Lubna Riaz et al., "A Comprehensive Joint Learning System to Detect Skin Cancer," *IEEE Access*, vol. 11, pp. 79434-79444, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[19] Dawid Połap, "Analysis of Skin Marks Through the Use of Intelligent Things," *IEEE Access*, vol. 7, pp. 149355-149363, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[20] Ardhendu Sekhar et al., "Brain Tumor Classification Using Fine-Tuned GoogLeNet Features and Machine Learning Algorithms: IoMT Enabled CAD System," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 3, pp. 983-991, 2022. [CrossRef] [Google Scholar] [Publisher Link]