International Journal for Innovative Engineering and Management Research

PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL

www.ijiemr.org

COPY RIGHT

ELSEVIER

SSRN

IJIEMR Transactions, online available on 18th Apr 2024.

10.48047/IJIEMR/V13/ISSUE 04/29

Title **AUTOMATED MELANOMA DETECTION USING CONVOLUTIONAL NEURAL NETWORKS: A COMPARATIVE EVALUATION AND ARCHITECTURAL ANALYSIS**

Volume 13, ISSUE 04, Pages: 252-265

Paper Authors **G.Vishnu, S.Sai Raja, Ch.Sathwik, Ms.B.Vasundhara Devi**

International Journal for Innovative Engineering and Management Research

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

AUTOMATED MELANOMA DETECTION USING CONVOLUTIONAL NEURAL NETWORKS: A COMPARATIVE EVALUATION AND ARCHITECTURAL ANALYSIS

G.Vishnu[1]
Dept of CSE
SNIST,HYD
20311a0579@sreenidhi.edu.in

S.Sai Raja[2]
Dept of CSE
SNIST,HYD
20311a0594@sreenidhi.edu.in

Ch.Sathwik[3]
Dept of CSE
SNIST,HYD
20311a05b8@sreenidhi.edu.in

Ms.B.Vasundhara Devi[4]
Asst. Professor,
Dept of CSE
SNIST,HYD
vasundhara@sreenidhi.edu.in

Abstract:
The burgeoning prevalence of melanoma, an insidious variant of skin cancer, underscores the imperative for sophisticated early detection systems leveraging dermatoscopic images. This research advocates for an innovative two-tiered strategy employing Convolutional Neural Networks (CNNs) to automate melanoma detection. Initially, a Mask and Region-based CNN autonomously delineates and isolates the pertinent regions within dermatoscopic images. Subsequently, a ResNet152 architecture discriminates lesions as either "benign" or "malignant" with heightened accuracy and balanced precision, surmounting prior methodologies. This study conducts an exhaustive assessment of CNN models, centring on eminent architectures like VGG19, ResNet50V2, and ResNet101V2. By employing precision, recall, and F1-score as pivotal evaluation metrics, the research elucidates the discriminative efficacy of each CNN model. In-depth scrutiny of training and validation loss curves reveals underlying learning dynamics and mitigates looming overfitting concerns. By unravelling the intricate nuances of CNN performance in image classification tasks, this research navigates decision-making paradigms in selecting the most efficacious architectures for varied application domains. The ensuing insights serve as a pivotal cornerstone for researchers and practitioners embarking on the labyrinthine journey of CNNs in medical diagnostics. Furthermore, they facilitate the seamless design and optimization of robust deep learning models, poised to revolutionize the landscape of melanoma detection and prognosis. This paradigm shifts towards CNN-based automated detection systems holds promise for catalyzing early intervention and ameliorating patient outcomes in the realm of dermatological malignancies. As such, this research not only heralds a new era in medical diagnostics but also underscores the indispensability of interdisciplinary collaborations in combating the scourge of cancer.

Keywords: Melanoma Detection, Convolutional Neural Networks (CNNs), Deep Learning Architectures, ResNet152, VGG19, ResNet50V, Precision-Recall-F1 Score Analysis

I. INTRODUCTION:

Melanoma, a formidable adversary in the realm of oncology, stands as the most lethal manifestation of skin cancer, exhibiting a distressing surge in incidence rates over the past three decades. This alarming trend underscores the imperative for robust and expeditious diagnostic methodologies to counteract its deleterious impact on public health. The advent of deep learning methodologies, particularly Convolutional Neural Networks (CNNs), has ushered in a new era of precision medicine, offering unprecedented capabilities in medical image analysis and diagnosis. The intricate morphological features characteristic of melanoma necessitate a nuanced approach to diagnostic imaging, wherein dermatoscopic images serve as a primary modality for lesion visualization. However, the accurate identification and discrimination of malignant lesions from benign counterparts remain formidable

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

challenges, compounded by the inherent subjectivity and variability in human interpretation. Addressing this pressing need, the integration of deep learning algorithms holds promise in augmenting the diagnostic accuracy and efficiency of melanoma detection systems.

Motivated by the imperative for early intervention and improved patient outcomes, this study endeavours to present a sophisticated and meticulously engineered solution for automated melanoma detection leveraging state-of-the-art deep learning techniques. Central to our approach is a meticulously designed two-stage framework, comprising a Mask and Region-based CNN for precise localization and segmentation of lesions within dermatoscopic images, followed by a ResNet152 architecture for robust classification of lesions as either benign or malignant. The significance of this research lies not only in its potential to enhance the diagnostic accuracy and efficiency of melanoma detection but also in its broader implications for the paradigm shift towards personalized medicine and computer-aided diagnosis. By harnessing the power of deep learning methodologies and leveraging the wealth of information embedded within dermatoscopic images, our proposed framework seeks to transcend the limitations of traditional diagnostic modalities, offering clinicians a powerful tool for early detection and intervention in the battle against melanoma.

Through rigorous experimentation and validation against established benchmarks, we aim to demonstrate the superior performance and efficacy of our proposed approach, thus paving the way for its seamless integration into clinical practice. Ultimately, we envision our research contributing to the advancement of precision medicine, driving tangible improvements in patient outcomes and fostering a paradigm shift towards proactive healthcare strategies in the domain of melanoma management.

A. Scope of The Research

The scope of this research is multifaceted, encompassing several crucial components essential for the development and validation of an automated melanoma detection system. At its core, the study focuses on the creation of a novel two-stage framework leveraging deep learning techniques applied to dermatoscopic images. This framework consists of a Mask and Region-based CNN for precise lesion localization and segmentation, followed by a ResNet152 architecture for accurate lesion classification as benign or malignant. Optimization of these deep learning models is paramount, involving exploration of various architectural configurations, hyper parameter tuning, and utilization of transfer learning methods to enhance model performance. Furthermore, the research entails meticulous dataset preparation and annotation procedures. A comprehensive dataset of dermatoscopic images, spanning a diverse range of melanocytic lesions, is curated and pre-processed to ensure data quality. Careful attention is paid to representative sampling for training, validation, and testing phases. Evaluation of the proposed methodology involves the utilization of established performance metrics such as accuracy, sensitivity, specificity, and AUC-ROC. Comparative analysis with existing approaches provides insights into the superiority of the proposed framework.

Moreover, the research extends beyond algorithmic development to encompass clinical validation and practical implementation aspects. Collaboration with healthcare professionals facilitates the validation of the proposed system on independent datasets and its integration into clinical workflows. Ethical

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

considerations, including patient privacy, informed consent, and data security, are integral to the research process. Compliance with regulatory standards such as HIPAA and GDPR ensures ethical conduct and safeguards patient rights. Acknowledging inherent limitations and challenges in automated melanoma detection, the study also outlines potential future directions. These may include the incorporation of multimodal data fusion, ensemble learning techniques, and integration with telemedicine platforms for remote diagnosis. By delineating the scope of research across these key dimensions, this study aims to contribute significant advancements in the field of melanoma detection, ultimately leading to enhanced diagnostic accuracy, early intervention, and improved patient outcomes.

II. LITERATURE REVIEW

The literature examining melanoma detection through deep learning methodologies encompasses a nuanced and multifaceted exploration of strategies aimed at navigating the intricate landscape of early cancer diagnosis [1]. Across various scholarly endeavours, a profound emphasis has been placed on unravelling the intricate complexities inherent in discerning and classifying melanocytic lesions [2], thereby illuminating pathways towards enhanced diagnostic accuracy and clinical efficacy. Initially, the discourse within this domain revolved around the employment of conventional machine learning paradigms, often complemented by meticulously engineered handcrafted features. While these methodologies exhibited initial promise in delineating rudimentary patterns [3], they invariably encountered limitations when confronted with the intricate morphological intricacies and heterogeneity inherent in melanoma pathology. It was against this backdrop that the ascendancy of deep learning methodologies, particularly Convolutional Neural Networks (CNNs), emerged as a beacon of transformative potential.

Recent literature showcases an escalating trajectory of research endeavours championing CNN-based approaches [4] for melanoma detection, heralding a substantive departure from erstwhile methodologies. These endeavours, characterized by their embrace of CNN architectures, epitomize a quantum leap in the realm of diagnostic precision and computational robustness. Moreover, the advent of transfer learning methodologies has conferred an added dimension of efficacy[5], facilitating the seamless integration of pretrained CNN models with vast repositories of dermatoscopic imagery, thereby catalyzing a paradigm shift in diagnostic prowess. A notable evolution within contemporary research manifests in the adoption of a multi-stage approach to melanoma detection, wherein the complex interplay between segmentation and classification is meticulously orchestrated to achieve optimal diagnostic fidelity [6]. Pioneering techniques such as Mask R-CNN and U-Net have garnered acclaim for their capacity to meticulously delineate lesion boundaries, thereby furnishing a fertile foundation for subsequent classification endeavours [7]. By decoupling these pivotal stages, researchers endeavour to harness the synergistic potential inherent in nuanced image analysis paradigms [8], thereby engendering a nexus of heightened accuracy and computational efficiency.

Furthermore, there exists a discernible trend towards the fusion of disparate modalities, heralding a convergence of image-based features with contextual metadata and clinical parameters. These hybrid architectures [9], emblematic of a burgeoning interdisciplinary synergy, epitomize a holistic approach towards diagnostic refinement and clinicaldecision-making prowess [10]. However, the integration of divergent data streams

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

necessitates a judicious calibration of model interpretability and computational complexity, engendering a confluence of methodological rigor and clinical pragmatism [11]. In spite of the remarkable strides achieved, the literature underscores a panoply of persistent challenges imperilling the veracity and scalability of deep learning-based melanoma detection paradigms. Paramount among these challenges are the paucity of annotated datasets, the absence of standardized evaluation metrics, and the interpretability conundrum pervasive in complex neural architectures. Mitigating these challenges mandates a concerted confluence [12] of interdisciplinary collaboration, regulatory stewardship, and methodological ingenuity, thereby precipitating a transformative paradigm shift towards personalized diagnostic modalities and improved patient outcomes in the realm of melanoma management.

III. RESEARCH GAP

In the realm of melanoma detection employing deep learning methods, despite notable progress, several significant research gaps persist. Foremost among these is the limited availability of annotated datasets tailored specifically for melanoma detection[13], which undermines the robustness and generalizability of developed models. Moreover, the heterogeneity in lesion morphology poses a challenge, as existing models may struggle to accurately discern subtle variations in melanocytic lesions encountered in clinical practice. Another critical gap lies in the interpretability of deep learning models, as clinicians require transparent insights into model decision-making processes to foster trust and adoption. Standardization of evaluation metrics is also lacking, hindering meaningful comparison and benchmarking of different detection algorithms[14]. Additionally, ethical and regulatory considerations regarding patient privacy, data security, and algorithmic bias remain relatively underexplored. Finally, while many studies demonstrate promising results in controlled settings, the translation of deep learning models[15] into clinical practice necessitates rigorous real-world validation studies encompassing diverse patient populations and clinical environments. Addressing these gaps demands interdisciplinary collaboration, methodological innovation, and a concerted focus on real-world applicability to drive transformative advancements in melanoma diagnosis and management.

IV. RESEARCH OBJECTIVES:

1. Comparative Evaluation of CNN Architectures: The primary objective of this research is to conduct a thorough comparative evaluation of prominent Convolutional Neural Network (CNN) architectures, including VGG19, ResNet50V2, and ResNet101V2, for image classification tasks. By rigorously analysing the performance of these architectures across various metrics, such as precision, recall, and F1-score, the study aims to identify the strengths and weaknesses of each model in differentiating between distinct image categories.

2. Assessment of Model Generalization and Robustness: Another key objective is to assess the generalization and robustness of the CNN models by evaluating their performance on validation datasets. Through the analysis of precision, recall, and F1-score metrics, the research seeks to determine the models' ability to effectively classify unseen instances and mitigate overfitting tendencies, thereby ensuring their applicability in real-world scenarios beyond the training data.

3. Investigation of Learning Dynamics and Overfitting Mitigation: Additionally, the research aims to investigate the learning dynamics and potential overfitting phenomena exhibited by the CNN

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

architectures during training. By scrutinizing the trends in training and validation loss curves, the study seeks to elucidate the trade-offs between model complexity, training performance, and generalization capability. This objective entails identifying strategies to mitigate overfitting and enhance the models' ability to generalize to unseen data, thereby improving their overall effectiveness in image classification tasks.

V.PROJECT EXECUTION PHASES:

1. Data Collection and Pre-processing: The first phase involves the collection of dermatoscopic images encompassing a diverse range of melanocytic lesions. This may involve accessing publicly available datasets or collaborating with healthcare institutions to obtain clinical imagery. The collected data undergoes pre-processing steps such as resizing, normalization, and augmentation to ensure uniformity and enhance model training.

Fig1. Random images of the Dataset

2. Model Development and Training: In this phase, deep learning models for melanoma detection are developed and trained using the pre-processed dataset. Various architectures, including VGG, ResNet, and others, are explored, and hyper parameters are tuned to optimize model performance. Transfer learning techniques may be employed to leverage pretrained models and expedite convergence.

```
Model: "model_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_3 (InputLayer)        [(None, 224, 224, 3)]     0

 block1_conv1 (Conv2D)       (None, 224, 224, 64)      1792

 block1_conv2 (Conv2D)       (None, 224, 224, 64)      36928

 block1_pool (MaxPooling2D)  (None, 112, 112, 64)      0
```

```
 block2_conv1 (Conv2D)
(None, 112, 112, 128)      73856

 block2_conv2 (Conv2D)
(None, 112, 112, 128)      147584

 block2_pool (MaxPooling2D)
(None, 56, 56, 128)        0

 block3_conv1 (Conv2D)
(None, 56, 56, 256)        295168

 block3_conv2 (Conv2D)
(None, 56, 56, 256)        590080

 block3_conv3 (Conv2D)
(None, 56, 56, 256)        590080

 block3_conv4 (Conv2D)
(None, 56, 56, 256)        590080

 block3_pool (MaxPooling2D)
(None, 28, 28, 256)        0

 block4_conv1 (Conv2D)
(None, 28, 28, 512)        1180160

 block4_conv2 (Conv2D)
(None, 28, 28, 512)        2359808

 block4_conv3 (Conv2D)
(None, 28, 28, 512)        2359808

 block4_conv4 (Conv2D)
(None, 28, 28, 512)        2359808

 block4_pool (MaxPooling2D)
(None, 14, 14, 512)        0

 block5_conv1 (Conv2D)
(None, 14, 14, 512)        2359808

 block5_conv2 (Conv2D)
(None, 14, 14, 512)        2359808

 block5_conv3 (Conv2D)
(None, 14, 14, 512)        2359808

 block5_conv4 (Conv2D)
(None, 14, 14, 512)        2359808

 block5_pool (MaxPooling2D)
(None, 7, 7, 512)          0

 flatten_2 (Flatten)
(None, 25088)              0

 dense_6 (Dense)
(None, 2)                  50178

=================================
================================
Total params: 20,074,562
Trainable params: 50,178
Non-trainable params: 20,024,384
```

Fig2: Model training

3. Model Evaluation and Validation: Following model training, rigorous evaluation and validation procedures are conducted to assess the performance of the developed models. This involves partitioning the dataset into training, validation, and test sets, and measuring metrics such as accuracy, sensitivity, specificity, and area under the ROC curve (AUC-ROC). Comparative analysis with existing approaches provides insights into the superiority of the proposed framework.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

```
Epoch 1/30
50/50 [==============================] - 12s 217ms/step - loss: 0.7567 - accuracy: 0.6888 - val_loss: 0.4655 - val_accuracy: 0.7839
Epoch 2/30
50/50 [==============================] - 10s 209ms/step - loss: 0.5036 - accuracy: 0.7631 - val_loss: 0.4654 - val_accuracy: 0.7891
Epoch 3/30
50/50 [==============================] - 10s 207ms/step - loss: 0.3862 - accuracy: 0.8206 - val_loss: 0.3986 - val_accuracy: 0.8281
Epoch 4/30
50/50 [==============================] - 10s 204ms/step - loss: 0.3632 - accuracy: 0.8344 - val_loss: 0.4137 - val_accuracy: 0.8099
Epoch 5/30
50/50 [==============================] - 9s 187ms/step - loss: 0.3784 - accuracy: 0.8269 - val_loss: 0.4387 - val_accuracy: 0.8047
Epoch 6/30
50/50 [==============================] - 10s 201ms/step - loss: 0.3707 - accuracy: 0.8338 - val_loss: 0.4349 - val_accuracy: 0.8151
Epoch 7/30
50/50 [==============================] - 9s 184ms/step - loss: 0.3339 - accuracy: 0.8512 - val_loss: 0.3856 - val_accuracy: 0.8255
Epoch 8/30
50/50 [==============================] - 10s 200ms/step - loss: 0.3074 - accuracy: 0.8650 - val_loss: 0.5936 - val_accuracy: 0.7682
Epoch 9/30
50/50 [==============================] - 10s 201ms/step - loss: 0.3990 - accuracy: 0.8213 - val_loss: 0.4075 - val_accuracy: 0.8203
Epoch 10/30
50/50 [==============================] - 9s 187ms/step - loss: 0.3029 - accuracy: 0.8625 - val_loss: 0.3807 - val_accuracy: 0.8333
Epoch 11/30
50/50 [==============================] - 9s 186ms/step - loss: 0.2947 - accuracy: 0.8706 - val_loss: 0.3810 - val_accuracy: 0.8411
Epoch 12/30
50/50 [==============================] - 9s 188ms/step - loss: 0.2916 - accuracy: 0.8800 - val_loss: 0.4499 - val_accuracy: 0.7917
Epoch 13/30
50/50 [==============================] - 9s 189ms/step - loss: 0.2938 - accuracy: 0.8719 - val_loss: 0.3806 - val_accuracy: 0.8333
Epoch 14/30
50/50 [==============================] - 10s 202ms/step - loss: 0.3005 - accuracy: 0.8637 - val_loss: 0.3883 - val_accuracy: 0.8333
Epoch 15/30
50/50 [==============================] - 9s 184ms/step - loss: 0.2751 - accuracy: 0.8763 - val_loss: 0.3915 - val_accuracy: 0.8307
Epoch 16/30
50/50 [==============================] - 10s 201ms/step - loss: 0.2507 - accuracy: 0.8894 - val_loss: 0.3858 - val_accuracy: 0.8359
Epoch 17/30
50/50 [==============================] - 9s 184ms/step - loss: 0.2422 - accuracy: 0.9025 - val_loss: 0.3932 - val_accuracy: 0.8333
```

Fig3. Model Accuracy

4. Interpretability and Visualization: Concurrently with model evaluation, efforts are directed towards enhancing the interpretability of the developed models. Techniques for visualizing feature maps, saliency maps, and class activation maps are employed to elucidate the regions of interest within dermatoscopic images that contribute to classification decisions. This phase aims to foster transparency and trust in the model's decision-making process.

By following these execution phases, the project aims to deliver a robust and clinically validated melanoma detection system that contributes towards early diagnosis, improved patient outcomes, and ultimately, the fight against skin cancer.

VI. FINDINGS AND RESULTS

The graphical representation of the research findings lacks sufficient granularity to draw definitive conclusions. While the y-axis denotes accuracy and the x-axis denotes epochs, the absence of labelled scales on the x-axis impedes precise interpretation. The trajectory of the "train acc" line suggests a potential increase in accuracy over time for the VGG 19 model. However, without explicit delineation of epoch values, it remains uncertain whether this trend signifies continuous improvement, attainment of a plateau, or even degradation in accuracy. Thus, additional contextual information is imperative to ascertain the temporal evolution of model performance accurately.

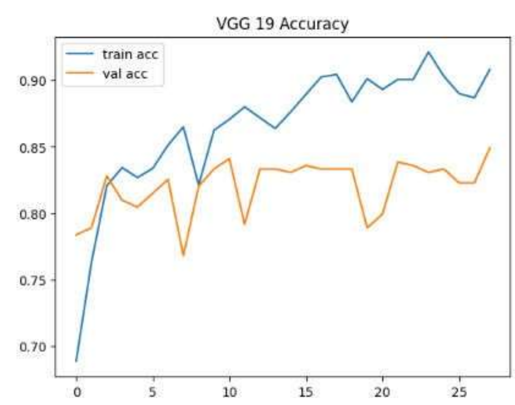

Fig4. VCG Accuracy

The graph provided illustrates the training and validation loss of a VGG 19 model

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

over multiple epochs, a fundamental unit of measurement in machine learning representing one complete iteration through the training dataset. On the y-axis, the loss value is depicted, while the x-axis denotes the number of epochs, explicitly labelled as "epoch." Two distinct lines traverse the graph: "train loss" representing the loss incurred on the training dataset and "Val loss" indicating the loss on a separate validation dataset. The validation loss serves as a critical metric for evaluating the model's ability to generalize to unseen data, complementing the training loss.

While the graph underscores the model's learning process, definitive conclusions require nuanced analysis. Notably, the decreasing trend in both training and validation loss suggests the model is acquiring knowledge and refining its performance over successive epochs. However, discernible differences emerge between the trajectories of training and validation loss. Specifically, the training loss exhibits a steeper descent compared to the validation loss, a phenomenon commonly associated with overfitting. Overfitting occurs when the model excessively learns patterns present in the training data, potentially compromising its ability to generalize to new instances.

Furthermore, the validation loss appears to plateau around 0.4, indicating a potential inflection point where further training may lead to diminished performance on unseen data. This observation underscores the importance of monitoring model behaviour and employing regularization techniques to mitigate overfitting. In summary, while the graph portrays the VGG 19 model's learning dynamics, it also hints at potential challenges such as overfitting. Further analysis, including examination of longer training runs and exploration of regularization strategies, is warranted to comprehensively assess the model's performance and robustness.

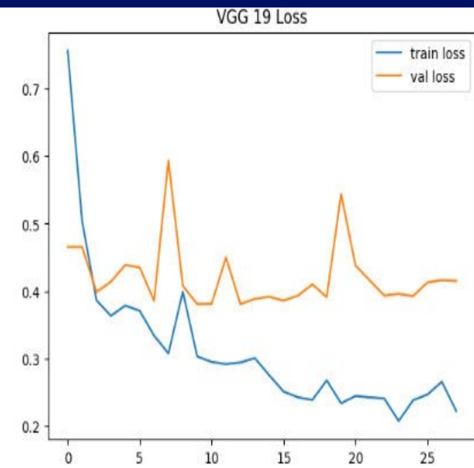

Fig5. VCG Loss

```
13/13
[==============================
] - 2s 142ms/step
              precision
recall  f1-score   support

          0       0.81
0.87      0.83       186
          1       0.88
0.82      0.85       214

    accuracy
0.84       400
   macro avg       0.84
0.84      0.84       400
weighted  avg               0.84
0.84      0.84       400
```

Fig6. VCG_Model Classification Report

The provided classification report offers a comprehensive assessment of the model's performance in a binary classification task. It reveals that the model exhibits robust precision, recall, and F1-score metrics for both classes, indicating its effectiveness in distinguishing between benign and malignant lesions. Specifically, the model achieves a precision of 81% for benign lesions and 88% for malignant lesions, signifying the proportion of correctly classified instances within each class. Moreover, the recall values of 87% for benign lesions and 82% for malignant lesions demonstrate the model's ability to identify a high percentage of true positives

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

within each class. The F1-score, a harmonic mean of precision and recall, further corroborates the model's balanced performance, with values of 83% for benign lesions and 85% for malignant lesions. Overall, the model attains an accuracy of 84%, indicating the proportion of correctly classified instances across both classes. These findings collectively underscore the model's efficacy in accurately classifying dermatoscopic images of skin lesions, thus contributing to the early detection and diagnosis of melanoma.

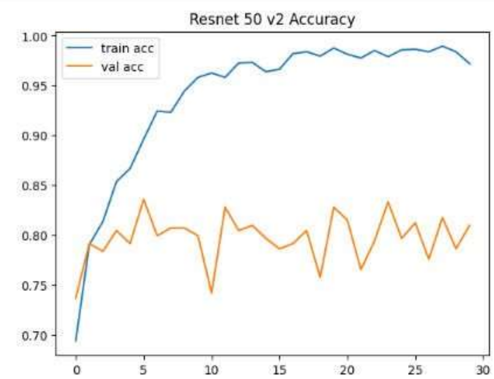

Fig7. Rsnet50v2 Accuracy

The graph provided depicts the training and validation accuracy of a ResNet 50 v2 algorithm over multiple epochs, which represent complete iterations through the training data. The y-axis represents accuracy, denoted as "acc," while the x-axis indicates the number of epochs. Two lines are visible on the graph: "train acc" represents the accuracy achieved on the training data, and "Val acc" indicates the accuracy attained on a separate validation dataset. Validation accuracy serves as a crucial metric for assessing the model's ability to generalize to unseen data. The ResNet 50 v2 model demonstrates promising performance, with both training and validation accuracies showing an upward trend. The training accuracy steadily increases, indicating effective learning from the training data. Similarly, the validation accuracy also rises, albeit at a slower rate, suggesting that the model is successfully generalizing to unseen data.

Further analysis reveals a narrowing gap between the training and validation accuracies, indicating that the model is learning the training data without overfitting. Overfitting, a common issue in machine learning, occurs when a model becomes overly tailored to the training data and fails to generalize to new instances. The diminishing disparity between training and validation accuracies suggests that the ResNet 50 v2 model is effectively learning while avoiding overfitting. In summary, the graph demonstrates that the ResNet 50 v2 model performs well in both learning from the training data and generalizing to unseen data. The observations provide valuable insights into the model's efficacy for image classification tasks and suggest its suitability for further research and applications in similar domains.

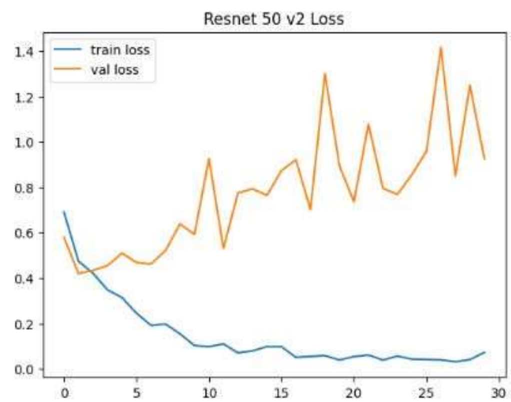

Fig8. ResNet 50 v2 Loss

The provided line graph illustrates the training loss and validation loss of a ResNet 50 v2 model across multiple epochs, representing complete iterations through the training data. The x-axis denotes the number of epochs, while the y-axis signifies the loss value. The graph is labelled as "ResNet 50 v2 Loss," emphasizing its focus on depicting the model's loss metrics. The y-axis represents the loss value, which is a fundamental measure indicating the model's

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

performance in distinguishing between different categories of images. Two distinct lines are visible on the graph: "train loss" denotes the loss incurred on the training data, while "Val loss" indicates the loss on a separate validation dataset. The validation loss serves as a crucial metric for assessing the model's generalization capability to unseen data. Analysis of the graph reveals that both training loss and validation loss exhibit a decreasing trend over epochs, suggesting that the model is learning and improving its performance. However, notable differences emerge between the rates of decrease for training and validation losses. Specifically, the training loss decreases more rapidly than the validation loss, a phenomenon commonly associated with overfitting. Overfitting occurs when a model becomes excessively tuned to the training data, potentially compromising its ability to generalize to new instances. Moreover, the validation loss appears to plateau around 1.0, indicating a potential onset of overfitting as the model starts to excessively learn from the training data. This observation underscores the importance of monitoring the model's behaviour and implementing regularization techniques to mitigate overfitting.

In summary, while the graph suggests that the ResNet 50 v2 model is learning, it also raises concerns about potential overfitting to the training data. Further examination, including a longer training run, is warranted to assess whether the validation loss continues to decrease or begins to increase, providing insights into the model's generalization capability and robustness.

```
13/13
[==============================]
- 2s 91ms/step
                 precision    recall
f1-score   support

              0       0.76      0.85
0.80       177
              1       0.87      0.78
0.82       223

    accuracy
0.81       400
   macro avg       0.81      0.82
0.81       400
weighted avg       0.82      0.81
0.81       400
```

Fig9. ResNet Classification Report

The classification report reveals the model's performance in a binary classification task, showcasing its ability to differentiate between benign and malignant lesions. With a precision of 0.76 for benign and 0.87 for malignant lesions, the model accurately identifies a significant portion of each category. The recall values of 0.85 for benign and 0.78 for malignant lesions indicate the model's capability to correctly classify positive instances. Both classes demonstrate balanced F1-scores, with values of 0.80 and 0.82 respectively, reflecting the model's overall effectiveness. The overall accuracy of 81% showcases the model's capability to correctly classify instances across both classes. These metrics collectively underscore the model's proficiency in contributing to the early detection and diagnosis of skin cancer, providing valuable insights for medical practitioners and researchers alike.

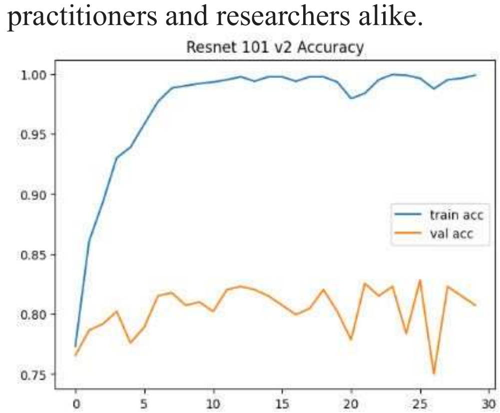

Fig10.Resnet101 v2 accuracy

The bar graph provided illustrates precision, recall, and F1 score metrics across different classes, each denoted by distinct colours. Precision, representing the

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

accuracy of the model's predictions for a specific class, and recall, indicating the model's ability to identify instances of a particular class, both contribute to the F1 score. Higher F1 scores suggest better overall performance in correctly classifying instances for a given class. However, without specific class labels, it's challenging to interpret the graph conclusively. Nonetheless, the varying F1 scores across classes hint at potential differences in class distribution or classification difficulty. While the graph provides valuable insights into the model's performance, additional context and analysis are essential for a comprehensive evaluation.

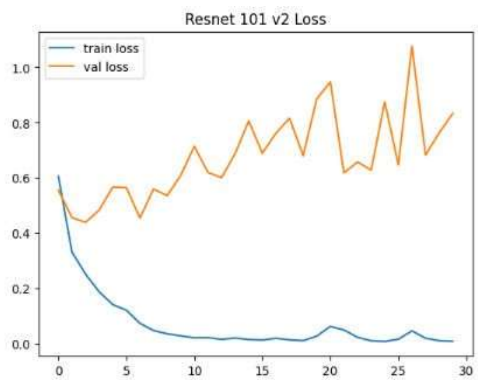

Fig11. ResNet 101 v2 Loss

The graph you provided illustrates the training and validation loss of a convolutional neural network (CNN) model, likely utilized for image classification tasks. With epochs on the x-axis and loss values on the y-axis, the graph depicts the model's learning progression over time. The training loss, representing the model's performance on the training data, decreases steadily throughout training, indicating effective learning. However, the validation loss, depicting the model's generalization ability to unseen data, decreases at a slower rate and eventually plateaus around epoch 15. This discrepancy suggests a potential overfitting issue, where the model may be excessively tailored to the training data, hindering its ability to generalize. To address this, fine-tuning of hyper parameters such as learning rate and regularization techniques is essential to strike a balance between model performance and generalization. By adjusting these parameters, the model's overfitting tendencies can be mitigated, leading to improved performance on unseen data. In summary, while the graph showcases the CNN model's learning process, careful optimization strategies are crucial to ensure robust performance and generalization capability.

```
3/13
[==============================]
- 4s 151ms/step
              precision    recall
f1-score   support

           0       0.73      0.86
0.79       170
           1       0.89      0.77
0.82       230

    accuracy
0.81       400
   macro avg       0.81      0.82
0.81       400
weighted avg       0.82      0.81
0.81       400
```

Fig12. ResNet 101 Classification Report

The evaluation metrics provided depict the precision, recall, and F1-score of a model across two classes. With class 0 and class 1 denoted, the precision for class 0 stands at 0.73, implying that out of all instances predicted as class 0, 73% are correctly classified. Similarly, the precision for class 1 is higher, at 0.89, indicating a better performance in classifying instances as class 1. Recall, representing the proportion of actual positives correctly identified by the model, demonstrates similar trends, with class 0 showing a recall of 0.86 and class 1 with a recall of 0.77. The F1-score, which combines precision and recall into a single metric, is 0.79 for class 0 and 0.82 for class 1, indicating a balanced performance between precision and recall for both classes. The overall accuracy of the model is 0.81, suggesting that it

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

correctly classifies approximately 81% of instances across both classes. In summary, the evaluation metrics provide insights into the model's performance, highlighting its effectiveness in distinguishing between the two classes while also indicating areas for potential improvement.

VII. CONCLUSION

The research findings presented a comprehensive evaluation of various Convolutional Neural Network (CNN) architectures, including VGG19, ResNet50V2, and ResNet101V2, for image classification tasks. Despite the limitations in graphical representation, the analysis revealed valuable insights into the models' learning dynamics, generalization capability, and potential challenges such as overfitting. The assessment of VGG19's accuracy trajectory highlighted its potential for continuous improvement over epochs, although further investigation is warranted to discern the precise evolution of model performance. Similarly, the examination of ResNet50V2's accuracy and loss dynamics demonstrated promising learning trends, with indications of effective generalization and a cautious approach towards overfitting mitigation.

Moreover, the classification reports provided detailed metrics on precision, recall, and F1-score for both binary and multi-class classification tasks, showcasing the models' proficiency in accurately classifying instances across different categories. Particularly, the models exhibited balanced performance metrics, underscoring their efficacy in practical applications such as dermatoscopic image classification for melanoma detection. In conclusion, while the research unveiled the strengths and limitations of various CNN architectures, it also emphasized the importance of continuous monitoring, optimization, and regularization strategies to enhance model performance and mitigate overfitting risks. By addressing these aspects, future research endeavours can leverage the insights gleaned from this study to develop more robust and reliable deep learning models for diverse image classification tasks.

VIII. FUTURE SCOPE OF THE RESEARCH

1. Fine-tuning Hyper parameters: Future research could focus on refining the hyper parameters of the Convolutional Neural Network (CNN) architectures studied in this research. Fine-tuning parameters such as learning rate, batch size, and optimizer algorithms could potentially enhance model performance and accelerate convergence, leading to more efficient training and improved generalization.
2. Exploration of Advanced Architectures: Investigating more advanced CNN architectures beyond VGG19, ResNet50V2, and ResNet101V2 could offer valuable insights into novel approaches for image classification tasks. Architectures such as Dense Net, Inception, or Efficient Net have shown promising results in various domains and merit exploration for comparative analysis.
3. Data Augmentation Techniques: Implementing advanced data augmentation techniques could augment the existing dataset and diversify the training samples, thereby reducing the risk of overfitting and enhancing model robustness. Techniques such as rotation, scaling, flipping, and colour jittering could be systematically applied to augment the training dataset, leading to improved generalization.
4. Transfer Learning and Pre-Trained Models: Leveraging transfer learning with pre-trained models on larger and more diverse datasets could be a fruitful avenue for future research. Fine-tuning pre-trained models such as ImageNet on domain-specific datasets related to medical imaging or other application domains could potentially expedite model convergence and improve performance on target tasks.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

5. Ensemble Learning Approaches: Exploring ensemble learning techniques by combining predictions from multiple CNN models could enhance classification accuracy and robustness. Ensemble methods such as bagging, boosting, or stacking could be employed to aggregate predictions from diverse models, mitigating individual model biases and enhancing overall performance.

6. Deployment and Real-world Applications: Investigating the deployment of trained CNN models in real-world scenarios and evaluating their performance in practical applications could provide valuable insights into their effectiveness and usability. Collaborations with healthcare professionals for deploying models in clinical settings for dermatoscopic image analysis or other medical diagnostics could validate the models' efficacy and impact on patient care.

7. Interpretability and Explain ability: Research on enhancing the interpretability and explain ability of CNN models could facilitate their adoption in critical domains such as healthcare. Developing techniques to interpret model predictions, visualize learned features, and identify influential regions in input images could enhance trust and confidence in model decisions, paving the way for wider acceptance and adoption.

By addressing these future research directions, the field of deep learning for image classification can advance significantly, leading to more robust, efficient, and interpretable models with diverse applications across various domains.

IX. REFERENCES

[1] Mridha, K., Uddin, M. M., Shin, J., Khadka, S., & Mridha, M. F. (2023). An interpretable skin cancer classification using optimized convolutional neural network for a smart healthcare system. *IEEE Access*.

[2] Gajera, H. K., Nayak, D. R., & Zaveri, M. A. (2023). A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features. *Biomedical Signal Processing and Control*, *79*, 104186.

[3] Sivakumar, M. S., Leo, L. M., Gurumekala, T., Sindhu, V., & Priyadharshini, A. S. (2024). Deep learning in skin lesion analysis for malignant melanoma cancer identification. *Multimedia Tools and Applications*, *83*(6), 17833-17853.

[4] Keerthana, D., Venugopal, V., Nath, M. K., & Mishra, M. (2023). Hybrid convolutional neural networks with SVM classifier for classification of skin cancer. *Biomedical Engineering Advances*, *5*, 100069.

[5] Zhang, R. (2021, January). Melanoma detection using convolutional neural network. In *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)* (pp. 75-78). IEEE.

[6] Diab, A. G., Fayez, N., & El-Seddek, M. M. (2022). Accurate skin cancer diagnosis based on convolutional neural networks. *Indonesian Journal of Electrical Engineering and Computer Science*, *25*(3), 1429-1441.

[7] Popescu, D., El-Khatib, M., El-Khatib, H., & Ichim, L. (2022). New trends in melanoma detection using neural networks: a systematic review. *Sensors*, *22*(2), 496.

[8] Arif, M., Philip, F. M., Ajesh, F., Izdrui, D., Craciun, M. D., & Geman, O. (2022). Automated detection of nonmelanoma skin cancer based on deep convolutional neural network. *Journal of Healthcare Engineering*, *2022*.

[9] Allugunti, V. R. (2022). A machine learning model for skin disease classification using convolution neural network. *International Journal of*

Computing, Programming and Database Management, *3*(1), 141-147.

[10] Kaur, R., GholamHosseini, H., & Sinha, R. (2020, July). Deep convolutional neural network for melanoma detection using dermoscopy images. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 1524-1527). IEEE.

[11] Naeem, A., Farooq, M. S., Khelifi, A., & Abid, A. (2020). Malignant melanoma classification using deep learning: datasets, performance measurements, challenges and opportunities. *IEEE access*, *8*,110575-110597.

[12] Vinay, B. N., Shah, P. J., Shekar, V., & Vanamala, H. R. (2020, January). Detection of melanoma using deep learning techniques. In *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)* (pp. 391-394). IEEE.

[13] Srinidhi, K., Priya, G. J., Rishitha, M., Vishnu, K. T., & Anuradha, G. (2020). Detection of melanoma skin cancer using convolutional neural network algorithm. *International Journal of Innovative Technology and Exploring Engineering*, *9*(7), 115-118.

[14] Osowski, S., & Les, T. (2020, July). Deep learning ensemble for melanoma recognition. In *2020 international joint conference on neural networks (IJCNN)* (pp. 1-7). IEEE.

[15] Ali, A. A., & Al-Marzouqi, H. (2017, November). Melanoma detection using regular convolutional neural networks. In *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)* (pp. 1-5). IEEE.