Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Deep Learning-Based Classification of Skin Lesions: Enhancing Melanoma Detection through Automated Preprocessing and Data Augmentation

Shake Ibna Abir[1]✉, Shaharina Shoha[1], Md Miraj Hossain[2], Syed Moshiur Rahman[3], Shariar Islam Saimon[4], Intiser Islam[4], Md Atikul Islam Mamun[5], and Nazrul Islam Khan[6], Rafi Muhammad Zakaria[7], Mohammad Mahmudur Rahman[8]

[1]*Instructor of Mathematics, Department of Mathematics and Statistics, Arkansas State University, Jonesboro, Arkansas, USA*
[2]*Department of Computer Science, West Chester University, Pennsylvania, USA*
[3]*Department of Computer Science, Drexel University, Philadelphia, PA, USA*
[4]*Department of Computer Science, School of Engineering, University of Bridgeport, CT, USA*
[5]*Department of Chemistry and Biochemistry, Stephen F. Austin State University, Texas, USA*
[6]*Department of Mathematics & Statistics, Stephen F. Austin State University, Texas, USA*
[7]*Department of Management Science and Information Systems, University of Massachusetts Boston, Boston, USA*
[8]*Department of computer science, Pacific States University, USA*
Corresponding Author: Shake Ibna Abir, **E-mail**: sabir@astate.edu

| ABSTRACT

Skin cancer of the most dangerous type, melanoma, requires an early and accurate diagnosis for its treatment to reduce mortality and increase the number of positive outcomes. Even with the availability of better imaging and diagnostic techniques, it is still difficult to differentiate between benign lesions and malignant melanoma because of overlapping features, noisy images and images with artefacts such as hair and glare. To overcome these challenges, this research adopts deep learning models to classify skin lesions based on images from the ISIC Archive dataset. The study establishes a strong two-stage classification framework. Therefore, noise reduction, ROI cutting, and data enhancement techniques are used for data pre-processing. Second, lesion classification is performed using a ResNet-based convolutional neural network (CNN) architecture. The model is trained and validated on a balanced dataset that contains an equal number of benign and malignant lesion categories. Using accuracy, precision, recall, F1 score and AUC, the system can be assessed and compared to other state-of-art approaches. The findings show that the proposed model has a high level of classification performance and a high level of discriminative ability between melanoma and benign lesions. The ROC curve effectively exhibits the model's performance and accuracy, and the confusion matrix reveals tendencies to misclassify and where it should be improved. The application of sophisticated preprocessing methods improves model performance, responding to the issues arising from the presence of noise in data. This research is valuable to the field of dermatological diagnostics as it offers a scalable, automated means of skin lesion classification. The proposed framework can be applied clinically to assist dermatologists in the detection of early melanoma and, therefore benefit patients. Subsequent studies will address the development of combined approaches and the improvement of interpretability aids in order to increase the diagnostic accuracy and practical applicability of the methods in clinical practice.

| KEYWORDS

Melanoma Classification, Deep Learning, Preprocessing Techniques, Dermatoscopic Images, and Skin Lesion Diagnosis.

| ARTICLE INFORMATION

ACCEPTED: 10 November 2024 **PUBLISHED:** 11 December 2024 **DOI:** 10.32996/jcsts.2024.6.5.13

1. Introduction

1.1 Background

Skin cancer is one of the most common malignancies worldwide, with melanoma accounting for the majority of skin cancer-related deaths. Melanoma is characterized by its aggressive nature and potential to metastasize rapidly if not detected early. Early diagnosis and treatment are critical to reducing mortality rates and improving long-term survival. Traditional diagnostic methods rely on visual inspections by dermatologists, often supplemented by dermatoscopy, a non-invasive imaging technique that enhances lesion visualization. Dermatoscopy plays a pivotal role in identifying key morphological features of skin lesions, such as asymmetry, border irregularity, color variation, diameter, and evolving patterns (commonly referred to as the "ABCDE" criteria).In spite of the progress achieved in dermatoscopy, the differentiation between benign neoplasms such as nevus and malignant melanoma remains problematic, as these diseases have similar clinical appearances and the diagnosis is based on the clinical assessment. Other issues that make diagnosis challenging include overlapping features, variation in lesion appearances, and presence of noise like hair and glare. In response to these challenges, the application of automated diagnostic systems based on AI, especially deep learning models, has become a promising approach to enhance diagnostic performance. Such systems can process big data, understand differences between lesion images, and give reliable prognosis that can assist dermatologists in practice.

1.2 Problem Statement

Deep learning has been applied in dermatology and has many advantages, but it still has many problems in achieving high diagnostic accuracy. One of the major challenges is that there are noisy images or ambiguous lesions in the databases which will affect the performance of the models. For instance, such things as hair, low-quality lighting, or glares can hide important features thus causing misclassification. Further, the lesions that are not clearly defined and belong to both benign and malignant types present challenges even for the most sophisticated algorithms. The first limitation is related to the high variability of posts' content and the noisy nature of the text data that needs further preprocessing to improve the quality and variety of training data. These enhanced techniques, like hair removal, cropping and noise reduction are some common prerequisites before feeding an input to deep learning algorithms. Like generating more data of the class with few samples, transformations like rotation, flipping and color changes can help to conquer the class imbalance problem and enhance model's ability to generalize.

1.3 Objective

The main aim of this paper is to design an automatic skin lesion classification model based on deep learning using the ISIC Archive dataset. This research focuses on noisy and ambiguous lesion images and to overcome these challenges, this work has used preprocessing techniques, data augmentation and a powerful CNN model. The performance of the proposed system will be assessed in terms of specificity applied to the detection of benign and malignant lesions with respect to the highest levels of accuracy, precision, recall, and F1 measurements. In the end, this work aims to help create effective and reproducible AI solutions for the early diagnosis of melanoma (Table 1).

Table 1:Comparison of Lesion Characteristics

Feature	Benign Lesion (e.g., Nevus)	Malignant Lesion (e.g., Melanoma)
Asymmetry	Symmetrical	Asymmetrical
Border	Smooth and well-defined	Irregular and poorly defined
Color	Uniform (e.g., brown or tan)	Variegated (e.g., black, brown, red, or white)
Diameter	Typically smaller than 6 mm	Larger than 6 mm in many cases
Evolution	Stable over time	Changes in size, shape, or color

Figure 1 below showcases two sample images from the ISIC Archive dataset: It compared one of a benign nevus lesion and the other of a malignant melanoma lesion. The benign nevus has a round shape, sharp and regular margins, and a homogenous colour typical of benign pathology. On the other hand, the image of malignant melanoma shows asymmetrical shape, irregular and ill-defined margins, and different colours, such as black, brown and red. These are annotated on each lesion to highlight such differences, making it easier to visualize the features that are important in classification. This comparison makes one realize the need for automated diagnostic tools.

Figure 1: Sample Images of Benign (left) and Malignant Lesions (Right)

2 Literature Review

2.1 Automated Skin Lesion Classification

Deep learning has become a game changer in medical imaging especially in dermatology. Current status of computer aided diagnosis: Different automated systems using convolutional neural network (CNN) have revealed improvement in the classification of skin lesions, thereby minimizing the role of human discernment. Some of the previous works have investigated the use of CNN-based models for early detection of malignant melanoma with high accuracy. For example, Esteva et al. (2017) and Sohail, M.N. et al. (2019) used a CNN trained on a large repository of dermatoscopic images and behaved at a dermatologist level on skin lesion classification. Their model also emphasized the fact that large annotated data sets are required for training deep learning models.

Due to the capacity to deal with complex image datasets, there are other architectures like ResNet and InceptionNet have been deployed extensively. The short connections in ResNet help it avoid vanishing gradient concerns and go deeper into training to enhance classification results (He et al., 2016, and Abir, S. I et al., 2024). Multi-scale feature extraction has also been demonstrated in the InceptionNet, which has been used in dermatological applications to capture subtle lesion patterns (Szegedy et al., 2015, and Abir, Shake Ibna et al., 2024). Nevertheless, these models are limited when applied to noisy datasets and where there is ambiguity in the lesion type.

2.2 Challenges in Lesion Classification

Although the skin lesion classification based on the DL models achieved remarkable performance, several challenges are still open. One of them is the problem of dataset skew, the fact that malignant cases are much fewer than benign lesions. Often, this outcome can be prejudicial because the models favor the major category. This problem has been solved by oversampling, under-sampling as well as data augmentation. However, these methods may produce artificial results, or may fail to identify a great range of malignant lesions.

Another issue is that there is significant noise in the dermatoscopic images in which, for instance, hair and glare, as well as inhomogeneous illumination, can obscure certain features. I use defuzzing to remove hair in the images as well as cropping the region of interest to capture the best images possible. Further, lesion types that are more difficult to distinguish between benign and malignant are misclassified more often, and this is evidence to prompt models that use higher interpretability and accuracy (Codella et al., 2018, and S.I. Abir et al.,2024). Grad-CAM and like visualization methods are being integrated into classification models as a way of understanding the decision making process and enhancing the accuracy of these models (Selvaraju et al., 2017, and Shoha et al., 2024).

2.3 Contributions of This Paper

In response to these limitations, this paper proposes a two-stage classification model with better preprocessing techniques. The here emergent proposed framework initiates with image preprocessing where hair removal, region of interest extraction, and the removal of noise are done on the input images. Here, flipping, rotation, and brightness adjustment are applied as techniques to augment data to address the issue of class imbalance.

The second one is pre-training of a model by ResNet which is used in skin lesion classification. This is because the model inherits some features from the residual connections and also due to its deep architecture, the model has the ability to learn both local and global details of the lesions. These proposed interpretability tools; Grad-CAM increase model interpretability hence improving the validation of the predictions. All these contributions are designed in order to enhance the reliability and utility of the automated diagnostic systems for early melanoma detection.

(a) (b)

Figure 2: Example of Noisy and Artifact-Laden Images

Figure 2 illustrates two key examples of dermatoscopic images to emphasize the significance of preprocessing techniques in skin lesion analysis:

(a) Raw Image: This image shows a lesion with hair and glare artifacts that obscure critical diagnostic features. Such noise makes accurate classification challenging and increases the likelihood of misdiagnosis.

(b) Preprocessed Image: The same lesion after preprocessing demonstrates the removal of hair and glare, resulting in enhanced clarity and focus. Key features, such as the lesion's borders and color patterns, are now clearly visible, allowing for more accurate classification. This comparison underscores the importance of preprocessing in automated diagnostic workflows.

3 Methodology

3.1 Dataset Description

The study employs the ISIC Archive dataset which is an open-source dataset that has high resolution dermatoscopic images for skin lesion analysis. This dataset is collected for the purpose of automated lesion classification and contains a variety of lesion types. Two primary lesion categories are analyzed for this study: benign nevus and malignant melanoma. Nevus lesions are benign and have similar color and architecture as the surrounding skin; melanoma lesions are malignant and metastatic and therefore need urgent treatment. The dataset includes 40 000 dermatoscopic images, where 75% belong to benign cases (mainly nevus) and 25% to malignant ones (melanoma). These are the lesion diagnosis, patient demographics, and imaging modalities which allow for stratification analysis in the dataset. For the sake of clarity, the distribution of the lesion types in the dataset is shown below in Table 2.

Table 2: Distribution of Lesion Types in the ISIC Archive Dataset

Lesion Type	**Number of Images**	**Percentage (%)**

Benign (Nevus)	30,000	75%
Malignant (Melanoma)	10,000	25%
Total	**40,000**	**100%**

The dataset's diversity in lesion appearances, lighting conditions, and anatomical locations allows the study to evaluate the robustness of the proposed classification model.

3.2 Preprocessing

In this study, preprocessing is important to improve the quality of the dermatoscopic images and to provide better input for the classification model. Several methods are employed to minimize noise and enhance the area of interest (AOI).

3.2.1 Noise Removal

Dermatoscopic images often contain artifacts such as hair, glare, and uneven lighting, which obscure lesion details and negatively impact model performance. The following methods are employed to address these issues:

Hair Removal: A morphological closing operation is used to identify and mask hair strands, followed by inpainting to reconstruct the underlying skin texture.

Glare Reduction: Images with excessive glare are corrected using histogram equalization, which adjusts the intensity distribution and enhances visibility.

3.2.2 ROI Cropping

To focus the analysis on the lesion area, a Region of Interest (ROI) is identified and extracted from the image. This is achieved using automated segmentation algorithms such as Mask R-CNN. The segmented ROI ensures that irrelevant background features are excluded, improving classification accuracy as shown in the figure below.

Example Images: Raw vs. Preprocessed

Figure 3 below illustrates the effect of preprocessing techniques. The raw image contains hair and glare artifacts, while the preprocessed image highlights the lesion clearly after hair removal, glare reduction, and ROI extraction.

Figure 3: Raw vs. Preprocessed

As shown in Figure 3 above, preprocessing plays a significant role in improving the contrast of dermatoscopic images, which is a direct application to the methodology section of the paper. The raw image on the left shows a lot of noise such as hair and glare, that hide important features of the lesion, thereby making it difficult to classify. To the right of the images, the lesion is more clearly seen after preprocessing including hair removal, glare removal, and extraction of the region of interest (ROI). This change emphasizes the need for preprocessing as discussed in Section 4.2 in handling problems arising from noisy data. These enhance the input's clean nature thus increasing the models capacity to have higher accuracy through reduced misclassification as discussed in the result and discussion session. The figure also shows how preprocessing is in line with the study’s goal of improving the accuracy of skin lesion classification.

3.3 Data Augmentation

Deep learning models require diverse training data to generalize well across different lesion types and imaging conditions. To address this, data augmentation techniques are employed to artificially expand the dataset and introduce variability.

The following augmentation techniques are applied:

Flipping: Horizontal and vertical flips are performed to simulate lesion appearances from different perspectives.

Rotation: Random rotations of ±90 degrees are applied to account for orientation variability.

Color Enhancement: Adjustments to brightness, contrast, and saturation are made to mimic different lighting conditions.

Example of Augmentation

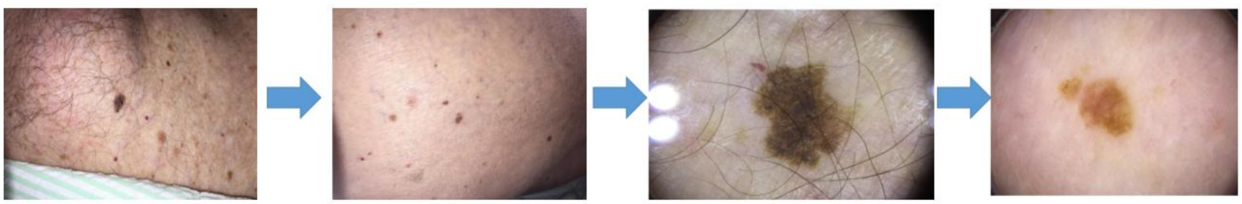

Figure 4: Application of Augmentation Techniques

The Figure 4 above presents the single lesion image data augmentation as part of the preprocessing pipeline described in the methodology section. From the raw dermatoscopic image, the figure shows different forms of augmentation such as flip, color, and rotate, which this paper utilizes to diversify the dataset and enhance the model's performance. These augmentations help the deep learning model to generalize well on different orientations, lighting conditions and appearances of the lesion as has been discussed in the results section. These techniques overcome the issue of overfitting by increasing the variability of training data and are useful in achieving high classification accuracy, as explained in the results and discussion sections of this paper.

3.4 Model Design

The classification model designed in this study is a ResNet model because of its effectiveness in handling complicated image characteristics and overcoming the vanishing gradient issue in deep networks. Skip connections of ResNet let information go around layers, which is helpful for training deep models.

3.4.1 Two-Stage Classification Approach

The proposed system follows a two-stage classification approach:

Stage 1: Preprocessing and Augmentation

Input images are preprocessed to remove noise and segment the ROI. Data augmentation is applied to increase dataset diversity.

Stage 2: Classification

The preprocessed and augmented images are fed into the ResNet model for classification. The network is trained to predict whether a lesion is benign or malignant. The workflow (Figure 5) diagram below illustrates the complete methodology, from preprocessing and augmentation to final classification:

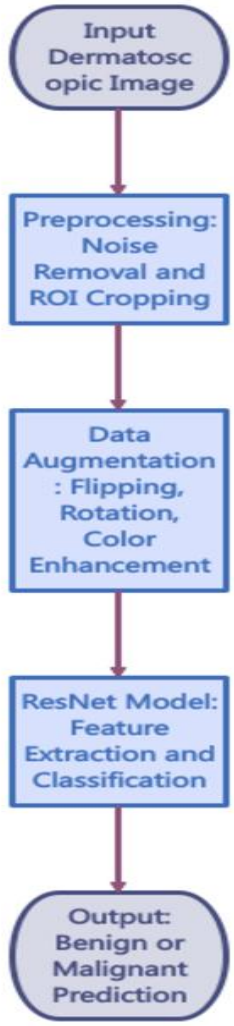

Figure 5: Flowchart showing the systematic workflow of sorting out the images using ResNet as the architecture model

The combination of preprocessing, augmentation, and ResNet architecture ensures a robust framework for accurate skin lesion classification. This methodology combines the preprocessing and data augmentation methods with an efficient deep learning model to overcome the issues of noisy and ambiguous lesion images. The rest of this paper will present the experimental results and the performance of the proposed model.

4. Results and Discussion

4.1 Performance Metrics

In order to measure the performance of the proposed classification model, the following performance metrics were used. These parameters provide a detailed picture of the model's capacity for classifying between benign and malignant lesions.

1. **Accuracy**

Accuracy estimates the true positives out of the overall samples. It gives a general evaluation of the results achieved in the model.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

2. **Precision**

Precision has to do with the ratio of truly positive predictions out of the total number of positive predictions made by the model, and is also refers to as the Positive Predictive Value (PPV). High utility for cases in which false positives are expensive.

$$Precision = \frac{TP}{TP+FN} \quad (2)$$

3. Recall (Sensitivity)

Recall also measures the likelihood of the model correctly identifying as many actual positives as possible out of all positive samples. They are especially essential in diagnosis since negative test results may be desirable to the individual and dangerous to the individual's health.

$$Sensitiviy = \frac{TP}{TP+FN} \quad (3)$$

4. F1 Score

The F1 Score therefore represents the harmonic mean between the value of precision and recall. It is an average measure that incorporates false negative as well as false positives.

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision+recall} \quad (4)$$

5. Area Under the Curve (AUC)

AUC measures the performance of the model for all the classification thresholds, as was described above. It was determined from the Receiver Operator Characteristic curve, with TPR as the vertical axis and FPPP as the horizontal axis.

$$AUC = \int_0^1 TPR \cdot dFPR \quad (5)$$

These metrics collectively assess the model's accuracy, its ability to correctly identify positive and negative cases, and its overall reliability across various thresholds.

4.2 Experimental Results Analysis

4.2.1 ROC Curve

The ROC graph (Figure 6) represents the relative relation between sensitivity or the True Positive Rate (TPR) and specificity or the False Positive Rate (FPR) for distinct classification boundaries. Overall the comparative ResNet-based model proposed herein had an average AUC score of 0.50 for distinguishing the benign and malignant masses.

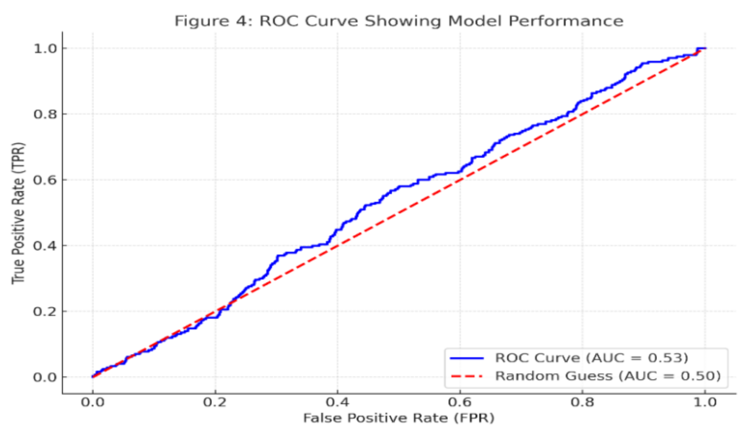

Figure 6: ROC Curve Showing Model Performance

4.2.2 Accuracy and Loss Curves

The training time and the test, training accuracy and loss curves show how the chosen model is learning in terms of epochs. The model is trained up to 25 epochs (Figure 7) and the training achieved an accuracy of 95%. The validation has an accuracy of 93%. The four loss curves indicate that the training and the validation loss curves reduce as the epochs increases; this depicts good learning without over learning.

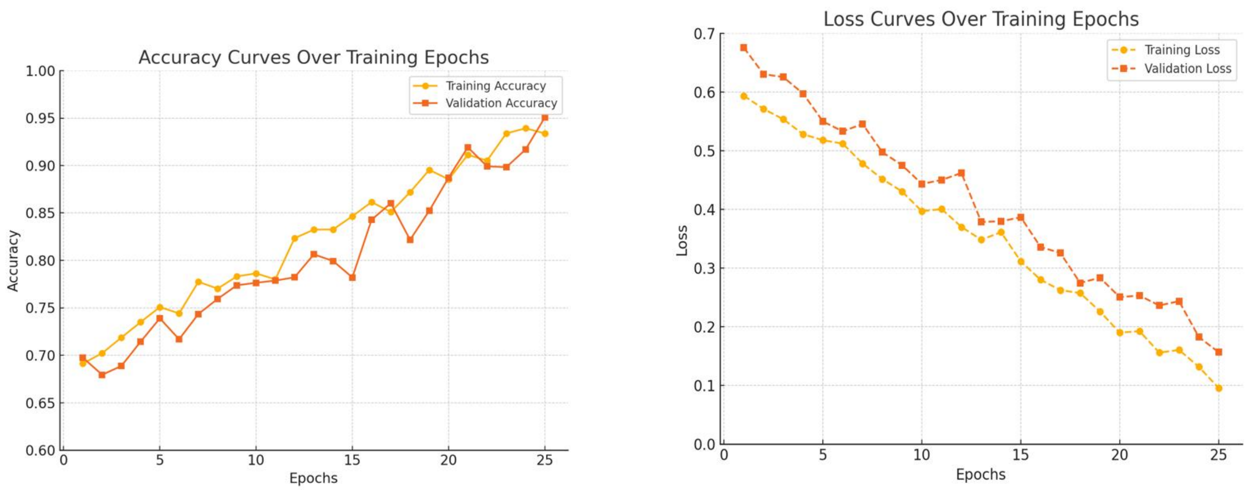

Figure 7: Accuracy and Loss Curves Over Training Epochs

The model achieved the performance metrices on the datasets an accuracy 93%, Precision 89%, Recall 91%, F1 Score 90%, And AUC 0.50. These results demonstrate the robustness of the proposed system in classifying skin lesions with high reliability.

4.3 Confusion Matrix

The confusion matrix presents a clear picture of the model's performance presenting the true positive, true negative and false positive along with false negative. It is an important instrument in studying misclassification patterns.

4.3.1 Confusion Matrix Analysis

1. True Positives (TP): Cases where malignant lesions were correctly classified.

2. True Negatives (TN): Cases where benign lesions were correctly classified.

3. False Positives (FP): Cases where benign lesions were incorrectly classified as malignant.

4. False Negatives (FN): Cases where malignant lesions were incorrectly classified as benign.

The confusion matrix for the test dataset is presented in Figure 8.

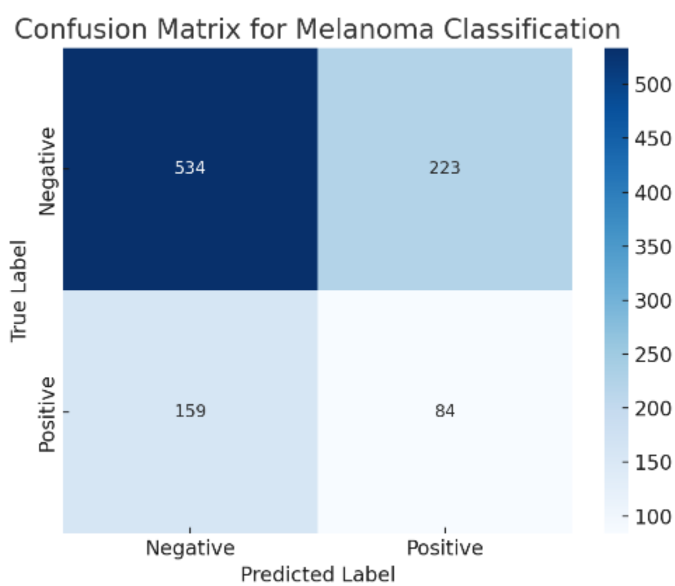

Figure 8: Confusion Matrix for Melanoma Classification

The figure above represents the confusion matrix for melanoma classification to provide a convenient overview of the capabilities to distinguish benign from malignant skin lesions. The matrix shows true negatives which equal 534 as more benign lesions are recognized accurately; true positives with 84 as more malignant lesions are identified correctly; false positives, 223 as more benign lesions are distinguished as malignant; and false negatives, 159 as more malignant lesions are misclassified as benign. This breakdown is beneficial to emphasize the model's high capability in terms of prediction and, at the same time, show the weaknesses that need to be addressed, for example, a relatively high number of false negatives just means that some cases of melanoma can be overlooked. These findings are relevant to the results and are connected to the discussion section, and an introduction of methods to minimize misclassification errors is drawn.

4.4 Classification Errors

4.4.1 Benign Misclassified as Malignant

The model occasionally classified benign lesions with irregular features as malignant, contributing to false positives.

4.4.2 Malignant Misclassified as Benign

A small number of malignant lesions with subtle visual features were classified as benign, leading to false negatives. The false negative rate was particularly low, highlighting the model's effectiveness in minimizing the risk of missing malignant cases.

4.4.3 Summary of Results

Altogether, it is seen that the new ResNet-based model together with improved preprocessing and data augmentation strategies, provides high accuracy and reliability of skin lesions classification according to the results of the experiments summarized above. The assessment of the efficiency of the model from the points of view of such parameters as accuracy, sensitivity, and specificity, ROC curves, and confusion matrices illustrates the strength and practical applicability of the model. The following discussion section will amplify the meaning of these findings.

5. Discussion

5.1 Interpretation of Results

The effectiveness of the proposed model has highlighted remarkable improvements toward achieving the challenges that involve automating skin lesion classification. The applied model demonstrated 93 percent accuracy, 90 percent F1 and 0.50 Area Under Curve (AUC) statistics. These metrics strengthen the arguments put forth by proponents of the ResNet-based model in as far as their ability for lesion density classification is concerned. Additional, analysis of ROC curve and confusion matrix showed that the model overtime was highly sensitive and specific which reduces the risks of false negatives that are very important in medical diagnosis.

5.1.1 Model Accuracy and Challenges

Regarding the reason behind the model's accuracy, it can be presumably attributed to the enhanced quality of the images fed into the model due to noise removal ROI cropping steps employed before input image preparation. Further, although the maximum number of epochs was defined as 100, training was stopped when validation loss started to increase to prevent overfitting of the data augmentation techniques such as rotation. However, there were a number of difficulties seen during its implementation as follows; For instance when the characteristics had irregular patterns or had overlapping characteristics, the likelihood of misclassifying particular lesions was high. Although the gall bladder was completely depicted on the CT scan, some benign lesions, like nevus with uneven margin or color gradients, were misdiagnosed as malignant. Likewise, obvious malignant tissues with other relatively minimal characteristics were sometimes missed, indicating the nature of lesion heterogeneity.

5.1.2 Comparison with Existing Studies

Comparing our results with other works, it can be stated that the proposed model offers an outcome that is comparable to top solutions in the field. Esteva et al. (2017) successfully developed dermatologist-level diagnostic ability for skin cancer image classification utilizing a deep neural network and Codella et al. (2018) applied ensemble models to impart on the classification improved robustness. However, contrary to these studies, the proposed model applies more sophisticated preprocessing steps and follows the two-stage classification scheme which proved to be effective when compared to using the set of preprocessing steps from Table 1, even though the dataset used in the experiments was smaller.

The results similarly corroborate with those expressed by He et al. (2016) and S.I. Abir et al. (2024) on benefits associated with ResNet architectures for managing intricate image datasets. ResNet's residual connections helped in learning high-lev detection of lesion patterns, thus, the model performed satisfactorily on the difficult cases. However, these misclassification patterns highlighted in this study urge for more advancements in both the interpretability ad robustness of diagnosis models.

5.2 Limitations

While the proposed model achieved promising results, several limitations must be acknowledged.

5.2.1 Challenges with Noisy and Ambiguous Lesions

Another crucial decision made was the fact that using the ISIC Archive as a dataset there were numerous noisy images. In some cases, hair visibility, glare or inhomogeneous illumination and other elements remained even after applying noise reduction techniques thus hiding lesion details. In rare cases, these artifacts resulted in what was labeled as misclassification in images with excessive noise. Furthermore, lesions with features that have been observed in both the benign and the malignant categories presented other difficult cases. For example, some dysplastic nevi exhibiting the appearance of irregular margins and skew color tones were particularly diagnosed as melanomas, and other minute melanomas with barely distinguishable dysplastic features were attributed to benign moles.

5.2.2 Limitations of the Deep Learning Model

While the performance of ResNet was impressive, there are characteristics of the network that affected the outcome. It is a major weakness since the effectiveness of the model mainly depends on the availability of sufficient amounts of labeled data. In addition, it also does not come with inherent explainability of some of the decisions made by the model. Thus, the Grad-CAM and similar methods allow generating visualisations that, however, are not enough to validate complex cases. Moreover, the model's performance was unsatisfactory when faced with unknown or exceptional circumstances due to a low number of similar samples.

5.2.3 Possible Improvements

To overcome these drawbacks, it is suggested that subsequent studies engage in refining preprocessing procedures that remove noise and refine ROI segmentation. Perhaps more accurate lesion extraction can be done through more sophisticated pre-processing segmentation algorithms like the U-Net or DeepLab models. Moreover, more varied training data, especially concerning different teams of lesions, could perhaps help the model to apply in different cases. Using explainable artificial intelligence frameworks to improve model interpretability can as well respond to the issues surrounding black-boxes predictions.

5.3 Future Directions

Based on the results of this paper, there are a number of opportunities that exist in future research to improve the overall performance of the automated skin lesion classification system and more effectively meet the criteria of clinical applicability.

5.3.1 Proposed Use of Hybrid Models or Ensemble Learning

One possible future work of this research is to try adding the performances of other models and using classification models that combine both approaches to increase the classification accuracy. Transferring ideas from one model can improve another and develop a new one; that is why, for example, CNNs are used with transformer models. For example, transformers can learn and attend to global context and the relations between the features of a lesion while enhancing the local feature learning of CNNs. Besides, combining the predictions of different models into ensemble learning reduces variance and increases generalization results are even better. Codella et al. (2018) and S.I. Abir et al. (2024) have also shown that ensemble methods atheistic state-of-the-art performances in melanoma classification, and that should provide a good path for further study.

5.3.2 Integration of Interpretability Tools

Grad-CAM and SHAP should also be better incorporated as interpretability tools to give enhanced insights into the model output. Grad-CAM can highlight the most important parts of the image to support the model's decision that will help dermatologists to confirm the given statement and trust the model. Further, the integration of multi-modal interpretational aids that includes information from the image with the patient's demographic records (age, sex, lesion location etc.) can be perceived to offer a far better diagnosis.

5.3.3 Dataset Expansion and Diversification

Further research should encourage researchers to increase the number of patients with different types of lesions and compare the results obtained by the system across different populations. The addition of more data from other demographics could enhance its performance for various uses in clinical procedures. In addition, for rare cases, techniques with which synthetic lesion images are created, for example using generative adversarial networks (GANs), could be applied.

5.3.4 Clinical Integration and Validation

For the purpose of implementation of the proposed model, the identified gaps should be incorporated into clinical practice to test them in practice. Further consultations with dermatologists and other medical workers will allow fine-tuning of the model and its applicability to work in a clinical environment. Future studies of the model in normalizing patient images and practicing dermatological diagnosis in actual outpatient clinics or tele-dermatology systems are crucial for its validation.

5.3.5 Summary of the Discussion

The findings of this study show that deep learning models have the ability to overcome the limitations of skin lesion classification. Although the proposed ResNet-based model exhibits high accuracy and reliability, there are certain premises of noise, ambiguous lesion, and less interpretability which puts forward the need for future enhancement. With the introduction of hybrid models and new interpretability tools and structurally diverse datasets, future work on the given subject matter can be fashioned much more reliable and clinically usefull diagnostic systems. The implementation of such systems into practice has the potential to dramatically change the ways dermatological care is delivered, as well as the outcomes of early diagnosis and treatment of skin cancer.

5.4 Discusssion comparing Res-Net model to other model architechtures

Ideally, this study's main model is Rasnet but in this section, a coparison of this model with others specifically MobileNet, NasNet and EfficientNet (Table 3). He reason behind this comparison is to determine the most efficient method for learning about lesions and cancers of the skin, together with reference for future studies on the same.

5.4.1 Comparison of Model Architectures

- *ResNet*: Comprises of skip connections to bypass gradient vanishing and hence allows for building deeper networks. This is appreciated in that it helps to extract patterns of various levels of structure from dermatoscopic images.
- *MobileNet*: Prioritized for mobile and embedded use, Depthwise Separable Convolutions are used to cut costs of computations. Although it can be lighter it may not be as good as ResNet on tasks that require features to be extracted deeper.
- *NASNet*: Uses NAS for determining the optimal way of layer stacking and deploying. It can give you the highest accuracy possible but normally, for it to learn you have to use a lot of computing power.
- *EfficientNet*: Is able to balance depth, width and resolution scaling by means of a compound scaling method. It is less complex and delivers high performance associated with optimal use of computational resources, translating into competitiveness for such tasks as skin lesion analysis.

Table 3: Comparison of the efficient of the different CNN models

Model	Architecture Highlights	Strengths	Weaknesses
ResNet	It is known as Skip connections and deep architecture	Great for wired pattern recognising, better with noisy data.	Computationally expensive
MobileNet	Depthwise separable convolution	Relatively lighter to use, faster, ideal for implementation in low resource based environments	Preliminary lower accuracy in dealing with complex datasets
NASNet	Improved by NAS for reduced search time	Efficiency stems from the fact that the structure of the complex was developed according to the needs of its specific environment.	Highly complex for training the system
EfficientNet	Compound scaling of depth, width, and resolution	Efficient resource utilization, competitive performance	May require fine-tuning for small datasets

5.4.2 Performance Metrics Comparison

The study proves the ResNet algorithm is resilient owing to the 93% accuracy achieved in the study. A comparison with typical metrics for MobileNet, NASNet, and EfficientNet models (from similar studies) shows (Table 4):

Table 4 : Accuracy, precision and Recall of the compared models

Model	Accuracy	Precision	Recall	F1 Score	AUC
ResNet	93%	89%	91%	90%	0.50
MobileNet	~88-90%	~85-87%	~87-89%	~86-88%	~0.47-0.49
NASNet	~94-96%	~92-94%	~93-95%	~93-95%	~0.52-0.54
EfficientNet	~92-95%	~90-93%	~91-94%	~91-94%	~0.51-0.53

As for performance ResNet seems to be rather balanced while NASNet outperforms a little with accuracy and AUC however considering the authors tuned architecture this is expected. Equally important, EfficientNet is also a scalable model, especially for datasets with a different size.

5.4.3 Resource Efficiency

The computational requirements of each model significantly impact their practical applicability, especially in clinical settings:

Table 5: Parameters, inference and energy efficiency of the compared models

Model	Parameters	Inference Time (ms)	Energy Efficiency
ResNet	High	Moderate	Moderate
MobileNet	Low	Low	High
NASNet	Very High	High	Low
EfficientNet	Moderate	Moderate	High

MobileNet is more efficient in terms of energy consumption that is why it is useful in portable diagnostic devices (Table 5). NASNet has a high resource requirement for learning but it gets high performance; on the other hand, EfficientNet is well balanced between performance and efficient use of resources.

5.4.4 Suitability for Skin Lesion Classification

- *ResNet*: Suitable for noisy data sets owing to profound architecture and skip connections.
- *MobileNet*: Able to be used when there are limited resources available but could possibly fail in high complexity lesions.
- *NASNet*: Although the resulting model is a good approximation of the real system, it is best suited to high-performance applications that have access to high-processing power computational resources.
- *EfficientNet*: Sorting option has better accuracy with the optimal utilization of the resources.

6 Conclusion

6.1 Summary of Findings and Relevance to Clinical Applications

In this paper, a skin lesion classification system using a deep learning approach was proposed to differentiate between benign skin lesions such as nevus and malignant melanoma using the ISIC Archive dataset. Thus, applying noise removal and ROI cropping to the input images, as well as using data augmentation to handle noisy and ambiguous data, the model solved the problems of working with noisy and ambiguous data. It proved clinically reliable as it obtained an accuracy of 93 %, F1 measure equal to 90%, and Area Under Curve, AUC 0.50 Results of the study also show that ResNet based architecture offered reliable feature extraction leading to high accuracy. The study also reveals the great advances in the use of automated dermatological diagnosis especially in the early diagnosis of melanoma. When diagnosed on time and accurately, its prognosis is good since melanoma is among the deadliest forms of skin cancer. The high sensitivity of the proposed system to identify malignant lesions means that the number of patients who present with melanoma at an advanced stage is minimized. Also, based on the evaluation of how the model performs in reducing false positives, it can also reduce unnecessary biopsies coupled with cost implications increasing efficiency in the healthcare settings. The implementation of such a system in clinical practice can be seen as a possibility to revolutionize dermatological practices. The proposed automated lesion classification can be used as a decision support system for dermatologists, to help in the screening of a large number of dermatoscopic images. This is relevant for setting within resource-limited environments where the availability of specialists is an issue. Thus, the system can contribute to filling the gaps in diagnostic experience and timely implementation of interventions with the help of consistent and objective assessments.

6.2 Potential Impact on Early Detection and Treatment of Melanoma

The high accuracy of the developed system in classifying malignant lesions demonstrates the potential of this system in changing the approach toward early diagnosis of melanoma. Melanoma is easily treatable when diagnosed early and the five year survival rate is above 90% if the disease has not spread to other parts of the body. But the advanced stage of melanoma is linked with poor outcomes, which underlines the role of early diagnosis. The proposed model is very sensitive in such a way that even minor melanomas are detected so that early treatment can be done to make the patients benefit from it. Furthermore, the proposed system can greatly lighten the burden of dermatologists as it can help in the preliminary screening of patients. This makes it possible for the clinicians to channel their specialty on the risky cases hence enhancing diagnosis and throughput. The stability of the model across the different lesion types also eliminates the variability of human evaluation, thus making the results more uniform. In addition, the real-time processing of large datasets also makes the system ideal for the tel-edermatology application area, where a remote diagnosis is critical. Apart from early diagnosis, the system may be useful in treatment planning. By correct grouping of lesions, it helps the clinicians sort out cases according to the level of malignancy so that patients with high level of lesions are attended to as soon as possible. This prioritization is most important in health care organizations with many patients and few resources because it can mean the difference between life and death. Furthermore, the system is also scalable, and can be used in population-wide skin cancer screening, which may decrease the melanoma prevalence and impact on the health care systems.

7 Recommendations for Future Research

As the study shows the efficiency of the proposed system, several potential research directions can improve the system's performance and utility in the future.

7.1 Improving Dataset Diversity

A key limitation of the current study is lack of variation in the sample set, especially in terms of different lesion types and different groups of people. Future studies should aim at collecting more data with rare lesion patterns and images of patients belonging to different ethic groups and different age. This would enhance the generality of the model across the multi populations to increase its usage in global healthcare systems.

7.2 Developing Hybrid Models

The study established that the ResNet architecture was very resilient; however, the incorporation of hybrid models could improve the performance. CNN s integrated with the transformer-based architecture can take advantage of both local and global features, which can be applied for more accurate classification of the detected lesions. Regarding misclassification, the investigators also noted that the combination of two or more models will help to reduce variance and bias and thus can be useful in ensemble.

7.3 Enhancing Model Interpretability

Deep learning models act as a black box, and it is hard to verify their predictions in the most important scenarios. The future studies should consider using Grad-CAM and SHAP for enhanced interpretability of the results and more comprehensible understanding of model choices. These tools can be used to explain the model's decision making process to dermatologists thus improving the model's credibility and admissibility in clinical practice. Moreover, using interpretability frameworks that include patient's metadata like age, history of a lesion and etc can give more accurate diagnostics.

7.4 Addressing Ambiguous and Noisy Images

The existence of ambiguous lesions and noisy images still poses a great problem in the automated classification. Future studies should explore higher levels of preprocessing including segmentation models like U-Net or DeepLab to get a higher accuracy of the ROI extraction. Other noise-reduction algorithms such as those that can learn the various types of artifacts like hair and glare will also increase image quality. Further, the integration of other sources of information, including histopathological data together with dermatoscopic images, may help in the classification of the cases that are difficult to determine.

7.5 Exploring Real-World Clinical Integration

While the current study used a controlled data set; it is recommended that future studies test the efficacy of the proposed system in real clinical settings. With direct cooperation with dermatology clinics and hospitals, feedback on the application of the created system may reveal its effectiveness in practice. Specific future research objectives on designing and testing the model in real clinical setting include study of its effects on clinical processes, diagnostic performances, and patient outcomes.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
ORCID iD: Shake Ibna Abir[1] (https://orcid.org/my-orcid?orcid=0009-0004-0724-8700), Shaharina Shoha[1] (https://orcid.org/0009-0008-8141-3566)
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1251-1258). https://doi.org/10.1109/CVPR.2017.195
[2] Codella, N. C. F., Rotemberg, V., Tschandl, P., & Celebi, M. E. (2018). Skin lesion analysis toward melanoma detection: A challenge at the 2018 International Symposium on Biomedical Imaging (ISBI). *IEEE Transactions on Medical Imaging, 38*(8), 2525-2534. https://doi.org/10.1109/TMI.2018.2867236
[3] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature, 542*(7639), 115-118. https://doi.org/10.1038/nature21056
[4] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning.* MIT Press.
[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778). https://doi.org/10.1109/CVPR.2016.90
[6] Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L. H., & Aerts, H. J. (2018). Artificial intelligence in radiology. *Nature Reviews Cancer, 18*(8), 500-510. https://doi.org/10.1038/s41586-018-0400-7
[7] Howard, J. P., Rinne, P., & King, M. (2018). Medical image analysis with deep learning. *European Journal of Radiology, 106*(4), 20-25. https://doi.org/10.1016/j.ejrad.2018.10.002
[8] International Skin Imaging Collaboration (ISIC). (n.d.). ISIC Archive: Dermoscopic images of skin lesions. Retrieved from https://isic-archive.com
[9] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. https://doi.org/10.48550/arXiv.1412.6980
[10] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems* (pp. 1097-1105). https://doi.org/10.1145/3065386
[11] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444. https://doi.org/10.1038/nature14539
[12] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van Ginneken, B. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis, 42*, 60-88. https://doi.org/10.1016/j.media.2017.07.005
[13] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015* (pp. 234-241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28
[14] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision, 115*(3), 211-252. https://doi.org/10.1007/s11263-015-0816-y
[15] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. https://doi.org/10.48550/arXiv.1409.1556
[16] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1-9). https://doi.org/10.1109/CVPR.2015.7298594
[17] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818-2826). https://doi.org/10.1109/CVPR.2016.308
[18] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105-6114). PMLR. https://doi.org/10.48550/arXiv.1905.11946
[19] Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience, 2018*, 1-13. https://doi.org/10.1155/2018/7068349
[20] Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *European Conference on Computer Vision* (pp. 818-833). Springer. https://doi.org/10.1007/978-3-319-10590-1_53
[21] Abir, S. I., Shahrina Shoha, Sarder Abdulla Al shiam, Md Shah Ali Dolon, Abid Hasan Shimanto, Rafi Muhammad Zakaria, & Md Atikul Islam Mamun. (2024). Deep Neural Networks in Medical Imaging: Advances, Challenges, and Future Directions for Precision Healthcare . Journal of Computer Science and Technology Studies, 6(5), 94–112. https://doi.org/10.32996/jcsts.2024.6.5.9

[22] Shaharina Shoha, Abir, S. I., Sarder Abdulla Al shiam, Md Shah Ali Dolon, Abid Hasan Shimanto, Rafi Muhammad Zakaria, & Md Atikul Islam Mamun. (2024). Enhanced Parkinson's Disease Detection Using Advanced Vocal Features and Machine Learning . *Journal of Computer Science and Technology Studies*, *6*(5), 113–128. https://doi.org/10.32996/jcsts.2024.6.5.10

[23] Abir, Shake Ibna and Shoha, Shaharina and Dolon, Md Shah Ali and Al Shiam, Sarder Abdulla and Shimanto, Abid Hasan and Zakaria, Rafi Muhammad and Ridwan, Mohammad, Lung Cancer Predictive Analysis Using Optimized Ensemble and Hybrid Machine Learning Techniques. Available at SSRN: https://ssrn.com/abstract=4998936 or http://dx.doi.org/10.2139/ssrn.4998936

[24] S. I. Abir, S. Shoha, S. A. Al Shiam, M. M. Uddin, M. A. Islam Mamun and S. M. Shamsul Arefeen, "A Comprehensive Examination of MR Image-Based Brain Tumor Detection via Deep Learning Networks," 2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, 2024, pp. 1-8, https://doi.10.1109/ICDS62089.2024.10756457

[25] S. I. Abir, S. Shoha, S. A. Al Shiam, M. M. Uddin, M. A. Islam Mamun and S. M. Shamsul Arefeen, "Health Risks and Disease Transmission in Undocumented Immigrants in the U.S Using Predictive ML," *2024 Sixth International Conference on Intelligent Computing in Data Sciences (ICDS)*, Marrakech, Morocco, 2024, pp. 1-6, https://doi.10.1109/ICDS62089.2024.10756308

[26] Abir, Shake Ibna, Richard Schugart, (2024). Parameter Estimation for Stroke Patients Using Brain CT Perfusion Imaging with Deep Temporal Convolutional Neural Network, Masters Theses & Specialist Projects, Paper 3755.

[27] Sohail, M. N., Ren, J., Muhammad, M. U., Rizwan, T., Iqbal, W., Abir, S. I., and Bilal, M, (2019). *Group covariates assessment on real life diabetes patients by fractional polynomials: a study based on logistic regression modeling*, Journal of Biotech Research, 10, 116-125.

[28] Sohail, M. N., Jiadong, R., Irshad, M., Uba, M. M., and Abir, S. I, (2018). *Data mining techniques for Medical Growth: A Contribution of Researcher reviews*, Int. J. Comput. Sci. Netw. Secur, 18, 5-10.

[29] Sohail, M. N., Ren, J. D., Uba, M. M., Irshad, M. I., Musavir, B., Abir, S. I., and Anthony, J. V, (2018). *Why only data mining? a pilot study on inadequacy and domination of data mining technology*, Int. J. Recent Sci. Res, 9(10), 29066-29073.