

Early Detection of Skin Cancer using CNN



Muhammad Yaseen

Registration No: 5-FET/MSBME/F23

Supervisor:

Prof. Dr. Muhammad Amir

**Department of Electrical and Computer Engineering
Faculty of Engineering and Technology International
Islamic University Islamabad**

2025

DISSERTATION

A dissertation submitted to the Department of Electrical Engineering, International Islamic University Islamabad as a partial fulfillment of the requirements for the award of the degree.

Department of Electrical Engineering

Faculty of Engineering and Technology

International Islamic University Islamabad

2025

DEDICATION

I dedicate research work to my beloved parents, respected teachers, siblings and all those who prayed for my success.

“Alhamdulillah for everything, we can never thank Allah enough for the countless bounties He blessed us with”

CERTIFICATE OF APPROVAL

Early Detection of Skin Cancer using CNN

Name of Student: Muhammad Yaseen

Registration No: 5-FET/MSBME/F23

Accepted by the Faculty of Engineering and Technology, International Islamic University Islamabad in partial fulfillment of the requirement of the MS Degree in Biomedical Engineering.

Viva voice committee:

Dean FET, IIU Islamabad

Assistant Professor, Chairman DEE, FET, IIUI

External Examiner

Internal Examiner

Supervisor:

Prof. Dr. Muhammad Amir

DECLARATION

I certify that research work titled “Early Detection of Skin Cancer using CNN” has been completed by me and it has not been done before and presented anywhere for evaluation. Furthermore, I have properly acknowledged the material taken from related sources.

ACKNOWLEDGEMENT

First of all, I would like to thank ALLAH (SWT) for giving me the great family, supportive teachers and co-operative friends. I couldn't finish my Research-work without His help.

This work is completed with the help of two people. I would like to express my gratitude to my supervisor Prof. Dr. Muhammad Amir for his continuous support, motivation, strength and valuable guidance.

In the last, I am very thankful to my beloved parents for their continuous support, unconditional love. They have always given me the moral and spiritual support and motivation to achieve my goals.

Table of Contents

Chapter 1	13
Introduction.....	13
1.1 Human Skin.....	13
1.2 Skin Cancer and its Types	14
1.3 Neural Networks	17
1.4 Feasible Models in Skin Cancer Detection	20
1.4.1 Convolutional Neural Networks (CNNs).....	20
1.4.2 Support Vector Machine (SVM)	23
1.5 CNN in Skin Cancer Diagnosis.....	25
1.5 Contributions Toward Skin Cancer Diagnosis Using CNN	26
1.6 Challenges in Dermoscopic Image Analysis	28
1.7 Our Contributions.....	32
Chapter 2	34
Literature Review.....	Error! Bookmark not defined.
Chapter 3	37
Methodology	37
3.1 Dataset Description	37
3.2 Model Architecture.....	39
3.4 Squeeze and Excitation Block.....	43

3.5	Channel Attention Mechanism	44
3.6	Spatial Attention Mechanism	45
3.7	Grad-CAM XAI Technique.....	46
3.8	Data Preprocessing.....	47
3.9	Simulation Environment	48
Chapter 4.....		49
Results and Analyses.....		49
4.1	Evaluation Metrics	49
4.2	Learning Behavior of the proposed model.....	49
4.3	Results	50
Chapter 5.....		54
Conclusions and Future Work.....		54
5.1	Conclusions	54
5.2	Future Work.....	54
References.....		55

List of Figures

Figure 1: Layers of Skin	14
Figure 2: Different types of Skin Cancer	16
Figure 3: Different types of Neural Networks	18
Figure 4: Application areas of Deep Learning	19
Figure 5: Benefits of using CNN	28
Figure 6: Challenges	32
Figure 7: Flowchart of the methodology	37
Figure 8: Some sample images of the dataset at random.....	39
Figure 9: Block diagram of proposed CNN architecture	40
Figure 10: Training and Validation Accuracy	50
Figure 11: Explanations obtained using Grad-CAM	52
Figure 12: Confusion Matrix	52

List of Tables

Table 1: Architectural summary of the proposed CNN architecture.....41

Table 2: Summary of evaluation metrics51

Table 3: Comparison53

ABSTRACT

Skin cancer is considered one of the most common types of cancer throughout the world, which requires an accurate and effective diagnostic method to improve the outcomes. Early detection of skin cancer is essential, which increases the survival rate with respect to time. This work proposes a Convolutional Neural Network (CNN) based model for detecting skin cancer from dermoscopic images, due to the model's strong image recognition feature/capability. The designed architecture of CS-CNN captures and differentiates subtle patterns and skin lesions texture, which in result enables the cancerous and non-cancerous cell/tissue classification. We have elevated the performance of the model on large, diverse datasets to examine the metrics based on specificity, sensitivity, and accuracy. From the result, we have extracted that our CS-CNN-based approach can achieve high diagnostic accuracy, and explainable artificial intelligence (XAI) techniques are incorporated to understand and build confidence in the decisions made by the model. Our research also mentions the CS-CNN-based approach, which is cost-effective and the solution is interpretable based on XAI techniques for the early detection of skin cancer, which in result supports healthcare providers while making data-driven diagnoses.

Chapter 1

Introduction

1.1 Human Skin

Skin is known as the largest organ of the human body and consists of three layers: Epidermis, Dermis, and Hypodermis shown in figure 1. The first layer, Epidermis includes stratum corneum, stratum lucidum, stratum granulosum, stratum spinosum, and stratum basale. Whereas the outermost layer of the skin is known as Stratum Corneum, which acts as a potential barrier. Thin, clear layer likely to be found on the palm is known as Stratum Lucidum. Stratum Spinosum is responsible for the flexibility and strength to the skin. The layer of the skin which contains all the stem cells is known as Stratum Basale. The second layer, the Dermis, consists of the Papillary layer and Reticular Layer. In the papillary layer, thin capillaries, sensory neurons, and thin fibers are located. Whereas the reticular layer is thicker and the deep layer includes large blood vessels and also includes deep-pressure receptors. The third layer, hypodermis, also known as the Subcutaneous Layer, consists of tissues and fats, which provide insulation and are capable of storing energy. Functions of the skin include protection, regulation, sensation, excretion, synthesis of vitamin D, and immunity. Skin appendages include hairs, nails, and glands (glands include Sebaceous Glands, Sweat Glands, Eccrine Glands, and Apocrine Glands)[1].

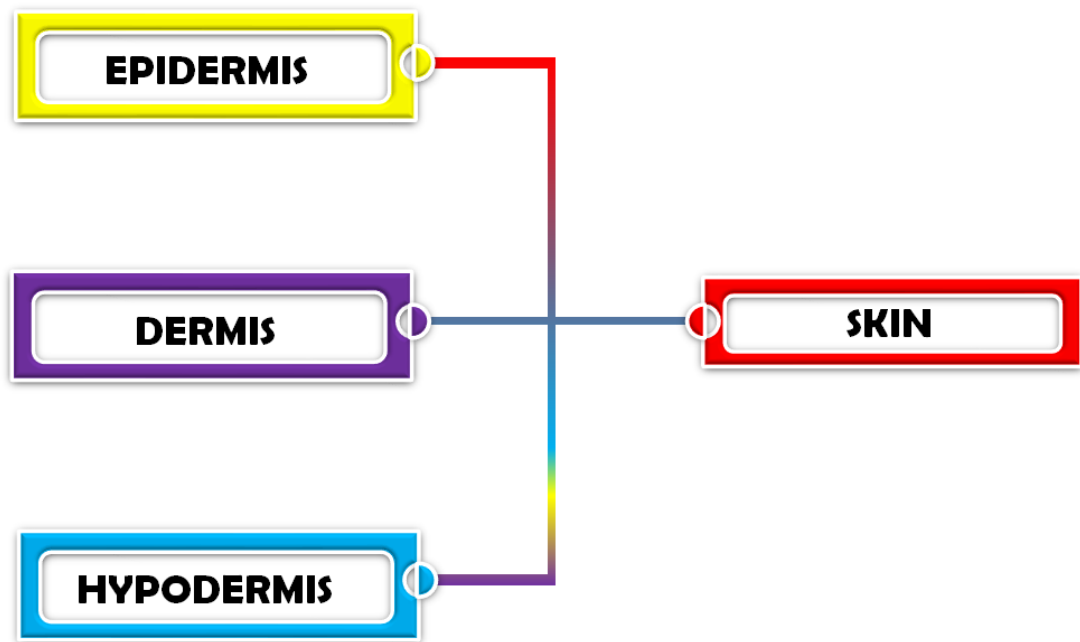


Figure 1: Layers of Skin

1.2 Skin Cancer and its Types

If the skin contains abnormal skin cells that are growing uncontrollably, which is often caused when DNA gets some kind of damage, which is often caused by ultraviolet radiations. Skin cancer is considered one of the most common types of cancer which can be further characterized into three main types which include Basal Cell Carcinoma (BCC), Squamous Cell Carcinoma (SCC), and Melanoma[2]. The most common type of skin cancer is called BCC, it often appears as translucent and pearly bump. Its settlement can be caused by sun-exposed areas such as the face, eyes, ears, nose and neck, etc. Its growth is typically slow but causes local damage if left untreated. The second type of skin cancer is SCC, which originates from the cell located in the epidermis. It appears as firm, red nodule and can be seen as a crusted surface. Its settlement can be caused by sun-exposed areas such as the face, eyes, ears, nose and

neck, etc. It can also occur on membranes and genitals. Its growth rate is much faster than BCC and has a high risk of spreading to other parts of the body. The third skin cancer type known as Melanoma, is considered the most dangerous type of skin cancer. Its origination is from those cells that produce pigments. If a mole exists on the skin, the melanoma can develop from there and appear as a new dark spot and can be characterized by multiple colors, irregular borders, diameter of 6mm (could be larger). It can appear anywhere on the body, it is not necessary to appear on those parts of the body that are exposed to the sun such as under nails, palms and soles of feet, etc. Melanoma is known for its faster growth rate and spreading to other parts of the body making the disease to be detected early and also making the treatment critical. Less common type of skin cancer includes Merkel Cell Carcinoma (MCC) and Kaposi's Sarcoma. The MCC is a rare and aggressive form of skin cancer that originates from the Merkel cell. It appears when the skin is exposed to sunlight and has a fast-growing rate. Whereas Kaposi's Sarcoma is the type that develops from the blood vessels. It appears as brown, red, or purple patches on the skin, commonly on the mouth, skin, and sometimes internal organs. Its settlement is linked to weak immune systems, especially in HIV/AIDS patients, etc. Different types of skin cancer are shown in figure 2. Skin cancer has a lot of risk factors which include UV radiation, weak immune system, age, family history, fair skin and previous skin cancer, etc. Treatment options include Surgical Removal, Targeted Therapy, Immunotherapy, Radiation Therapy, and Chemotherapy. Sun protection, Awareness, and regular skin checkups come if we talk about the prevention and detection of skin cancer[3].

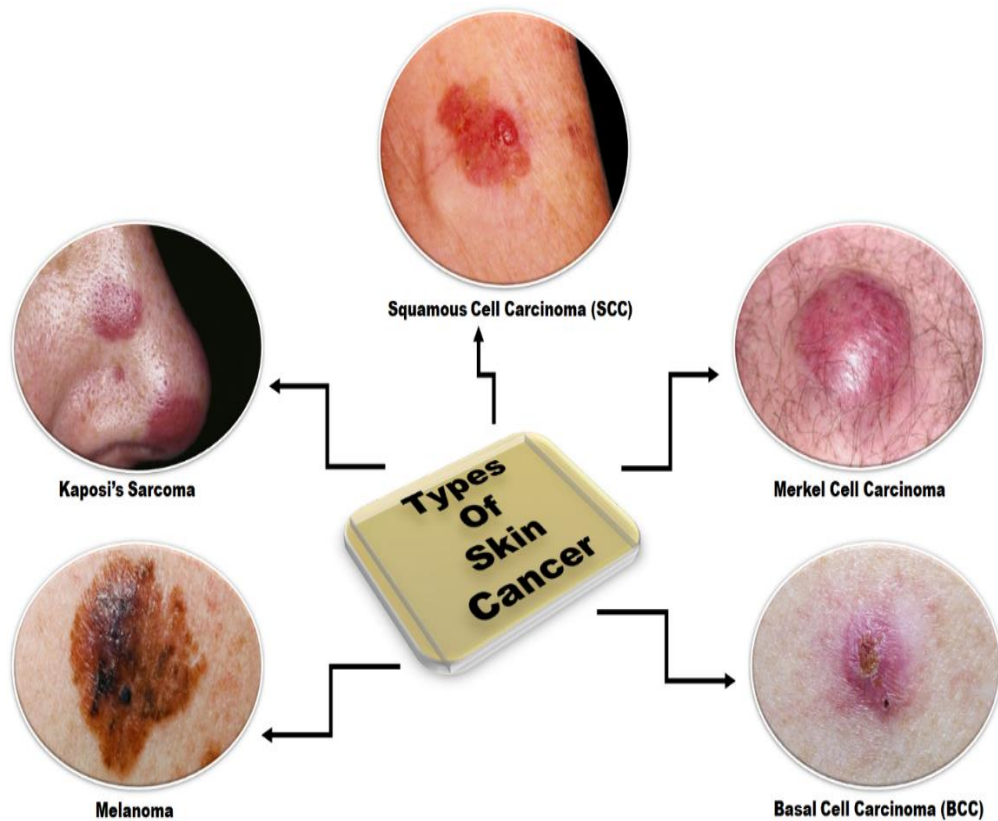


Figure 2: Different types of Skin Cancer

The rate of skin cancer has been rising with time. In the year 2024, the expected cases of melanoma are 200,240 in the United States, with 100,640 being invasive and 99,700 being noninvasive. The number of cases has been raised numerous over the past years. In 2023, the total cases reported were 186,680 in number and 197,700 in 2022. The number of deaths caused by melanoma in the year 2024 is estimated to the number of 8290 deaths, the prevalence of men (5,430) is higher than women (2,870). The spreading rate of melanoma has increased by 16% over the past five years and 32% over the past few decades. In the United Kingdom, skin cancer cases reached the highest record of 17,500 in number. Estimations suggest the number of cases could increase by 50% in the next 20 years.

1.3 Neural Networks

Machine Learning[4] has a subset called Deep Learning[5] which has full concentration on using Neural Networks which consist of many layers and are also capable of understanding data complex patterns. Deep learning is a handy tool to carry out various tasks like natural language processing, image and speech recognition, and autonomous driving. The core concept of deep learning includes Neural Networks and Training Process. Neural networks include Artificial Neurons (These are the basic units of neural networks that take the input and then apply weights and pass the result as an activation function) and Layers (Layers are organized in the network which includes the input as well as the output). The training process consists of Forward Propagation, Loss Function, and Backward Propagation. In forward propagation, data passes through layer by layer in the network. The difference between predicted and estimated values is measured in the loss function. Optimization techniques like gradient descent are applied for loss function minimization after the adjustment of weights, Back propagation is responsible for this process. Neural Network consists of six types: Feed-forward Neural Network (FNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Generative Adversarial Networks (GAN), Autoencoders, and Transformer Networks. FNN is the simplest type of neural network where cycles are not formed with the connection. Used for basic pattern recognition like image classification. Different types of neural networks currently being utilized in the field are shown in **Figure-3**.

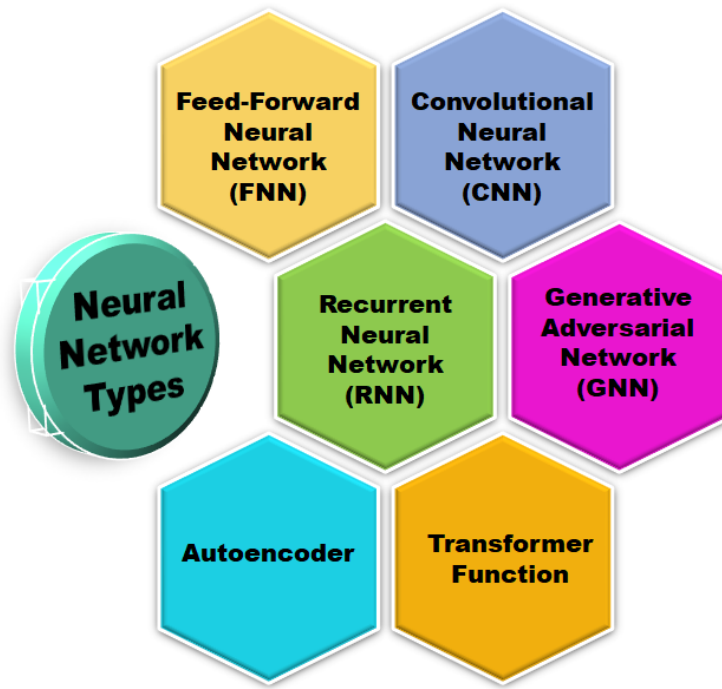


Figure 3: Different types of Neural Networks

CNN[6] is used for the processing of data consisting of grids such as images and is also used to detect features. Components of CNN include Convolutional Layers, Pooling Layers, and Fully Connected Layers. GAN consists of two networks, generator and discriminator which further compete in zero-sum games. Used in case of generation of images and style transfer etc. Autoencoder is specified for tasks based on supervised learning tasks. Components of autoencoder include Encoder and Decoder. Variants of autoencoders include Variational Autoencoders (VAE), and Denoising Autoencoders (DAE). For sequential data handling, Transformer functions are used which consist of an architecture that allows parallel processing. Components of the transformer function include Attention Mechanism and Encoder-Decoder Structure. Its variants include Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT). Advanced topics in deep learning include Transfer Learning, reinforcement learning (Components include Reward Signal, Environment,

and Agents), Self-supervised Learning, Semi-supervised Learning, and Meta-Learning. Applications of Deep Learning (DL) include Intelligent Fault Diagnosis of Rotating Machinery[7], Supply Chain Management[8], Recognizing and Classifying Culinary Dishes in Images[9], Fault Diagnosis of Rotating Machinery using Vibrations Signals[10], Diagnosis of COVID-19 [11] and Medical Image Reconstruction and Enhancement[12] etc. The application areas of deep learning are shown in **Figure-4**.

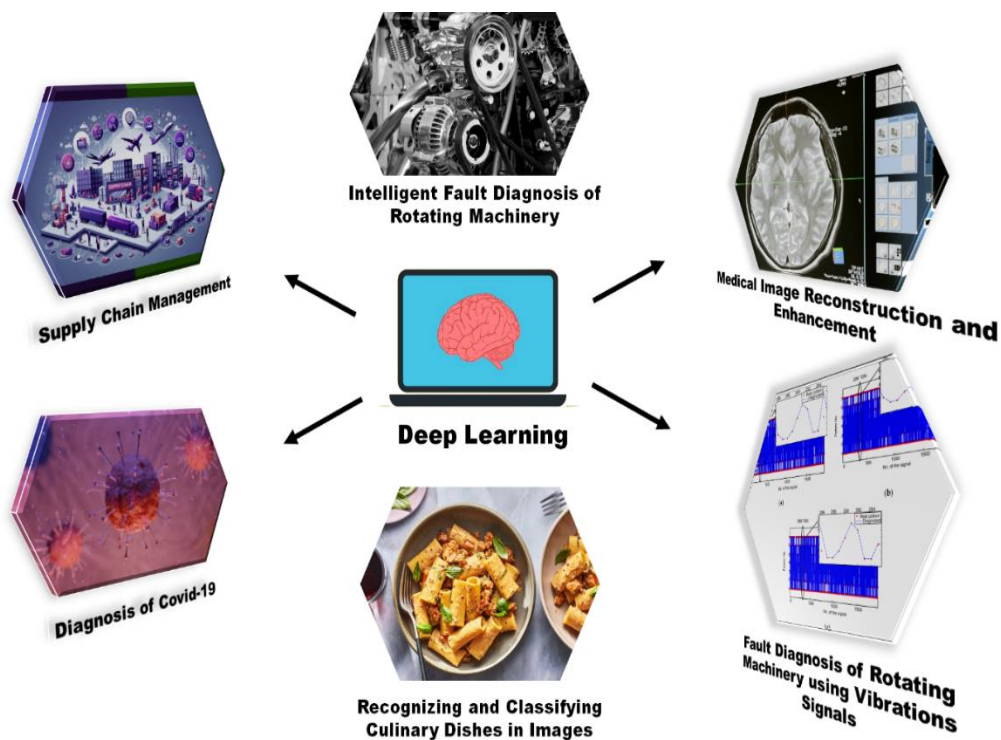


Figure 4: Application areas of Deep Learning

There are neural networks that can examine dermoscopic images of skin lesions and identify the specific patterns and characteristics that may indicate the possibility of skin cancer. This encompasses features such as shape, color, texture, and other visible aspects of the lesions. CNNs[13] extract relevant information from images automatically. Those properties allow to distinguish between benign and malignant lesions without having to choose them manually. Massive data sets containing labeled

pictures (images with known diagnoses) are used in training these neural networks. Such learning process is aimed at making a network sensitive to some visual cues about different types of skin diseases. It has been demonstrated that sometimes, studies could show how neural networks can diagnose cancer better than or as well as experts would do it. For places, where there is limited access to specialists, this is highly important. The early identification of skin cancer is critical for successful treatment. Neural networks can help detect worrisome growths before they become dangerous, improving patient outcomes. In addition to automating screening for skin lesions in clinical settings, the use of these models reduces healthcare professional burnout and increases resource efficiency. As more data becomes available over time, neural networks will get enhanced diagnostic capabilities through exposure to new inputs, which have made them improve in recent years while retaining this trend.

1.4 Feasible Models in Skin Cancer Detection

Many of the machine learning model are quite feasible to use for the detection of skin cancer, each of them has their own significances, features, strength and challenges.

1.4.1 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are most widely used for carrying out image-based tasks because it has the ability to take out features automatically such as texture, edge, and pattern which can be taken out from dermoscopic images. Key features of CNNs includes convolutions (require to extract important features from the image), pooling layers (use to preserve key features from the image while reducing dimensionalities), attention mechanism (which focuses on the most relevant part of image) and advanced architectures which includes ResNet, DenseNet, EfficientNet.

While carrying out tasks using Convolutional Neural Networks (CNNs), which requires a large labelled dataset.

The general mathematics of Convolutional Neural Network (CNN) consist of convolutional operation, activation function, pooling and fully connected layer.

- CNN can be mathematically represented as,

$$P(a,b) = \sum_i \sum_j (a+i, b+j) \cdot K(i, j) \quad (1)$$

Where Z and K represent 2D image and kernel respectively, output position is represented (a, b).

CNN contains strides and padding in which the stride controls the step size of the convolution and padding is responsible for adding extra pixels around the input for controlling the output size.

- Activation function is responsible for adding non-linearities to the network. Commonly used activation functions are,

1. **ReLU:** $f(p) = \max(0, p)$

2. **Sigmoid:** $f(p) = \frac{1}{1 + e^{-p}}$

3. **Tanh:** $f(p) = \frac{e^p - e^{-p}}{e^p + e^{-p}}$

- Pooling layer is used to reduce the spatial elimination of feature map.

1. Max Pooling

$$X(m, n) = \max S(m+a, n+b) \quad (2)$$

2. Average Pooling

$$X(m,n) = \frac{1}{x} \sum_{a,b} S(m+a, n+b) \quad (3)$$

- Fully Connected Layer

$$p = f(Ax + t) \quad (4)$$

Where A represents weight of the matrix and t is for bias.

- Loss function takes out how efficiently prediction of network matches the ground truth.

1. Cross-Entropy Loss (Classification)

$$Loss = - \sum_{x=1}^A b_x \log(K_x) \quad (5)$$

Where b_x and K_x represents true label and predicted probability respectively for class x.

2. Mean Square Error MSE (Regression)

$$Loss = \frac{1}{x} \sum_{a=1}^x (b_x - \hat{b}_x)^2 \quad (6)$$

- Back propagation is used for the training purpose of CNN.

$$X \leftarrow X \leftarrow \eta \nabla p$$

$x, \eta, \nabla p$ represents weights, learning rate and gradient respectively.

- Spatial Attention

$$B = \text{soft max} (f(P)) \quad (7)$$

$f(P)$ generates spatial attention maps.

- Channel Attention

$$D = \sigma(A_2 \cdot \text{ReLU}(A_1 \cdot \text{GAP}(E))) \quad (7)$$

Where GAP is Global Average Pooling, A_1 and A_2 represents learning weights.

1.4.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is widely used to carry out those tasks which are based on binary classification for example distinguishing between benign lesion and malignant. SVM key features includes is its kernel functions (linear, RBF), margins between the classes can be maximized, SVM is quite effective while handling small datasets when they are paired with handcrafted features like texture descriptors etc. SVM is not scalable for large datasets and its performance depends of manual feature engineering.

Support Vector Machine (SVM) belongs to supervised machine learning algorithm which is used for classification and regression purpose. They are used to find the optimal hyperplane which do the task of separating data points in to classes. General mathematics of SVM is shown below:

- SVM do the job of finding a hyperplane $x^a y + b = 0$, which maximizes the margins between classes and other minimizing misclassifications.
- In mathematical formulation, a dataset is given $\{(a_i, b_i)\}_{i=1}^x$, where a_i features vector of the i^{th} sample and $b_i \in \{-1, +1\}$ represents class label of the i^{th} sample.

The decision function is,

$$f(y) = \text{sign}(x^A y + b) \quad (8)$$

Where x is the weight vector and b represents bias term.

- Once the data is linearly separable,

$$b_i(x^A y_i + b) \geq 1, \forall i \quad (9)$$

Equation 9 can be formulated as a convex optimization problem,

$$\min_{x,b} \frac{1}{2} \|x\|^2 \quad (10)$$

Equation 10 can be subjected as,

$$b_i(x^A y_i + b) \geq 1, \forall i \quad (11)$$

- If the data is not perfectly separable, a new variable called slaked variable $\xi_i \geq 0$ is introduced.

$$\min_{x,b,\xi} \frac{1}{2} \|x\|^2 + B \sum_{i=1}^x \xi_i \quad (12)$$

Where B represents regularization parameter which controls the trade-off between maximization and minimizing margins and classification errors.

Equation 12 can be subjected to,

$$b_i(x^A y_i + b) \geq 1 - \xi_i \quad (13)$$

- After the integration of Lagrange Multiplier (β_i),

$$\max_{\beta} \sum_{i=1}^x \beta_i - \frac{1}{2} \sum_{i=1}^x \sum_{j=1}^x \beta_i \beta_j h_i h_j k(a_i, b_i) \quad (14)$$

Where $k(a_i, b_i) = a_i b_i$ represents the Kernel version.

Equation 14 will become,

$$0 < \beta_i \leq B, \sum_{i=1}^x \beta_i h_i = 0 \quad (15)$$

- For the data which is non-linear separable data, the trick kernel maps the input feature features into a higher-dimensional space. Common kernel features are listed below,

1. **Linear Kernel:** $B(a_i, b_i) = a_i^A b_i$
 2. **Polynomial Kernel:** $B(a_i, b_i) = (\alpha a_i^A b_i + c)^e$
 3. **Radial Basis Function (RBF):** $B(a_i, b_i) = \exp(-\alpha \|a_i - b_i\|^2)$
 4. **Sigmoid Kernel:** $B(a_i, b_i) = \tanh(\alpha a_i^A b_i + c)$
- Once the designed model is trained, the function for decision for a new point y is trained,

$$f(y) = \text{sign}\left(\sum_{i=1}^x \beta_i h_i(k(a_i, a) + b)\right) \quad (16)$$

- For penalizing misclassification points the hinge loss function is used in SVM,

$$X = \frac{1}{n} \sum_{i=1}^n \max(0, 1 - a_i(a^A y_i + b_i)) \quad (17)$$

- The methods which are commonly used for solving SVM optimization problems includes Sequential Minimal Optimization (SMO) and Gradient Descent (used for large scale SVM problems).

Summary of key features of SVM includes,

1. Linear SVMs which directly uses $x^A y = b$.
2. Kernelized SVMs is used to map data to higher dimensions.
3. Regularization is responsible for balancing the margins maximization and error minimization.
4. Hinge loss is used to measure classification error during the training duration.

1.5 CNN in Skin Cancer Diagnosis

Convolutional Neural Networks (CNNs) have emerged as potent instruments in the realm of medical image analysis, especially for diagnosing skin cancer. Among numerous deep learning algorithms available, CNNs are preferable for tasks involving

image data because they can learn spatial hierarchies of features automatically. In skin cancer diagnosis, CNNs are trained on large datasets of dermoscopic images labeled with their corresponding diagnoses that can consist of a variety of different types of skin lesions including benign moles and malignant melanomas. This process starts with preprocessing whereby images are standardized to ensure that their sizes and resolutions match; sometimes color normalization is used so that all the images in the dataset have similar appearance. The architecture of the CNN usually entails several layers such as convolutional layers responsible for detecting edges, textures, and shapes and pooling layers reducing computational load through dimensionality reduction to avoid overfitting; fully connected layers bring together these characteristics to obtain final classification. The CNN learns how to differentiate between benign and malignant tumors by recognizing patterns and features during training. This learning occurs through backpropagation and optimization algorithms e.g., stochastic gradient descent which tune the network[13].

1.5 Contributions Toward Skin Cancer Diagnosis Using CNN

The Convolutional Neural Network (CNN)[14] plays an important role in revolutionizing dermatology, because it offers several significant contributions towards the diagnosis of skin cancer. By the use of CNN, a remarkable accuracy is seen while skin cancer detection. By the analysis of dermoscopic and non dermoscopic images, CNN has the capability to differentiate between benign and malignant lesion with high precision. The next significance is the feature extraction, where convolutional neural networks detect important features automatically such as color, shape, texture etc, this feature also reduces the manual feature engineering. By the incorporation of advance architectures with feasible modules such as separable convolutional networks and attention mechanism can further contributes in the incrementation of accuracy and the

ability to capture intricate patterns. Early detection/diagnosis is an important step for treating skin cancer. CNN can detect signs of skin cancer in the early stages, which can contribute a lot in the treatment process also reduces the mortality rate.

Convolutional Neural Network (CNN) has the capability to do analysis of different types of images which includes dermoscopic images, histopathology slides, and clinical photography. Due to which CNN can be considered as an important tool for the detection purpose of skin cancer. CNN contains a feature called Spatial Attention Mechanism[15] which contains a technique named as Squeeze and Excitation Network and Spatial Attention layer, these techniques help convolutional neural network to focus on those regions of images which is most important to extract features, which in result greatly improves interpretability and reduce false positive or negative. CNN also contain handling imbalance data significance in which strategies like data augmentations, transfer learning, and synthetic data generation is including in order to address the imbalance data in the class, which is considered a very common issue in skin cancer detection. By the development of tools based on convolutional neural network on web platforms and mobile can make the easy access toward dermatological services particularly in those regions which are remote, under development region or which has very less access towards healthcare professionals.

Convolutional Neural Networks (CNN) is flexible which means it can be integrated with various other models which includes Long Short-Term Memory (LSTM)[16] for sequential imaging data etc. CNN can be incorporated with technology like Internet of Thing (IoT) for monitoring continuously etc. which comes handy while creating robust diagnostic systems. We can extract performance matrices from Convolutional Neural Networks (CNN) based models like specificity, accuracy, sensitivity, and AUC-ROC.

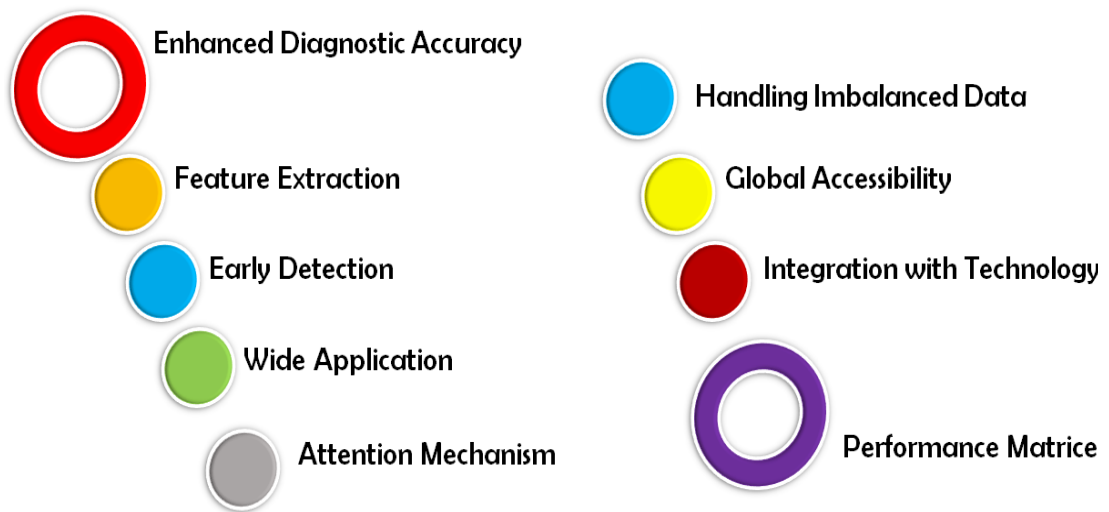


Figure 5: Benefits of using CNN

If we manage to achieve high specificity and high sensitivity provides us reliable predictions which also minimizes the diagnostic predictions mentioned in the figure 5.

1.6 Challenges in Dermoscopic Image Analysis

The challenges in the analysis of dermoscopic images rises due to the complexities of characteristics contained skin lesion, image condition variations and high diagnostic accuracy need.

- **Variability in Skin Lesions**

1. **Diverse Lesion Types:** Skin lesion shows different important variations in different aspects which includes color, size, shape, and texture, which makes the lesion type features generalization more difficult.
2. **Overlapping Features:** Both malignant and benign lesions can show same type of visual characteristics which often leads to misclassification.
3. **Subtle Feature:** Early-stage melanoma which is type of malignant lesions, which contains only subtle differences as compared to benign lesions, also which is quite difficult to detect.

- **Artifacts in Dermoscopic Images**

1. Human hairs often contain important parts of lesion, analysis of which is quite complicated.
2. Dermoscopic images captures under a specific condition might contains light reflections which has the interference with lesion visibility, may also contains bubbles.
3. Marks made by the use of ink or ruler in order to refer during clinical procedures also lesion boundaries can be obscure.

- **Inconsistent Image Quality**

1. There are variations caused by non-uniform condition of lights in contrast and image brightness, which leads toward the feature extraction.
2. Images of different resolutions produced by different dermoscopic image capturing devices results in impacting of analysis of feature.
3. Variations in angle and focus leads can result in blurred or distorted images.

- **Limited Availability of Annotated Data**

1. Require annotation regarding datasets of dermoscopic images, which is expensive and time consuming.
2. Most of the datasets are imbalanced which significantly contains more benign lesions and less malignant lesions, which leads to biased models.

- **High Intra-class and Inter-class Variabilities**

1. In Intra-class variabilities only a single type of lesion can appear in different terms of shape, texture and color.
2. In Inter-class similarities many different type of skin lesions appears visually similar, which leads towards complicating classifications.

- **Complex Lesion Boundaries**

1. Irregular, blurred, or indistinct boundaries often contained by malignant lesions which makes the segmentation process more complicated.
2. Those lesions which are overlapping with healthy skin or other lesion types may diverts the focusing of segmentation algorithms.

- **Diverse Skin Type Generalization**

1. Those models which are trained on limited datasets might not generalize well while dealing with different skin colors.
2. Some datasets refer features of specific ethnic groups which leads towards biased results in under representable populations.

- **Real-time Processing Requirement**

1. For integrating of analysis of dermoscopic images into clinical workflows, real-time analysis is essential in order to require optimization of algorithm.
2. Resource-intensive models might not work properly especially on those devices which are used in clinics which are portable.

- **Lack of Explainability**

1. Clinicians need an interpretable result in order to trust diagnosis automatically. Many of the designed models act as black boxes, which makes complications while understanding why a specific decision is made.
2. Creating of heatmaps or attention maps which gets align with clinical reasoning is challenging.

- **Noise and Augmentation Challenges**

1. Noise can be introduced by poor image conditions which reduces the accuracy of feature extraction.
2. To make the dataset limited, while using data augmentation the images might not be captured fully which competes to real world scenarios.

- **Integration into Clinical Practice**

1. Those models which are trained in a controlled environment might not show the best performance in real world clinical environment.
2. If the automated system conducts misdiagnosis, leads towards serious complications which requires stringent validation and approved processes.

Challenges are shown in figure 6. For addressing these challenges which require robust preprocessing techniques, advance model architectures and incorporation of specific domain knowledge in deep learning model architectures like proposed CS-CNN.

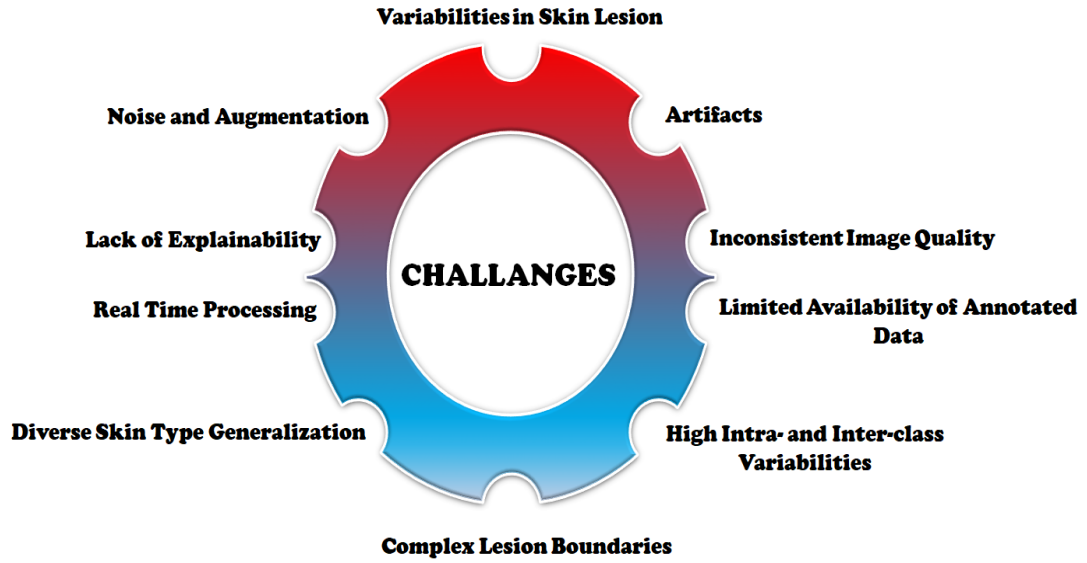


Figure 6: Challenges

1.7 Our Contributions

The contributions of this research work are summarized as follows:

1. Design of an enhanced CS-CNN model for the classification of skin lesions as benign or malignant.
2. Integration of unexplored feature extraction mechanisms based on the latest features such as separable convolution, depth-wise convolution, spatial attention, and channel attention techniques into CS-CNN architecture to enhance accuracy.
3. Focus on maintaining diagnostic speed alongside enhanced accuracy hence neglecting any sort of trade-offs.
4. Integration of the model's interpretability based on the GradCam XAI technique which enhances the confidence of the user in the model's decision.

5. Achieving 96.74% accuracy, surpassing all the standard counterparts mentioned in the literature.

Chapter 2

Related Works

Convolutional Neural Networks can learn directly from the data available, also CNN is used widely for image classification. CNN is considered one of the efficient machine learning algorithms for the analysis of data which has the grid-like structure e.g. images.

Syed Qasim Gillani[17] incorporated deep spiking neural networks based on the VGG-13 model, because of their power-efficient behavior and they utilized the surrogate gradient descent (SGD) method for the classification of melanoma and non-melanoma 3670 and 3323 images respectively. ISIC 2019 is the dataset that took part in the proposed work. By the use of the spiking neural network, they were able to achieve 89.57% accuracy and with 90.07% F score. Syed Inthiyaz[18] has proposed a Convolutional Layer Network automated method on images for the detection of skin problems by the use of machine learning classification. Res-net 50 is a type of neural network, which consist of 50 layers which is then used for the classification of images. Support Vector Machine (SVM) classifier is incorporated in this study. Their proposed model extracts the best features from the images of skin after that the image is then classified using the SoftMax classifier which is considered a highly accurate classifier. Their proposed model is capable of acquiring an accuracy of 87%.

Duggani Keerthana[19] they have proposed two novel hybrid models based on CNN. For the classification purpose of dermoscopic images (either benign or melanoma lesions), they have applied the SVM classifier at the output layer. The first CNN model is there for the extraction of features and the second CNN feeds the features extracted to the SVM classifier for further classification. The dataset used in the proposed work is ISBI 2016 and attained an accuracy of 87.43%. Maryam Tahir[20] has proposed a

skin care classification model based on deep learning (DSCC-Net) which is based on CNN, three datasets are taken into consideration during this work ISIC 2020, DermIS, and HAM10000. They have taken Inception V-3, EfficientNet-B0, Vgg-16, MobileNet, Vgg-19, and ResNet-152 as a baseline model which has an accuracy of 91.28%, 89.46%, 89.12%, 91.46%, 92.51%, and 89.68% respectively. However, the proposed network DSCC-Net has an accuracy of 94.17%, 93.93% F1 score, precision of 94.28%, and recall of 93.76% while categorizing four different types of skin cancer. Three of the datasets were used, ISIC2020, HAM-10000, and Dermis.net. To tackle the issues of minority classes that have an existence in the dataset they utilized the SMOTE Tomek.

Ghadah Alwakid[21] has proposed a deep learning model for the extraction purpose of lesion zone with precision. At first, for the improvement of image quality, they used Enhanced Super Resolution Generative Adversarial Network (ESRGAN)[22]. From the full image for the segment Region of Interest (ROI), the segmentation process is used. For the analysis of the image, ResNet-50 (modified) along with CNN is used. They targeted seven different kinds of skin cancer taken from the dataset of HAM10000, which contains the total images of 6705 that are benign, 2197 unknown lesions images, and 1113 images that are malignant. Their proposed deep learning model achieved an accuracy of 86%, precision of 84%, F-1 score of 86%, and recall of 86%. Md Shahin Ali[23], they have introduced a DL model based on CNN for the classification with accuracy between malignant skin and lesions and benign. At first, they applied a kernel filter for noise removal then they focused on feature extraction and image normalization for the classification to be carried out accurately. At last, for the increment in the number of images, the augmentation process is applied for the improvement in classification rate in case of accuracy. They have also compared their proposed model with other transfer learning models which include: VGG-16, MobileNet, ResNet, AlexNet, and

DenseNet, etc. HAM10000 dataset is used through which they have achieved an accuracy of 91.93%.

Jairo Hurtado[24] proposed a novel approach and came up with a new skin cancer classification system. Their system works on the images that are taken from a standard camera and is able to study the impact on results based on smooth bootstrapping, which is then utilized in the augmentation of the original dataset. They utilized eight different classifiers along with different topologies which include SVM, ANN, and KNN, further, they were compared with and without the data augmentation. They have used the ABCD rule (A for Asymmetry, B for Borders, C for Color, and D for Dermoscopic Structures). According to the results, they got the outcome that ANN is the classifier with the highest performance rate and is considered the most balanced one with data augmentation. By the results, they achieved an accuracy of 87.1% which is considered the improved accuracy from 84.3% of ANN when trained on the original dataset. Gaouda et al.[25] has utilized ESRGAN[22] for synthetic image generation to increase the size of datasets for the training purpose of CNN to classify skin images. They have trained the CNN on the dataset called ISIC 2018 and were able to achieve 83.2% accuracy, which is quite comparable with the achievements of complex networks e.g. ResNet-50, Inception V-3, and Inception Res-net[26]. They have also tested their work on small datasets by using 3533 images from ISIC 2018. They have acquired an accuracy of 85%, which they have obtained by using Inception-50 which is quite low. The target is the improvement of diagnostic accuracy of skin cancer that's why we use machine in case of skin cancer disease.

Chapter 3

Methodology

This section provides detailed information on the methodology employed during the research work. This includes all the aspects from the dataset description in detail along with sample dataset images, the detailed architecture of the proposed model along with the block diagram of the model and parameter details, the data preprocessing which includes normalization and resizing, the detail on hyperparameters of model training. The flowchart of the methodology is shown in **Figure 7**.

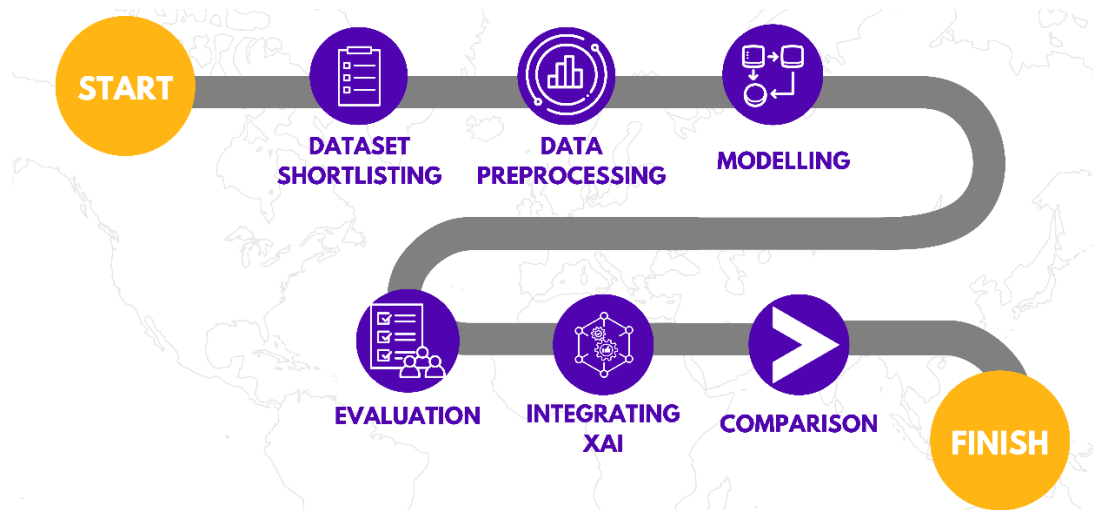


Figure 7: Flowchart of the methodology

3.1 Dataset Description

A top-notch collection of dermoscopic images, the HAM10000 dataset (Human Against Machine with 10,000 training images) is frequently used for the creation and assessment of machine-learning models in the classification of skin lesions. Melanocytic nevi (NV), melanoma (MEL), benign keratosis-like lesions (BKL), basal cell carcinoma (BCC), actinic keratoses intraepithelial carcinoma (AKIEC), vascular

lesions (VASC), and dermatofibroma (DF) are among the seven common diagnostic categories represented in the 10,015 images of pigmented skin lesions. The dataset is appropriate for both binary and multiclass classification problems because these categories encompass a variety of skin disorders commonly seen in dermatology. The photos, which were gathered from a variety of sources, such as clinics in Austria and Australia, represent a range of ethnic backgrounds and imaging circumstances, which improves the generalizability of the dataset. For training and testing artificial intelligence models, each image is labeled with expert dermatological validation, guaranteeing diagnostic accuracy and dependability. The HAM10000 is perfect for creating CNNs that can generalize well because its high-resolution images include natural imperfections like hair and shadows, which present difficulties typical of real-world dermoscopic imaging. Because HAM10000 contains a balanced mix of benign and malignant lesions, researchers commonly use it to train segmentation algorithms like U-Net or classification models like ResNet and EfficientNet. Due to its open availability under an academic license, the dataset has become a standard for classifying skin lesions, with clinical applications emphasizing ethical concerns about patient privacy and data security. As a result, the HAM10000 dataset is essential to the development of dermatological AI, early skin cancer diagnosis, and life-saving diagnostics through medical technology innovation. Some sample images of the HAM10000 dataset are shown in **Figure 8** below.

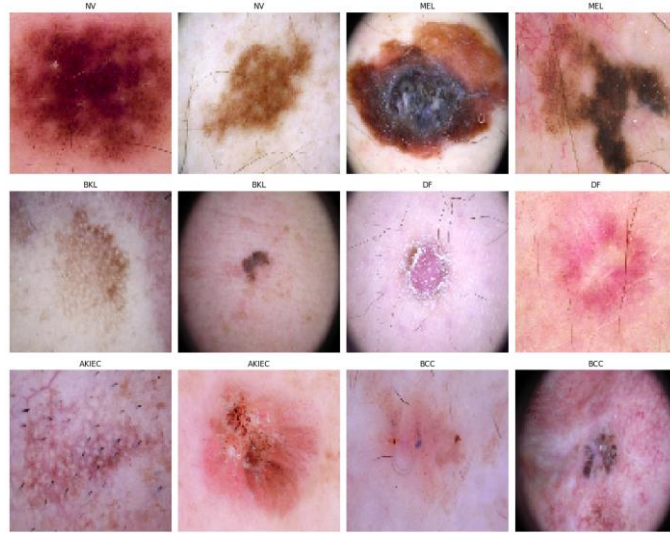


Figure 8: Some sample images of the dataset at random

3.2 Model Architecture

The proposed CS-CNN model is designed using several sophisticated parts to improve its capacity to classify images. It begins with an input layer that takes in images at dimension $96 \times 96 \times 3$ to represent RGB images. The architecture is designed to comprise several convolution blocks with certain types of convolution layers, such as separable and depth wise. These layers are more or less optimized for extracting highly detailed features from images. The architectural block diagram is shown in **Figure-9**.

For the first block, Separable Convolutions are deployed, after which is Batch Normalization to help stabilize the learning process as it normalizes the layer's inputs. Next, a channel attention block is incorporated to further improve the model's learning of key features as it re-scales the channels of the output of the convolution process. This block ends with MaxPooling, which is a technique that properly down samples out spatial dimensions but extracts important features.

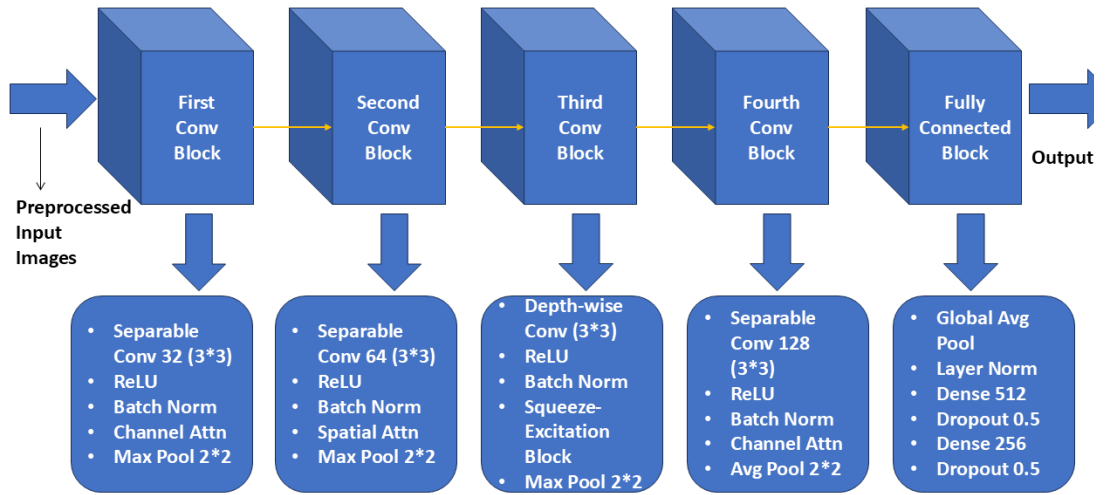


Figure 9: Block diagram of proposed CNN architecture

The second block is also built in the same way, but it has a spatial attention mechanism in its structure. There are two parts in this spatial attention layer. To highlight some important parts of the image, both the average pooled and the maximally pooled features are used. Just like the second block, the third block also includes MaxPooling, DW Convolution, and Squeeze-and-Excitation (SE) blocks. The SE block provides scale factors to different fields within the feature maps and resets weights to amplify helpful image features.

In the fourth convolutional block, the model reverts to Separable Convolutions with channel attention and Average Pooling to further lessen dimensions down to the final fully connected layers. A Global Average Pooling layer then integrates spatial information which is compact to dense layers. This is then passed through Layer Normalization to help training stability, two fully connected ReLU layers, each of which is then passed through Dropout layers to help counter overfitting. Finally, to determine the output of the model, the output layer makes use of a SoftMax activation for binary classification.

3.3 Model Training

This model is trained using a Nadam optimizer characterized by the features of the adaptive learning rate, and categorical cross entropy as the loss function because of the classification problem. This architecture offers low latency or in other words, has the best efficiency when accuracy is counted as a performance measure, uses attention mechanisms, and offers several types of normalizations, which makes it desirable for demanding image classification tasks. The architectural detail is summarized in the following **Table 1**.

Table 1: Architectural summary of the proposed CNN architecture

Layer Type	Details	Output Shape
Input	Input layer with shape (96, 96, 3).	(96, 96, 3)
Separable Convolution 1	32 filters, kernel size (3x3), activation: ReLU, padding: "same".	(96, 96, 32)
Batch Normalization 1	Normalizes activations.	(96, 96, 32)
Channel Attention Block	Enhances feature maps using channel attention.	(96, 96, 32)
MaxPooling 1	Pool size: (2x2).	(48, 48, 32)
Separable Convolution 2	64 filters, kernel size (3x3), activation: ReLU, padding: "same".	(48, 48, 64)
Batch Normalization 2	Normalizes activations.	(48, 48, 64)

Spatial Attention Block	Highlights spatially significant features.	(48, 48, 64)
MaxPooling 2	Pool size: (2x2).	(24, 24, 64)
Depthwise Convolution	Kernel size (3x3), activation: ReLU, padding: "same".	(24, 24, 64)
Batch Normalization 3	Normalizes activations.	(24, 24, 64)
Squeeze-and-Excitation Block	Global average pooling, dense layers, and feature scaling.	(24, 24, 64)
MaxPooling 3	Pool size: (2x2).	(12, 12, 64)
Separable Convolution 3	128 filters, kernel size (3x3), activation: ReLU, padding: "same".	(12, 12, 128)
Batch Normalization 4	Normalizes activations.	(12, 12, 128)
Channel Attention Block	Enhances feature maps using channel attention.	(12, 12, 128)
Average Pooling	Pool size: (2x2).	(6, 6, 128)
Global Average Pooling	Reduces feature maps to a single vector per channel.	(128,)
Layer Normalization	Normalizes the global feature vector.	(128,)
Dense 1	512 units, activation: ReLU.	(512,)
Dropout 1	Dropout rate: 50%.	(512,)
Dense 2	256 units, activation: ReLU.	(256,)

Dropout 2	Dropout rate: 50%.	(256,)
Output Layer	Dense layer with 2 units, activation: SoftMax (for binary classification).	(2,)

3.4 Squeeze and Excitation Block

For representation power enhancement in CS-CNN, the SE block is used for the designing of the model by keeping under consideration the important feature channel. At first, they came in Squeeze and Excitation Network (SENet), the ImageNet 2017 competition winner. The goal of SE block is its ability to concentrate on relevant features by utilizing its feature known as adaptive recalibrating feature maps in CNN. In simple words, it learns the important features contained by each channel and then enhances them accordingly. This enhances the useful feature channel by focusing on an important one other than the one with less importance. The SE block consists of Squeeze and Excitation steps. In the first Squeeze step which is also called Global Information Embedding, in which a feature map is taken as an input by the block for example from a layer of CNN a 3D tensor that has the dimensions of $H \times W \times C$, in which H and W refer to height and width of the map featured where C refers to the number of channels. Squeeze (Global Information Embedding) attempts to perform global average pooling which incorporates height and width which produces a single number for each channel. In result, a vector is generated by utilizing information globally whose dimension will be $1 \times 1 \times C$, which provides summary of information for each feature channel. The second step is called Excitation also known as Self Gating, in which the vector obtained in the first step is then passed through a two-layer fully connected (FC) network which also has non-linear activation, normally ReLU is used which is also followed by sigmoid activation. For describing the importance of each

channel, the output is taken in the form of sets consisting of weights, which range from 0 and 1 which provides a vector for scaling for each channel. In the end, the original feature map of each channel is then multiplied by its weight from the excitation vector, which in result calibrates the feature map effectively to suppress channels which will be based on their importance. The SE block's importance is to focus on those relevant channels which will be based on their importance which will come in handy to enhance the network ability which will be helpful to make predictions more accurate. By the use of the SE block, the overall performance of the model is enhanced with minimal computations, as it adds up only a few of the parameters from the fully connected layers. As a result, we can say that by adding SE block to the CS-CNN, the designed model will set its priority on those channels that contain the critical information, which will help boost the overall performance while carrying out tasks like image classification, detection of objects, etc.[27].

3.5 Channel Attention Mechanism

The Channel Attention Mechanism technique is mainly used in deep learning architecture, most importantly in CNNs to make the model/network focusing feature while dealing with different channels in the feature map. The CNN's channels contain different views of the input data in which it captures various aspects of the image and signal which contains color, texture and edges, etc. The addition of a channel attention mechanism in the model architecture helps to emphasize the model selectivity or to suppress the channel to improve performance while carrying out tasks like detection, segmentation, and image classification. There are three important/key steps in the channel attention mechanism which include Feature Aggregation, Channel-Wise Excitation, and Rescaling. In the first step, the mechanism gathers spatial information across each channel. This goal is commonly achieved through global max pooling

(GMP) and global average pooling (GAP), through which the value of each channel is reduced to a summarized single value. By doing so, we achieve a compact descriptor of each of the channel's overall importance. In the second step, the gathered features are then passed through a set of fully connected dense layers, which is followed by non-linear activations like ReLU, which is then used for the channel-wise important score computation. These extracted scores are then responsible for showing how relevant each of the channels is for the tasks that are carried out. After that, the model proceeds to the third and final step, the rescaling in which the channel scores that are taken out during channel-wise excitation are used to assign weights to each channel in the feature map. To bring each score between 0 and 1, a sigmoid activation is applied and each channel is then multiplied by its respective score, which highlights the important channel and reduces the influence of the less important one. Benefits of the channel attention mechanism include feature emphasis, improved accuracy, and efficient learning. Examples of channel attention mechanisms include SE Block[28]. The general mathematics is shown below,

3.6 Spatial Attention Mechanism

To carry out tasks in computer vision, we use the Spatial Attention Mechanism which helps our model to focus on the relevant spatial locations inside an input such as image. This is done by assigning different levels of importance to different regions. This type of mechanism is used especially in CNN, this mechanism is used for object detection, segmentation, and image classification tasks. The main task of the spatial attention mechanism is to boost the model's ability to attend to parts of an image taken as input which is the most important for a given task, this helps the model to focus on those areas that have more useful information such as the face in a portrait for recognition/authentication instead of treating all the parts of image. The spatial

attention mechanism takes a feature map as input from the convolutional layer, and that input will be based on parameters like width, height, depth, etc. After taking the input, the mechanism moves toward the channel-wise summation in which the mechanism normally merges the channel dimensions which was taken as input before, and this is done by using max pooling or average pooling like operations to generate a 2D attention map. After the map generation, it is then passed through one or multiple layers of convolutional network (normally 1×1 convolution) to purify the weight's attention across spatial locations. The output is the map of attention which shows the importance of each spatial location. After that, the product of the map element-wise with the original featured map is carried out to emphasize specific spatial locations. After that, the feature-weighted map with special location is enhanced, which is then used in the network consisting of subsequent layers, which also helps the model to focus on those critical areas that are more informative. The benefits of the spatial attention mechanism include improved focusing ability, efficient learning, and flexibility. Applications include detection of objects, image segmentation, and image classification. In other words, spatial attention mechanisms boost CNN by focusing on those spatial features that are most important to carry out a certain task[29].

3.7 Grad-CAM XAI Technique

Grad-CAM stands for Gradient Weighted Class Activation Mapping which is an explainable artificial intelligence technique, which is responsible for explaining visual predictions which is made by deep learning models, particularly convolutional neural networks. In the first step, Grad-CAM identifies the target layer which uses the last layer of CNN since it extracts useful information for finding the location. In the second step, the gradient of the output related to the featured maps is calculated. These gradients represent each pixel in the featured map for prediction. After that, the average

of the gradients is carried out to produce a weight for each map featured. This map represents how important the feature map is to class prediction. After that, the sum of weights is calculated which will be called a heatmap which in result marks the area in the image that is taken as input which impacts the model decision most. At the end of processing, a heat map is generated which is then superimposed on the image taken as input, which shows different parts of the image that are more feasible to carry out the predictions. Advantages of the Grad-CAM technique include interpretability, debugging, trustworthiness, etc. Applications include medical imaging, autonomous vehicles, and content moderation, etc. For the detection of skin cancer using CNN, the Grad-CAM technique is quite useful. By applying it in the trained CNN model, it can be verified that the model's ability to focus on the relevant region in skin cancer, like lesions instead of searching irrelevant areas, which will result in the enhancement of the designed model[30].

3.8Data Preprocessing

The HAM10000 skin cancer dataset was pre-processed in this study for skin cancer classification precisely for the 'Malignant' and 'Benign' classes of the diseases. These two classes were the only ones of interest in this analysis so the pre-processed dataset was then filtered to only contain these classes thus forming a binary classification problem. As a bit of pre-processing, the 'Malignant' and 'Benign' classes were given numeric values of 0 and 1 in a new column. Finally, each image was then loaded, resized to a 96x96 pixels image, and normalized so that the input into the model has equal dimensions. To avoid misinformation, some images that were not retrievable were not included in an attempt to work with reliable data.

From the filtered dataset, a random sampling technique stratified as above was adopted to form the training set, validation set, and test set in the ratios of 8:1:1 respectively. Since the nature of the dataset was imbalanced between the two classes, to overcome the class imbalance problem in the training data, the Random Over Sampler technique was used, which creates a completely new set of samples out of the minority category to balance the feature. The above-resampled images were resized, renamed, and changed back to their original format as required for other processing.

Last but not least, to validate our resampling, we applied transformation on a randomly selected subset of the images so that all the classes are represented in the final training dataset. The dataset has been prepared for training and balanced preprocessed for better accuracy in the Identification of malignancy from the processed images.

3.9 Simulation Environment

In this study, the CS-CNN model for the classification of skin lesion images is designed and validated in a simulated environment. TensorFlow v2.14.0 and the Keras API in Python are used to implement the simulation framework. Google Collaboratory is used as a coding environment and is outfitted with a 12th Gen Intel(R) Core (TM) i5-1235U 1.30 GHz processor, 8GB of RAM, and an NVIDIA-SMI 525.105.17 Driver Version: Version 12.0 (CUDA) graphics processing unit.

Chapter 4

Results and Analyses

4.1 Evaluation Metrics

Our proposed CS-CNN model is evaluated using a wide range of assessment criteria to gain a comprehensive understanding of the model's efficacy and performance. The accuracy and other characteristics that clearly convey the information about the model's performance are included in these evaluation measures. The evaluation metrics employed are accuracy, precision, recall, F1 Score, AUC-ROC, Sensitivity, and Specificity.

4.2 Learning Behavior of the proposed model

The proposed CS-CNN model is trained in a simulated environment. After extensive hyperparameter tuning, experimentation, and observations, the optimized hyperparameters turn out to have a learning rate of 0.001, batch size of 32, loss monitored is categorical cross entropy and the training process is carried out for 100 iterations. The accuracy and loss are monitored at every iteration and plotted in the figure which shows the optimal learning behavior of the model. The bias demonstrated by the model is 0.559 and the variance comes out to be 0.154 so there is no suspicion of overfitting or underfitting. The training accuracy is 0.9845, validation accuracy is 0.9550, and validation loss is 0.2162. The epoch-wise learning behavior of the model is shown in **Figure 10**.

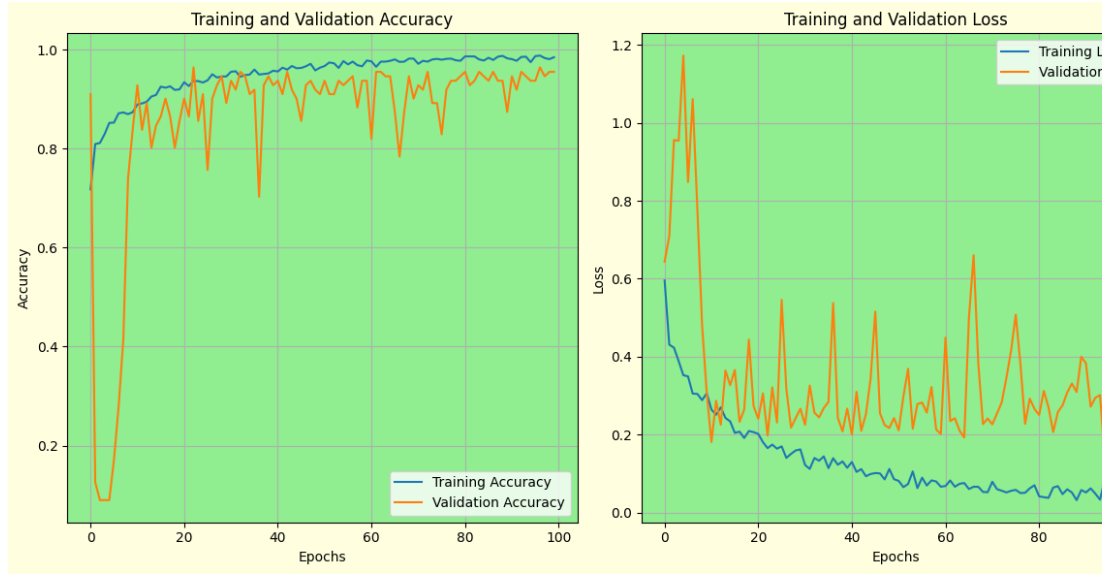


Figure 10: Training and Validation Accuracy

4.3 Results

As mentioned in the methodology section, various benchmark features such as separable convolution, Depth-wise convolution, Squeeze Excitation attention, spatial attention, and global average pooling techniques are incorporated to develop a sophisticated feature extraction mechanism in the CS-CNN architecture. Moreover, for the classification part, the Nesterov Momentum-based optimizer that concatenates the adaptability of the learning rate based on the momentum to find the global minima, the Nadam optimizer is exploited to compile the model and effectively enhances the accuracy. Overall, the advancements in the feature extraction module and the classification module have led to the best performance on the benchmark dataset for the classification of skin lesions. The accuracy of 0.9674 is achieved which surpasses the standard counterparts in the literature, sensitivity of 0.8374, AUC-ROC of 0.6417, and specificity of 0.84 showing the excellent classification ability of our proposed model. All the results are summarized in **Table 2** shown below.

Table 2: Summary of evaluation metrics

Metric	Value
Training Accuracy	0.9845
Validation Accuracy	0.9550
Test Accuracy	0.9674
Training Loss	0.0559
Validation Loss	0.2162
Test Loss	0.1542
Bias	0.0559
Variance	0.1542
AUC-ROC	0.6417
Sensitivity	0.8374
Specificity	0.84
F1 Score	0.8374

As discussed above, the Grad-CAM technique was used to incorporate the explainability or the interpretability of the model. The region of interest in which the malignancy lies is highlighted in purple color. The explanations obtained using Grad cam are shown in **Figure-11**. Finally, the confusion matrix of the model is presented in **Figure-12**. A comparison of the proposed technique with the other state-of-the-art is shown in **Table 3**.

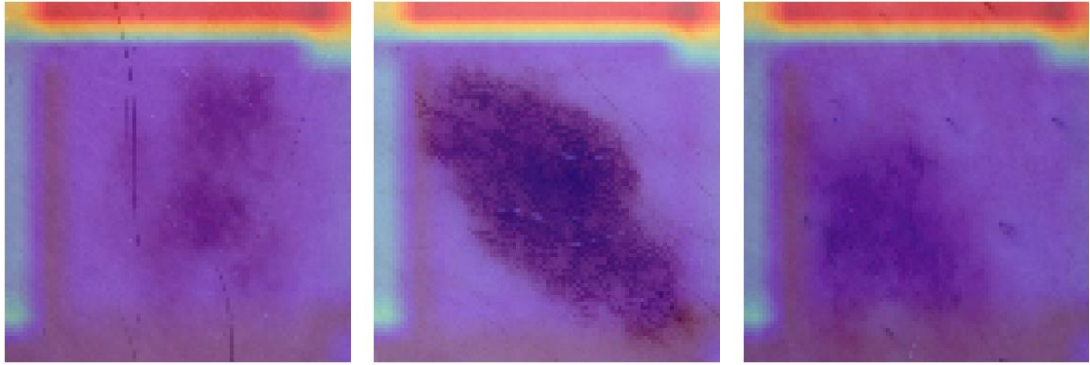


Figure 11: Explanations obtained using Grad-CAM

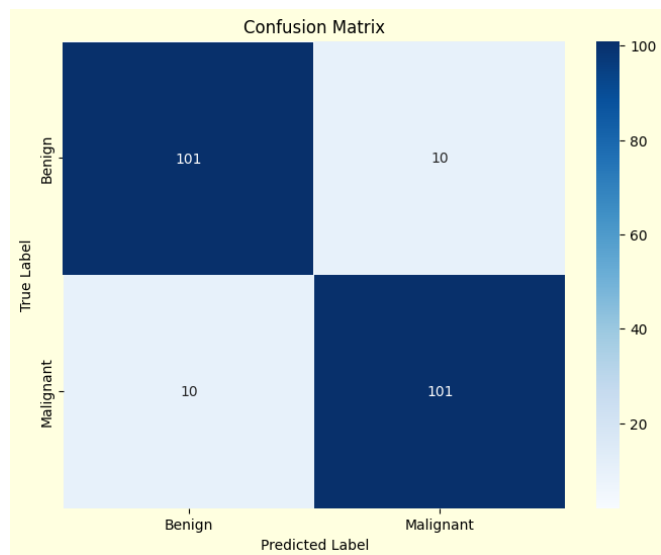


Figure 12: Confusion Matrix

Table 3: Comparison

Ref	Technique	Accuracy	CS-CNN Accuracy 96.74%
14	DSNN with GDM	89.57%	
15	Image Classification CNN	87%	
16	HCNN with SVM	87.43%	
17	DL (Multi Classification)	94.17%	
18	ESRGAN, CNN	86%	
20	DCNN	91.93%	
21	KNN, ANN and SVM	87.1%	
22	DL, CNN and ESRGAN	83.2%	

Chapter 5

Conclusions and Future Work

5.1 Conclusions

The proposed CNN-based model for the detection of skin cancer explains a significant improvement in the classification and feature extraction, which is capable of achieving high accuracy and robustness compared to existing benchmarks. The involvement of squeeze excitation attention, spatial attention mechanism, and global average pooling has contributed a lot in the case of feature extraction capability. The proposed model achieved an accuracy of 0.9674, sensitivity of 0.8374, AUC-ROC of 0.6417, and specificity of 0.84. These metrics prove the efficiency of the proposed model which will emphasize its potential as an important diagnostic aid in the field of dermatology.

5.2 Future Work

Future work can be explored in case of dataset expansion, real-time implementation, hybrid architecture, and cross-domain testing, by addressing these areas future work can further be taken out, particularly in skin cancer detection.

References

- [1] J. A. McGrath, R. Eady, and F. Pope, "Anatomy and organization of human skin," Rook's textbook of dermatology, vol. 1, pp. 3.2-3.80, 2004.
- [2] K. Lacy and W. Alwan, "Skin cancer," Medicine, vol. 41, no. 7, pp. 402-405, 2013.
- [3] U. Leiter, U. Keim, and C. Garbe, "Epidemiology of skin cancer: update 2019," Sunlight, Vitamin D and Skin Cancer, pp. 123-139, 2020.
- [4] E. Alpaydin, Introduction to machine learning. MIT press, 2020.
- [5] J. Kufel et al., "What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine," Diagnostics, vol. 13, no. 15, p. 2582, 2023.
- [6] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE transactions on neural networks and learning systems, vol. 33, no. 12, pp. 6999-7019, 2021.
- [7] Z. Zhu et al., "A review of the application of deep learning in intelligent fault diagnosis of rotating machinery," Measurement, vol. 206, p. 112346, 2023.
- [8] F. Hosseinnia Shavaki and A. Ebrahimi Ghahnavieh, "Applications of deep learning into supply chain management: a systematic literature review and a framework for future research," Artificial Intelligence Review, vol. 56, no. 5, pp. 4447-4489, 2023.
- [9] I. Tvoroshenko, V. Gorokhovatskyi, O. Kobylin, and A. Tvoroshenko, "Application of deep learning methods for recognizing and classifying culinary dishes in images," 2023.
- [10] B. A. Tama, M. Vania, S. Lee, and S. Lim, "Recent advances in the application of deep learning for fault diagnosis of rotating machinery using vibration signals," Artificial Intelligence Review, vol. 56, no. 5, pp. 4667-4709, 2023.

- [11] Y. H. Bhosale and K. S. Patnaik, "Application of deep learning techniques in diagnosis of covid-19 (coronavirus): a systematic review," *Neural processing letters*, vol. 55, no. 3, pp. 3551-3603, 2023.
- [12] Y. Gong, H. Qiu, X. Liu, Y. Yang, and M. Zhu, "Research and Application of Deep Learning in Medical Image Reconstruction and Enhancement," *Frontiers in Computing and Intelligent Systems*, vol. 7, no. 3, pp. 72-76, 2024.
- [13] M. S. Junayed, N. Anjum, A. Noman, and B. Islam, "A deep CNN model for skin cancer detection and classification," 2021.
- [14] L. O. Chua and T. Roska, "The CNN paradigm," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 40, no. 3, pp. 147-156, 1993.
- [15] Q. Chu, W. Ouyang, H. Li, X. Wang, B. Liu, and N. Yu, "Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4836-4845.
- [16] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural computation*, vol. 31, no. 7, pp. 1235-1270, 2019.
- [17] S. Qasim Gilani, T. Syed, M. Umair, and O. Marques, "Skin cancer classification using deep spiking neural network," *Journal of Digital Imaging*, vol. 36, no. 3, pp. 1137-1147, 2023.
- [18] S. Inthiyaz et al., "Skin disease detection using deep learning," *Advances in Engineering Software*, vol. 175, p. 103361, 2023.

- [19] D. Keerthana, V. Venugopal, M. K. Nath, and M. Mishra, "Hybrid convolutional neural networks with SVM classifier for classification of skin cancer," *Biomedical Engineering Advances*, vol. 5, p. 100069, 2023.
- [20] M. Tahir, A. Naeem, H. Malik, J. Tanveer, R. A. Naqvi, and S.-W. Lee, "DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images," *Cancers*, vol. 15, no. 7, p. 2179, 2023.
- [21] G. Alwakid, W. Gouda, M. Humayun, and N. U. Sama, "Melanoma detection using deep learning-based classifications," in *Healthcare*, 2022, vol. 10, no. 12: MDPI, p. 2481.
- [22] X. Wang et al., "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0-0.
- [23] M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," *Machine Learning with Applications*, vol. 5, p. 100036, 2021.
- [24] J. Hurtado and F. Reales, "A machine learning approach for the recognition of melanoma skin cancer on macroscopic images," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 19, no. 4, pp. 1357-1368, 2021.
- [25] W. Gouda, N. Sama, G. Al-Waakid, M. Humayun, and N. Jhanjhi, "Detection of Skin-Cancer Based on Skin Lesion Images Using Deep Learning. *Healthcare* 2022, 10, 1183," ed: s Note: MDPI stays neu-tral with regard to jurisdictional claims in ..., 2022.

- [26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in Proceedings of the AAAI conference on artificial intelligence, 2017, vol. 31, no. 1.
- [27] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 7132-7141.
- [28] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," *Neurocomputing*, vol. 452, pp. 48-62, 2021.
- [29] M.-H. Guo et al., "Attention mechanisms in computer vision: A survey," *Computational visual media*, vol. 8, no. 3, pp. 331-368, 2022.
- [30] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," *International journal of computer vision*, vol. 128, pp. 336-359, 2020.