Care Vision AI: Revolutionizing Disease Detection through Artificial Intelligence

Tapan Mahata, Diganta Diasi, Tarun Kumar, Major Vaibhav Mishra, Anjali Jaiswal

May 8, 2024

GitHub Repository
Supervisor: Dr. Debanga Raj Neog

Abstract

The Care Vision AI project aims to develop an AI-powered diagnostic assistant to revolutionize disease detection processes. Traditional diagnostic methods face challenges such as time-consuming procedures, high costs, and potential errors. Leveraging artificial intelligence and machine learning, Care Vision AI automates and streamlines diagnostics. Using diverse medical image datasets, including those for brain tumors, nail diseases, lung cancer, and breast cancer, the project trains AI models to accurately identify disease patterns and features. The algorithmic approach integrates advanced image processing techniques with Convolutional Neural Networks (CNNs), with transfer learning enhancing efficiency. Image augmentation and exploration of alternative architectures like RCNN ensure robustness. The project culminates in a user-friendly web portal for disease screening, offering rapid, accurate assessments. Additionally, the project conducts a comprehensive investigation into image processing and machine learning techniques for disease detection, focusing on skin cancer, brain tumor, pneumonia, and lung cancer. Methodologies, datasets, model architectures, and evaluation metrics are explored, with model performance analysis and reflections on implications and future research directions concluding the study. Care Vision AI endeavors to transform disease diagnosis, improving patient outcomes and advancing healthcare technology.

Keywords: Artificial Intelligence, Machine Learning, Disease Detection, Medical Imaging, Convolutional Neural Networks, Transfer Learning, Image Processing

1 Introduction

The realm of medical diagnostics stands on the precipice of a transformative era, propelled by advancements in image processing and machine learning. Early detection of diseases such as skin cancer, brain tumor, pneumonia, and lung cancer holds the key to improved treatment outcomes and patient survival rates. Leveraging the power of medical imaging datasets and sophisticated algorithms, this project endeavors to develop and evaluate diagnostic models for these life-threatening conditions.

2 Problem Statement

Timely and accurate disease diagnosis is paramount for effective treatment and enhanced patient outcomes in healthcare. However, traditional diagnostic methods pose significant challenges. Labor-intensive procedures, such as manual interpretation of medical images and diagnostic tests, consume considerable time and expertise from healthcare professionals. Moreover, these methods incur high healthcare costs, further straining healthcare systems.

The project aims to tackle these challenges by addressing the pressing need for accurate and efficient disease detection in healthcare. Specifically, the project focuses on the detection of skin cancer, brain tumors, and pneumonia, where early diagnosis is crucial for timely intervention and improved patient prognosis. Through the integration of advanced technologies, such as artificial intelligence and machine learning, the project seeks to develop innovative solutions that streamline the diagnostic process, reduce healthcare burdens, and ultimately enhance patient care.

3 Dataset Description

3.1 Brain Tumor

The Brain MRI Images dataset contains a collection of 253 brain MRI images, categorized into "yes" and "no" folders. The "yes" folder comprises 155 images depicting tumorous cases, while the "no" folder contains 98 images representing non-tumorous cases. These images provide valuable insights into the diverse manifestations of brain tumors, aiding in the development of accurate diagnostic models.

3.2 Pneumonia

The dataset is organized into 3 folders (train, test, val) and contains subfolders for each image category (Pneumonia/Normal). There are 5,863 X-Ray images and 2 categories (Pneumonia/Normal). Chest X-ray images were selected from retrospective cohorts of pediatric patients of one to five years old from Guangzhou Women and Children's Medical Center, Guangzhou. All chest X-ray imaging was performed as part of patients' routine clinical care. For the analysis of chest x-ray images, all chest radiographs were initially screened for quality control by removing all low quality or unreadable scans.

3.3 Skin Cancer

The HAM10000 dataset serves as the foundation for skin cancer detection models in this project. Comprising 10,015 dermatoscopic images, it covers a wide spectrum of skin lesions classified into seven diagnostic categories. These categories include Actinic keratoses and intraepithelial carcinoma (akiec), Basal cell carcinoma (bcc), Benign keratosis-like lesions (bkl), Dermatofibroma (df), Melanoma (mel), Melanocytic nevi (nv), and Vascular lesions (vasc). More than 50% of the lesions in the dataset have been confirmed through histopathology, while others have been diagnosed through follow-up examination, expert consensus, or in-vivo confocal microscopy confirmation.

3.4 Lung Cancer

Chest X-ray images serve as the basis for lung cancer classification models. The dataset includes images depicting different subtypes of lung cancer, such as adenocarcinoma, large cell carcinoma, squamous cell carcinoma, and normal cases. These images provide a comprehensive view of lung cancer pathology, enabling the development of robust classification models capable of accurately identifying and classifying diverse lung cancer subtypes.

Through the utilization of these diverse and rich medical imaging datasets, this project aims to harness the power of image processing and machine learning techniques to develop accurate and reliable diagnostic models for skin cancer, brain tumor, pneumonia, and lung cancer.

4 Image Processing and Data Preprocessing

4.1 Skin Cancer

Image preprocessing for skin cancer involves several steps to enhance image quality and standardize them for analysis:

- Resize images to a standard size, typically 28x28 pixels, to ensure uniformity and compatibility with the model architecture.
- Normalize pixel values to a common scale, typically ranging from 0 to 1, to facilitate model training and convergence.
- Apply augmentation techniques such as rotation, flipping, and zooming to increase the diversity of the training dataset and improve model generalization.
- Address class imbalance through oversampling techniques to ensure balanced representation of different diagnostic categories.
- Split the dataset into training and testing sets for model training and evaluation, respectively.

4.2 Brain Tumor

Image preprocessing for brain tumor detection involves a series of steps tailored to MRI images:

- Convert the input image from BGR to grayscale to simplify the image representation and focus on relevant features.
- Apply Gaussian blur with a kernel size of (5x5) and a standard deviation of 0 to reduce noise and enhance image clarity.
- Perform binary thresholding with a threshold value of 45 to segment the brain region from the background.
- Execute two iterations of erosion followed by two iterations of dilation to further refine the segmented brain region.
- Detect contours using the processed thresholded image to identify the boundaries of the brain region.
- Find the contour with the maximum area, representing the brain region of interest.
- Determine extreme points (leftmost, rightmost, topmost, and bottommost) to define a bounding box around the brain region and crop the original image accordingly.
- Obtain the cropped image containing only the brain region for further analysis.

4.3 Pneumonia

- Normalization: Pixel values were rescaled to a [0, 1] range to facilitate model convergence.
- Splitting: The training data included a validation subset (20%) for model performance monitoring during training.

4.4 Lung Cancer

Image preprocessing for lung cancer detection focuses on standardizing the images and preparing them for model input:

- Resize images to a fixed size to ensure uniformity and compatibility with the model architecture.
- Convert images to grayscale to simplify the representation and reduce computational complexity.
- Normalize pixel values to a common scale, typically ranging from 0 to 1, to facilitate model training and convergence.
- Split the dataset into training, validation, and testing sets for model training, validation, and evaluation, respectively.
- Implement data augmentation techniques such as rotation, flipping, and zooming to increase the diversity of the training dataset and improve model robustness.

By applying these image processing and data preprocessing techniques, the project aims to enhance the quality of the medical images, standardize their representation, and prepare them for effective utilization in machine learning models for disease detection.

5 Model Architecture for Each Disease

5.1 Brain Tumor Detection

Model 1

The model begins with an input layer expecting images with dimensions of 240x240 pixels and 3 color channels (RGB). The input images undergo zero-padding to maintain spatial information during subsequent operations. Following this, a convolutional layer applies 32 filters of size 7x7 to extract features from the input images. Batch normalization is then applied to stabilize the training process, followed

by rectified linear unit (ReLU) activation to introduce non-linearity. Subsequently, two max-pooling layers sequentially downsample the feature maps, reducing their spatial dimensions. The flattened feature maps are then passed through a fully connected layer with a single neuron and sigmoid activation, which outputs a probability indicating the likelihood of the input image containing brain tissue. This model architecture is designed to effectively process 240x240 RGB images for binary brain detection classification tasks.

Model 2

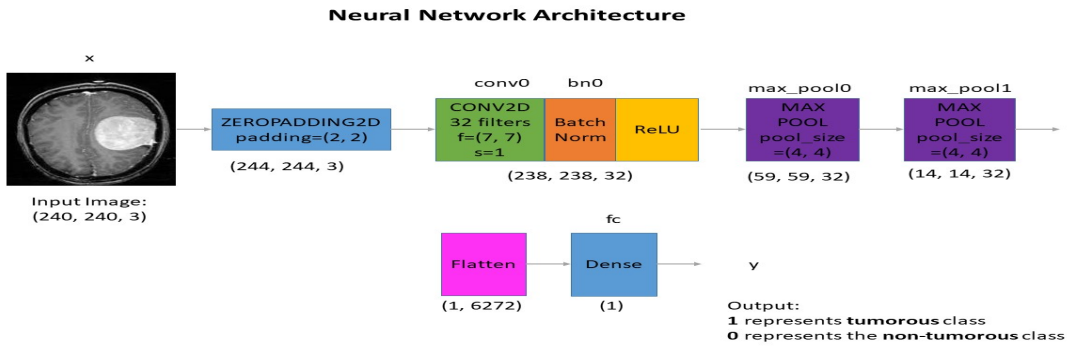


Figure 1: Model Architecture for Brain Tumor Detection

The model uses a pre-trained ResNet50 base model to process input images of 240x240 pixels with three color channels. The base model's layers are frozen to retain learned representations. A global average pooling layer reduces the spatial dimensions of the feature maps, followed by a dense layer with ReLU activation for feature extraction. The output layer, with a sigmoid activation function, predicts the probability of a brain tumor presence. The model is compiled with Adam optimizer, binary cross-entropy loss, and accuracy metric. Trained for 10 epochs, it's evaluated on a test set for performance assessment. Overall, the architecture leverages transfer learning for effective brain tumor classification.

5.2 Skin Cancer Detection

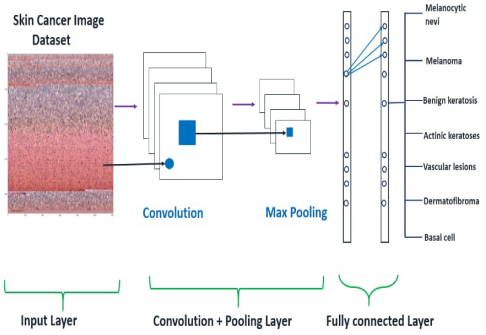


Figure 2: Model Architecture of Skin Cancer Detection

The model architecture for classifying skin cancer types using images is designed to effectively process and extract features from input images for accurate classification. The architecture begins with an input layer that accepts RGB images of size (28, 28, 3). This is followed by two convolutional layers, each utilizing ReLU activation and padding to maintain spatial dimensions. Subsequently, a max-pooling layer is applied to downsample the feature maps. Two additional convolutional layers further extract features, followed by another max-pooling layer. The feature maps are then flattened into a 1D vector before passing through two fully connected dense layers with ReLU activation. Finally, the output layer consists of 7 units, corresponding to the number of skin cancer types, and employs softmax activation for multiclass classification. This hierarchical architecture enables the model to learn and integrate features at different levels of abstraction, resulting in accurate classification of skin cancer types.

5.3 Pneumonia Detection

5.3.1 CNN Architecture

Model Development

Architecture: The CNN model architecture comprises a series of convolutional layers followed by max-pooling operations to extract features and reduce spatial dimensions. The initial convolutional layer, `conv2d`, processes the input data with 32 filters, resulting in an output shape of $(148, 148, 32)$. Subsequent max-pooling reduces the spatial dimensions by half. The following `conv2d_1` layer applies 64 filters to the feature maps, producing an output shape of $(72, 72, 64)$, which is again downsampled by max-pooling. The pattern continues with `conv2d_2`, which increases the filter count to 128, resulting in an output shape of $(34, 34, 128)$. After another max-pooling operation, `conv2d_3` further processes the features with 128 filters, resulting in an output shape of $(15, 15, 128)$. Finally, the feature maps are flattened to a vector representation and fed into two fully connected dense layers with 512 and 1 units, respectively, for classification. The total number of parameters in the model architecture is 3,452,545.

Hyperparameter Tuning: Random Search from Keras Tuner optimized the model's hyperparameters.

Training: The best hyperparameter settings trained the model for 50 epochs, ensuring sufficient learning without overfitting.

Hyperparameter Tuning Details

- Filters in convolutional layers varied from 32 to 128.
- Kernel sizes for convolutional layers were 3x3 and 5x5.
- Dense layer units ranged from 32 to 512.
- Learning rates tested were 1e-2, 1e-3, and 1e-4.

5.3.2 SVM Classifier

Model Specifications

- Model Type: Support Vector Machine (SVM)
- Kernel Type: Linear
- Feature Descriptor: Histogram of Oriented Gradients (HOG)
- Image Dimensions (for HOG): 150 x 150 pixels

Feature Extraction Parameters

- Orientations: 8
- Pixels per Cell: (16, 16)
- Cells per Block: (1, 1)
- Block Normalization: L2-Hys

5.4 Lung Cancer Classification

CNN Architecture:

- Input Layer: Accepts images of size (64, 64, 3).
- Convolutional Layers: Three convolutional layers with increasing filter sizes (16, 32, and 64) and ReLU activation.
- Max Pooling Layers: Downsample the feature maps.
- Flatten Layer: Transition from convolutional layers to fully connected dense layers.
- Dense Layers: Three dense layers with 256, 128, and 64 units, respectively, using ReLU activation.
- Output Layer: Softmax activation for multiclass classification.

6 Results

6.1 Brain Tumor

Model 1 The model achieved promising results in the task of binary brain tumor detection. With a total of 11,137 parameters, of which 11,073 are trainable, the model demonstrates its ability to effectively learn and extract relevant features from input images. Utilizing the Adam optimizer and binary cross-entropy loss function, the model underwent training for 24 epochs. On the test dataset, the model attained a test loss of 0.24 and a test accuracy of 91%, indicating its robustness in classifying brain tumor images. Furthermore, the model's performance was evaluated on both the validation and test sets, resulting in accuracies of 91% and 89%, respectively. Additionally, the F1 scores for the validation and test sets were calculated at 0.91 and 0.88, demonstrating the model's balanced precision and recall1a. These results underscore the efficacy of the proposed model architecture in accurately detecting brain tumors from input images.

Model 2

The model, equipped with a total of 24,112,513 parameters, of which 524,801 are trainable, showcased its performance in binary classification for brain tumor detection. Employing the Adam optimizer and binary cross-entropy loss function, the model underwent rigorous training. Upon evaluation, the model achieved a test accuracy of 62.47% with a corresponding test loss of 0.6471, indicating moderate performance on unseen data. Furthermore, the model's performance was assessed on both the validation and test sets, yielding accuracies of 71% and 72%, respectively. Additionally, the F1 scores for both validation and test sets were computed at 71% 1b, showcasing a balanced precision and recall. These results underscore the model's capability in discerning brain tumor presence from input images, albeit with room for further improvement.

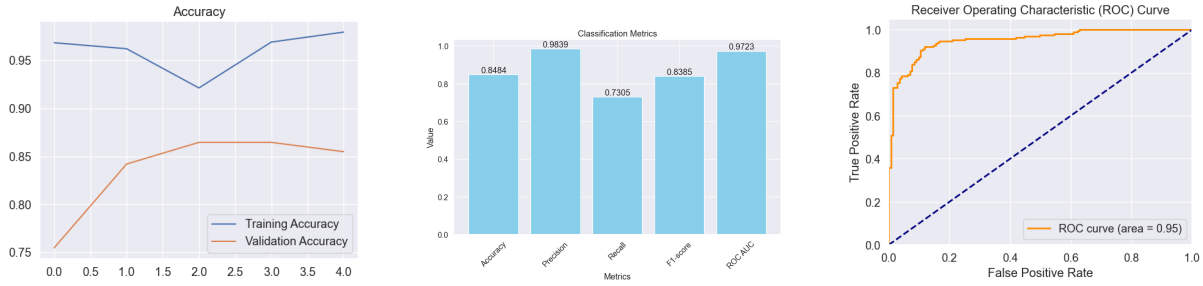
	Validation Set	Test Set
Accuracy	91%	89%
F1 Score	0.91	0.88

(a) Model 1

	Validation Set	Test Set
Accuracy	71%	72%
F1 Score	71%	72%

(b) Model 2

Table 1: Performance Comparison on Validation and Test Sets of Brain Tumor



(a) BT1 Accuracy (b) BT1 Classification Metrics (c) ROC Curve

Figure 3: Results of Model 1 in Brain Tumor Detection

6.2 Pneumonia Detection

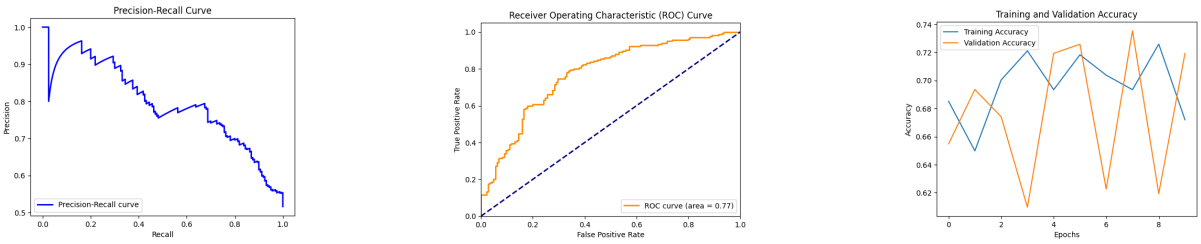
6.2.1 CNN Architecture

Model Training Performance

The validation accuracy peaked at 97.39% at epoch 16.

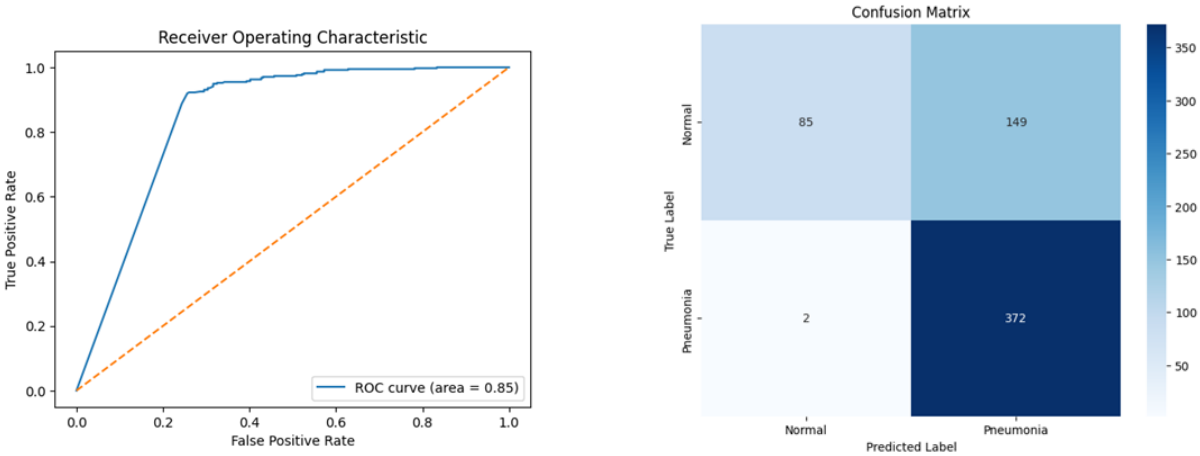
F1 scores for each class fluctuated, indicating the variability in model performance over epochs.

The high recall is indicative of the model's sensitivity to pneumonia presence, prioritizing the detection of as many positive cases as possible. The implications of the precision-recall trade-off and the potential impact on clinical practice are examined.



(a) Precision-Recall Curve (b) ROC Curve (c) Accuracy

Figure 4: Results of Model 2 in Brain Tumor Detection



(a) ROC Curve (b) Confusion Matrix

Figure 5: Results of CNN Architecture in Pneumonia Detection

Table 2: Result of CNN Architecture in Pneumonia

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.98	0.36	0.53	234
1	0.71	0.99	0.83	374
Accuracy	NaN	NaN	0.75	608
Macro Avg	0.85	0.68	0.68	608
Weighted Avg	0.82	0.75	0.72	608

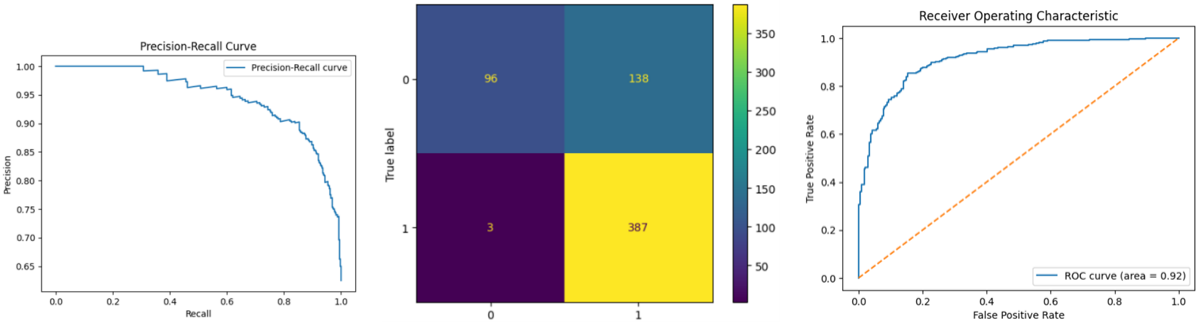
6.2.2 SVM Classifier

Model Training and Performance

The model was trained using the extracted HOG features from the training set. Post-training, the model was evaluated on a separate test set to measure its generalization capabilities.

Table 3: Result of SVM with HOG Architecture in Pneumonia

Class	Precision	Recall	F1-Score	Support
0	0.96	0.49	0.65	234
1	0.75	0.99	0.86	374
Accuracy	Nan	Nan	0.79	608
Macro Avg	0.86	0.74	0.75	608
Weighted Avg	0.83	0.79	0.77	608



(a) PneuSVM (Image 1) (b) PneuSVM (Image 2) (c) PneuSVM (Image 3)

Figure 6: Results of CNN Architecture in Pneumonia Detection

These results demonstrate the performance of both CNN and SVM with HOG architectures for pneumonia detection. Both models achieve similar accuracy, precision, recall, and F1 score values, indicating comparable performance in classifying pneumonia cases from chest X-ray images.

6.2.3 CNN vs. SVM for Chest X-Ray Classification

Table 4: Comparison of SVM and CNN for Chest X-Ray Classification

Metric	SVM	CNN
Precision (Class 0)	0.97	0.98
Recall (Class 0)	0.41	0.36
F1-Score (Class 0)	0.58	0.53
Support (Class 0)	234.0	234.0
Precision (Class 1)	0.74	0.71
Recall (Class 1)	0.99	0.99
F1-Score (Class 1)	0.85	0.83
Support (Class 1)	390.0	374.0
Accuracy	0.774	0.75
AUC (ROC Curve)	0.92	0.85

Based on the criteria that prioritize minimizing false negatives (FN), and then false positives (FP) if the false negatives are equal, the SVM model is recommended. Both the SVM and CNN models have a high recall for pneumonia (Class 1), which stands at 0.99, indicating that they are both very good at identifying positive cases of pneumonia. However, the secondary criterion requires us to consider the count of false positives. The SVM model has fewer false positives (138) compared to the CNN model (149), which implies that the SVM model is less likely to incorrectly label normal cases as pneumonia. This could be particularly important in a medical setting where the cost of false alarms is high. Moreover, the slightly higher overall accuracy (0.774) and AUC of the ROC curve (0.92) for the SVM suggests that

it has better overall performance in terms of both sensitivity and specificity. These factors combined provide a strong justification for the recommendation of the SVM model for this particular task.

6.3 Skin Cancer Detection

Table 5: Disease Names and Referred Short Names

Disease Name	**Short Name**
Melanocytic Nevi	nv
Melanoma	mel
Benign Keratosis-like Lesions	bkl
Basal Cell Carcinoma	bcc
Pyogenic Granulomas and Hemorrhage	vasc
Actinic Keratoses and Intraepithelial Carcinomae	akiec
Dermatofibroma	df

Table 6: Results of Skin Cancer Detection

Class	**Precision**	**Recall**	**F1 Score**	**Support**
akiec	0.97	1.00	0.99	1359
bcc	0.99	0.99	0.99	1318
bkl	0.94	0.97	0.96	1262
df	0.99	1.00	1.00	1351
nv	0.89	0.88	0.88	1374
vasc	1.00	1.00	1.00	1358
mel	0.95	0.91	0.93	1365
Accuracy			0.96	9387

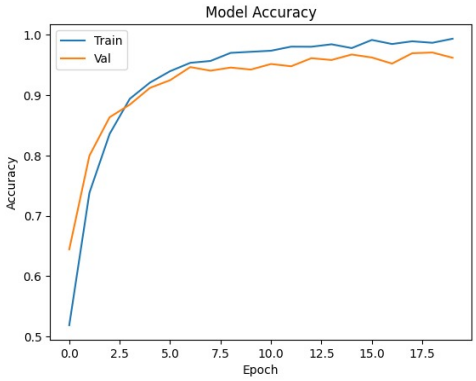


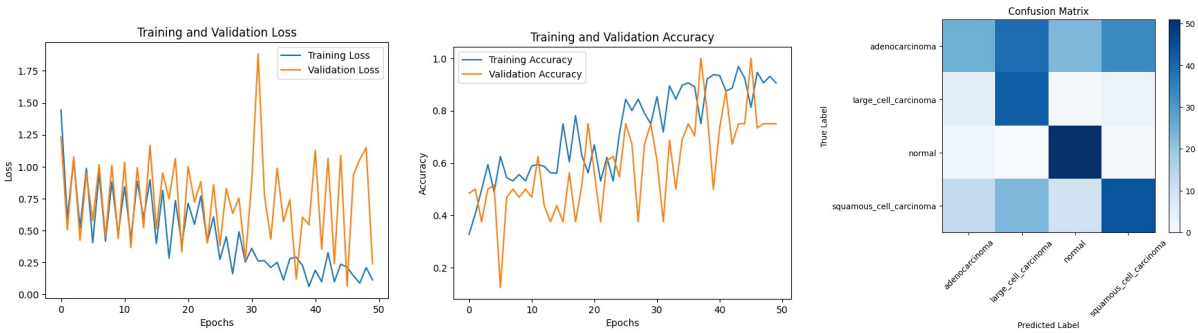
Figure 7: Accuracy of Skin Cancer Model

The skin cancer detection model achieves an impressive accuracy of 96%, demonstrating its effectiveness in classifying dermatoscopic images into seven diagnostic categories. The precision, recall, and F1-score values are high across all classes, indicating reliable performance in identifying various types of skin lesions.

6.4 Lung Cancer

The lung cancer detection model achieved an accuracy of 51% on the test data, with an AUC (Area Under the Curve) of 0.76. Despite these results, several factors contribute to the suboptimal performance of the model. Firstly, the dataset suffers from class imbalance, with significant variations in the distribution of data across the four classes. Additionally, the dataset is limited in size, particularly in terms of

image data, which could be addressed by augmenting the dataset to potentially improve performance. Furthermore, the quality of images in the dataset is subpar, which likely impacted the model's ability to accurately detect lung cancer. These factors collectively contribute to the challenges faced by the model in achieving higher accuracy and performance metrics.



(a) Loss (b) Accuracy (c) Confusion Matrix

Figure 8: Results of Lung Cancer Model

7 Conclusion

The Care Vision AI project represents a significant advancement in the field of disease diagnosis, leveraging artificial intelligence and machine learning to develop innovative solutions for skin cancer, brain tumor, pneumonia, and lung cancer detection. By harnessing the power of medical imaging datasets and sophisticated algorithms, the project aims to revolutionize disease detection processes, improve patient outcomes, and advance healthcare technology. Through meticulous image processing, data preprocessing, and model architecture design, the project endeavors to develop accurate and reliable diagnostic models capable of rapid and precise disease identification. By bridging the gap between cutting-edge technology and clinical practice, Care Vision AI paves the way for a future where disease diagnosis is streamlined, accessible, and efficient.

References

1. World Cancer Research Fund. (2020). *Skin Cancer Statistics*. Retrieved from `https://www.wcrf.org/dietandcancer/cancer-trends/skin-cancer-statistics`. Accessed on March 30, 2020.
2. AW, K., TG, S., J, S., AA, M., & RS, B. (2021). *Techniques of Cutaneous Examination for the Detection of Skin Cancer. PubMed*. Retrieved from `https://pubmed.ncbi.nlm.nih.gov/7804995/`. Accessed on April 10, 2021.
3. Mahmud, M. I., Mamun, M., & Abdelgawad, A. (2023). *A Deep Analysis of Brain Tumor Detection for MRI Images Using Deep Learning Networks. Algorithms*, MDPI. (Preprint)
4. Hany, M. (2020, August 20). *Chest CT-scan Images Dataset*. Kaggle. Retrieved from `https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images`. Accessed on November 13, 2022.
5. Mamun, M., et al. (2023). *Lcdctcnn: Lung Cancer Diagnosis of CT Scan Images Using CNN Based Model*. In *10th International Conference on Signal Processing and Integrated Networks (SPIN)*. IEEE.
6. *Brain Tumour Detection Using Deep Learning*. IEEE Xplore. Retrieved from `https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9445185`.
7. *A Deep Analysis of Brain Tumor Detection from MR Images Using Deep Learning Networks*. (2021). *MDPI*. Retrieved from `https://www.mdpi.com/1999-4893/16/4/176`.
8. He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep Residual Learning for Image Recognition. arXiv preprint arXiv:1512.03385*.