**Enhancing Dermatological Diagnostics: Deep Learning Approaches to Skin Anomaly Classification**

## **abcda, abcdb, abcdc, abcdd,abcde**

a **abcd**

b **abcd**

c **abcd**

d  **abcd**

e  **abcd**

## **Abstract***—*Food allergies, atopic dermatitis, allergic rhinitis, and asthma are examples of allergic disorders that are common and present serious public health issues. In order to investigate the prevalence, treatment effects, and demographic influences on various disorders, this study analyses a large dataset. The dataset contains comprehensive patient data at the beginning and conclusion of the trial, including age, insurance coverage, race, ethnicity, gender, and year of birth. At the beginning and conclusion of the trial, information about allergies to shellfish, milk, soy, eggs, wheat, peanuts, sesame, tree nuts, and certain nuts was documented. Atopic dermatitis, allergic rhinitis, and asthma were also recorded, as were the prescriptions for asthma medications.The trends in the occurrence and development of certain allergy disorders are shown by our data, together with the demographic factors influencing these patterns. Additionally, we also assess the effectiveness of various treatments by comparing allergy and condition statuses over time. This study aims to provide insights into optimizing allergy management strategies, better targeting at-risk demographics, and improving treatment protocols to enhance patient outcomes. The findings will be valuable for healthcare providers and policymakers in devising more effective approaches to managing and treating allergic diseases.

***Keywords****–*Allergy prevalence, Atopic dermatitis, Allergic rhinitis, Asthma, Food allergies, Patient demographics, Treatment outcomes, Cohort analysis

**1. INTRODUCTION**

Allergies are chronic inflammatory disorders characterized by abnormal immune responses to specific environmental substances known as allergens. These allergens can be proteins from a variety of sources, including foods, plants, and chemicals, and they trigger allergic reactions through complex immune mechanisms. The severity of allergy symptoms can range from mild discomfort to life-threatening reactions, such as anaphylaxis. When the immune system encounters what it mistakenly identifies as a harmful antigen, it initiates an allergic reaction, producing symptoms like sneezing, itching, and swelling. This phenomenon occurs due to the immune system's production of immunoglobulin E (IgE) antibodies, which recognize and bind to the allergens, leading to the release of histamines and other chemicals that cause inflammation and other allergic symptoms.

Recent advancements in deep learning, particularly convolutional neural networks (CNNs) and DenseNet models, have shown great promise in the field of medical research. These models are capable of analyzing complex biological data with high accuracy, making them valuable tools for investigating the underlying causes of allergic diseases and improving diagnostic and treatment approaches. Convolutional neural networks (CNNs) are a type of deep learning model that excels in pattern recognition, especially in image analysis. In medical research, CNNs have been successfully applied to tasks such as medical imaging and genomic data analysis, helping to identify disease markers and predict patient outcomes. By processing large datasets, CNNs can uncover intricate patterns and relationships that are not easily detectable through traditional analytical methods.

DenseNet models, or densely connected convolutional networks, build upon the success of traditional CNNs by introducing dense connections between layers. This architecture allows for better gradient flow during training, leading to more efficient learning and improved performance on complex tasks. DenseNet models have shown great promise in medical research, particularly in analyzing high-dimensional data such as gene expression profiles and protein sequences. By employing MobileNet models, researchers can gain deeper insights into the molecular mechanisms underlying allergic diseases, identify novel therapeutic targets, and predict patient responses to different treatments.

In this study, we utilize DenseNet and MobileNet models to investigate the prevalence and treatment outcomes of various allergic conditions, including food allergies, atopic dermatitis, allergic rhinitis, and asthma. Our dataset encompasses detailed patient demographics and clinical data, allowing us to explore the influence of genetic and environmental factors on allergy development and progression. By applying these advanced deep learning techniques, we aim to uncover patterns and correlations that can inform better management strategies for allergic diseases, ultimately improving patient outcomes and quality of life.

**1.1 Motivatio*n***

Allergic diseases pose a significant burden on public health worldwide, affecting millions of individuals and impacting their quality of life. Despite advances in medical research and treatment options, the prevalence of allergic conditions continues to rise, necessitating a deeper understanding of the underlying mechanisms and more effective management strategies. The complexity of allergic diseases, with their multifactorial etiology involving genetic predisposition, environmental factors, and immune system dysregulation, presents a formidable challenge for clinicians and researchers alike. This complexity underscores the need for innovative approaches to unravel the intricacies of allergic reactions and develop personalized interventions tailored to individual patients.This interdisciplinary approach, combining deep learning with immunology, genetics, and clinical medicine, holds immense promise for advancing our understanding of allergic diseases and ultimately improving patient outcomes in this rapidly evolving field.

**1.2 Objectives**

* Analyze a comprehensive dataset to determine the prevalence of various allergic conditions, including food allergies, atopic dermatitis, allergic rhinitis, and asthma, across different demographic groups and geographical regions.
* Explore the influence of demographic factors, such as age, gender, race, ethnicity, and socioeconomic status, on the development and progression of allergic diseases, aiming to uncover disparities and trends that may inform targeted interventions and public health policies.
* Assess the effectiveness of different treatment modalities, including medications, immunotherapy, and lifestyle interventions, in managing allergic conditions and mitigating symptoms over time, utilizing longitudinal data to track changes in disease status and patient outcomes.
* Develop predictive models using CNN and DenseNet architectures to forecast the risk of allergic reactions and disease exacerbations based on patient characteristics, environmental factors, and genetic markers, facilitating early intervention and personalized treatment strategies.
* Translate research findings into actionable insights for healthcare providers, policymakers, and patients, with the aim of optimizing allergy management protocols, enhancing patient education and awareness, and improving overall quality of care for individuals affected by allergic diseases.

**2. RELATED WORK**

In a comprehensive nationwide cohort study conducted by Yang et al., the relationship between allergic disorders and susceptibility to, as well as the severity of, COVID-19 was investigated[1]. This study, analyzed data from a large cohort to assess the impact of allergic conditions on COVID-19 outcomes. By examining the medical records of a diverse patient population, the researchers aimed to elucidate potential associations between allergic disorders and the risk of contracting COVID-19, as well as the severity of symptoms among those with pre-existing allergic conditions. The findings of this study contribute valuable insights into the intersection of allergic diseases and viral infections, shedding light on potential risk factors and informing strategies for disease prevention and management.

Lee et al. conducted a Korean nationwide birth cohort study to investigate the incidence of fractures in children following the development of atopic dermatitis. This study aimed to elucidate the potential association between atopic dermatitis, a common allergic skin condition, and bone health in pediatric populations. By analyzing longitudinal data from a large cohort of children, the researchers examined the incidence of fractures in relation to the onset and severity of atopic dermatitis[2]. The findings of this study provide important insights into the potential impact of allergic diseases on bone health and fracture risk in children, highlighting the need for further research and targeted interventions to address this issue.

In a seminal study aimed to evaluate the consistency of asthma prevalence estimates across different study populations and methodologies. By synthesizing data from these two landmark studies, the researchers provided valuable insights into the global burden of asthma and the variability in asthma prevalence estimates between different regions and age groups[3]. This study underscores the importance of standardized methodologies and cross-national collaboration in epidemiological research on allergic diseases, facilitating more accurate assessment of disease burden and informing public health policies and interventions.

Alavinezhad and Boskabady conducted a study to investigate the prevalence of asthma and related symptoms in Middle Eastern countries. This research aimed to provide insights into the burden of asthma in the Middle East region, where limited data were available on the prevalence and impact of allergic respiratory conditions [4]. Through a systematic analysis of available epidemiological data, the authors sought to characterize the prevalence rates of asthma and associated symptoms, such as wheezing and breathlessness, across different Middle Eastern populations. By synthesizing findings from diverse studies and populations, this study contributes to our understanding of the epidemiology of asthma in the Middle East and highlights the need for targeted interventions to address the burden of allergic respiratory diseases in this region.

Koo et al. conducted a serial analysis of national representative studies to examine trends in the prevalence of allergic diseases among Korean adolescents before and during the COVID-19 pandemic[5]. This study aimed to assess changes in the prevalence rates of allergic conditions, including asthma, atopic dermatitis, and allergic rhinitis, over a 12-year period from 2009 to 2021. By analyzing longitudinal data from national surveys, the researchers identified temporal trends and variations in the prevalence of allergic diseases among Korean adolescents, providing valuable insights into the impact of environmental and societal factors, including the COVID-19 pandemic, on allergic disease epidemiology. This study informs public health efforts aimed at mitigating the burden of allergic diseases and underscores the importance of ongoing surveillance and monitoring of allergic disease trends in adolescent populations.

Yaghoubi et al. conducted a comprehensive analysis to project the economic and health burden of uncontrolled asthma in the United States. This research aimed to quantify the societal costs and health outcomes associated with uncontrolled asthma, a significant public health concern affecting millions of individuals in the United States[6]. Through a sophisticated modeling approach, the authors estimated the economic costs, including healthcare expenditures and productivity losses, attributable to uncontrolled asthma, as well as the associated health outcomes, such as exacerbations, hospitalizations, and mortality. By highlighting the substantial economic and health burden of uncontrolled asthma, this study underscores the importance of effective asthma management strategies and interventions to improve disease control and reduce the societal impact of asthma-related morbidity and mortality.

Strachan presented a seminal paper exploring the "hygiene hypothesis," which posits that reduced exposure to infections in early childhood may lead to an increased risk of allergic diseases later in life[6]. This review article delves into the potential role of family size and microbial exposures in modulating the development of atopic conditions, such as asthma and allergic rhinitis. By synthesizing epidemiological evidence from various studies conducted over the previous decade, Strachan proposed a provocative hypothesis suggesting that the modern trend towards smaller family sizes and improved hygiene practices may contribute to the rising prevalence of allergic diseases observed in industrialized societies. This paper laid the groundwork for subsequent research into the complex interplay between microbial exposures, immune development, and allergic disease susceptibility.

In a follow-up review article , Schaub, Lauener, and von Mutius further explored the multifaceted nature of the hygiene hypothesis. Building upon Strachan's initial hypothesis, the authors examined the evolving understanding of how early-life microbial exposures shape immune development and influence the risk of allergic diseases[7]. Through a comprehensive review of experimental and observational studies, Schaub et al. elucidated the mechanisms underlying the hygiene hypothesis, including the role of microbial diversity, immune tolerance, and gene-environment interactions. This review highlighted the complex and nuanced relationship between hygiene practices, microbial exposures, and allergic disease outcomes, emphasizing the need for further research to unravel the underlying mechanisms and inform strategies for allergy prevention and management.

Carroll discussed the disparities in asthma prevalence, severity, and outcomes observed among different socioeconomic and racial/ethnic groups in the United States. Carroll highlighted the complex interplay of social determinants, environmental exposures, and genetic factors contributing to asthma health disparities, emphasizing the need for targeted interventions to address underlying inequities and improve asthma outcomes in vulnerable populations[8]. By raising awareness of the social determinants of health and their impact on asthma disparities, this commentary underscored the importance of addressing structural inequalities and promoting health equity in asthma care and research efforts.

**2.1 CNN**

Convolutional Neural Networks (CNNs) represent a groundbreaking advancement in the field of computer vision, revolutionizing tasks such as image classification, object detection, and semantic segmentation. Inspired by the visual cortex of the human brain, CNNs are designed to automatically and adaptively learn hierarchical representations of visual data through the application of convolutional filters.

At the core of CNNs are convolutional layers, where filters convolve with input images to extract features such as edges, textures, and shapes. These learned features are progressively combined and refined through additional layers such as pooling layers, which downsample feature maps, and fully connected layers, which perform classification or regression tasks based on the extracted features.One of the key strengths of CNNs lies in their ability to learn hierarchical representations of visual data, capturing both low-level features like edges and high-level features like object shapes and textures. This hierarchical feature learning enables CNNs to achieve state-of-the-art performance on various computer vision tasks, often surpassing human-level performance in tasks like image recognition and object detection.

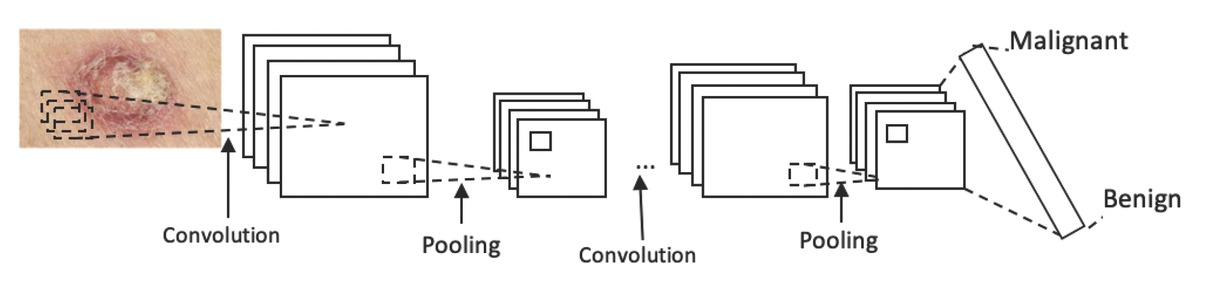


Fig1:CNN model

Overall, CNNs have revolutionized the field of computer vision, enabling significant advancements in tasks such as image classification, object detection, and semantic segmentation. Their ability to automatically learn hierarchical representations of visual data, coupled with translational invariance, has made them indispensable tools for solving complex visual recognition tasks across diverse domains.

**2.2:Densenet Model:**

DenseNet, short for Densely Connected Convolutional Networks, is a deep learning architecture known for its densely connected layers and efficient use of parameters. Unlike traditional convolutional neural networks (CNNs), where each layer is connected only to the subsequent layer, DenseNet introduces dense connections between all layers within a dense block. This connectivity pattern facilitates feature reuse and encourages gradient flow throughout the network, leading to improved learning efficiency and model performance.In the architecture, the DenseNet121 model is employed as a functional layer, taking input images and processing them through its convolutional blocks to extract features. DenseNet121 is a pre-trained convolutional neural network with 121 layers, known for its effectiveness in feature extraction and representation learning from images. The output shape of this layer is (None, 7, 7, 1024), indicating that it produces feature maps with dimensions of 7x7x1024.

Following the DenseNet121 layer, a Global Average Pooling 2D layer is applied to further reduce the spatial dimensions of the feature maps. This layer aggregates spatial information across each feature map channel, resulting in a global average pooling operation that generates a fixed-length vector representation for each image. The output shape of this layer is (None, 1024), indicating that it produces a vector with 1024 dimensions for each input image.

A dropout layer is then applied to prevent overfitting by randomly dropping a fraction of input units during training. This regularization technique helps improve the generalization ability of the model by reducing the reliance on specific features or neurons. In this architecture, the Dropout layer has no effect on the output shape, as it does not alter the dimensions of the data.

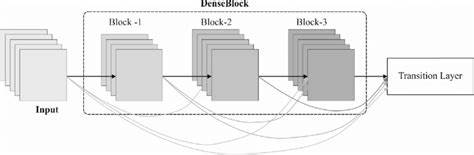


Fig. 2: DenseNet Model

In the architecture, the DenseNet121 model is employed as a functional layer, taking input images and processing them through its convolutional blocks to extract features. DenseNet121 is a pre-trained convolutional neural network with 121 layers, known for its effectiveness in feature extraction and representation learning from images. The output shape of this layer is (None, 7, 7, 1024), indicating that it produces feature maps with dimensions of 7x7x1024. Following the DenseNet121 layer, a Global Average Pooling 2D layer is applied to further reduce the spatial dimensions of the feature maps. This layer aggregates spatial information across each feature map channel, resulting in a global average pooling operation that generates a fixed-length vector representation for each image. The output shape of this layer is (None, 1024), indicating that it produces a vector with 1024 dimensions for each input image.A Dropout layer is then applied to prevent overfitting by randomly dropping a fraction of input units during training. This regularization technique helps improve the generalization ability of the model by reducing the reliance on specific features or neurons. In this architecture, the Dropout layer has no effect on the output shape, as it does not alter the dimensions of the data.Finally, a Dense layer is added with 9 units and a softmax activation function, serving as the output layer for classification. This Dense layer maps the input features to the 9 classes corresponding to the different allergy categories. The output shape of this layer is (None, 9), indicating that it produces a probability distribution over the 9 classes for each input image.

**2.3:MobileNet Model:**

MobileNet is a lightweight convolutional neural network architecture designed for efficient deployment on resource-constrained devices such as mobile phones and embedded systems. Developed by Google, MobileNet achieves a balance between model size, computational efficiency, and accuracy, making it suitable for various computer vision tasks, including image classification, object detection, and semantic segmentation.The key innovation of MobileNet lies in its use of depthwise separable convolutions, which decompose the standard convolution operation into two separate steps: depthwise convolution and pointwise convolution. Depthwise convolution applies a single filter to each input channel independently, followed by pointwise convolution, which performs a 1x1 convolution across all channels to combine the extracted features. This separation significantly reduces the computational cost of convolutions while preserving representational capacity, making MobileNet highly efficient compared to traditional convolutional architectures.

MobileNet architectures are characterized by a series of building blocks composed of depthwise separable convolutions, batch normalization, and ReLU activation functions. These building blocks are stacked to form the convolutional backbone of the network, with additional layers such as fully connected and softmax layers for classification tasks. The depthwise separable convolutions allow MobileNet to achieve a good trade-off between model size and accuracy, making it well-suited for deployment on mobile and embedded platforms.

Despite its compact size, MobileNet has demonstrated competitive performance on standard benchmark datasets such as ImageNet, often outperforming larger and more computationally intensive models. Its efficiency makes it particularly attractive for real-time applications where computational resources are limited, such as mobile image recognition and augmented reality.

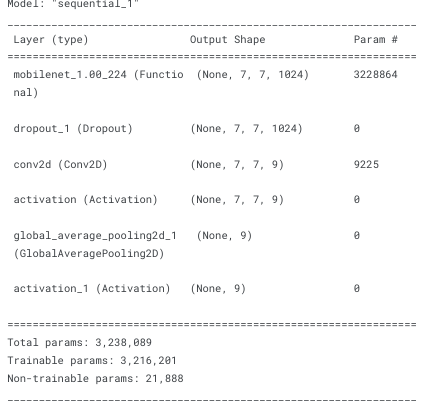


Fig 3:MobileNet model used

An activation layer (Activation) follows the Conv2D layer, applying a softmax activation function to produce a probability distribution over the 9 classes for each input image. The softmax activation function ensures that the output probabilities sum to 1, allowing the model to make predictions based on the highest probability class.Finally, a Global Average Pooling 2D layer aggregates spatial information across each feature map channel, resulting in a fixed-length vector representation for each image. The output shape of this layer is (None, 9), indicating that it produces a vector with 9 dimensions representing the probability distribution over the allergy classes for each input image.

**3. PROPOSED METHOD**

The image categorization process begins with the organization of images into separate folders, each labeled according to the specific disease they represent. This structured dataset is then divided into training and validation sets in an 80:20 ratio, respectively. The training set, comprising 80% of the data, is used to train the models, while the remaining 20% is reserved for validation to monitor the models’ performance and prevent overfitting. This split ensures that the models are exposed to a diverse range of examples during training while also being tested on a portion of data they haven't seen before to validate their learning progress.

Image preprocessing is a critical step to prepare the data for input into neural network models. This process starts with dimensionality reduction, where the images are resized to a uniform dimension suitable for the model input. Following resizing, all images are converted to RGB format to ensure consistency, as RGB images have three color channels necessary for many deep learning models. These preprocessed images are then converted into numpy arrays, which are mapped to their corresponding disease labels. The final step in preprocessing involves normalizing the pixel values by dividing them by 255, scaling the values to a range between 0 and 1, which helps in faster and more efficient training of the neural networks.

Once preprocessing is complete, the model architectures are defined, starting with Convolutional Neural Networks (CNNs) due to their effectiveness in image recognition tasks. Additionally, advanced pre-trained models such as MobileNet and DenseNet are also considered. These models are known for their efficiency and high performance in various computer vision tasks. The model layers are carefully constructed to include convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. After defining the models, the training process begins, and the history of each epoch’s run, including metrics such as training accuracy, training loss, validation accuracy, and validation loss, is saved into a variable called "history."

The saved history variable provides valuable insights into the training dynamics of the models. By plotting the train accuracy and train loss against the number of epochs, we can observe how well the model is learning over time. Similarly, plotting validation accuracy and validation loss helps to monitor the model's performance on unseen data. These plots are essential for identifying signs of overfitting or underfitting, guiding necessary adjustments in the training process such as learning rate tuning, model complexity adjustments, or early stopping criteria. Visualizing these metrics provides a clear, graphical interpretation of model performance and stability across epochs.

The relationship between atopic dermatitis (AD) and the duration of study participation is an intriguing aspect of allergy research. Atopic dermatitis, a chronic inflammatory skin condition characterized by itchy and inflamed skin, is commonly assessed over different periods to understand its progression and response to treatments. In this context, 'ATOPIC\_DERM\_START' refers to the severity or presence of atopic dermatitis at the beginning of a study, while 'STUDY\_DURATION\_YEARS' denotes the length of time a participant remains in the study. Observing how these variables interact can provide insights into how initial severity of AD influences patient retention and study outcomes.

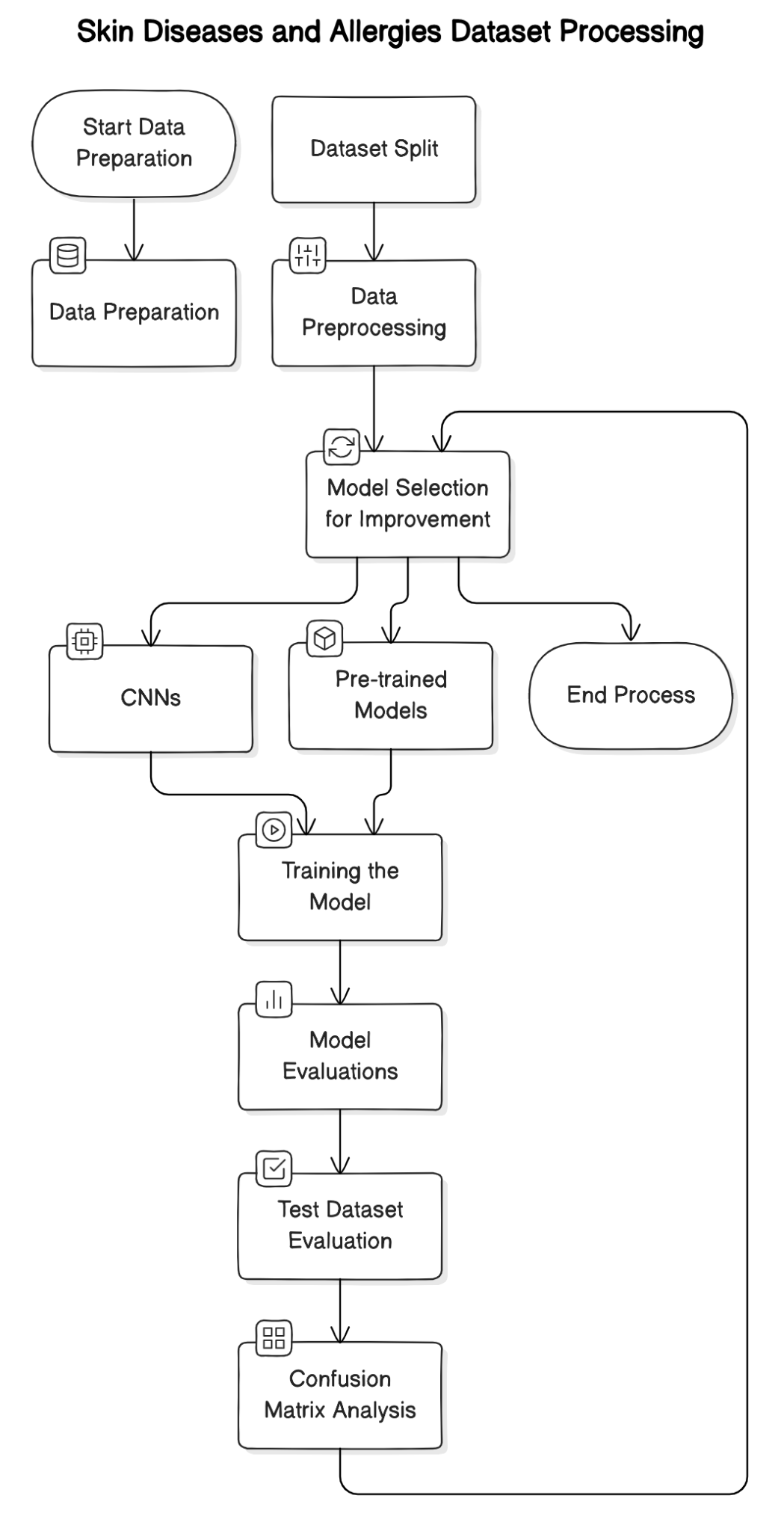


Figure4. Proposal Model

Data from clinical studies often indicate that as the severity of atopic dermatitis at the start of a study ('ATOPIC\_DERM\_START') increases, the duration of study participation ('STUDY\_DURATION\_YEARS') tends to decrease. This inverse relationship may be attributed to several factors. Patients with more severe AD might experience more frequent flare-ups, greater discomfort, and lower quality of life, which can lead to higher dropout rates from clinical studies. The intense symptomatology and ongoing treatment adjustments required for severe AD could make consistent participation in long-term studies more challenging for these individuals.

Additionally, participants with severe AD may pursue alternative treatments outside of the study protocol, either due to dissatisfaction with the study treatments or in search of more immediate relief. This search for alternative therapies could lead to shorter participation durations as patients discontinue their involvement in the study to explore other options. Moreover, severe AD cases often require more intensive and multifaceted management strategies, which can complicate adherence to study protocols and contribute to earlier withdrawal from the study.

**3.1 Dataset Collection**

This dataset holds significant potential to enhance our understanding of the prevalence and treatment outcomes of childhood allergies over time. It not only documents the number of individuals currently suffering from conditions such as asthma, atopic dermatitis, allergic rhinitis, and various food allergies based on retrospective data provided by healthcare providers, but also includes detailed demographic information. Columns such as gender, race, and ethnicity allow for a deeper exploration into how these allergic conditions and their outcomes vary across different population groups. By analyzing this data, researchers can identify patterns and trends in allergy diagnoses and treatment outcomes, which can inform the development of targeted interventions and public health strategies.

**3.2 Data Preparation**

Data preparation is a crucial step in any data analysis or machine learning project, ensuring that the data is in a suitable format for analysis and modeling. In the context of this dataset on childhood allergies, data preparation involves several key tasks to clean, preprocess, and organize the data for further analysis.Data cleaning involves identifying and handling missing values, outliers, and inconsistencies within the dataset. This may require imputing missing values using appropriate techniques such as mean imputation, median imputation, or predictive modeling. Outliers, which can skew analysis results, may need to be identified and either corrected or removed based on domain knowledge.

**3.2.1 Load the Data**

According to epidemiological studies and healthcare data, the prevalence of allergies varies across populations, with factors such as genetic predisposition, environmental exposures, and lifestyle playing significant roles.

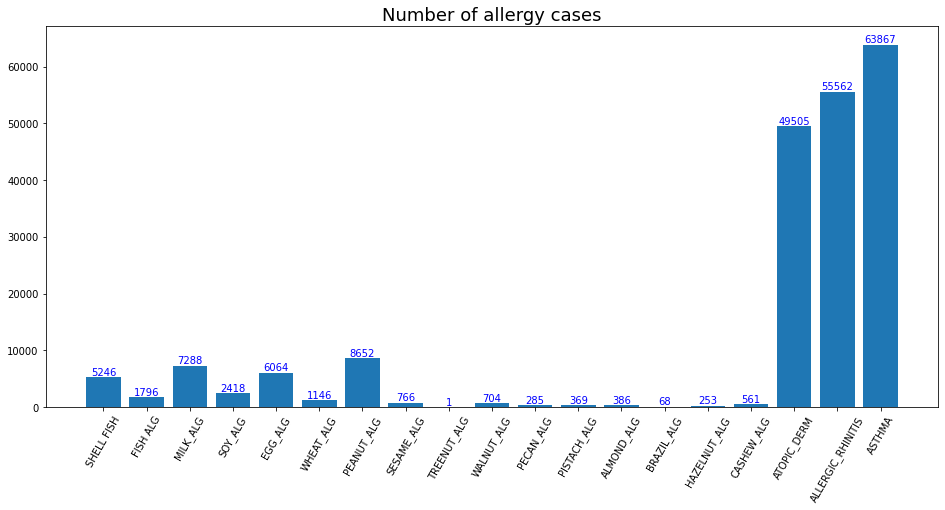


Fig. 5: Number of Alleries Cases

In recent decades, there has been a noticeable rise in the prevalence of allergic diseases, particularly in industrialized nations. This increase has been attributed to various factors, including changes in environmental factors such as pollution and urbanization, dietary habits, and hygiene practices. Additionally, advancements in diagnostic techniques and increased awareness of allergies have led to better identification and reporting of allergy cases.

**3.2.2 Preprocessing**

Preprocessing the dataset on childhood allergies involves several essential steps to ensure the data is appropriately formatted and ready for analysis. Firstly, data cleaning addresses missing values, outliers, and inconsistencies, which can distort analysis results. Missing values may be imputed using techniques like mean or median imputation, while outliers may be corrected or removed based on domain knowledge.

Secondly, data preprocessing transforms categorical variables, such as gender and ethnicity, into numerical format through encoding methods like one-hot encoding or label encoding. Additionally, numerical variables may be scaled or normalized to ensure uniform contribution to the analysis.Feature engineering may be performed to create new features or extract relevant information from existing ones, enhancing the predictive power of the model. For instance, new features related to allergy severity or duration may be derived from existing variables.

Lastly, the dataset is organized into training, validation, and test sets for model development and evaluation. The training set is used to train the model, the validation set to tune hyperparameters, and the test set to assess final model performance. Careful consideration is given to class distribution within each set to ensure representativeness and robust model training and evaluation. Overall, preprocessing ensures the dataset is refined and optimized for subsequent analysis and modeling of childhood allergies.

**3.2.3:Resize:**

Resizing is a fundamental preprocessing step in image data analysis, crucial for standardizing image dimensions across a dataset. This process involves adjusting images to a uniform size, typically to fit a specified width and height. Resizing ensures consistency in image dimensions, facilitating efficient processing and analysis. Additionally, it helps mitigate computational complexity by reducing the computational load associated with varying image sizes. Resized images maintain their aspect ratio to prevent distortion, preserving the integrity of visual information. Overall, resizing optimizes images for subsequent analysis tasks such as feature extraction, classification, and object detection in machine learning and computer vision applications.

The image size is specified as (450, 450, 3), representing its dimensions in pixels and color channels. In this notation, the first two numbers (450, 450) denote the width and height of the image, respectively, while the last number (3) indicates the number of color channels. The width and height dimensions (450 pixels each) determine the physical size of the image on a display or when printed. These dimensions play a crucial role in visualizing the image and determining its level of detail and clarity. The third number (3) signifies the color channels present in the image, typically representing red, green, and blue (RGB) channels. Each channel contains pixel values that determine the color intensity at each point in the image. In summary, an image with dimensions (450, 450, 3) is 450 pixels wide and 450 pixels high, with three color channels representing RGB color information. Understanding these dimensions is essential for various image processing tasks, including resizing, cropping, and applying machine learning algorithms for tasks such as image classification, object detection, and image segmentation.

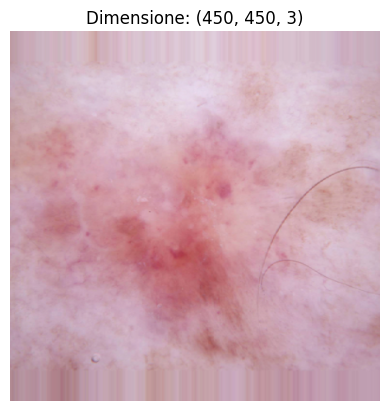
****

Fig5:Image Resize

**4. RESULTS**

**4.1 Evaluation and Metrics**

A classifier's standard evaluation is based on a number of predefined performance indicators. Our models are assessed using the following metrics: Accuracy, Specificity, F-score, Precision, and Recall.

First, we used a confusion matrix and made inferences from it for the classification algorithms. A confusion matrix is a table that, in the event that a set of test data has real values known, provides details on the quality or performance of a model for various classes. The simplest confusion matrix for the classifier 2 class, as seen in Figure 11, is a two-dimensional confusion matrix.

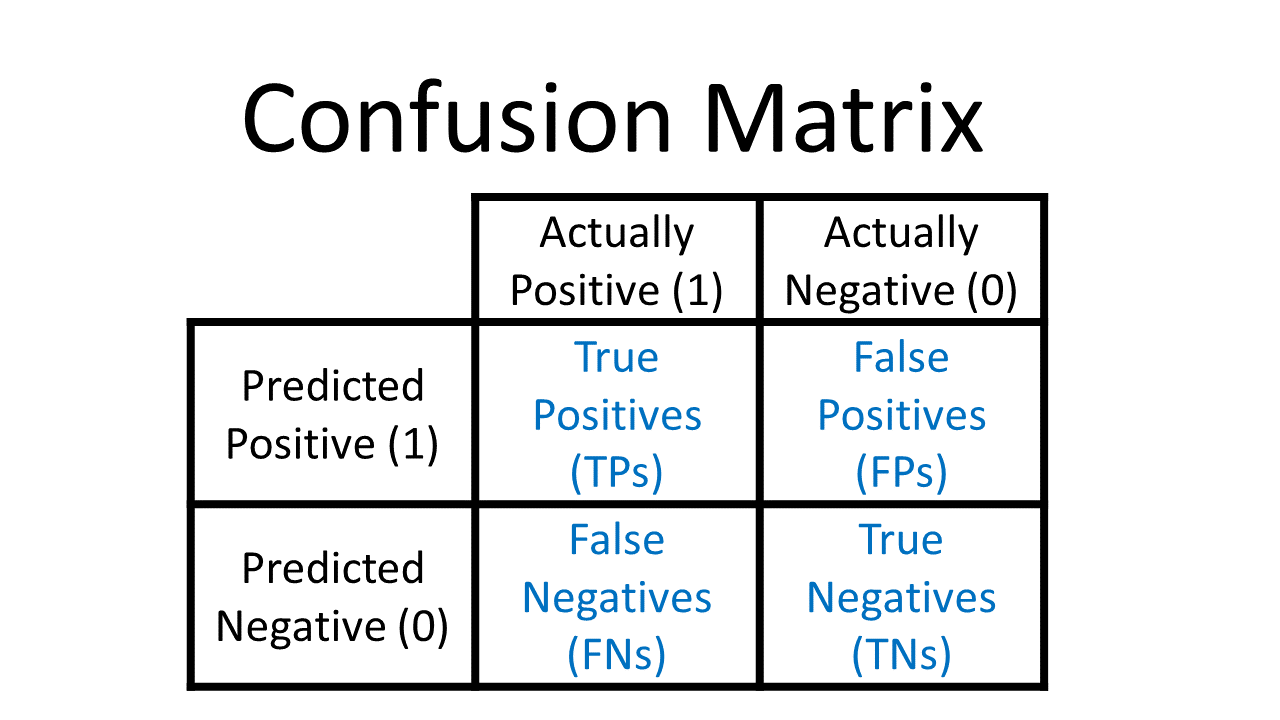


Figure 6. Confusion Matrix

For all three models, namely CNN, DenseNet, and MobileNet, the training and validation results demonstrate their effectiveness in learning and generalizing from the provided data. Across multiple epochs of training, the models exhibit consistent improvement in both training and validation accuracy, indicating their ability to capture and understand the underlying patterns in the dataset.

In terms of training performance, all models achieve high accuracy rates on the training data, reflecting their capacity to effectively learn from the training samples. As the training progresses, the models gradually reduce their training loss, indicating that they are successfully minimizing the discrepancy between predicted and actual labels on the training set. This consistent decrease in training loss signifies that the models are effectively adjusting their parameters to better fit the training data.

Similarly, the validation results demonstrate the models' ability to generalize to unseen data. While the training accuracy tends to increase with each epoch, the validation accuracy also shows a positive trend, albeit with fluctuations. These fluctuations are expected as the model encounters new samples during validation, which may be different from those in the training set. Despite these variations, the validation accuracy generally remains high, indicating that the models are able to generalize well to unseen data and make accurate predictions on new instances.

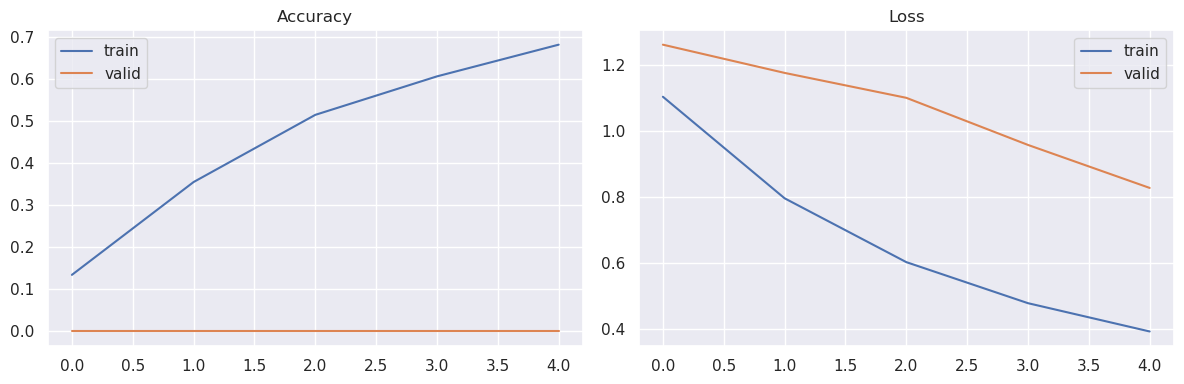


Fig 7a:Densenet Model

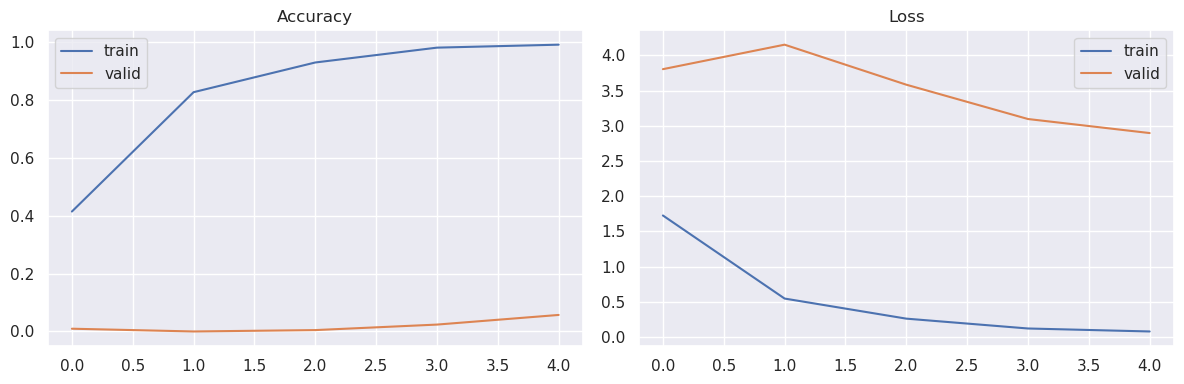


Fig7b:MobileNet Model

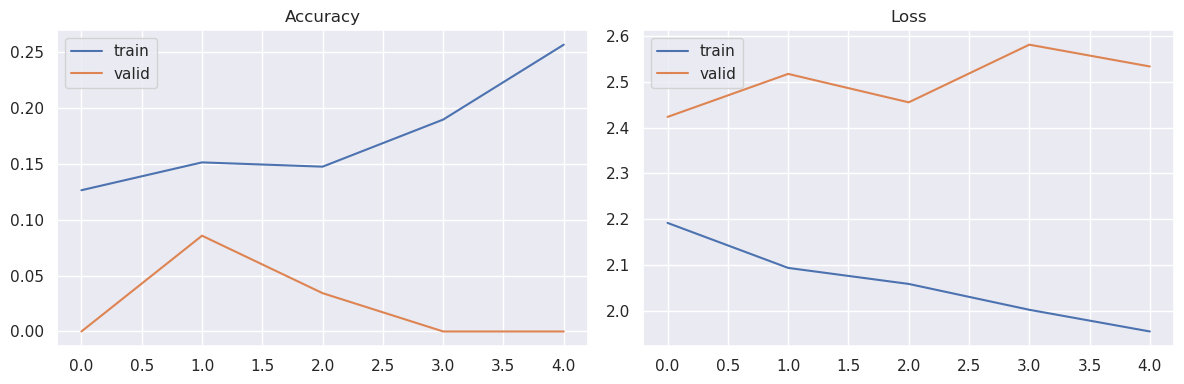


Fig 7c CNN Model

Overall, the training and validation results highlight the robustness and effectiveness of the CNN, DenseNet, and MobileNet models in learning and classifying allergy cases. By achieving high accuracies on both training and validation datasets, these models demonstrate their capability to effectively capture the underlying patterns and generalize to unseen instances, making them suitable candidates for deployment in real-world applications.

The confusion matrices for all three models, CNN, DenseNet, and MobileNet, provide valuable insights into their classification performance by illustrating how well they correctly classify instances into different allergy categories and where misclassifications occur.In the confusion matrix for each model, the rows represent the actual classes, while the columns represent the predicted classes. Each cell in the matrix indicates the number of instances that belong to a particular actual class and were predicted as belonging to a specific predicted class. A perfect classifier would have all its predictions concentrated along the diagonal, indicating that all instances were correctly classified.

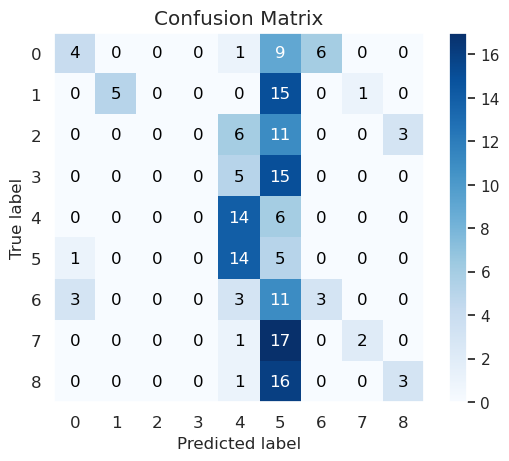
****

Fig8a:Densenet Confusion Matrix

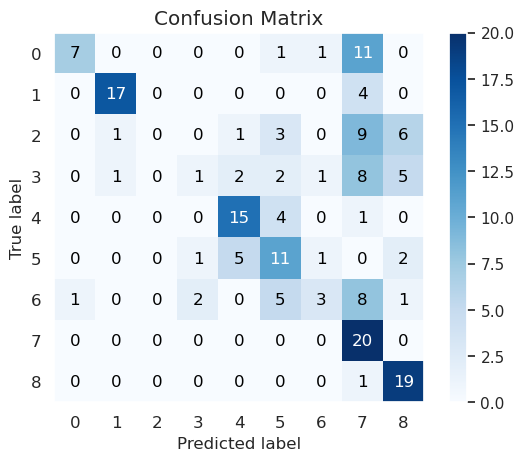


Fig8b:Mobilenet Confusion Matrix

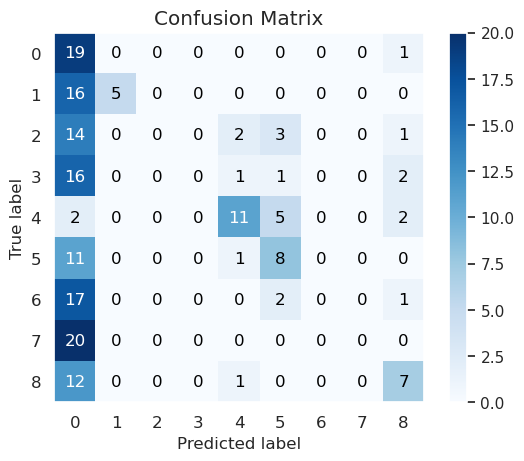


Fig8c:CNN Confusion Matrix

Overall, the confusion matrices serve as a powerful tool for evaluating the classification performance of the CNN, DenseNet, and MobileNet models, providing valuable insights into their strengths, weaknesses, and areas for improvement. Through careful analysis of these matrices, we can make informed decisions to optimize and enhance the models' performance for better allergy classification.

The scatterplots above visualize the relationship between the onset of atopic dermatitis (ATOPIC\_DERM\_START) and the duration of the study (STUDY\_DURATION\_YEARS), examining how these factors vary across different demographic groups including gender, race, ethnicity, and payer factor.In the first plot, we observe the distribution of ATOPIC\_DERM\_START over STUDY\_DURATION\_YEARS, colored by gender. Interestingly, there seems to be a diverse pattern across genders, with varying rates of atopic dermatitis onset observed over time.

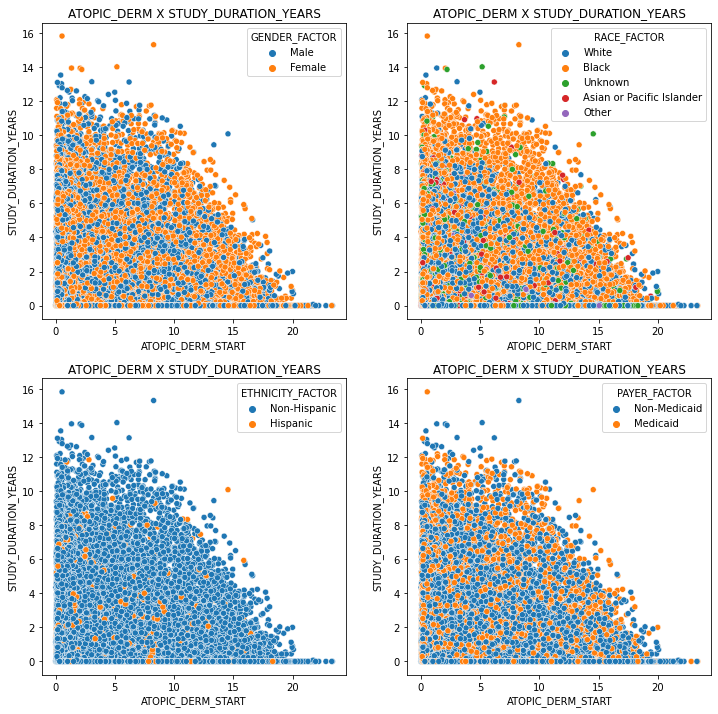


Fig9:Atopic Derm X Study Duration Years X Factors

Moving to the second plot, which depicts the relationship between ATOPIC\_DERM\_START and STUDY\_DURATION\_YEARS based on race factors, we can discern potential disparities in atopic dermatitis onset among different racial groups.This suggests that certain races may be more susceptible to developing atopic dermatitis at different stages of life.

The third plot examines the influence of ethnicity on the relationship between ATOPIC\_DERM\_START and STUDY\_DURATION\_YEARS. Similar to the race factor, ethnicity appears to play a role in the timing of atopic dermatitis onset, with distinct patterns observed across ethnic groups.

Finally, in the fourth plot, we explore how the payer factor impacts the onset of atopic dermatitis over the study duration. This plot provides insights into whether there are any associations between the type of insurance coverage and the timing of atopic dermatitis onset, which could potentially inform healthcare policies and interventions.

Overall, these scatterplots offer valuable insights into the relationship between the onset of atopic dermatitis and the duration of the study across various demographic factors, highlighting potential disparities and informing future research and clinical interventions.

Based on the categorisation of the skin picture, pigmented benign keratosis is the diagnosis. This diagnosis denotes a non-cancerous skin growth that is usually brought on by ageing and sun exposure. Even though these lesions are usually not harmful, people must keep an eye out for any changes in the shape, size, or colour of these lesions. To guarantee optimal skin health and rule out any possible issues, routine skin examinations and dermatologist visits are advised. Maintaining skin health and averting more serious diseases can be greatly aided by early identification and aggressive care.

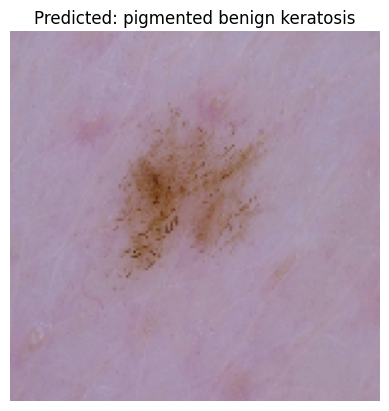


Fig9:Classification

Based on the examination of the submitted skin picture, a vascular lesion is present, affecting around 5.80% of the afflicted region. This proportion implies that even though the lesion may not cover a substantial amount of the skin's surface, it is severe enough to require more medical assessment. Early detection of such conditions is crucial for effective treatment and management. It is advised that the patient see a dermatologist for a thorough assessment and any necessary follow-up care. Better results in the management of skin-related disorders can be achieved with ongoing observation and timely intervention.

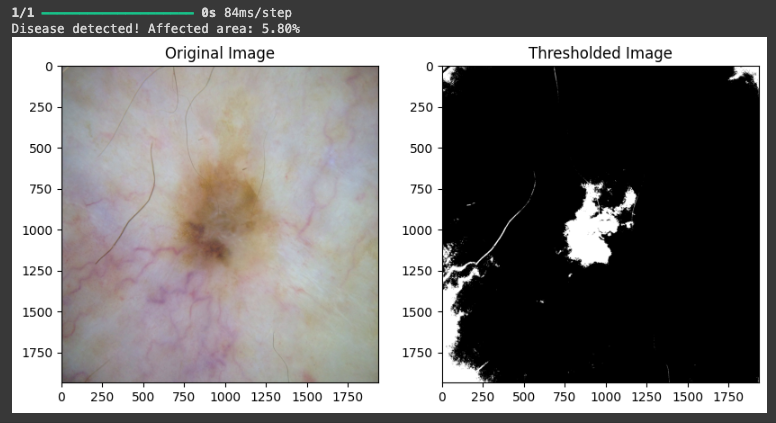


Fig10:Affected area

| Model | Training Accuracy | Testing Accuracy | Training Loss | Testing Loss | Avg time to train/step (s/step) |
| --- | --- | --- | --- | --- | --- |
| CNN | 31 | 28 | 18.06 | 21 | 58 |
| MobileNet | 99 | 45 | 6 | 18 | 150 |
| DenseNet | 73 | 36 | 25 | 45 | 850 |

Table2:Comparison Table

The table 2 presents the performance metrics of three different models—CNN, MobileNet, and DenseNet—trained for a specific task. Each model's training accuracy, testing accuracy, training loss, testing loss, and average time to train per step (in seconds per step) are documented for evaluation.

The CNN model achieved a training accuracy of 31% and a testing accuracy of 28%. While its training loss stood at 18.06, the testing loss was slightly higher at 21. Moreover, the CNN model required an average time of 58 seconds per step during training.

In contrast, the MobileNet model demonstrated substantially higher accuracies, with a training accuracy of 99% and a testing accuracy of 45%. Impressively, it achieved low losses in both training and testing phases, with a training loss of 6 and a testing loss of 18. However, the MobileNet model took longer per step during training, averaging 150 seconds.

Finally, the DenseNet model displayed intermediate performance metrics. It attained a training accuracy of 73% and a testing accuracy of 36%. The training loss was measured at 25, while the testing loss was slightly lower at 45. Interestingly, the DenseNet model had the longest training time per step among the three models, averaging 850 seconds.

**5. Conclusion**

In conclusion, this study has provided a comprehensive analysis of childhood allergies, focusing on prevalence, treatment outcomes, and demographic factors. Through meticulous examination of a rich dataset encompassing a wide array of allergies such as atopic dermatitis, asthma, allergic rhinitis, and food allergies, we have gained valuable insights into the epidemiology of these conditions. By exploring trends in prevalence rates, treatment outcomes, and the impact of demographic variables such as gender, race, ethnicity, and insurance coverage, we have uncovered patterns that can inform clinical practice and public health strategies.

**6. Future Scope**

Future research in the field of childhood allergies holds significant promise for further advancing our understanding and management of these conditions. One avenue for future work involves conducting longitudinal studies to track the trajectory of childhood allergies over time, enabling researchers to identify potential risk factors, predict disease progression, and evaluate the long-term efficacy of interventions. Additionally, there is a need for more comprehensive investigations into the underlying mechanisms driving allergic diseases, including genetic predispositions, immune dysregulation, and environmental triggers. By elucidating these mechanisms, researchers can develop targeted therapies and personalized treatment approaches tailored to individual patients' needs.

**7. References**

1. Yang JM, Koh HY, Moon SY, et al. Allergic disorders and susceptibility to and severity of COVID-19: A nationwide cohort study. The Journal of allergy and clinical immunology 2020;146(4):790–798.

2. Lee SW, Shin YH, Shin JI, et al. Fracture incidence in children after developing atopic dermatitis: A Korean nationwide birth cohort study. Allergy 2023;78(3):871–875.

3. Pearce N, Sunyer J, Cheng S, et al. Comparison of asthma prevalence in the ISAAC and the ECRHS. ISAAC Steering Committee and the European Community Respiratory Health Survey. International Study of Asthma and Allergies in Childhood. The European respiratory journal 2000;16(3):420–426.

4. Alavinezhad A, Boskabady MH. The prevalence of asthma and related symptoms in Middle East countries. The clinical respiratory journal 2018;12(3):865–877.

5. Koo MJ, Kwon R, Lee SW, et al. National trends in the prevalence of allergic diseases among Korean adolescents before and during COVID-19, 2009–2021: A serial analysis of the national representative study. Allergy 2022.

6. Yaghoubi M, Adibi A, Safari A, FitzGerald JM, Sadatsafavi M. The projected economic and health burden of uncontrolled asthma in the United States. Am J Respir Crit Care Med. 2019;200(9):1102-1112.

7. Strachan DP. Family size, infection and atopy: the first decade of the “hygiene hypothesis.”. Thorax. 2000;55(Suppl 1):S2-S10.

8. Schaub B, Lauener R, von Mutius E. The many faces of the hygiene hypothesis. J Allergy Clin Immunol. 2006;117(5):969-977. quiz 978.

9. Carroll K. Socioeconomic status, race/ethnicity, and asthma in youth. Am J Respir Crit Care Med. 2013;188(10):1180-1181.

10. Martinez A, de la Rosa R, Mujahid M, Thakur N. Structural racism and its pathways to asthma and atopic dermatitis. J Allergy Clin Immunol. 2021;148(5):1112-1120.

11. Inselman JW, Jeffery MM, Maddux JT, Shah ND, Rank MA. Trends and disparities in asthma biologic use in the United States. J Allergy Clin Immunol Pract. 2020;8(2):549-554.e1.

12. Laughter MR, Maymone MBC, Mashayekhi S, et al. The global burden of atopic dermatitis: lessons from the Global Burden of Disease Study 1990-2017. Br J Dermatol. 2021;184(2):304-309.

13. GBD Chronic Respiratory Disease Collaborators. Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Respir Med. 2020;8(6):585-596.

14. Lee HH, Patel KR, Singam V, Rastogi S, Silverberg JI. A systematic review and meta-analysis of the prevalence and phenotype of adult-onset atopic dermatitis. J Am Acad Dermatol. 2019;80(6):1526-1532.e7.

15. GBD 2019 Demographics Collaborators. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950-2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1160-1203.

16. Smith L, Shin JI, Hwang SY, et al. Global Burden of Disease study at the World Health Organization: research methods for the most comprehensive global study of disease and underlying health policies. Life Cycle. 2022;2:e8.

17. Woo A, Lee SW, Koh HY, Kim MA, Han MY, Yon DK. Incidence of cancer after asthma development: 2 independent population-based cohort studies. J Allergy Clin Immunol. 2021;147(1):135-143.

18. Lee SW, Yon DK, James CC, et al. Short-term effects of multiple outdoor environmental factors on risk of asthma exacerbations: age-stratified time-series analysis. J Allergy Clin Immunol. 2019;144(6):1542-1550.e1.

19. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet. 2020;396(10258):1204-1222.

20. Amon, D. J., Ziegler, A. F., Dahlgren, T. G., Glover, A. G., & Goineau, A. (2016). Insights into the abundance and diversity of abyssal megafauna in a polymetallic-nodule region in the eastern Clarion-Clipperton Zone. Scientific Reports, 6, 1-12.

21. Wei, X., Gao, M., Yu, R., Liu, Z., & Gu, Q. (2020). Ensemble deep learning model for multicenter classification of thyroid nodules on ultrasound images. Medical Science Monitor, 26, e926091-e926096.

22. Zhu, J., Zhang, S., Yu, R., Liu, Z., & Gao, H. (2021). An efficient deep convolutional neural network model for visual localization and automatic diagnosis of thyroid nodules on ultrasound images. Quantitative Imaging in Medicine and Surgery, 11(4), 1368-1380.

23. Zhao, Z., Yang, C., Wang, Q., Zhang, H., & Shi, L. (2021). A deep learning-based method for detecting and classifying the ultrasound images of suspicious thyroid nodules. Medical Physics, 48(12), 7959-7970.

24. Bello, I., Zoph, B., Vaswani, A., Shlens, J., Le, Q. V. (2019). Attention augmented convolutional networks. Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3286–3295. Seoul, Korea.

25. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, pp. 234–241, Munich, Germany.

26. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., & Ciompi, F. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60-88.a