ABSTRACT

The Pretrained Large Language Models (LLMs) have demonstrated outstanding performance across a range of domains. Assisting with medical diagnosis with accurate symptom descriptions is one possible use. To analyze the practical application of LLMs in healthcare, this thesis explores the arena of skin disease diagnosis, a classic topic within AI on medicine. Conventional AI-based techniques for diagnosing skin diseases rely on deep network-driven image classification models, such as ResNet, VGG, Dense Net, etc., which frequently provide single-dimensional capabilities and lack mechanistic knowledge. We choose to incorporate the advantages of both paradigms because, despite their present instability in reasoning, LLMs are inherently flexible. In this thesis, we present a multimodal big language chat-based interactive skin disease diagnosis system. 93% validation accuracy was attained by Visual GLM, a model for image classification that was trained on the HAM10000 dataset. In order to improve the automated diagnostic process, this system interacts with users dialogically, clarifying the reasoning behind the diagnosis and allowing users to offer more context during the chat session. Our research demonstrates the unrealized potential of LLMs in this vital subject and serves as an investigation into the practical application of LLMs in healthcare.

CHAPTER 1

INTRODUCTION

Skin diseases and dermatological issues affect millions of people's quality of life and pose serious difficulties to worldwide healthcare systems. Effective management requires a timelyand accurate diagnosis, but healthcare practitioners frequently experience excessive demand, which causes delays in the start of therapy. With its new technologies that have the potential to transform dermatological care, artificial intelligence (AI) has emerged as a promising answer to these difficulties. Artificial intelligence (AI)-based technologies provide quick and precise evaluations of skin diseases by analyzing large datasets, including pictures and clinical data. This enables medical personnel to effectively make decisions based on information, even in cases where capability is limited. The creation and application of an AI-based tool for first dermatological diagnosis is described in this project report. It discusses the pressing need for enhanced diagnostic skills and looks into how AI

might supplement the knowledge of healthcare professionals in this critical field.

CHAPTER 2

OBJECTIVE

This project endeavors to design, construct, and deploy an advanced artificial intelligence-driven solution tailored specifically for the initial diagnosis of dermatological ailments. The overarching goal is to revolutionize the diagnostic landscape within dermatology by harnessing the power of AI technologies. Through the development of a sophisticated AI tool, we aim to streamline and expedite the diagnostic process, thereby addressing the challenges of delayed therapy initiation and resource constraints faced by healthcare professionals. By leveraging large datasets comprising diverse clinical information and images, the AI-based system will facilitate rapid and precise evaluations of skin diseases. The ultimate objective is to empower medical personnel with an invaluable tool that complements their expertise,

thereby enabling more effective decision-making and ultimately improving patient outcomes in dermatological care.

2.1 Problem Statement

Skin diseases are a major global health concern, ranking as the 4th leading cause of nonfatal diseases. They often signal underlying health issues, impacting well-being. Yet, access to timely and accurate dermatological care remains a challenge, especially in resource-limited regions, leading to delayed treatments. In these underserved areas, skin diseases' impact is worsened by the lack of effective diagnostic tools, connectivity issues, and inadequate labs, making dermatological care tough to improve. Our solution tackles this issue by merging AI and telehealth. We use large language modules (LLM) advanced Artificial Intelligence (AI), like Convolutional Neural Networks (CNNs), for quick and precise skin condition diagnoses. Plus, we add AI-driven Natural Language Processing (NLP) and Natural Language.Understanding (NLU) for easy symptom descriptions during telehealth chats.

2.2 Background

Skin problems represent a significant global health concern, impacting an estimated 1.8 billion individuals worldwide, as outlined by the World Health Organization (WHO) [1]. Addressing these issues effectively necessitates timely and accurate diagnosis to manage conditions and enhance patient quality of life.

Contemporary diagnostic approaches heavily rely on deep learning image classification models such as VGG, ResNet, and Dense Net. These models have demonstrated remarkable success in identifying various skin conditions from image data, leveraging intricate patterns often imperceptible to the human eye [5].

However, despite their efficacy, challenges persist in the development of fully automated diagnostic platforms. Notably, the interpretability of deep learning models poses a significant obstacle, as they often operate as "black boxes," offering results devoid of context. This lack of transparency undermines trust among both patients and clinicians [6]. Furthermore, these models are primarily adept at image categorization and struggle with complex, multimodal tasks or adapting to individual patient circumstances.

Conversely, Large Language Models (LLMs) like BERT and GPT-3 have emerged as prominent tools in natural language processing. These models, exemplified by OpenAI's ChatGPT [9], offer conversational AI capabilities, processing text data and generating contextually relevant responses with human-like interactions. Leveraging the flexibility and diversity of LLMs could significantly enhance the functionality of automated diagnostic platforms.

Despite the potential benefits of LLM integration, challenges persist. LLMs exhibit limitations in understanding the physical world, particularly visual information crucial for skin disease diagnosis. Additionally, their reasoning abilities may be inconsistent, posing challenges in accurately interpreting and diagnosing skin conditions based on images [7, 8].

In summary, while deep learning models have shown promise in skin disease diagnosis, significant hurdles remain in developing fully automated diagnostic platforms. Integrating the capabilities of LLMs holds potential for enhancing diagnostic functionality, but addressing their limitations in visual understanding and reasoning is paramount for effective application in dermatology.

2.3 Aims & Objectives

Our study is motivated by three primary factors in the current state of the field:

1. Utilization of Large Language Models (LLMs) in Medicine:

Despite the widespread use of image classification models in skin disease detection applications, the application of large language models (LLMs) in specialized fields such as skin disease detection is still in its infancy. This research gap presents an opportunity to leverage the language skills of LLMs to enhance the development of platforms for detecting skin diseases.

2. Drawbacks of Current AI-based Medical Models:

While AI-based medical models are receiving increasing attention, they suffer from drawbacks such as knowledge delusion and poor multi-modal capabilities, rendering them unsuitable for direct use in diagnosing skin diseases. Therefore, it is imperative to enhance the capabilities of these models and tailor them to meet the unique requirements of diagnosing skin

diseases.

3. Limitations of LLMs in Real-world Applications:

Most recent AI in medicine projects involving LLMs provide only basic conversational functionalities once trained on specific contextual data. Additionally, there are gaps in these models' real-world application, such as the inability to swiftly fine-tune them in the context of rapidly iterating datasets. Our goal is to investigate methods to bridge these gaps and apply the advantages of LLMs to real-world scenarios in skin disease diagnosis.

CHAPTER 3

LITERATURE SURVEY

In this chapter, we provide a comprehensive review of the literature relevant to our study. The chapter is structured into four main sections: Machine Learning Systems, Large Language Models, Datasets, and Image Classification Models. We begin by discussing the datasets utilized for diagnosing skin diseases, followed by an exploration of large language models and image classification models. Finally, we delve into the current machine learning frameworks, with a focus on their application to large language models and skin disease detection.

Datasets:

Deep learning research relies heavily on datasets, particularly those for skin illness picture classification, essential for developing automated systems for diagnosing skin disorders. Fortunately, numerous publicly available datasets are now accessible for studying the image categorization of skin diseases. We highlight two widely used datasets in this context:

3.1 PH2 Dataset:

The PH2 dataset comprises a series of dermoscopic images focusing on melanocytic lesions. Intended for educational and scientific purposes, this dataset consists of 200 dermoscopic images covering a range of benign lesions and melanomas. It is frequently utilized in studies and initiatives related to the identification and categorization of melanoma [10].

3.2 HAM 10000 Dataset

Developed by the International Skin Imaging Collaboration (ISIC), the HAM10000 dataset serves as a training resource for neural network models for automated diagnosis of skin lesions. It contains over 10,000 dermatoscopic images representing various pigmented skin lesions, including melanoma, basal cell carcinoma, vascular lesions, benign keratoid lesions, melanocyte nevi, actinic keratoses, and dermatofibroma. These images are sourced from diverse populations and captured using multiple modalities, ensuring broad representation [11]

These datasets provide invaluable resources for training and validating machine learning models for skin disease diagnosis, facilitating advancements in automated diagnostic systems.

In subsequent sections, we will explore the utilization of large language models and image classification models in the context of skin disease detection, drawing insights from existing literature to inform our research methodology and approach.

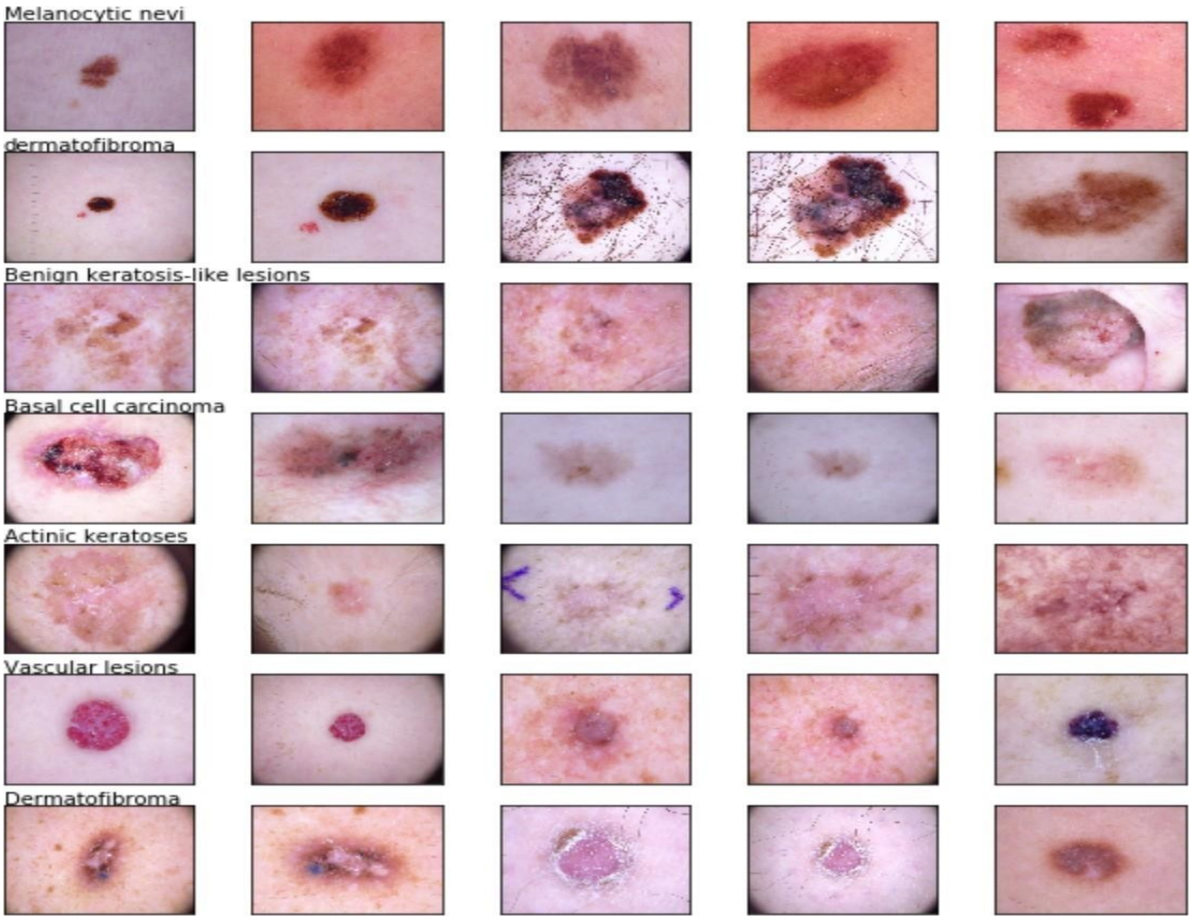

Fig:3.2.1 HAM10000 samples.

3.3 BCN20000 Dataset

The BCN20000 dataset comprises 19,424 dermoscopic images of skin lesions collected at the Hospital Clínic in Barcelona between 2010 and 2016. This meticulously curated dataset is checked for diagnostic accuracy and is suitable for various tasks such as lesion detection,

segmentation, and classification. It has obtained all necessary institutional ethics approvals and is based on high-resolution images [12].

3.4 ISIC Archive

The ISIC Archive is an open-access collection of dermatoscopic images, serving as a valuable resource for dermatologists developing and validating image analysis algorithms. This archive contains an extensive collection of skin lesion photos, updated yearly, captured using various tools and settings, reflecting real-world challenges in skin lesion identification. As of June 2023, it encompasses 76,295 dermatoscopic images, each accompanied by comprehensive metadata including diagnosis (if available) and expert commentaries on many photographs [13].

3.5 Comparison of Commonly Used Skin Disease Datasets

DATASET	NO. OF IMAGES	DISEASE TYPES	YEAR
PH^2	200	Melanoma and benign lesions	2013
HAM10000	10015	7 Types	2018
BCN20000	19424	8 Types	2021
ISIC Archive	76295	Various	2018

Table 3.5: Comparison of Commonly Used Skin Disease Datasets

3.6 Deep Learning on Skin Disease Detection:

3.6.1 Deep Learning:

Deep learning has revolutionized various fields including computer vision, natural language processing, and healthcare. In computer vision, deep learning has enabled computers to comprehend visual data with accuracy comparable to or better than human levels. Deep learning models, also known as deep neural networks, are designed to extract higher-level

features from raw input data through multiple layers, allowing them to identify intricate patterns. Key concepts in deep learning include:

Linear Models: Linear models form the foundation of deep learning, relying on linear combinations of input variables for predictions or judgments.

While straightforward, they have limited expressive power and struggle with complex interactions.

Activation Functions: Neural networks employ activation functions to carry out non-linear transformations, enabling them to recognize complex patterns. Common activation functions include Sigmoid, tanh, and ReLU (Rectified Linear Unit).

Fully Connected Networks: Fully connected networks, also known as multi-layer perceptions, consist of neurons connected to every other neuron in adjacent layers. While capable of learning complex functions from input features, they are prone to overfitting and a large number of parameters.

Training deep learning models involves optimization algorithms and loss functions. Backpropagation, a technique used in model training, modifies parameters based on loss function. Optimization algorithms, such as gradient descent, update model parameters iteratively to minimize loss.

Loss Functions: Loss functions measure the difference between actual and predicted values, with common types including Mean Squared Error and Cross-Entropy Loss.

Optimization Algorithms: Techniques like gradient descent update model

parameters to minimize loss, scanning the entire training set at each parameter update step.

These foundational concepts underpin the training and deployment of deep learning models, including those used in skin disease detection applications.

3.6.2 CNN:

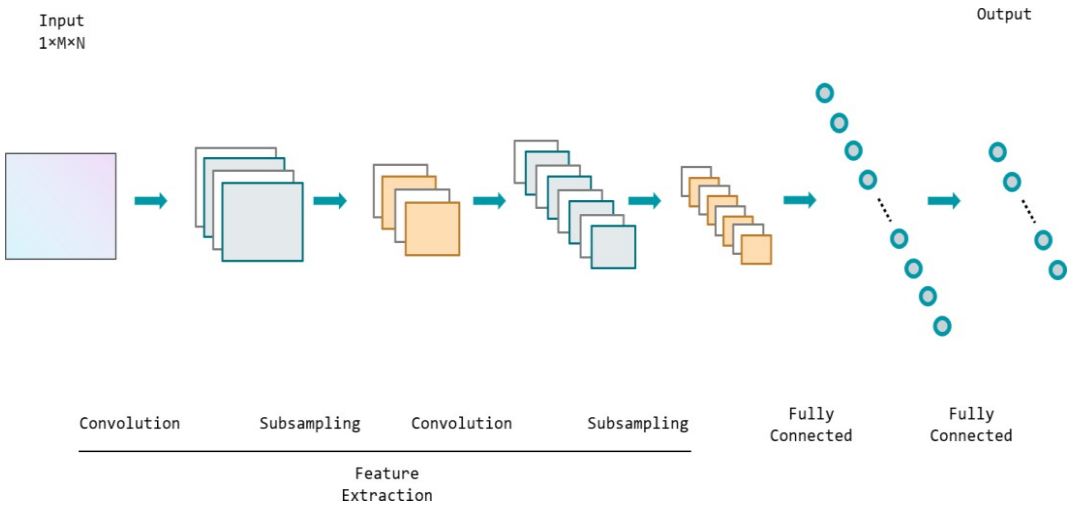

Fig 3.6.2 An illustration of CNN structure.

Convolutional Neural Networks (CNNs) have become a cornerstone in deep learning models, particularly in computer vision and pattern recognition applications. Yann LeCun et al. (1998) introduced CNNs, which have gained significant traction due to their efficacy in processing grid-like

data such as images and audio. The key features of CNNs include pooling layers, shared weights, and local connections, enabling them to identify local patterns in data efficiently while minimizing computing costs and preventing overfitting.

The fundamental architecture of CNNs consists of convolutional and pooling layers. Convolutional layers utilize kernels or filters to extract features from input data, such as edges and textures, through element-wise multiplication and summation. Pooling layers, like max pooling, then reduce the spatial dimensions of the feature maps, preserving essential features while reducing computational complexity.

Fully connected layers further process high-dimensional feature maps, converting them into a probability distribution corresponding to different classifications. During training, CNNs utilize backpropagation to adjust model parameters, optimizing them to minimize the difference between expected and actual labels through techniques like gradient descent.

In various computer vision tasks including semantic segmentation, object detection, and image classification, CNNs have demonstrated remarkable performance due to their ability to abstract and memorize visual features effectively.

Many studies have utilized CNN-based models for skin disease diagnosis, incorporating architectures such as VGG, ResNet, Inception, and Dense Net:

VGG: Known for its simple and uniform design, VGG has been widely used in skin disease classification tasks. Studies like Thao et al. (2017) employed VGG-16 to improve the accuracy of skin disease classification by learning powerful representations from dermatology photos.

ResNet: ResNet addresses the vanishing gradient problem and enables training of extremely deep networks through skip connections and residual blocks. Studies like Mendes et al. (2018) utilized ResNet-152 to effectively categorize various skin disorders with high accuracy.

Inception: Characterized by its Inception modules, Inception drastically reduces the number of model parameters while capturing multi-scale features efficiently. Studies like Devries et al.

(2017) demonstrated the effectiveness of Inception v3 in skin cancer classification using images from the ISIC 2017 archive.

Dense Net: Dense Net introduces dense connections between layers, promoting feature reuse and gradient flow. Studies like Gessert et al. (2018) achieved commendable results in classifying skin lesions using Dense Net.

Overall, these studies showcase the effectiveness of CNN-based models in automating the detection of skin diseases, achieving high accuracy and demonstrating the potential for improved diagnostic capabilities.

3.7 Large Language Models:

Large Language Models (LLMs) are at the forefront of artificial intelligence, offering groundbreaking capabilities in text prediction and generation. Leveraging artificial neural networks and transformer-based attention mechanisms, LLMs have revolutionized natural language processing (NLP) tasks. In the healthcare sector, their potential for transforming medical assistance, diagnosis, and treatment is particularly promising.

The essence of LLMs lies in their expansive structure, incorporating millions to billions of adjustable weights. This allows them to capture complex patterns and generate coherent text, making them invaluable tools for various applications. The transformer architecture, introduced in 2017, has played a pivotal role in advancing LLMs, surpassing older specialized models and setting new standards in linguistic tasks.

Notable LLMs such as GPT-4 and Llama exemplify the power and versatility of these models.
They serve as the backbone of AI search engines and chatbots, showcasing their immense

promise in healthcare. Open-source projects like Alpaca and ChatGLM-6B have further expanded the capabilities of LLMs, demonstrating significant advancements in NLP techniques.

In healthcare, LLMs have been instrumental in tasks ranging from medical advice to diagnosis. Projects like Chat Doctor and Visual-Med-

Alpaca showcase how LLMs can provide medical assistance beyond text-only interactions. Visual GLM, a multimodal big language model with both picture and text processing capabilities, has emerged as a promising tool for medical applications.

Despite the presence of robust general-purpose language models like GPT-4, fine-tuning remains necessary for adapting LLMs to specific healthcare tasks. Parameter-efficient fine-tuning techniques such as distillation, adapter training, and Low-Rank Adaption (LoRA) minimize computational resources while optimizing pre-trained models for specialized

applications. These techniques empower researchers and developers to customize LLMs for
various healthcare challenges.

Overall, the adaptability and effectiveness of LLMs in healthcare underscore their immense potential for revolutionizing medical assistance. Fine-tuning techniques further enhance their utility, making them invaluable tools for addressing specific challenges and improving patient outcomes in the medical field.

3.7.1 Chat GLM:

Chat GLM, built on the General Language Model (GLM) framework, represents a significant advancement in natural language understanding (NLU). Unlike traditional pretraining frameworks, GLM excels in tasks such

as natural language generation and understanding, unconditional generation, and conditional generation.

The key innovation of GLM lies in its autoregressive blank infilling approach, a subset of masked language modeling. Unlike conventional masked language modeling, which treats missing tokens as unrelated, autoregressive blank infilling simulates the relationships between missing tokens. This is achieved through a mixed attention mask that allows the model to attend to both left and right contexts of missing tokens. Additionally, GLM employs a unique 2D position encoding system, encoding the location of each token within a phrase and each missing token's location relative to its context.

ChatGLM-6B, a specific instance of the GLM framework, boasts 6.2 billion parameters and integrates quantization techniques for efficient implementation on consumer-grade graphics cards. The model has been refined through techniques such as feedback bootstrap, reinforcement learning with human input, and supervised fine-tuning. Despite its complexity, ChatGLM-6B achieves results aligned with human preferences.

Access to ChatGLM-6B for academic research is unrestricted, while commercial use requires completion of a questionnaire, ensuring accessibility for both research and business purposes.

In evaluations against previous state-of-the-art models like BERT, T5, and GPT, ChatGLM consistently outperforms in tasks such as text categorization, sentiment analysis, and natural language inference. For

example, on the GLUE benchmark, ChatGLM scores 90.5, surpassing T5's prior record of 89.6. Similarly, on the LAMBADA dataset, which tests long-range dependency representation, ChatGLM achieves a lower perplexity score of 8.6 compared to GPT Large's 10.5.

CHAPTER 4
SYSTEM DESIGN

4.1 Architectural Overview:

The Skin Disease Chat system is meticulously designed to offer efficient and effective skin disease diagnosis to users. The system architecture is divided into three main components: Skin Disease Net, Diagnosis Chat Model, and User Interface system. Each component serves a distinct purpose and contributes to the overall functionality of the system.

1. Skin Disease Net: At the core of the system lies Skin Disease Net, a robust picture classification model dedicated to identifying potential skin diseases from user-submitted images. Leveraging various deep learning networks such as VGG, ResNet, and DenseNet, Skin Disease Net performs the computationally intensive task of image classification. By employing state-of-the-art algorithms, it ensures accurate and reliable analysis of skin images.

2. Diagnosis Chat Model: Acting as the cognitive engine of the system,

Diagnosis Chat Model is a sophisticated multimodal Large Language Model (LLM) essential for facilitating human-computer interaction. This component processes the output from Skin Disease Net, engages in conversations with users, and translates technical classification results into user-friendly language. Its primary function is to provide users with relevant information about identified skin ailments, answer queries, and offer assistance throughout the diagnosis process.

3. User Interface System: Serving as the entry point for users, the User Interface system provides a seamless and intuitive platform for users to interact with the Skin Disease Chat system. Through an easy-to-use interface, users can upload skin photos, communicate with Diagnosis Chat Model, and receive predictions for skin diseases. This component ensures accessibility and enhances user experience by offering a user-friendly interface for navigation and interaction.

The modular design of the Skin Disease Chat system enables efficient management and facilitates updates or enhancements to individual components without compromising the system's overall performance. By leveraging cutting-edge technologies and adopting a user-centric approach, the system aims to provide users with accurate, timely, and user-friendly skin disease diagnosis services.

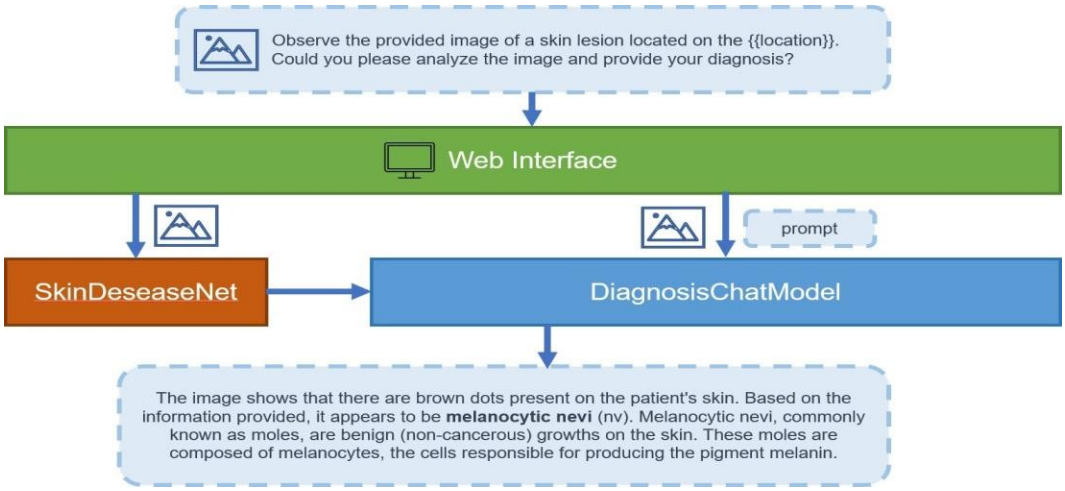

Fig 4.1 System workflow of Derm-AI Chat.

The combination of Skin Disease Net, Diagnosis Chat Model, and the User Interface system results in a complete and easily navigable tool for diagnosing skin diseases. Through a modular design approach, each component specializes in its respective function, ensuring the system's overall strength, effectiveness, and efficiency. The system's workflow, illustrated in Figure 3.1, showcases the seamless interactions between the components, with the User Interface system managing user inputs, forwarding them to the AI models for inference, and providing real-time streaming of results. Skin Disease Net employs advanced deep learning networks to identify skin conditions from image inputs, while DiagnosisChatModel handles user interactions, generating personalized conversational responses. Together, these components create a user-friendly experience that enhances the diagnostic process.

4.2 Skin Diseases Net:

Neural network architectures such as VGG, ResNet, and DenseNet are viable options for constructing Skin Disease Net. While our system supports multiple models, we will primarily discuss our implementation with the ResNet-50 model to avoid excessive detail. The implementation process involves four key steps: model construction, training, evaluation, and data analysis and preprocessing. These steps are crucial for building an effective and accurate skin disease classification system.

MODEL	DEPTH	NUMBER OF PARAMETERS	VALIDATION ACCURACY
ResNet-50	50	25.6 million	0.90
ResNet-101	101	44.6 million	0.92
ResNet-152	152	60.3 million	0.93
VGG-16	16	138 million	0.87
DenseNet-121	121	8 million	0.88

TABLE 4.2.1: Comparison of pre-trained Skin Disease Net models.

4.2.1 Data Preprocessing and analysis:

Data quality and management are fundamental to the success of machine learning models, and our approach begins with leveraging the extensive HAM10000 dataset. This dataset comprises a diverse collection of dermato-scopic images of pigmented lesions sourced from various

channels. To ensure stable and effective training of Skin Disease Net, we prioritize data normalization. This process involves computing the mean and standard deviation of the RGB channels across the entire dataset. Initially, the mean (μ) is calculated by summing up all individual data points (xi) and dividing the sum by the total number of data points (N). Subsequently, the standard deviation (σ) is determined by computing the differences between each data point and the mean, squaring these differences, summing up the squared differences, dividing by the total number of data points, and taking the square root of the result.

After data preprocessing, the dataset is divided into training and validation sets. From the total of 9,187 records in the dataset, 828 records are allocated to the validation set. This allocation is conducted randomly to ensure a consistent distribution of classes across both sets, maintaining the integrity of the dataset. Addressing class imbalances within the dataset is crucial to prevent model bias and ensure accurate predictions. Our analysis identifies significant class imbalances across the seven classes in the training data. To mitigate this issue, we adopt a dual approach involving equalization sampling and a class-balancing loss function.

Equalization sampling is implemented by duplicating records from classes with fewer instances, effectively creating a balanced distribution of instances across all seven classes. This approach helps to rectify the class imbalances within the dataset and ensures that the classifier does not become biased towards dominant classes during training. Additionally, incorporating a class-balancing loss function further enhances the model's

ability to learn from imbalanced data by adjusting the contribution of each class to the overall loss function based on their frequency in the training set. These preprocessing techniques collectively contribute to the robustness and effectiveness of Skin Disease Net in accurately identifying skin diseases from input images.

4.2.2 Model Building:

We proceed to the model building step once the data has been properly prepared. We use the PyTorch framework [45] for this, which has the benefits of dynamic computing graphs, streamlined model creation, and a helpful community. The ResNet-50 architecture is the one we have chosen for this purpose. ResNet, also known as Residual Networks, uses skip connections or shortcut connections to allow the network to skip layers and avoid disappearing gradients, a typical issue with deep neural networks. In ResNet-50, the "50" indicates

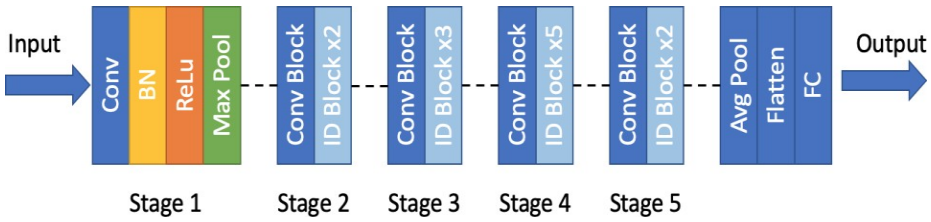

Fig4.2.1: ResNet-50 network architecture.

ResNet-50, a widely adopted convolutional neural network architecture, is meticulously designed to identify intricate patterns in data. The architecture's composition is structured to facilitate the recognition of complex features within images.

Input Layer:

The initial layer of ResNet-50, the input layer, receives image data with dimensions of 224x224x3. This configuration signifies that the input comprises images with a resolution of 224 pixels in height and width, with three channels representing the RGB color space.

Initial Convolutional and Max-Pooling Layers:

Following the input layer, ResNet-50 begins with a solitary convolutional layer featuring a 7x7 sized kernel, a stride of 2, and generating 64 output channels. Subsequently, a batch normalization layer, a ReLU activation layer, and a max pooling layer work synergistically to reduce dimensionality and extract primary features from the image.

Convolutional Layers:

ResNet-50 is characterized by its 48 convolutional layers, which play a pivotal role in capturing intricate patterns in the data. These layers are organized into Convolutional Blocks and Identity Blocks, distributed across four stages. Each stage comprises a different number of blocks, with each block containing three layers.

Fully Connected Layer:

As the network progresses towards its final layers, it encounters a global average pooling layer, followed by a fully connected layer. The fully connected layer condenses the extracted information to generate final predictions. Subsequently, the softmax layer assigns probabilities to each class in the classification problem, with the model's prediction being determined by selecting the class with the highest probability.

4.2.3 Model Training:

The first stage in which the model learns to differentiate between various skin conditions is the training phase. The ImageNet dataset [46], a massive database containing over a million tagged images in a thousand categories, is used to pre-train our ResNet-50 network. Pre-training, or transfer learning, is an essential approach that helps us jump-start learning using patterns previously learnt from a related task, resulting in improved performance and faster convergence.

We use a cross-entropy based loss function during the training procedure. The effectiveness of our classification model is assessed by this function, which produces a probability value between 0 and 1. The size of the cross-entropy loss increases in proportion to the projected probability's deviation from the real label, which motivates our model to produce accurate predictions. Let n be the total number of classes—7 in our case—and let yi be the true probability (ground truth) of the i-th class. Let yˆi be the predicted probability (output) of the i-th class by the model.

The following is the definition of the Cross-Entropy Loss for a singledata point:

For training, we employ the Adam optimizer. Popular optimization algorithm Adam (Adaptive Moment Estimation) combines the advantages of AdaGrad [16] and RMSprop.

The model parameters to be optimized are represented by θ. The learning rate is denoted by α. The gradient of the objective function with respect to θ at time step t is represented by gt. The exponential decay rates for the moving averages are represented by β1 and β2.

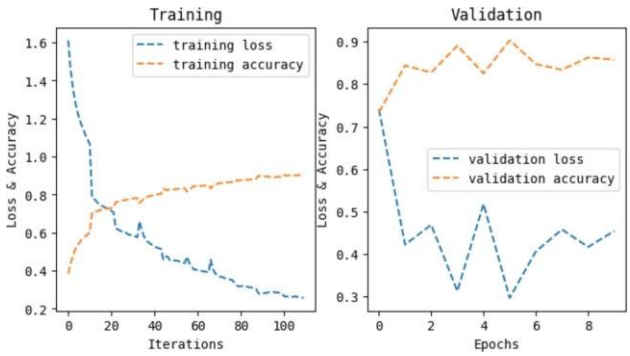

Fig 4.2.2: The fine-tuning procedure

4.2.4 Model Evaluation:

Any model's effectiveness is determined by how effectively it functions with unknown data. Using our validation set as a testing ground, our ResNet-50 model produced an average 90%accuracy over ten epochs in all seven classes. It illustrates how the model may apply generalization to fresh data and produce accurate predictions. The accuracy and loss during epochs and iterations for training and validation are displayed in

Figure 4.2.2.

4.3 Diagnosis Chat Model:

The conversational AI part of our system that manages the language part of our application is called Diagnosis Chat Model. Its duties include communicating with users, answering their inquiries, posing pertinent queries, and helping to interpret the Skin Disease Net model's predictions of skin diseases. It assists the user in comprehending the possible skin disorders identified by translating the image categorization results into comprehensible and educational text. Additionally, it directs users to take additional steps, such getting expert medical counsel as needed.

4.3.1 Model Architecture

Three different forms of input are accepted by Diagnosis Chat Model: user-uploaded images,Skin Disease Net results, and user-inputted text prompts. The system responds to the user with a stream of text created after processing these inputs. Diagnosis Chat Model uses the VisualGLM-6B model as the multimodal LLM for user interaction. An open-source, multimodal conversation language model that supports both Chinese and English text and images is called VisualGLM-6B. It can be locally installed on GPUs intended for consumer usage and

doesn't require a lot of resources. At an INT4 quantization level, 8.7GB of GPU RAM is all that is required. This feature demonstrates how VisualGLM-6B is useful and within reach for a wide range of users. Its linguistic component comes from encompassing 6.2 billion parameters in ChatGLM-6B. By means of the visual and linguistic models, resulting in a total of 7.8 billion parameters for the entire model. This is achieved by the use of BLIP2-Qformer training [47] in the image section. The HealthcareMagic-100k Datasets [48] and 100 diagnostic conversations produced by the HAM10000 dataset were the firsttwo sets of data used to fine-tune the model.

4.3.2 Fine Tuning

We use a special technique called P-tuning v2 in the training phase of our Diagnosis Chat Model to fine-tune our LLM. Over a range of tasks and scales, P-tuning v2, an improved version of the deep prompt tuning technique, performs comparably to classical fine- tuning [40]. Deep prompt tuning, which applies continuous prompts to each layer input of thepretrained transformer model, is the fundamental technique used in P -tuning v2. This approach greatly improves continuous prompting and effectively closes the performance difference with traditional fine-tuning in a variety of scenarios, especially with smaller models and harder jobs. Put differently, P-tuning v2 is an enhanced version of Deep Prompt Tuning that has been tailored and refined for use in natural language

understanding (NLU) applications.

Let's examine the particular parameters that were applied during the model training stage in order to put this into context. Table 3.2 displays the training parameters: To determine how long the token sequence is before the model processes it, we first set the soft prompt length (pre seq len) to 128. 2e−2 is the model's learning rate. Each repetition brings it closer to the loss function's minimum while controlling the step size. The maximum length of the source and target sequences used for training is indicated by the values max source length

and max target length, which we employ as 64. We have 4 and 1 device-per-batch sizes for training and evaluation. The number of gradient steps is four, which we have set for the gradient ac- cumulation stages buildup prior to carrying out a reverse propagation. Finally, we use a quantization bit value of 4 for the model, which helps to optimize its performance by lowering the computational intensity and model size, which makes the model faster and lighter to execute.

Parameter	Value
pre seq len	128
max source length	128
max target length	64
batch size	4
gradient accumulation steps	4
learning rate	2e-4

quantization bit	4

TABLE 4.2.2: Fine-tuning parameters of Diagnosis Chat Model.

INPUT	OUTPUT
ISIC 0029041.jpg	The image shows that there are brown dots present on the patient's scalp. Based on the in- formation provided, it appears to be Actinic keratosis (AK). Actinic keratosis, also known as so- lar keratosis, is a common skin condition characterized by rough, scaly patches on the skin. These patches are typically found on parts of the body that are most exposed to the sun, such as the face, ears, back of the hands, fore- arms, scalp, and neck. AK is considered a precancerous condition, which means that if left un- treated, it has the potential to progress into squamous cell car-cinoma, a type of skin cancer.

TABLE 4.2.3: Inference example 1 of Diagnosis Chat Model.

We provide a range of customizable options in the inference step of our Diagnosis Chat Model, enabling customers to customize the AI models to their own requirements. The user has the ability to modify each of these variables to suit their requirements and tastes. Among these choices are:

- **Temperature:** A crucial component of the AI model, this parameter regulates how unpredictable the model's outputs are. Whereas a lower temperature value yields more deterministic and concentrated outputs, a higher temperature value yields more varied and creative output. Users are able to customize the trade-off between precision and diversity by adjusting the temperature to suit their preferences.

INPUT	OUTPUT
ISIC 0028790.jpg	The image depicts a reddish-brown bump on the patient's back, consistent with Dermatofibroma. Dermatofibroma is a benign skin condition resulting from fibrous tissue overgrowth, often triggered by minor skin trauma or insect bites. Typically found on the extremities, the lesion is painless but might be tender or itchy. When squeezed, it exhibits a characteristic "dimple sign." While Dermatofibromas are generally harmless and do not necessitate treatment, medical evaluation is essential to confirm the diagnosis and rule out other conditions. Consulting a dermatologist is advised for further assessment and guidance on any concerning skin changes or discomfort.

TABLE 4.2.4: Inference example 2 of Diagnosis Chat Model.

Repeat Penalty: The resulting text's word repetition penalty is adjusted by this language model option. Raising the repeat penalty can have a particularly positive effect by decreasingthe frequency of repeated information.

Max Tokens: The maximum length of the model's response, expressed in terms of tokens, is determined by this parameter. Users can regulate how long the model responds by changing this variable, which will guarantee that the responses are succinct and appropriate for the situation.

Top P: This text generation parameter helps manage the output generation unpredictability of the model. Top P sampling strikes a compromise between diversity and relevance in the model's output by taking into account the smallest

group of words whose cumulative probability is greater than a given threshold (P). In healthcare settings whereby confusionmay result from repeated repetition.

Top K: Language models employ top-k sampling as a decoding technique to generate text. Itlimits the set of tokens available for the next prediction to the top k most likely tokens. We can modify the resulting text's balance between diversity (higher k) and reliability

(lower k) by varying the parameter k. It is employed to guarantee that the result has a balance betweenrelevancy and unpredictability, producing more believable and cogent writing.

INPUT	OUTPUT
What should the patient do next	The patient should seek a consultation with a dermatologist for further evaluation and guidance. A dermatologist will be able to confirm the diagnosis of Dermatofibroma and assess if any additional tests or treatments are necessary. If the lesion causes dis- comfort, pain, or cosmetic concerns, the dermatologist may discuss potential treatment options, such as surgical excision or laser therapy. Additionally, the dermatologist can provide adviceon skin care and preventive measures to minimize the risk of further skin issues. It's crucial for the patient to prioritize their skin health and follow the dermatologist's recommendations for proper management and monitoring of the condition.

TABLE 4.2.5: Inference example 3 of Diagnosis Chat Model

4.4 User Interface

A user submits an image of a skin condition to the system in a typical use scenario. This user could be a patient or a healthcare practitioner. After applying inference to the image, the Skin Disease Net classifies the skin illness and produces a preliminary result. The Visual GLM then receives this outcome and the image as input. The user is presented with a chat-like interface via Visual GLM, which creates a text-based interpretation of the classification and an explanation of the reasons. Users can submit queries or further data, which the machine will utilize to revise or elucidate its interpretation. Skin Disease Chat's user interface is intended for simplicity and usability. It has three features: a text input box where users may communicate with the system, an image upload option for the skin illness image, and a chat window displaying the diagnosis and interpretation. Because the dialogue with the

system is meant to be as interactive and natural as possible, users without any medical background can still use it. The Skin Disease Chat system's user interface, interactivity, and diagnostic examples are displayed.

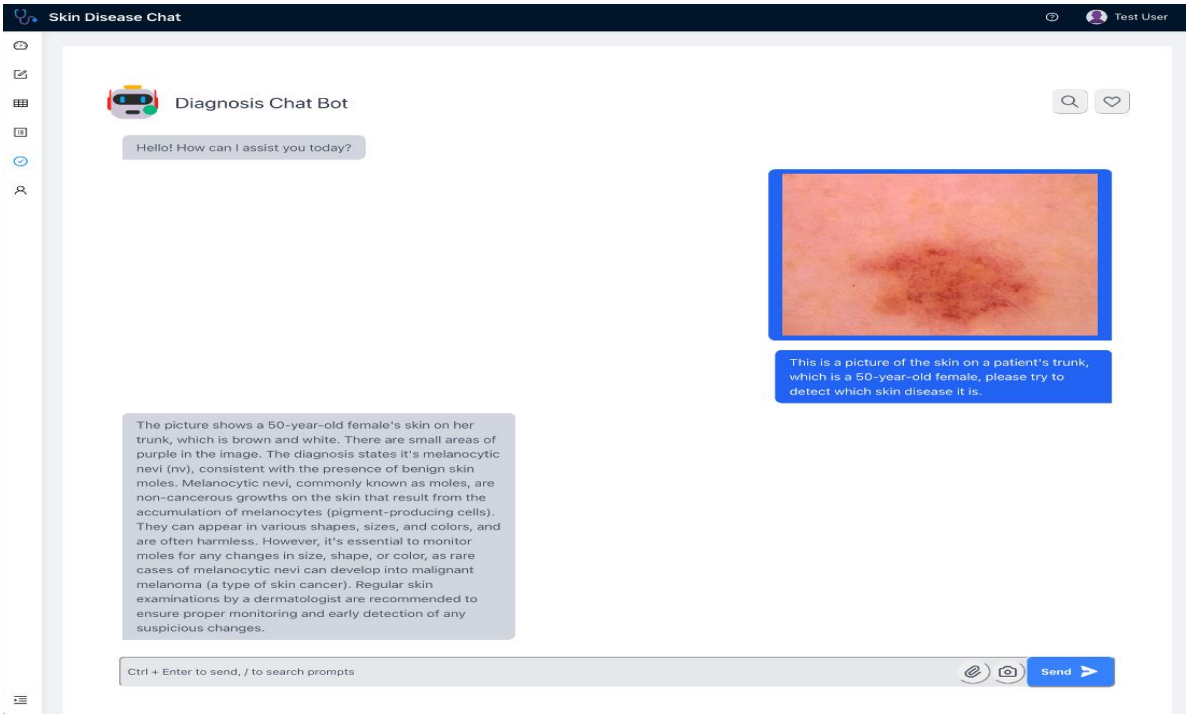

Fig 4.4.1 Demonstration of Skin Disease Chat diagnosing benign skin disease

We created our user interface as a web application. The system architecture is made to be highly scalable, dependable, and performant. It is made up of a number of interconnected parts, each intended to carry out a particular function inside the system. Figure 3.7 displays the system design.

Front-end: To construct our web interface, we use Tailwind , CSS [51], Ant Design UI Framework [50], and React.js [49] on the client-side. With the help of the well-known open-source JavaScript package React.js, we can construct a dynamic and responsive user interface

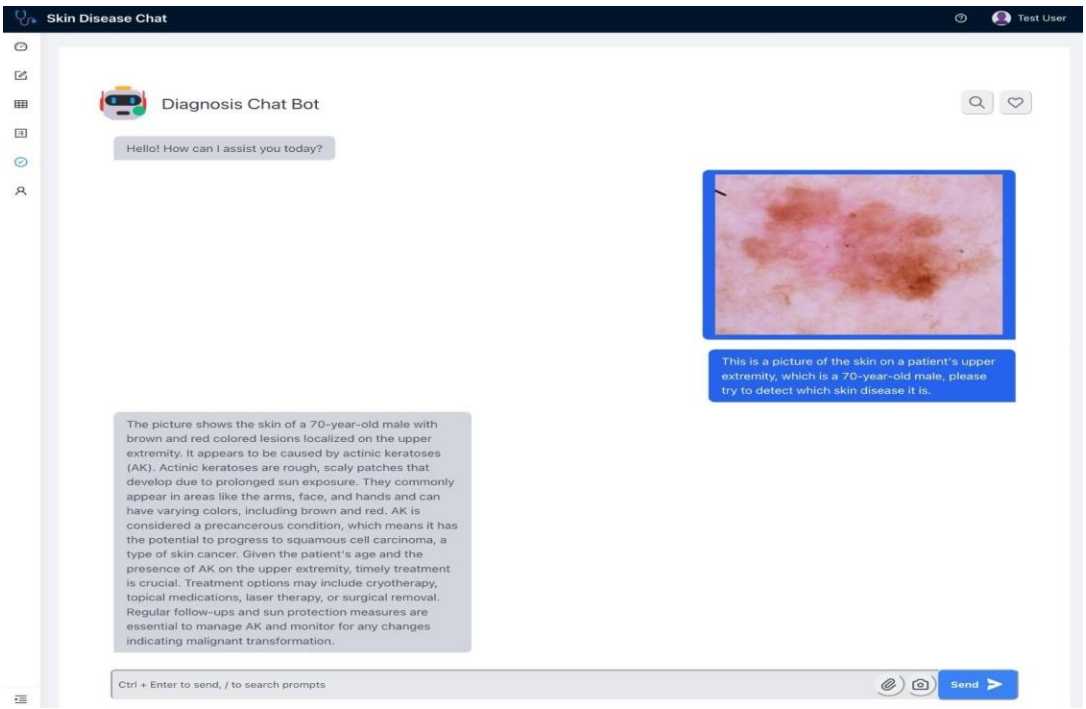

Fig 4.4.2: Demonstration of Skin Disease Chat diagnosing malignant skin diseases.

Back-end and Server: We use Fast API [52], a popular, high-performing, and modern web backend for Python, on the server side. Fast API is made to provide great performance and be user-friendly at the same time. It offers a simple method for configuring pathways and managing HTTP requests and replies. The robust web server software Nginx [53], which doubles as a reverse proxy, is used to further encapsulate our web server. By efficiently managing client connections, Nginx improves the efficiency of our web application and lets us handle several concurrent user requests without experiencing any performance reduction.

Data Storage: A PostgreSQL database [54], well-known for its feature set, dependability, and resilience, is used to manage and store all of the data in our system. We can manage a variety of data kinds and keep our data intact with PostgreSQL.

AI Model Inference: Our AI models, which are operating on a different GPU server, are communicated with by the web server during the inference process. Remote Procedure Call

(RPC) is the protocol used to conduct the communication. It allows computer programs to run a routine or procedure in a separate address space without requiring the programmer to explicitly code the details of the distant interaction. With this configuration, we can maintain the web server running while using the GPU server's processing capacity for demanding AI model inference work. sensitive to the needs of the user.

Model Checkpoints Storage: We use MinIO, an open-source object storage platform that is compatible with the Amazon S3 cloud storage service through APIs, to store the AI model checkpoints. This high-capacity storage option offers a practical means of in order to access and control model checkpoints. Additionally, as the system develops, its scalability enables us to manage bigger model sizes and rising storage requirements. To sum up, we have meticulously crafted our web application architecture to guarantee a superior user experience, all the while effectively handling the computing demands of our artificial intelligence models. Through the utilization of contemporary technology and adherence to industry best practices, we have created a scalable and highly performant system.

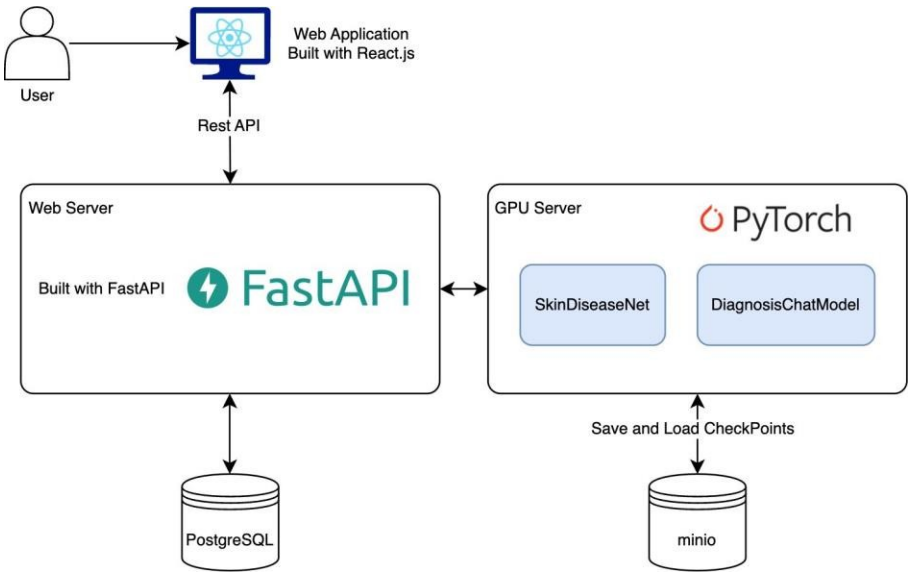

Fig 4.4.3: System design of the User Interface

We introduce our diagnostic system, Skin Disease Chat, which aims to investigate how to combine Large Language Models (LLMs) with image categorization to diagnose skin diseases. This system combines an LLM, Diagnosis Chat Model, with a powerful image classification model, Skin Disease Net, to potentially help with skin problem detection and understanding. Diagnosis Chat makes use of LLMs to simulate human interactions by providing explanations and responding to user queries, while Skin Disease Net uses deep learning techniques to assess visual data and diagnose possible skin illnesses. These two components work together to produce an interactive and educational platform for diagnosing skin diseases. But we also need to emphasize that the technology we are using have inherent limitations. LLMs Despite their sophistication, are not immune to the uncertainties associated with probabilistic

randomness, which could result in inaccurate results and a vulnerability to being misled by unclear or inaccurate data. We have tried to take advantage of the advantagesof both image classification and LLMs in light of these limitations. While Diagnosis Chat Model is tasked with producing understandable and helpful responses based onuser input and the output from Skin Disease Net, Skin Disease Net concentrates on the precise analysis of visual data.

CHAPTER 5
SYSTEM ARCHITECTURE

Artificial Intelligence-based diagnostics is an ever-evolving field. Rapid developments in AI technology and the creation of fresh datasets mean that diagnostic systems must always be flexible and adjust as necessary. Nevertheless, the difficulty is in allowing users—especially those without deep experience with AI models—to take advantage of these developments without sacrificing system efficacy or ease of use. In order to solve these issues, we present in this chapter the additions we made to our Skin Disease diagnostic system, the Model Management System. We specifically present enhanced model management features for the Diagnosis Chat and Skin Disease Net models. These features enable users to fine-

tune current models using data in defined formats, choose among pre-trained model checkpoints, and transition between models with ease. We've also improved system administration. features to offer a more flexible and user-friendly platform, such as user management. The management system offers a user-friendly method for operating the deep learning model and LLM.

5.1 Skin Disease Net Model Management:

Model Selection: Users may effectively manage and load the preferred model and historical checkpoints for Derm AI via an intuitive interface. This feature gives customers the freedom to select the most appropriate or desired model for their purposes by giving them direct access to a variety of model configurations and their corresponding checkpoints. For example, users can adjust to changing resource availability with the help of the model selection tool. Should GPU resources be scarce, consumers may use the ResNet-50 model because of its reduced processing requirements. However, users can opt to employ the ResNet-152 model in order to obtain better inference results when enough resources are available. As a result, this feature allows for a dynamic balance to be struck between diagnostic accuracy and computing efficiency, enhancing system performance under various circumstances.

Model Training: Users can choose a particular model and start training right away from the model training page, as shown in Fig. 4.3. This functionality allows the model to be updated quickly in response to sudden changes in the dataset. Without requiring complex actions, it offers customers a fully automated model training interface. Therefore, updating and refining the Derm AI model is simple even for users who lack substantial experience with AI model training. enhancing the usability and flexibility of the program to accommodate changing datasets. The results page (Figure 4.4) offers a clear view of the model training progress and is

developed with the user experience in mind. Through the online interface, users can follow the training process' progress in real-time once it has started. To provide you with the most recent information, the page is constantly updated. The page shows the accuracy and training loss as of right now. Additionally, the website offers details on the "Current Checkpoint Hash," which gives users a unique identification for the model's current state.

5.2 Diagnosis chat Model:

Users can adjust the LLM's temperature and Top-P, among other factors. Users can adjust the model's performance to suit their needs and tastes by adjusting these parameters, which regulate the output's diversity and unpredictability during inference. The model's output's randomness is determined by the temperature parameter; a higher temperature produces more diverse responses, while a lower temperature produces more deterministic replies. However, the Top-P parameter, sometimes referred to as the nucleus.

New dialogue datasets can be uploaded by users to quickly fine-tune the model. In order to tailor the previously trained model to the particular task at hand, fine-tuning is an essential process that usually calls for a high level of technical skill. Nevertheless, our approach streamlines this procedure so that users with limited experience with AI model training can easily improve the chat model. A real-time user interface called the fine-tuning results page was created to give users information about the fine-tuning procedure.

Users can access the page to track the operation's progress and status once the fine-tuning procedure has started. The "Running Log" is located in the middle of the page. The fine-tuning process's current state is shown in detail in the running log. It

records and captures any noteworthy occurrences, modifications, or mistakes, and it serves as a monitor for troubleshooting any possible problems that can come up during fine-tuning.

Through these features, the Diagnosis Chat Model Management not only allows users to control the performance of the LLM but also enables them to adapt the model to the evolving datasets and tasks efficiently. This user-friendly design enhances the overall user experience and effectiveness of our diagnostic system.

CHAPTER 6
CONCLUSION

The interactive chat-based system for diagnosing skin diseases, called Derm AI,

was showcased in this project. By engaging consumers in discussion and providing an explanation of the diagnosis, Derm AI adds a new level of complexity to automated skin condition diagnosis by using advancements in big language models and image classification models. Although multimodal big language models have obvious drawbacks, such as worse perceptual and cognitive capacities, they significantly increase the system's adaptability and user friendliness. We augment the large language model (LLM) with a conventional CNN-based image classification technique to improve automatic diagnosis accuracy, attaining a 93% validation accuracy. We intend to make the system available on GitHub as an open-source project.

6.1 Future Work:

As we look towards the future, several potential improvements and new directions present themselves. For instance, the accuracy and reliability of AI model outputs remain challenging due to their probabilistic nature. Hence, investigating techniques to mitigate the risk of models being misled and to bolster system robustness against adversarial attacks is of importance. Plans are also in place to integrate more comprehensive databases and expand the system’s diagnostic range. Specifically, potential areas of application include the management of chronic diseases such as diabetes and cardiovascular diseases, where accurate and timely diagnosis is critical for effective treatment and disease control. The methodology of augmenting traditional AI models with LLMs to enhance accuracy shows substantial promise, extending beyond skin disease diagnosis. It can be adapted to a broader range of medical domains, providing crucial support in automatic disease detection and understanding.

Deep learning technology has shown significant promise and success in various medical fields, including: X-ray Analysis, Tumor Detection, Retinal Disease Diagnosis, Brain Imaging Analysis, Cardiac Image Analysis and Cancer Treatment Planning. These applications also can be integrated with large language models to improve the result analysis. Lastly, the LLM in our system relies heavily on the quality of prompts used during both fine-tuning and inference stages. Therefore, our future endeavors with LLM will prioritize enhancing prompt quality and exploring innovative techniques to boost the system's overall performance and reliability.

REFERENCES

[1] Skin diseases disable, stigmatize and cause suffering and mental health conditions, 2023. URL https://www.who.int/news/item/31-03-2023-who- first-global-meeting-onskin-ntds-calls-for-greater-efforts-to- address-their-burden.

[2] Gao Huang, Zhuang Liu, Laurens van der Maaten, and Kilian Q. Weinberger. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017. 1, 11, 12, 18

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 770–778, 2016. doi: 10.1109/CVPR. 2016.90. 1, 11, 12, 18

[4] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for largescale image recognition, 2015. 1, 11, 18

[5] Mehwish Dildar, Shumaila Akram, Muhammad Irfan, Hikmat Ullah Khan, Muhammad Ramzan, Abdur Rehman Mahmood, Soliman Ayed Alsaiari, Ab- dul Hakeem M Saeed, Mohammed Olaythah Alraddadi, and Mater Hussen Mahnashi. Skin cancer detection: A review using deep learning techniques. International Journal of Environmental Research and Public Health, 18(10), 2021. ISSN 1660-4601. doi: 10.3390/ijerph18105479. URL https://www.mdpi.com/1660-4601/18/10/5479. 1

[6] Fleur W. Kong, Caitlin Horsham, Alexander Ngoo, H. Peter Soyer, and Monika Janda. Review of smartphone mobile applications for skin can- cer detection: what are the changes in availability, functionality, and costs to users over time? International Journal of Dermatology, 60(3):289–308, 2021. doi: https://doi.org/10.1111/ijd.15132. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ijd.15132. 1

[7] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Ka- plan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCan- dlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners,
2020. 2

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina N. Toutanova. Bert: Pretraining of deep bidirectional transformers for language understanding, 2018. URL https://arxiv.org/abs/1810.04805. 2, 16

[9] Chatgpt, 2023. URL https://openai.com/blog/chatgpt. 2, 13

[10] Teresa Mendon¸ca, Pedro M. Ferreira, Jorge S. Marques, Andr´e R. S. Marcal, and Jorge Rozeira. Ph2 - a dermoscopic image database for research and benchmarking. In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 5437–5440, 2013. doi: 10.1109/EMBC.2013.6610779. 5

[11] Philipp Tschandl. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions, 2018. URL https://doi.org/10.7910/DVN/DBW86T. 6

[12] Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023. 13

[13] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timoth´ee Lacroix, Baptiste Rozi`ere, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guil- laume Lample. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971, 2023. 13, 14

[14] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. Stanford al- paca: An instruction following llama model. [https://github.com/tatsu- lab/stanford alpaca