iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83869

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

Asclepius AI: A Thorough Multi-modal Machine Learning Framework for Disease Prediction from Symptoms and Images

Prafulla B.Bhand, Aniruddha N.Rane, Ayush K.Khairnar, Soham S.Phatak, Punam A.Shinde, Dr. Emmanuel M
Information Technology Pune Institute of Computer Technology Pune, India

Abstract: ***The increasing digitization of healthcare hasled to an urgent need for scalable, accurate, and reliable com-putational tools that can meaningfully interpret heterogeneous patientdata.Inclinicalinformatics,traditionaldiagnosticmodels are usually built around a single modality of data, such as laboratory blood panels, isolated dermatological photographs, or individualpatient-reportedsymptoms.Thisinherentlyconstrains their clinical utility and generalizability to real-world situations. These unimodal systems are ill-suited to capture the complex cross-modal physiological realities underlying systemic diseases. To overcome these limitations, this paper introduces Asclepius AI: a comprehensive, multi-modal digital health platform that connects patient-reported symptom data with high-accuracy Artificial Intelligence (AI) based diagnostic reasoning. To tackle the intrinsic complexity and high dimensionality of the medical inputs,theproposedframeworkemploysathree-layerdiagnostic architecture. The first tier, designated the Basic Engine, applies an ensemble Random Forest classifier for rapid, interpretable screening of straight forward symptom presentations. Thesecond tier,theAdvancedEngine,leveragesacarefullyoptimizedArtificialNeuralNetwork(ANN)capableofhigh-dimensionalsystemic disease prediction spanning 41 distinct disease categories. The thirdtier,theSkinEngine,employsafine-tunedEfficientNet-V2-S convolutional model for the classification of dermatological images. The validation accuracy of the Skin Engine reaches 96.25% on a heterogeneous dataset of 22,057 augmented dermatological images and wide symptom vectors, achieving state-of-the-art results. In addition to its predictive engines, the platform is deployed as an integrated full-stack ecosystem built on Next.js and FastAPI, with a Large Language Model (LLM)-powered conversational agent for interactive health guidance and secure Electronic Health Record (EHR) management supported by PostgreSQL and Prisma ORM. Together, these results show that thesynergyofmulti-modaldeeplearningandpersistentconversationalcontextcandramaticallyshortendiagnosticworkflowsand develop a transparent and clinically actionable decision support system.***

Index Terms: ***Multi-Modal AI, Disease Prediction, Efficient-Net, Artificial Neural Networks, Electronic Health Records, Explainable AI, Clinical Decision Support Systems, Full-Stack Architecture.***

I. INTRODUCTION

The application of Artificial Intelligence (AI) and Machine Learning (ML) in clinical medicine is one of the most impor-tant technological advances of this century. Across the world, medicineisincreasinglyturningtodata-drivenapproachesthat promise to improve efficiency and to improve the accuracy of identifying diseases at their earliest and most treatable stages [1], [4]. Theoretical advances in algorithmic design have been rapid and far-reaching, with evidence that well-trained ML systems can match or even exceed human experts in some narrow diagnostic tasks [1].

However, a real and practical gap persists between labora-toryperformanceandclinicalutility.Thereliabilityandsafety of any ML model are determined by the type, quality and contextual relevance of the data used to train and test the model [1]. Structural inefficiencies which plague healthcare systems in many parts of the world are well documented. Extendeddiagnosistimes,potentialforphysicianbias,absenceof integratedrecordmanagement,andlimitedaccesstospecialist care, particularly for rural or underserved groups, contributeto poorer patient outcomes [1].

In many instances, people may wait weeks for a derma-tology or internal medicine consultation, allowing an acute condition to become chronic or life-threatening. Exacerbating this issue is the current state of healthcare institutions, which produce a plethora of heterogeneous data such as structured clinical measurements, laboratory data, unstructured patient narratives, and high-resolution diagnostic images. AI systems designed for a single data modality can perform impressively in controlled experimental settings, but often fail to deliver real-world performance.

They tend to overlook the subtle associations between systemic signs like prolonged fever or joint inflammation and localphysicalsignslikeskinlesionsoroculardiscoloration.

There is a growing body of empirical evidence supporting the view that multimodal AI architectures that integrate and learn acrossdiverseinputtypessimultaneouslydelivermeaningfully stronger predictive performance across a wide spectrum of diagnostic and prognostic applications [2]. Multimodal tech-nologies that exploit complementary data sources to enable more comprehensive and accurate medical judgments are transforming clinical practice [2].

However, even systems that manage to integrate some modalities, such as imaging together with biomarker data, typically fail to account for the wider cross-modal relation-ships that experienced clinicians intuitively synthesize when evaluating patients [2], [5]. Another barrier to clinical uptake often cited is the opacity of complex deep learning networks. Clinicians want more than a probability score. They want a coherent rationale, a transparent account of how the model reachedaparticularconclusionandhowthatconclusionrelates to the patient's clinical presentation.

Models that operate as 'black boxes' pose a major friction point in clinical trust and serious ethical and medico-legal concerns[6].AsclepiusAIisofferedasadirectresponse to these interrelated challenges. The platform uses a hybrid machine learning architecture to predict systemic diseases fromsymptomdescriptionsenteredbypatients,andtoconcur-rently classify dermatological conditions from clinical images uploaded.

Additionally, the system utilises the nascent reasoning capabilities of modern multimodal large language models, which have recently demonstrated significant promise in au-tonomously directing intricate end-to-end diagnostic work-flows [3]. Asclepius AI is not meant to act as an independent diagnostic oracle, but as a first-response screening tool that coordinates a systematic clinical workflow and provides its outputs in the context of a larger patient care pathway [3].

The platform integrates its prediction engines into a secure and responsive web environment that allows patients to main-tain their longitudinal medical records, request AI-assisted health advice in natural language, and securely transfer di-agnostic summaries to qualified physicians, thereby creating a coherent and closed-loop clinical network.

II. REVIEW OF LITERATURE

The intersection of deep learning, computer vision and clinical informatics has grown rapidly over the last decade, resulting in a rich literature on which the current framework builds. We describe a multi-modal methodology that expands on seminal contributions and recent advances in three related themes: (i) the context of data modality, (ii) architectural advances in medical image analysis and (iii) the need for both multimodal integration and model interpretability.

A. Data Modalities and Contexts

ThepredictivepowerofanyMLmodelusedinmedicine is highly dependent on the selection of the data modality, the modelling approach taken, and the particular clinical context beingaddressed[1].Inasystematicevaluationofavariety of machine learning approaches on four different medical datasets, Mohamed et al. demonstrated that the predictive accuracy varied hugely, depending on contextual factors.

The strongest symptom-input models were those combined with other streams of physiological data. When only separate laboratory results were used in cases of clinical uncertainty, classification results were inconsistent [1]. Their work high-lights the critical role of meticulous dataset selection and context-sensitive feature engineering [1]. Similar results were reported by Mohan et al., who proposed a symptom-based diseasepredictionframeworkusingensemble(RandomForest with Gradient Boosting) and hybrid classifiers (Naïve Bayes with AdaBoost) [4].

Theirhybridmodelhad85%accuracy,whiletheensem-ble configuration received 92%. This demonstrates the value of combining several weak learners together for structured, categorical symptom inputs [4]. Bhanuteja et al. also showthat while Random Forest classifiers can predict multi-disease symptoms with accuracies up to 98.3%, this level of per-formance is maintained only through heavy preprocessing, including aggressive feature reduction that may leave only 95 out of 132 available symptoms to avoid overfitting [14].

This bottleneck is addressed by the Advanced Engine in Asclepius AI, which uses a carefully regularised ANN withthepotentialtoutilisethefull132-dimensionalsymptomspace without extensive manual feature elimination. This prevents the predictive signal from rare or infrequent co-occurring symptomsthatwouldotherwisebelost.Anotherchallenge in real clinical settings is data incompleteness. Austin et al. highlight the need for principled imputation strategies when missing data arises in clinical research [15].

Our system copes with this problem through solid vectori-sation procedures and a tiered prediction strategy that can provide partial outputs when the symptom lists are sparse, by sending short symptom arrays to the Random Forest engine, andleavingtheANNfordenser,morecompletepresentations.

B. Advances in Medical Imaging Architecture

In the field of image based diagnosis, traditional Convolu-tional Neural Networks (CNNs) have been the foundation for today's automated image analysis, but they usually involve a largeamountofcomputation.AseminalsurveybyLitjens et al. extensively reported how deep learning, and CNNs in particular, revolutionised the interpretation of medical images by replacing hand-crafted feature engineering with networks that learn discriminative filters directly from raw pixel data [13].

Based on this work, Alruwaili and Mohamed proposed a deeplearningarchitecturebasedonthefusionofEfficientNet-B0, EfficientNet-B2 and ResNet50 to improve the training stability and the accuracy of multi-class skin disease classifi-cation[5].Theirexperimentsshowedthattheadditionofdeep residualconnectionscaneffectivelysolvethevanishinggradi-ent problem, enabling robust propagation of features through many layers of the network [5].

EfficientNet-family archi-tecturesareparticularlywell-suitedforresource-constrained deployment scenarios that require a good performance-to-efficiency trade-off, due to their principled compound scaling strategy that uniformly scales network depth, width, and resolution [5].

InAsclepiusAI'sSkinEngine,thisdesignphilosophyisim-plementedusingtheEfficientNet-V2-Svariant.Incomparison with its earlier series, the V2 series adopts Fused-MBConv blocks in the initial part of the network, where standard convolutionisusedinsteadofdepthwiseseparableconvolution to improve hardware utilisation and accelerate training. The model was extensively augmented during training to achieve robust performance on clinical photographs taken in the real-world setting, incorporating variable lighting and consumer-grade smartphone cameras [2], [5].

C. Multimodal Fusion and Explainability

The finding that models operating across multiple comple-mentary modalities outperform unimodal counterparts in both diagnostic accuracy and prognostic power is consistently sup-portedbyempiricalevidence[2].Zhangetal.[2]observedthat the combination of heterogeneous data types, such as time-series physiological signals, clinical imaging, and free-text medical narratives, requires architectures with strong cross-domain generalisability and built-in interpretability mecha-nisms.

Taking this further, Meenalochini proposed an Explainable AI (XAI) framework based on combining medical imaging andclinicaltextwithVisionTransformersandLargeLanguage Models.Thisledtoimproveddiagnosticaccuracyandreduced opacity that can alienate clinical users [6]. Attention-based visualisation and natural language explanation help clinicians understand model reasoning and build the trust needed for practical adoption. Both Stiglic et al. and Alkhanbouli et al. contend that accountability and interpretability need to be considered if clinical decision support is to emerge from raw algorithmic outputs as a truly trustworthy tool.

Guleria et al. validated a XAI-augmented ensemble classi-fierforcardiovasculardiseaseprediction,findingthatsurfacing feature importance values for variables such as age and sex concretelyimprovesclinicianconfidenceinmodeloutputs[7]. Moreover,Meinckeetal. demonstratedthatmultimodalLLMs areabletoindependentlymanagecomplexclinicalworkflows, including ophisticated prioritisationbehaviourssuchasfront-loading the most diagnostically informative tests early in the patient encounter [3].

The integration of the Groq-powered conversational agent within Asclepius AI as an intelligent guide, contextualisingthe diagnostic outputs and supporting the follow-up clinical reasoning [3] was the result of these findings.

III. SYSTEM ANALYSIS AND SYSTEM DESIGN

The Asclepius AI platform is conceived as a multi-tiered service-oriented ecosystem, where deep learning diagnostic engines, a persistent conversational interface, and a clinical record management subsystem act in close coordination but are functionally decoupled.

A. System's Topology

Overall the architecture is divided into a client-facing pre-sentation layer and a machine learning processing backend. Thesetwodomainstalkthroughawell-definedAPIboundary, and are organised into four core functional blocks:

1) *User Interaction Layer (Frontend):* Built using Next.js 14andstyledwithTailwindCSS,thislayergathersmulti-modal patient input. It handles application state well and provides a responsive, clinical application feel in mobile anddesktopsettings.

Thislayeristaskedwithdoing the lightweight client-side preprocessing of images and structuring of symptom arrays for transmission to the backend.

2) Diagnostic Routing Module (API Gateway): The cen-tral orchestrator of diagnostic requests is the FastAPI service written in Python. The gateway takes the json payload, identifies the input type and directs the request totheinferenceengineaccordingly.Theinferenceengine can be a Random Forest (Basic tier), an ANN (Advanced tier) or EfficientNet (Skin tier). Asynchronous Uvicorn workers avoid any thread blocking during computation-ally intensive tensor operations.
3) Conversational Layer (Ask AI): This layer employsthe Groq LPU (Language Processing Unit) inference architecture to interface with open source large language models (LLaMA-3, in particular) to provide the userwithnon-diagnostichealthadvice,simpleexplanationsof medical terms and contextual analysis of their diagnostic reports.
4) Persistence and Storage Layer: The main relational store is a managed PostgreSQL instance provisioned via Supabase, which includes encrypted diagnostic records, orderedconversationhistories,andphysicianprofiledata. Using Prisma ORM to build type-safe queries reduces exposure to SQL injection vulnerabilities substantially. Cloudinary offers secure, scalable hosting for clinical image assets.

B. Logical Design & Workflow

Thesystemhasadualplatforminteractionlogic,serving as a patient-facing diagnostic screening tool and as a secure clinical records environment for registered physicians. Clear state transitions in the diagnostic workflow maintain data integrityastheuser progressesfrominitialsymptomscreening to specialist referral recommendations. For each diagnostic sessionsecure OAuthauthenticationisstarted.Theuserenters the diagnostic loop after login.

For image-based submissions, we perform a quality check ontheclientsidetomakesuretheuploadedphotographmeets minimum clarity standards prior to sending the image buffertothebackend.Thebackendperformsinferenceandreturns a ranked list of predictions with confidence scores. The result is immediately persisted to the PostgreSQL database such that thepatient'slongitudinalmedicalhistoryremainsuptodate.

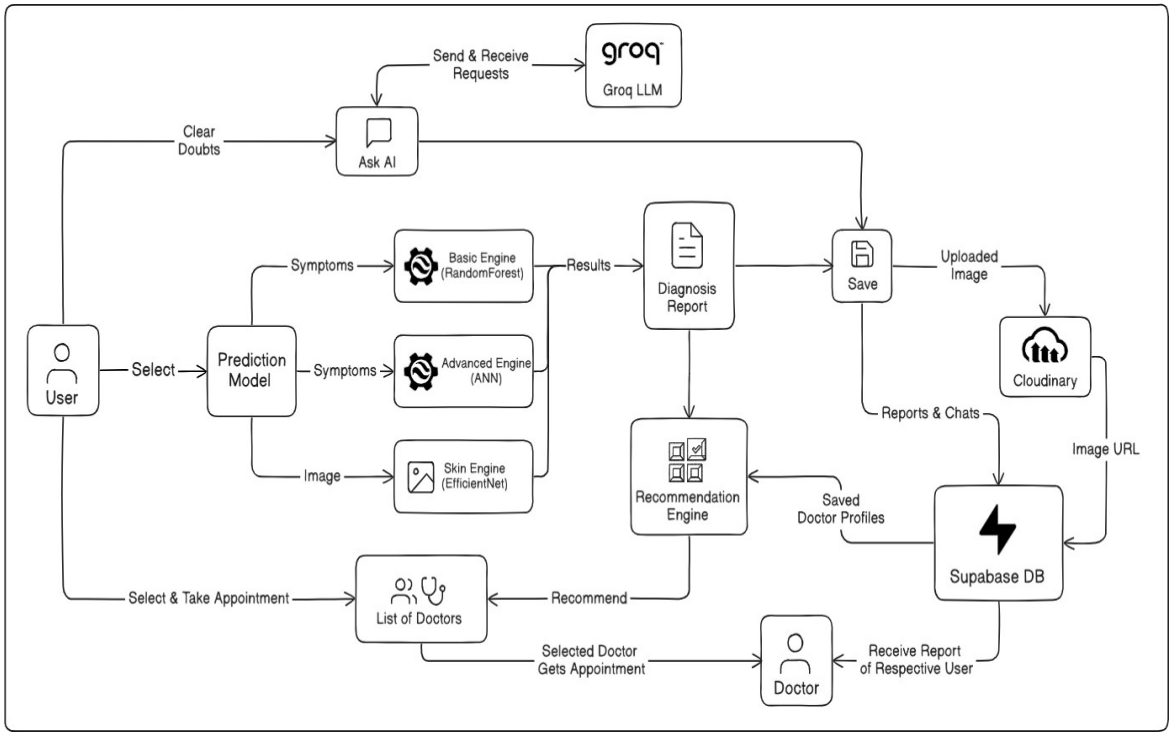


Figure1:AsclepiusAISystemArchitecture.Thismulti-tierservice-orientedecosystemintegratesspecializeddiagnosticengines (Basic, Advanced, Skin) with a persistent conversational Groq LLM layer and secure Cloudinary/PostgreSQL storage. The decoupled nature ensures high-dimensional feature extraction and image processing occur asynchronously via the FastAPIgateway.

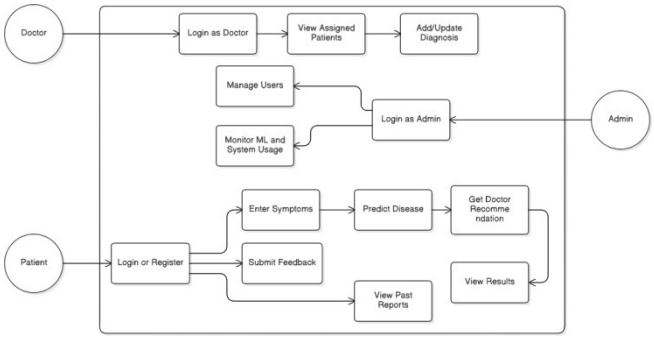


Figure 2: Comprehensive Use Case Diagram illustrating the distinctinteractionpathwaysforPatients,Doctors,andSystem Administrators.

IV. DATA ENGINEERING AND PREPROCESSING

Sound data engineering is the backbone of any well per-forming machine learning system. We built individual, but complementary, preprocessing pipelines for the structured symptom data, and the unstructured dermatological image data, respectively, according to the needs of each downstream model.

A. Symptom Data Vectorization

Thetrainingdatasetforthesystemicdiseaseisacategorical description of symptoms. In order to convert these into a form appropriate for neural computation, One-Hot Encoding was applied over a master vocabulary of 132 clinically verified, uniquesymptoms.Theprocessgeneratesa132-dimensional binaryvectorforeachpatientinteraction. Anentryof1meansthe symptom is present, andanentryof0meansthesymptomisabsent.Thisrepresentationwaschosenintentionally to prevent the model from learning an unintended ordinal relationship between symptom categories.

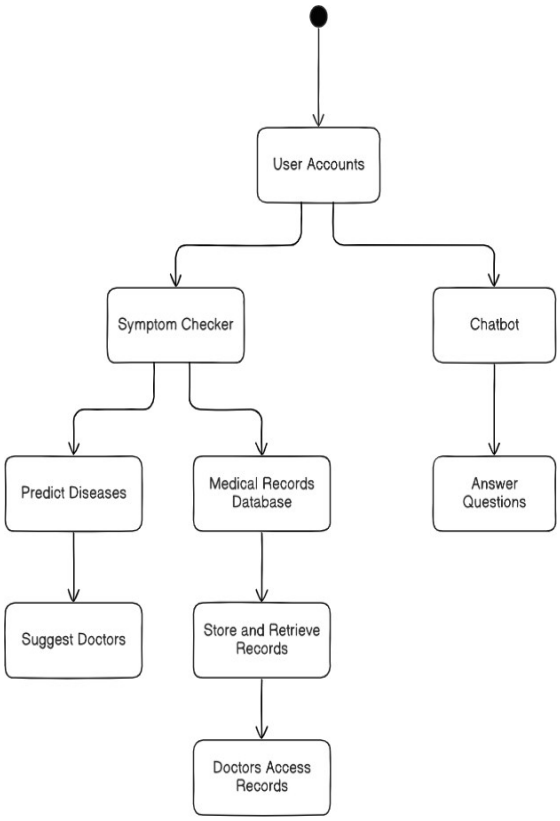


Figure 3: System State Diagram detailing the sequential life-cycle of a diagnostic session, from initialization to result persistence and specialist routing.

B. Processing of Dermatological Images

Clinical photos for Skin Engine were preprocessed via a multi-step preprocessing pipeline following the input specs of the EfficientNet-V2-S architecture:

- Dimensionality standardisation: All incoming images were uniformly resized and center-cropped to produce a 224×224×3 tensor, regardless of their original aspect ratio or resolution.
- Pixel normalisation: Raw pixel intensities were first scaled to an interval of continuous values from the orig-inal integer range. Next, we performed Z-score normali-sation using the per-channel mean values and standard deviationsobtained fromtheImage Netdataset.This step stabilises the magnitude of gradient, and speeds up convergence.
- Data Augmentation with RandAugment: Dermato-logical datasets are often characterised by great class imbalance.Controlledsyntheticvariancewasusedin the training pipeline to avoid memorisation of class-specificphotographicartefacts,usinghorizontalandverti-cal flipping, random rotations, and moderate contrast and brightnessperturbations.Thesetransformationsallowthe modeltolearnpathologicalfeaturesthatarespatiallyand texturally meaningful, rather than accidentally imaging features.

V. METHODOLOGY FOR MACHINE LEARNING

Asclepius AI's predictive intelligence is distributed across three distinct but complementary machine learning models, each selected and fine-tuned to best process a given combina-tion of data type, input dimensionality and inference speed.

A. Basic Engine: Ensemble Majority Vote

The Basic Engine is initiated when a patient reports a short list of symptoms with less than five entries. This tier is based on the Random Forest classifier. The random forest classifier is an ensemble learning method that trains a large collectionof independent decision trees and combines their outputs at prediction time.

The algorithm uses feature bagging to decorrelate the individualtrees.Thefinalpredicteddiagnosisforaninput symptomvectorxisthemajorityvoteover$N$=100trees:

$$\hat{y}=\text{mode}\{T_1(x),T_2(x),\ldots,T_N(x)\} \quad (1)$$

In each tree, the Gini impurity criterion is used for splitting nodes:

$$G=1-\sum_{i=1}^{C}(p_i)^2 \quad (2)$$

The ensemble averages predictions over many different trees, which effectively dampens the high variance that makes indi-vidual decision trees susceptible to overfitting.

B. Advancedengine–Optimisedneuraltopology

For dense and symptom-rich inputs the Advanced Engine takes control. This tier is powered by a carefully regularised ANNthatisintendedtomaptheentire132-dimensionalbinary symptom vector to one of 41 systemic disease classes.

The network has an input layer and a few fully connected hidden layers with Rectified Linear Unit (ReLU) activations. We use two complementary mechanisms to address overfit-ting: L2 weight regularisation, which penalises large weight magnitudes, and Dropout layers with a dropout rate of 30%. Dropout randomly disables a proportion of neurones on each forward pass in training, forcing the remaining units to learn more robust, distributed feature representations instead of co-adaptingtoparticularco-occurrencesofsymptoms.Theoutput layer applies the Softmax function to return a normalised probability distribution over all 41 disease classes:

$$\sigma(\mathbf{z})_i=\frac{e^{z_i}}{\sum_{j=1}^{K}e^{z_j}} \quad (3)$$

Trainingminimisesthecategoricalcross-entropylossfunction:

$$L=-\sum_{i=1}^{K}y_i\log(\hat{y}_i) \quad (4)$$

C. SkinEngine:EfficientNet-V2-S&AdamW

TheSkinEnginecarriesoutthemostcomputationallyinten-sivepredictiontask:classificationofdermatologicalconditions fromclinicalphotographs.WeoptedforEfficientNet-V2-S instead of heavier alternatives like ResNet152 due to its better parameter efficiency without losing accuracy. The V2 architecturediffersfromearlierEfficientNetmodelsinthat it uses Fused-MBConv layers in the early stages, where the depthwise and pointwise operations are fused into a single normal convolution, instead of depthwise separable convolu-tions.Thisresultsinmuchfastertrainingthroughputandbetter hardware utilisation.

Thetrainingwasconductedintwosequentialtransferlearn-ing phases:

Phase 1 — Warmup: The pre-trained convolutional base, initialised from ImageNet weights, was completely frozen. In this phase, only the newly appended classification head was trained,whichallowedtherandomlyinitialisedweightsofthis head to fit into the dermatological feature distribution before the base network was modified.

Phase 2 — Fine-Tuning: The top convolutional blocks were unfrozen and the whole network was trained end-to-end using the AdamWoptimiser. AdamW is an improvement to Adam that decouples the weight decay term from the adaptive gra-dient update to obtain better-regularized solutions and better generalisation:

$$w_{t+1}=w_t-\eta(\nabla f_t(w_t)+\lambda w_t) \quad (5)$$

D. The Multi-Modal Routing Algorithm

Algorithm1AsclepiusAIMulti-ModalRoutingLogic

Require:Patient Input I, Modality $M \in \{Symptom, Image\}$
Ensure:Diagnosis$D$,ConfidenceScore$C$
if$M==Image$then
$Img_{std}\leftarrow$Resize(I,224,224)
$Img_{norm}\leftarrow$ Normalize(Img_{std}, ImageNet)
$D, C\leftarrow$ EfficientNet-V2-S.predict(Img_{norm})
else
$V\leftarrow$OneHotEncode(I,MasterSymptomList)
iflen(I)<5then
$D,C\leftarrow$RandomForest.predict(V)
else
$D,C\leftarrow$TunedANN.predict(V)
endif
endif
returnD,C

VI. IMPLEMENTATION AND TEST CONFIGURATION

Translating the proposed models to a production-ready de-ploymentrequiredcarefulconsiderationofhardwareselection, softwareconfigurationandinferenceoptimisationtechniques.

A. Requirementsonsoftwareandhardware

All deep learning models were trained on Google ColabPro instances with 16 GB VRAM and 32 GB system RAM with an NVIDIA Tesla T4 GPU. The ML software stack included Python 3.10.12, TensorFlow 2.15.0 for the systemic disease engines and PyTorch 2.1.0 for the visual classification architecture.

B. ModelserialisationandONNXexport

One practical challenge of deploying deep learning models onthewebisthememoryoverheadofthefull-precisionmodel weights, which can limit the server concurrency and increase the latency.

To fix this, the PyTorch weights of the Skin EnginewerequantisedtoFP16(16-bitfloatingpoint)precision and NetworkExchange)format.ThisallowstheFastAPIbackend to perform inference on images using highly optimised CPU-based onnxruntimeand removes the need of having a dedicatedGPUontheproductionserver.Thesystemcanserve multiple concurrent requests efficiently.

C. Datalayer and persistence

Therelationaldatabaseschemadescribesacleanandex-tensibledatamodel.TheprimaryUserobjecthasone-to- manyrelationships to both Predictionrecords and Chat sessions.EachChatMessagerowhasanorderinteger fieldthatmaintainsthechronologicalorderofeachpatient–AI interaction to ensure the integrity of the conversation. When theuserinvokesthe“ResumeChat”feature,theLLMreceives thefullreconstructedconversationhistoryintheoriginalorder,

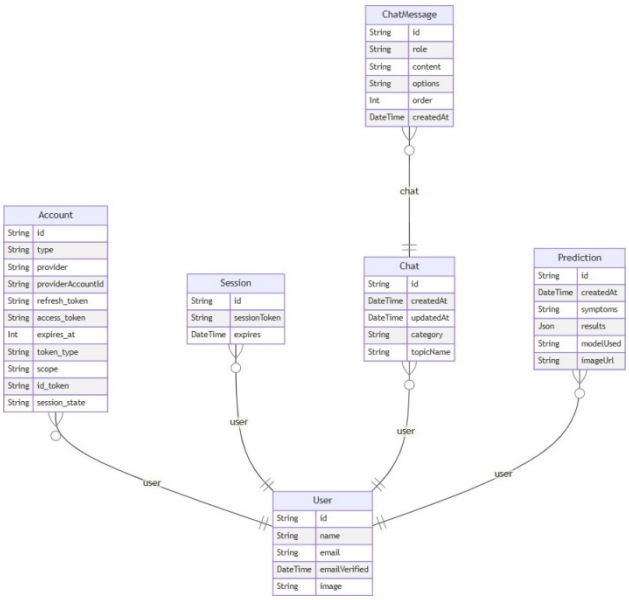


Figure 4: Entity Relationship Diagram (ERD). Detailed schema generated via Prisma illustrating the relational map-pingbetweenUsers,sequentialChatMessages,diagnosticPre-dictions, and secureSessions withinthe PostgreSQL database.
generatingtheexperienceofcontinuous,context-awarepatient engagement instead of disjointed, stateless exchanges.

VII. DISCUSSION AND RESULTS

The results of the experiments testify to the effectiveness of the multi-level diagnostic strategy adopted in Asclepius AI.

A. Metrics for Evaluation

We evaluated the performance of each diagnostic engine using a standard set of classification metrics calculated from the entries of the confusion matrix, namely True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives(FN).ThestandardmetricsareAccuracy,Precision, Recall (Sensitivity) and F1-Score:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (7)$$

$$\text{Recall(Sensitivity)} = \frac{TP}{TP+FN} \quad (8)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}} \quad (9)$$

B. Layered predictive accuracy

TheBasicEngine(RandomForest)achievedanaccuracyof 95.12% on the structured symptom data, confirming its value as a quick and reliable screener for patients with a limited set of symptoms.

The Advanced Engine (ANN) was trained on a corpus of 246,945rowsandtrainedtomapsparsesymptomvectorsonto 41diseaseclasses.ThevalidationaccuracyfortheANNmodel was 86.06%. Batch Normalisation layers were critical in the hidden stack, keeping internal activation distributions stable duringtrainingandallowingthenetworktoreliablydistinguish diseases with largely overlapping symptom profiles, such as severe viral influenza and early-stage pneumonia, based on slight differences in the combination and frequency of symp-toms.

```
Fine-Tuning - Epoch 17/20
---------------
Train Loss: 1.1139 Acc: 0.8910 | LR: 3.1e-06
Val   Loss: 0.9499 Acc: 0.9600

Fine-Tuning - Epoch 18/20
---------------
Train Loss: 1.1103 Acc: 0.8923 | LR: 5.7e-06
Val   Loss: 0.9537 Acc: 0.9585

Fine-Tuning - Epoch 19/20
---------------
Train Loss: 1.1102 Acc: 0.8930 | LR: 9.1e-06
Val   Loss: 0.9441 Acc: 0.9599

Fine-Tuning - Epoch 20/20
---------------
Train Loss: 1.1005 Acc: 0.8962 | LR: 1.3e-05
Val   Loss: 0.9421 Acc: 0.9626

Fine-Tuning complete in 80m 32s
Best Val Acc: 0.962594
```

Figure5:SkinEnginePerformanceMetrics.Confusionmatrix andclassificationreportdemonstratingafinalvalidationaccu-racy of 96.25% across 25 skin disease categories. The strong diagonal axis proves the model's high discriminative power.

The most interesting result was The Skin Engine. The EfficientNet-V2-S model was fine-tuned for 20 epochs and obtained a multi-class validation accuracy of 96.25% on 25 dermatologicalconditions.Themodelachievedgoodper-class discrimination with F1-scores higher than 0.95 for high-risk classes like Melanoma and Basal Cell Carcinoma. False negatives in these classes are of particular clinical importance because of the dire consequences of delayed detection in oncological contexts.

Table I: Comparative Performance Summary of Diagnostic Engines

ModelTier	CoreAlgorithm	Epochs	Accuracy
Basic	RandomForest	N/A	95.12%
Advanced	TunedANN	42	86.06%
SkinEngine	EfficientNet-V2-S	20	96.25%

C. Comparison of Optimisation Strategies

An important factor for the strong convergence results acrossallthreetierswasthecarefuluseofdifferentiated learning rate schedules, appropriate for the loss landscape of eachmodel.TheANNwastrainedwithanadaptivestep-decay policy, where the learning rate was halved if the validationloss did not improve over three consecutive epochs. This conservative strategy kept the weight updates stable in the presenceofahighdimensional,relativelysmoothlosssurface. For the Skin Engine, we used a Cosine Annealing schedule however.

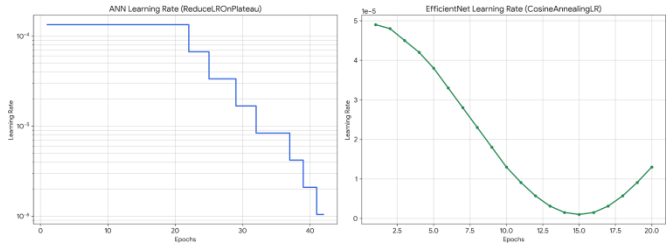


Figure 6: Learning Rate Schedules. Adaptive Step Decay (ANN) vs. Cosine Annealing (Skin Engine). The plots il-lustrate the smooth weight decay required for EfficientNet convergenceversusthestaircasedescentofthesystemicANN.

The learning rate decays smoothly and continuously in a cosine fashion across the training epochs. The relatively high learning rate during the first phase of fine-tuning enabled the network to explore the complex, non-convex loss landscapeof clinical image classification. The decreasing rate, which steadily decreased, allowed the optimiser to settle into a deep and stable local minimum as training progressed, and this is widely credited for the final 96.25% accuracy achieved by the model.

D. Robustness of the System and Real-Time Delay

Along with classification accuracy, the platform was also analysed for end-to-end responsiveness and load stability. Black-box API testing has confirmed that all RESTful end-points on the FastAPI backend are within acceptable time bounds. Both symptom vectorisation and image tensor infer-encereturnstructuredJSONpredictionobjectsconsistently in under 2.5 seconds, meeting the real time responsiveness expectations of a consumer-facing medical web application. Continuous execution monitoring confirmed the stability of memory behaviour in multi-user simulation, with no observ-able memory leaks due to the ONNX runtime during long bursts of concurrent requests.

E. Ablation Experiment

In the validation phase, targeted ablation experiments were conductedtoisolatethecontributionofindividualarchitectural choices. Turning off the RandAugment pipeline for the Skin Engine dropped the validation accuracy by approximately 4.8%andresultedinalargeincreaseinoverfittingintheearly phases of training, indicating that aggressive augmentation is not only useful but crucial when classifying visually similar lesion types that differ by subtle pathological features.

The Dropout regularisation layers were removed from the Advanced Engine, and this had a severe effect: validation accuracydroppedto79.2%,asthenetworkquicklymemorised training samples without learning generalisablerepresenta-tions. This result emphasises the need for explicit regularisa-tionwhenworkingwithhigh-dimensional,sparsebinaryinputs in a heterogeneous patient population.

VIII. ETHICAL CONSIDERATIONS AND LIMITATIONS

Despite the strong performance metrics reported above, the responsible deployment of an AI-based clinical screening tool requires an honest recognition of its current limitations and associated ethical considerations.

A. Dataset Bias & Generalisability

A major limitation of the current Skin Engine is the demo-graphic representation of the training data. The image corpus used in this work is, like most publicly available dermatologi-caldatasets,severelyskewedtowardslighterskintones.Therefore,themodeldiscriminativeperformancemightbeconsider-ablydecreasedwhenappliedtoclinicalimagesofpatientswith darker skin tones, possibly exacerbating healthcare disparities instead of mitigating them. Future dataset development cycles will need to deliberately oversample and actively collect data from under-represented demographic groups in order to scale this platform responsibly.

B. HallucinationsofLLMsandPatientSafety

The Groq-powered conversational agent is explicitly con-strained via system-level prompting to only provide general health information and is prohibited from issuing definitive diagnoses, but the risk of model hallucination cannot be fully eliminated. As Meincke et al. caution, the use of LLMs in clinical-adjacent contexts requires robust guardrails against unwarranted diagnostic confidence [3].

If a patient who does not know the limitations of AI is provided with a spurious medical explanation, it may induce undue anxiety or unsafe self-treatment decisions. Future versions of Asclepius AI will include Retrieval-Augmented Generation (RAG) to signifi-cantly decrease this risk.

C. Hardware Dependency

Currently, sub-2 second inference latency during peak load is dependent on access to cloud hosted computing infras-tructure. Although ONNX conversion has greatly reduced the computational footprint of the Skin Engine, a remaining open challenge is reliably deploying the system in environments with intermittent connectivity or highly constrained hardware, such as rural primary care clinics. Future work will explore edge-computing adaptations, possibly replacing the Efficient-Net backbone with lighter architectures like MobileNetvari-ants, for offline-capable deployment in resource-constrained settings.

Conclusion and Future Directions Here we introduce As-clepius AI, a fully realised end-to-end multi-modal diagnostic platformthatcombinesthreecomplementarymachinelearning paradigms into a single coherent clinical application. The systematic integration of ensemble Random Forests for fast interpretablescreening,regularisedArtificialNeuralNetworks forhigh-dimensionalsystemicreasoningandEfficientNet-V2-Sforrobustdermatologicalimageclassificationcreates a system that can provide clinically meaningful screening accuracy over fundamentally different diagnostic modalities.

The intentional embedding of an always-on, LLM-driven conversational layer directly tackles two long-standing short-comings of traditional digital health products: the lack of longitudinal patient engagement and the data silos that result from diagnostic data and clinical communication being spread across various systems. Asclepius AI is not meant to replace physician judgement, but to extend and amplify it through a secure and transparent conduit from automated screening to professional clinical review, making it a useful first response tool.

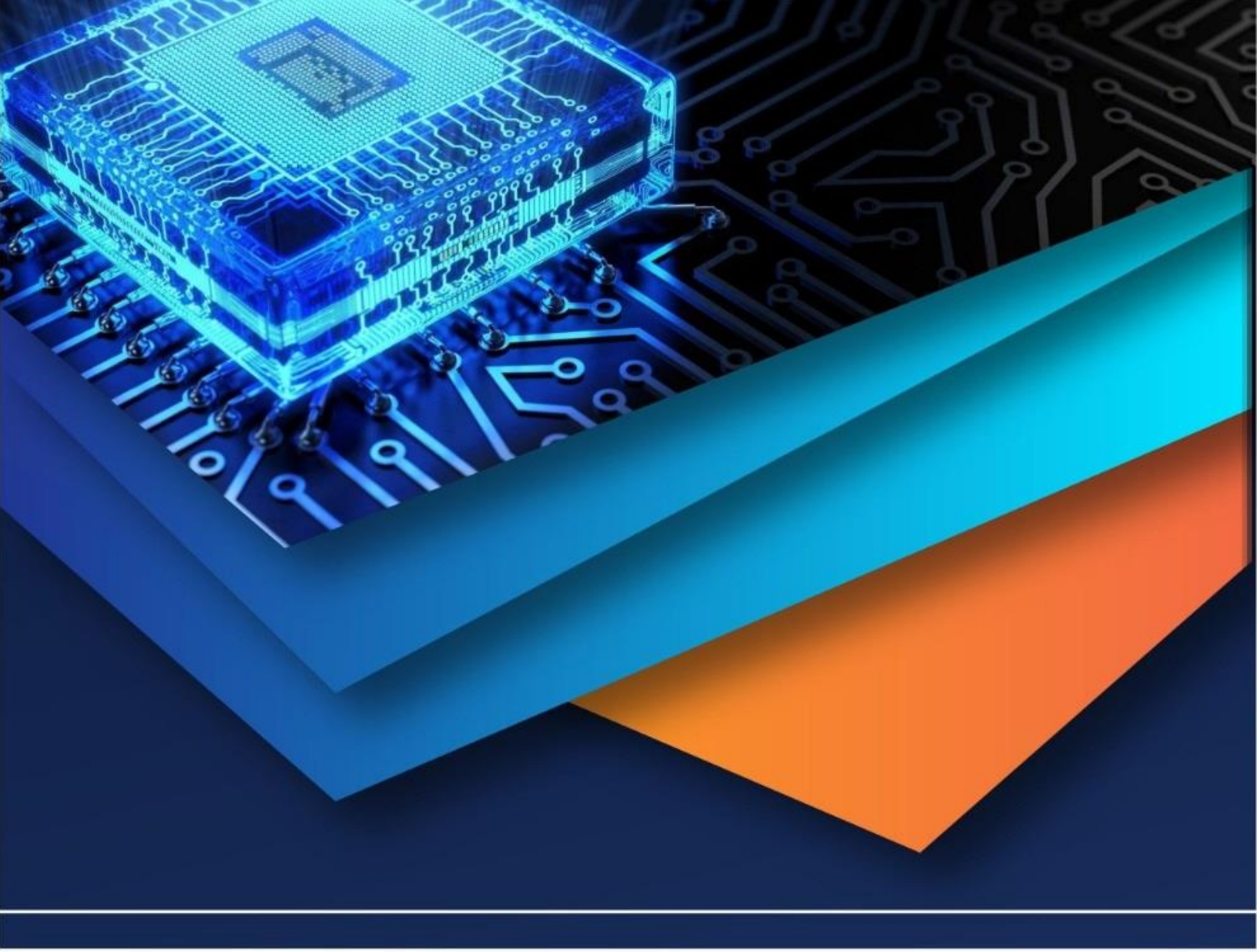
There are many directions for further research. From an architectural perspective, we will explore true cross-modal fusion in future versions (rather than sequential modality routing). Symptom vectors and image tensors will be em-bedded into a common latent representation space to capture inter-modaldependenciesthatsequentialprocessinginherently misses. Concurrently, we plan to integrate rigorous Explain-able AI tooling, notably Gradient-weighted Class Activation Mapping (Grad-CAM) for the generation of visual saliency overlays on skin predictions and SHAP (SHapley Additive exPlanations) for symptom-level feature attribution in the systemic engines.

Finally, by adopting HL7 FHIR interoperability standards, AsclepiusAIwillbeabletonativelycommunicatewithhospi-tal information systems and electronic health record platforms aroundtheworld,realisingitspotentialasawidelydeployable, institution-agnostic clinical decision support utility.

REFERENCES

[1] Mohamed, M. Abdelrehim, and R. Al-Barazie, “Context matters inmachine learning based disease prediction with insights from diverseclinical and symptom data,” Scientific Reports, vol. 15, p. 42669, 2025.
[2] R. Zhang et al., “Multimodal artificial intelligence in medicine: a task-oriented framework for clinical translation,” Frontiers in Medicine, vol.12, p. 1736272, Jan. 2026.
[3] L.Meincke,C.Terwiesch,andA.Huchzermeier,“EvaluatingLLMsforDynamic,MultimodalClinicalDecision-Making,”MackInstituteforIn-novationManagement,TheWhartonSchool,UniversityofPennsylvania,Jan. 2026.
[4] G. Mohan et al., “Disease Prediction Based on Symptoms UsingEnsemble and Hybrid Machine Learning Models,” Amrita School ofComputing, 2023.
[5] M. Alruwaili and M. Mohamed, “An Integrated Deep Learning Modelwith EfficientNet and ResNet for Accurate Multi-Class Skin DiseaseClassification,” Diagnostics, vol. 15, no. 5, p. 551, Feb. 2025.
[6] P. Meenalochini, “Explainable AI for Healthcare Diagnostics: A Multi-modal Approach Using Vision and Language Models,” Journal ofArtificial Intelligence and Cyber Security (JAICS), vol. 9, no. 1, 2025.
[7] P. Guleria, P. N. Srinivasu, S. Ahmed, N. Almusallam, and F. K.Alarfaj, “XAI framework for cardiovascular disease prediction usingclassification techniques,” Electronics, vol. 11, no. 24, Art. no. 4086,2022.
[8] M. Wojtara, E. Rana, T. Rahman, P. Khanna, and H. Singh, “Artificialintelligence in rare disease diagnosis and treatment,” Clin. Transl. Sci.,vol. 16, no. 11, pp. 2106–2111, 2023.
[9] M. M. Ahsan and Z. Siddique, “Machine learning-based heart diseasediagnosis: A systematic literature review,” Artif. Intell. Med., vol. 128,Art. no. 102289, 2022.
[10] N. G. Nia, E. Kaplanoglu, and A. Nasab, “Evaluation of artificialintelligence techniques in disease diagnosis and prediction,” Discov.Artif. Intell., vol. 3, no. 1, Art. no. 5, 2023.
[11] M. M. Ahsan, S. A. Luna, and Z. Siddique, “Machine-learning-baseddisease diagnosis: A comprehensive review,” Healthcare (Basel), vol.10, no. 3, Art. no. 541, 2022.
[12] G.Stiglic,P.Kocbek,N.Fijacko,M.Zitnik,K.Verbert,andL.Cilar, “Interpretability of machine learning-based prediction models inhealthcare,”WileyInterdiscip.Rev.DataMiningKnowl.Discovery,vol.10, no. 5, Art. no. e1379, 2020.

[13] G.Litjensetal.,"Asurveyondeeplearninginmedicalimageanalysis,"Med. Image Anal., vol. 42, pp. 60–88, 2017.
[14] T. Bhanuteja, K. V. N. Kumar, K. S. Poornachand, C. Ashish, and P.Anudeep, "Symptoms-based multiple disease prediction model usingmachine learning approach," Int. J. Innov. Technol. Explor. Eng., 2021.
[15] P. C. Austin, I. R. White, D. S. Lee, and S. van Buuren, "Missing datain clinical research: a tutorial on multiple imputation," Can. J. Cardiol.,vol. 37, no. 9, pp. 1322–1331, 2021.
[16] R. Alkhanbouli, H. M. A. Almadhaani, F. Alhosani, and M. C. E.Simsekler, "The role of explainable artificial intelligence in diseaseprediction:asystematicliteraturereviewandfutureresearchdirections,"BMC Med. Inform. Decis. Making, vol. 25, no. 1, Art. no. 110, 2025.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY