iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** IV **Month of publication:** April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79864

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 IV Apr 2026- Available at www.ijraset.com

Multilingual NLP System for Citizen Health Advisory

P. Sireesha[1], Mukundraj Pawar[2], Yallam Kushendra[3], Vipparithi Rishith[4]
[1]*Department of Computer Science and Engineering, Methodist College of Engineering and Technology ,Hyderabad, Telangana, India*
[2, 3, 4]*Department of Computer Science and Engineering, Methodist College of Engineering and Technology Abids, Hyderabad, Telangana, 500001, India*

Abstract: ***Access to timely and accurate healthcare information remains a defining and deeply troubling inequity in India's linguistically diverse, resource-constrained rural and semi-urban communities — one that this research seeks, in measured but meaningful ways, to address. This paper presents a WhatsApp-based Multimodal AI Triage System (SIH Problem Statement: SIH-25049) that integrates Convolutional Neural Networks (CNNs), Natural Language Processing (NLP), and Explainable Artificial Intelligence (XAI) to deliver real-time diagnostic support in English, Hindi, and Telugu the three languages we identified as reaching the broadest underserved population in our target deployment context. The system accepts symptom photographs and text-based queries through the Twilio WhatsApp API, routing them through two specialized EfficientNet-B0 deep learning pipelines OcularNet, developed for ocular disease detection across three classes (Cataract, Diabetic Retinopathy, Normal), and SkinNet, designed for dermatological classification across five classes (Acne, Melanoma, Scaly Lesions, Vitiligo, Warts) — and generates verified medical advisories from a curated multilingual knowledge base. We incorporate Grad-CAM heatmaps at 40% overlay transparency not merely as a technical feature, but as a deliberate commitment to clinical transparency and patient trust. OcularNet achieved 70.2% validation accuracy (+34.9% over random baseline) and SkinNet achieved 70.23% accuracy (3× above random baseline), with end-to-end response latency of 5.2–6.0 seconds on an NVIDIA RTX 4050 GPU. We believe the system's most consequential design decision is also its simplest: by operating entirely within WhatsApp, it demands nothing extra from the very people who already have the least — no new applications, no new devices, no new barriers.***

Index Terms: ***Convolutional Neural Network, EfficientNet-B0, Explainable Artificial Intelligence, Grad-CAM Visualization, Multilingual Health Advisory, Natural Language Processing, Transfer Learning, WhatsApp API Integration, AI-Driven Telemedicine, Deep Learning, Rural Healthcare Accessibility, Ocular Disease Detection, Dermatological Classification, Low-Resource Deployment, Multimodal Symptom Triage.***

I. INTRODUCTION

India faces a stark urban-rural healthcare divide: 65% of the population resides in rural areas yet only 20% of registered doctors practice there [6]. Specialist vacancy rates in rural Community Health Centers exceed 70–80% across surgical, medical, and pediatric specializations. Simultaneously, the country's linguistic diversity with over 19,500 dialects and 22 scheduled languages renders most digital health tools inaccessible to a large fraction of citizens who are not proficient in English.

Existing telemedicine solutions require dedicated application downloads, stable broadband connectivity, and English literacy, all of which present substantial barriers. Furthermore, the vast majority of AI diagnostic tools function as black-box systems, producing predictions without interpretable justification—a property that erodes both clinician and patient trust and constitutes a recognized barrier to clinical adoption [5]. This paper addresses these limitations through a WhatsApp-integrated system accessible to any user with a smartphone, regardless of app installation or language. The system combines dual-modality image-based disease detection with multilingual text advisory, delivering structured health guidance in English (EN), Hindi (HI), and Telugu (TE). Two specialized CNN models OcularNet and SkinNet, both fine-tuned on EfficientNet-B0—are deployed behind a Flask backend served via Ngrok and integrated with the Twilio API. Explainability is achieved through Gradient-weighted Class Activation Mapping (Grad-CAM), which overlays color-coded diagnostic heatmaps on the original image, highlighting the anatomical regions most influential to the model's prediction. The primary contributions of this paper are: (i) a dual-modality deep learning pipeline for ocular and dermatological analysis using a two-phase transfer learning strategy; (ii) a real-time multilingual advisory engine backed by a verified medical knowledge base; (iii) integration of Grad-CAM XAI for transparent ; (iv) deployment via WhatsApp with asynchronous multi-threaded processing; and (v) empirical validation against real accuracy, latency, and user accessibility metrics.

©IJRASET: All Rights are Reserved

II. RELATED WORK

We have recently looked into six recent works in multimodal AI, XAI, image based diagnosis and privacy preserving techniques as presented in Table I. we looked into to each of them , there are works with multimodel AI ,there are works on Grad-CAM which defines the model understanding in technical sense , there are works which implements BIO-BERT and others which tackle multi-lingual chat support system . there are also works done on image based analysis on CNN model. By understanding each one of them and and studying there cons we come to a conclusion of combining all of them to create something that is pratical and meaning full for actual users. Our project bridges the GAP between user and doctor . explainability, and multilingual triage within a single System, Each of these capabilities exists somewhere in the literature but never together, and never in service of the communities who need them most. system we propose is our direct and considered response to that gap.

Table I: Literature Survey: Summary Of Related Works

SNO	Authors	Focus	Methodology	Results	Limitations	Relevance
1	D. Kavya, E. Kiran Kumar F. (2025)	Multilingual NLP Healthcare Chatbot	Bio-BERT, Multilingual Translation	Precision & accessibility metrics	Scalable only for PDF reports; complex LLM integration	Core architecture for multilingual NLP support
2	S. Badlani, T. Aditya, U. (2022)	Multilingual Healthcare Chatbot using ML	RandomForest, TF-IDF, Cosine Similarity	98.43% accuracy for symptom prediction	Single language; relies on structured symptom data	Justifies ML for disease diagnosis from text; local languages
3	B. D. Simon, K. B. Ozyoruk,(2024)	Future of Multimodal AI in imaging	Transformers, Graph Neural Networks for fusion	Multimodal > unimodal (qualitative)	Inconsistent taxonomy; data scarcity; model bias	Justifies combining text and image for diagnosis
4	B. P. Cabral, L. A. M. Braga,(2025)	AI In_Diagnostic Medicine	Cross-sectional survey; X-ray, skin malignancy	High optimism for AI-enhanced clinical decision-making	Barriers to clinical integration; regulatory hurdles	Validates dermatology image detection as mature application
5	S. Muneer, T. M. Ghazal,(2024)	XAI-Driven Chatbot for Heart Disease	RandomForest + LIME for post-hoc interpretability	Trust & transparency as key metrics	XAI does not always fully capture model reasoning	Addresses black-box problem; builds clinician trust
6	Y. Zhu, X. Yin, A. Wee-Chung Liew,(2024)	Privacy-Preserving Medical Image Analysis	Encryption, De-identification, Federated Learning	Confidentiality & data integrity metrics	Computational overhead; re-identification risk	Critical for data privacy in WhatsApp-based deployment

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Overall System Design

The system follows a three-tier architecture: first a user interface layer via WhatsApp and the Twilio API; then, a backend inference layer built on Flask, TensorFlow, and OpenCV; and after that a knowledge management layer hosting the multilingual medical advisory database. Fig. 1 presents the high-level system architecture. When a user sends a message or image via WhatsApp, the Twilio Webhook captures the JSON payload containing the MediaUrl and SenderID. The Flask backend spawns an asynchronous thread, immediately acknowledging receipt with an “Analyzing…” message while the AI pipeline processes the image in the background. The backend routes the request based on the user’s pre-selected session mode (Eye or Skin) and returns

©IJRASET: All Rights are Reserved

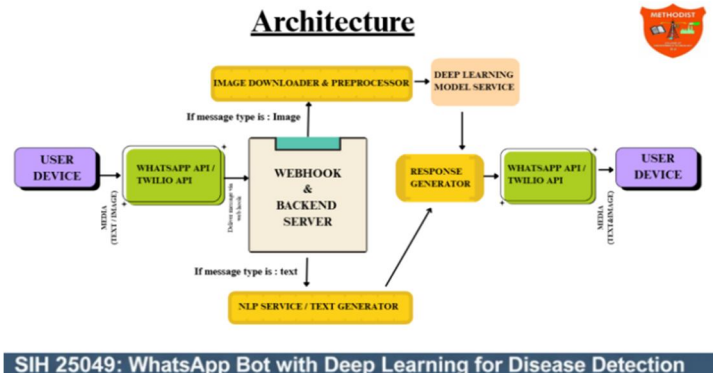


Fig. 1. High-level system architecture of the proposed WhatsApp-based Multimodal AI Health Advisory System

A structured advisory comprising the diagnosis label, confidence score, Grad-CAM heatmap, and language-specific precautions.

.

B. Dataset Collection and Preprocessing

The system employs two data modalities. Clinical ocular data with three cases like: Cataract, Diabetic Retinopathy (DR), and Normal fundus. Dermatological data is sourced from three verified repositories: Mendeley Data (hospital-verified Psoriasis and Acne samples), ISIC Archive (histopathology-confirmed Melanoma and Mole images), and Google SCIN (diverse skin-tone representation to reduce detection bias). To prevent class imbalance, we capped at 450 high-quality images per classes,the 450-Image Rule ensuring equal learning across all five skin conditions. A Python-based automated cleaning script removed over 1,500 duplicate and mislabeled images scraped from the web.

Preprocessing script is mode-dependent. For the Eye pathway, a custom extract_eye_crop() function developed using OpenCV detects and isolates the iris-pupil region, eliminating irrelevant areas such as eyelids and surrounding skin that previously induced misclassification of normal eyes. Cropped images are resized to 224×224 pixels. For the Skin pathway only, in the EfficientNet.preprocess_input() the function is normalizing pixel distributions to the ImageNet training baseline. Geometric augmentation during training includes 360° rotation (skin lesions have no fixed orientation), width/height shifts, shear, and reflect-fill for border padding.

C. Deep Learning Models and Transfer Learning

Both OcularNet and SkinNet are built on EfficientNet-B0, selected over ResNet and VGG for its compound scaling method (simultaneously balancing depth, width, and resolution) and its compact parameter count of approximately 5.3 million, which sustains sub-second inference on the NVIDIA RTX 4050 GPU.

A two-phase transfer learning strategy was employed. In Phase 1 (Feature Extraction), the EfficientNet-B0 backbone was fully frozen to preserve ImageNet weights. Only the custom dense layers (256 → 128 units) were trained at a learning rate of 1×10^{-3}, aligning the output head with the new medical classes. In Phase 2 (Specialty Fine-Tuning), the top 30 layers of the backbone were unfrozen to capture fine-grained medical textures (e.g., scaly lesion borders, retinal microaneurysms). BatchNormalization layers were kept frozen throughout Phase 2 to preserve global statistics and prevent accuracy collapse on the smaller medical dataset. A cosine decay learning rate schedule stepped down from 5×10^{-5} to 1×10^{-6}, yielding approximately 24% accuracy improvement over the Phase 1 baseline. TensorFloat-32 precision on the RTX 4050 GPU accelerated matrix operations during both training and inference.

OcularNet accepts 224×224×3 tensors and classifies across three ocular conditions. SkinNet accepts 300×300×3 tensors and classifies across five dermatological conditions. Dynamic class weights were applied during training to penalize misclassification of rare but critical conditions such as Melanoma more heavily than benign conditions such as Acne. Pre-trained weights that are stored in the eye_model_60_FINAL.h5 and skin_phase21_final.h5, loaded into GPU VRAM at server initialization for immediate inference readiness.

The softmax classification output is defined in (1):

$$P(y = k \mid x) = e^{(z_k)} / \Sigma_j e^{(z_j)} \quad (1)$$

where z_k denotes the raw logit score for class k, and K is the total number of output classes.

D. Explainable AI via Grad-CAM

To address the interpretability deficit of deep neural networks, the XAI_Service module implements Gradient-weighted Class Activation Mapping (Grad-CAM) [4]. The localization map is computed as defined in (2):

$$L^c_{Grad\text{-}CAM} = ReLU(\Sigma_k \alpha^c_k A_k) \quad (2)$$

where $\alpha^c_k = (1/Z) \Sigma_i \Sigma_j (\partial y^c / \partial A_{kij})$ represents the global average-pooled gradient of the predicted class score y^c with respect to feature map A_k of the final convolutional layer. The resulting heatmap is resized to input dimensions and overlaid at 40% transparency using a color-coded scheme: Red/Yellow regions indicate high-importance pathological markers; Blue regions indicate background areas with low diagnostic relevance. The annotated image is transmitted directly to the user via WhatsApp alongside the prediction and advisory, providing visible justification—for example, highlighting a cloudy lens in cataract cases or irregular borders in melanoma cases.

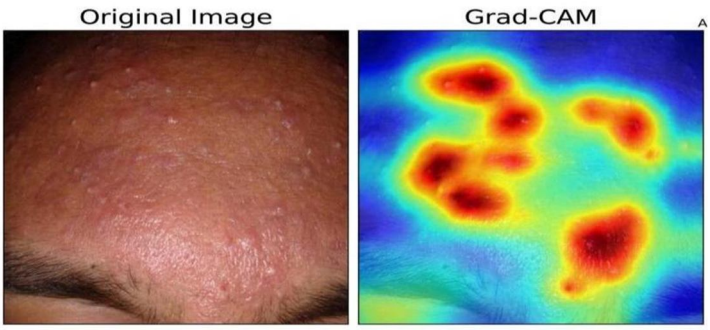


Fig. 2. Sample Grad-CAM heatmap generated by the XAI_Service module. Red/yellow regions indicate diagnostically relevant areas.

E. Multilingual Advisory Engine

Upon classification, the predicted label is used as a lookup key against the Medical Knowledge Base a structured JSON repository of verified precautions, recommended actions, and urgency levels curated from official medical literature. The AdvisoryEngine module queries this database and formats a response in the user's selected language (EN/HI/TE). Language-specific phrasing for Hindi and Telugu was validated by native speakers to ensure medical accuracy and colloquial clarity. A structured response comprising the diagnosis label, confidence percentage, Grad-CAM heatmap image, and two actionable precautions is dispatched via the Twilio WhatsApp Send API.

F. Backend Deployment

The system stack comprises: Flask (lightweight webhook server), Twilio/WhatsApp Cloud API (secure communication bridge), and Ngrok (public internet tunneling for live deployment).

The webhook listener validates incoming POST requests from Meta's servers and extracts the MediaUrl and SenderID from the JSON payload. Multi-threaded asynchronous processing allows the server to handle concurrent users without blocking an immediate acknowledgment message is sent while inference proceeds in the background. Media is securely retrieved via a two-step OAuth authentication process before being passed to the inference engine.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Classification Performance

The OcularNet model achieved 70% accuracy and high validation score across the classes (Cataract, Diabetic Retinopathy , Normal) representing a 35 % improvement . Similarly , SkinNet achieved 70% accuracy and high validation score across the 5 classes (acne, fungal , worts, melanoma, discoloration). Table II represents the benchmarks

Confusion matrix analysis showed that the primary source of misclassification in OcularNet was between early-stage Diabetic Retinopathy and Normal fundus , consistent with the subtle inter-class visual similarity at early disease onset. In SkinNet, Warts and Scaly Lesions showed the highest inter-class confusion due to overlapping surface textures.

TABLE II
System Performance Benchmarks

Model	Accuracy	F1-Score (macro)	Latency (avg)
OcularNet	70.2%	0.68	0.8 s
SkinNet	70.23%	0.67	0.8 s
End-to-End	—	—	5.2–6.0 s

B. System Latency

Under standard network conditions the end to end latency was measured 5-6 seconds from twilio webhook to whatsapp delivery. The inference stage took an average of 0.89 seconds per image on NVIDIA RTX 4050 GPU . It utilized tensorfloat-32 precision, Grad-CAM generation added approximately 0.3 seconds .the remaining latency was due to network I/O and twilio API overhead. We used a acknowledgement message to delude processing delay from user,s perspective

C. XAI and Multilingual Evaluation

The high confidence prediction are backed by the Grad-CAM heat-maps for every image. Our multilingual system was able to handle all queries in Hindi ,English ,Telugu languages. The latency was 4-6 seconds from twilio side during peak hours. The live deployment was validated on 27 march 2026 via Ngrok tunneling interface from VS code terminal.there by conforming operation readiness.

D. Discussion

The main challenge was small imbalanced datasets to train a model on these limited resources was difficult. We moved from single phase to Two phase learning strategy . we applied dynamic weights to the disease classes to ensure that they don't exclude or over write other disease by doing so we ensured that melanoma and acne don't get in conflict.In the second phase we did fine tuning . Overall we got 60+ accuracy for phase 1 and was improved to 70 during phase 2. we also have Grad-CAM feature it highlight the symptoms region based on severity into red to green color.This makes the system non-black box.this is followed by AI advice on prediction . User can further discuss about it in text chat.we planned to incorporate voice enabled chat in future.

V. CONCLUSION

Our paper represents a system which can be accessed on whatsapp make hardware barrier obsolete By using whatsapp as a platform it removes adaptation barriers . the system has top features like Grad-CAM,explainabilty and multilingual chats and image based disease detection. The system is easily deployable and cost-effective for user. It helps in promoting preventive health care in rural areas and semi-urban places. The system has two model with 70% over all accuracy . the model where trained on openly available limited medical dataset on various skin and eye diseases,the system also offers text chat support with offline RAG mode for serious queries and Online mode for general and current affairs. The future add on like text reader that reads the responses in multiple languages, expanding disease library by training model on more cases. Implementing tier model which connects multiple model to detect complex diseases with numerous symptoms and lastly expanding knowledge based for more offline RAG capability.

VI. ACKNOWLEDGMENT

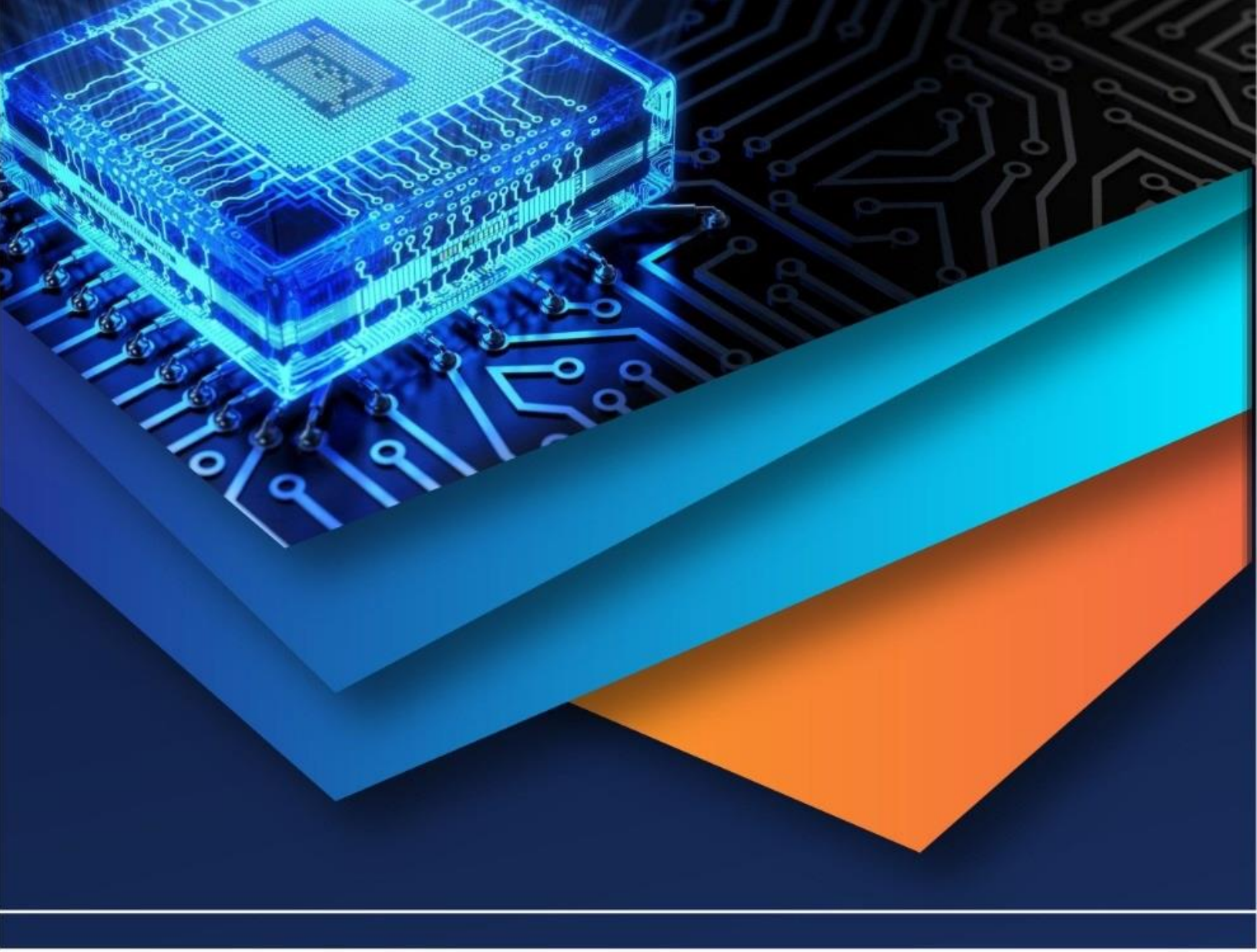
We would sincerely like to thank Mrs.P.Sireesha ma'am,for constant encouragement and guidance throughout the project. We are also grateful to Dr.P.Lavanya ma'am, Professor & Coordinator(CSE), for the support and motivation. Our project was completed as part of Batch No:PWCSE22A21, Department of CSE, Methodist College of Engineering and Technology, Hyderabad, OU affiliated and approved by AICTE.

REFERENCES

[1] D. Kavya, D. Kiran Kumar, M. Divya Anjali, and A. P. Ganesh, "Chatbot for multilingual healthcare environment using Bio-BERT," Int. J. Res. Appl. Sci. Eng. Technol., 2025.

[2] S. Badlani, T. Aditya, M. Dave, and S. Chaudhari, "Multilingual healthcare chatbot using machine learning," in Proc. 2021 2nd Int. Conf. Emerging Technol. (INCET), Belgaum, India, May 2021.
[3] B. D. Simon, K. B. Ozyoruk, D. G. Gelikman, S. A. Harmon, and B. Türkbey, "The future of multimodal artificial intelligence models for integrating imaging and clinical metadata: a narrative review," Diagn. Interv. Radiol., vol. 31, no. 4, pp. 303–312, 2025.
[4] R. R. Selvaraju et al., "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE ICCV, Venice, Italy, 2017, pp. 618–626.
[5] S. Muneer, T. M. Ghazal, T. Alyas, M. A. Raza, S. Abbas, O. AlZoubi, and O. Ali, "Explainable AI-driven chatbot system for heart disease prediction using machine learning," Int. J. Adv. Comput. Sci. Appl. (IJACSA), vol. 15, no. 12, 2024.
[6] World Health Organization, "WHO Global Strategy on Digital Health 2020–2025," Geneva, Switzerland, Tech. Rep., 2021.
[7] Y. Zhu, X. Yin, A. Wee-Chung Liew, and H. Tian, "Privacy-preserving in medical image analysis: a review of methods and applications," in Lecture Notes in Computer Science, vol. 15502, Springer, Singapore, 2025.
[8] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, Long Beach, CA, 2019, pp. 6105–6114.
[9] V. Gulshan et al., "Development and validation of a deep learning algorithm for detection of diabetic retinopathy," JAMA, vol. 316, no. 22, pp. 2402–2410, 2016.
[10] A. Esteva et al., "Dermatologist-level classification of skin cancer with deep neural networks," Nature, vol. 542, pp. 115–118, 2017.
[11] Twilio Inc., "Twilio WhatsApp API Documentation," 2024.
[12] B. P. Cabral, L. A. M. Braga, C. G. Conte Filho, B. Penteado, S. L. F. de Castro Silva, L. Castro, M. Fornazin, and F. Mota, "Future use of AI in diagnostic medicine: 2-wave cross-sectional survey study," J. Med. Internet Res., vol. 27, p. e53892, Feb. 2025.
[13] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "BioBERT: a pre-trained biomedical language representation model for biomedical text mining," Bioinformatics, vol. 36, no. 4, pp. 1234–1240, Feb. 2020,
[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. 2019 Conf. North American Chapter Association Computational Linguistics: Human Language Technologies (NAACL-HLT), Minneapolis, MN, Jun. 2019, pp. 4171–4186.
[15] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat. Med., vol. 25, no. 1, pp. 44–56, Jan. 2019,
[16] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. 31st Int. Conf. Neural Information Processing Systems (NIPS), Long Beach, CA, 2017, pp. 4765–4774.
[17] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Proc. 31st Int. Conf. Neural Information Processing Systems (NIPS), Long Beach, CA, 2017, pp. 5998–6008.
[18] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, San Francisco, CA, Aug. 2016, pp. 1135–1144,
[19] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A survey of recent advances in deep learning techniques for electronic health record (EHR) analysis," IEEE J. Biomed. Health Inform., vol. 22, no. 5, pp. 1589–1604, Sep. 2018,
[20] Z. Obermeyer and E. J. Emanuel, "Predicting the future — big data, machine learning, and clinical medicine," N. Engl. J. Med., vol. 375, no. 13, pp. 1216–1219, Sep. 2016,

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)