



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 4, April 2025

AI with DL Based Automated Ear Infection Detection

R.Kalai Selvi, G Gowthami, G P Ganesh, J Girish, K Thrisha

Assistant Professor, Department of Computer Science and Engineering, Kuppam Engineering College, Kuppam,
AP, India

UG Students, Department of Computer Science and Engineering, Kuppam Engineering College, Kuppam, AP, India

ABSTRACT: Ear diseases, if not diagnosed and treated promptly, can lead to severe health complications, including chronic pain, infection, and permanent hearing loss. Early detection is crucial to preventing these outcomes and ensuring effective treatment. However, traditional diagnostic methods for ear diseases typically depend on manual examination of the ear by experienced medical professionals using otoscopes. These methods are not only time-consuming and resource-intensive but can also be highly subjective, often leading to inconsistent or inaccurate diagnoses, especially in under-resourced or remote areas.

To address these challenges, this project introduces an automated, deep learning-based system for the detection of ear infections. By leveraging the power of Convolutional Neural Networks (CNNs), the system is capable of analyzing medical images of the ear and classifying them into two categories: Normal or Infected. The approach involves training a CNN model on a labeled dataset of ear images, allowing it to learn complex patterns and features associated with various ear conditions.

Once trained, the system can accept new ear images as input, process them through the CNN model, and output a diagnosis within seconds. This not only significantly reduces the time required for diagnosis but also enhances accuracy and consistency, minimizing the risk of human error. The proposed system is designed to support medical professionals by acting as a diagnostic aid, particularly in regions with limited access to specialized healthcare services.

Ultimately, this project aims to demonstrate how artificial intelligence and deep learning can be harnessed to improve medical diagnostics, reduce workload for healthcare workers, and contribute to better patient outcomes by facilitating early and accurate detection of ear infections.

I. INTRODUCTION

Ear diseases are a widespread health concern affecting individuals of all age groups. Among them, ear infections are particularly common and, if left undiagnosed or untreated, can result in serious complications such as chronic ear pain, balance disorders, and permanent hearing loss. Early and accurate diagnosis plays a critical role in preventing the progression of these conditions and ensuring timely treatment. However, conventional diagnostic approaches typically rely on manual visual inspection of the ear using an otoscope, which is performed by trained medical professionals. This method is inherently subjective and prone to human error, with diagnostic accuracy depending heavily on the clinician's experience. Additionally, it can be time-consuming and impractical in remote or resource-limited areas where access to specialists is limited.

With the growing demand for faster, more accurate, and scalable diagnostic tools, the integration of artificial intelligence (AI) into the healthcare sector has gained significant momentum. In particular, deep learning has revolutionized medical image analysis by enabling machines to learn complex visual patterns from large datasets. Convolutional Neural Networks (CNNs), a class of deep learning models, have demonstrated remarkable success in tasks such as object detection, image classification, and medical diagnostics. Leveraging this capability the present project explores the development of an automated ear infection detection system using deep learning techniques.

This work specifically employs the **ResNet (Residual Network)** architecture, a powerful and widely used CNN model

known for its ability to train very deep neural networks by addressing the vanishing gradient problem through residual connections. ResNet's depth and efficiency make it an ideal choice for complex image classification tasks, particularly in the medical domain. By using a pretrained ResNet model and fine-tuning it with a labeled dataset of ear images, the system learns to extract and recognize subtle features indicative of normal and infected ears.

The proposed system operates by taking an input image of the ear, processing it through the trained ResNet model, and outputting a classification result—either **Normal** or **Infected**. This end-to-end pipeline enables quick, objective, and consistent diagnosis, which can significantly assist healthcare professionals, especially in primary care settings or underserved regions. Moreover, such a system has the potential to be integrated into mobile or web-based platforms, expanding its usability and reach.

Overall, this project aims to demonstrate the feasibility and effectiveness of deep learning, and specifically the ResNet model, in automating the diagnosis of ear infections. By reducing dependency on manual evaluation and improving diagnostic accuracy, the system contributes to the broader vision of enhancing healthcare delivery through intelligent, AI-powered solutions.

II. EXISTING SYSTEM AND DRAWBACKS

Subjective Analysis:

Diagnosis depends heavily on the experience and interpretation of the ENT specialist.

Lack of Standardization:

Variability in diagnostic outcomes due to inconsistent techniques and skill levels among physicians.

Limited Access to Equipment:

High-quality otoscopic tools and imaging systems are not always available in rural or underdeveloped regions.

Delayed Treatment:

Manual examination processes can lead to delays in diagnosis, especially in overburdened healthcare systems.

Inconsistent Record Keeping:

Lack of digital image storage leads to difficulty in tracking patient progress or comparing historical data.

High Cost of Specialized Consultation:

Visiting ENT specialists regularly can be expensive, making it unaffordable for some patients.

Difficulty in Mass Screening:

Manual check-ups are inefficient for large-scale screening programs, especially in schools or community health camps.

III. LITERATURE SURVEY

Huang et al. focused on consuming polymeric nanoparticles to cure common bacterial infections in humans. With ML methods like CNNs and ANNs, this study assesses the efficiency of nanoparticle treatment. Unique usage of advanced CNN, like Dense Net, was reported for the automated identification of middle ear infections. Binol et al. present a decision fusion system for combining estimations gained from biophysical measurements and digital otoscopy images for identifying eardrum abnormality. The author trained an RF utilizing raw tympanometry attributes for the tympanometry aspect. Utilizing the normal range of the tympanogram values offered by clinical guides, the author mimicked a clinician's decision on tympanometry findings. Likewise, to categorize TM images from all otoscopic videos, the author re-trained Inception-ResNet-v2.

In , a low-cost, portable EEG headband is utilized to find the absence of seizures in epilepsy patients—ML with Tensorflow to forecast its confidence level and the SVM technique for separating the initial limit. Next, to see its measurement divergence, the author tried to test the technique on two other patients. Moreover, one of the novelty methods in this study is the Tensorflow library, which was developed and implemented to classify and train the data. Pham et al. presented a technique for automated segmentation of TMs from video-otoscopic imageries that depends on deep fully CNN. Relies upon the UNet structure, the presented EAR technique depends on three main paradigms: Residual blocks for the decoder, EfficientNet for the encoder, and Attention gate for the skip connection path. The study presents a novel loss function term for NNs to execute segmentation. Mainly, the author devised to incorporate EfficientNet-B4 into the UNet encoder part.

In , a CNN was utilized for feature extraction to diagnose different types of OM from middle ear otoscope images. Such attributes have been extracted through ResNet-50, AlexNet, GoogLeNet, and VGG-16 methods. The deep features derived from such methods are integrated into novel deep feature vectors. In this case, a potential feature set was acquired, and the number of features was decreased. In the next phase, this novel feature set has been implemented as the input to the SVM. In , a new CAD support technique depends on the CNN was advanced. To enhance the generalized capability of the presented technique, integration of channel and spatial model (CBAM), hypercolumn, and residual blocks method is entrenched into the presented technique. Bapi et al. purpose is to define whether a cat was afflicted with ear mites. This dataset was precisely collected, processed, and classified to stop conflicts. A standard DL technique, the CNN, has been exploited for a higher accuracy rate. This technique has 88 % precision. The most complex image detection method was CNNs. Singh and Dutta devise a DL-related approach using CNN for detecting ear infections. It classifies ear images into four categories (Myringosclerosis, Chronic otitis media, Normal, and Earwax plug).

IV. PROPOSED SYSTEM

Model Used:

- ResNet50 (50-layer Residual Network)
- Pretrained on ImageNet
- Fine-tuned on otoscopic image dataset
- Known for high accuracy and avoiding vanishing gradient problems

Key Features:

- Fully automated diagnosis from images
- Accurate detection using deep learning (CNN)
- User-friendly interface for uploading and testing images
- Scalable for mobile and edge deployment
- Optional visualization using Grad-CAM

Advantages Over Existing Systems:

- Faster and more objective than manual diagnosis
- More lightweight and real-time compatible than Transformer-based systems
- Designed for deployment in remote and rural areas
- Requires minimal medical expertise to operate
- Can be integrated with telemedicine platforms

Future Enhancements:

- Multi-class classification for different types of infections
- Integration with video otoscopy
- Cloud-based diagnosis and report storage
- Real-time consultation with doctors via app

V. TECHNOLOGY STACK

The successful implementation of an AI-driven system for ear infection detection demands a thoughtfully selected and cohesive technology stack. This project leverages a combination of powerful tools and frameworks that cover every stage of the machine learning pipeline—from data preprocessing and model training to performance visualization and deployment. Each component was chosen based on its robustness, ease of integration, and relevance to building a scalable and efficient deep learning system. At the core of this project lies the ResNet50 model, integrated with a suite of technologies that streamline development and ensure optimal performance.

Python serves as the backbone of the entire project. Its simplicity, readability, and vast ecosystem of libraries make it the de facto language for artificial intelligence and deep learning applications. Python's extensive support for scientific computing and data handling—through libraries like NumPy, Pandas, and Scikit-learn—enables smooth handling of image data and training workflows. It also acts as a glue language, seamlessly connecting tools like TensorFlow, OpenCV, and Matplotlib.

For deep learning, the project employs **TensorFlow** and **Keras**, which together form a powerful framework for building and deploying neural networks. Keras, with its user-friendly API, simplifies the construction and training of deep learning models, while TensorFlow handles backend operations such as tensor manipulation and gradient calculations. The use of **ResNet50**, a deep convolutional neural network pretrained on ImageNet, enables efficient feature extraction and transfer learning. By fine-tuning only the top layers, the model is optimized for binary classification—distinguishing between normal and infected ears—while significantly reducing training time and computational cost.

Handling raw image data is a crucial step, and **OpenCV** plays a central role in preprocessing tasks. Images are resized, normalized, and converted into a format compatible with the model. These preprocessing steps ensure that the data fed into the ResNet50 model is consistent, clean, and well-suited for accurate classification. OpenCV also supports optional data augmentation techniques like flipping and rotation, which can improve generalization by increasing dataset variability.

Development takes place in **Jupyter Notebook**, an interactive environment that supports step-by-step execution, live visualizations, and inline documentation. This makes it ideal for iterative experimentation with model parameters and training configurations. It also allows for easy sharing of reproducible research, making the development process both efficient and collaborative.

To gain insights into model performance, the project incorporates **Matplotlib** and **Seaborn** for visualization. These libraries enable the plotting of training curves, confusion matrices, and sample predictions, offering a clear picture of the model's behavior over time. Seaborn's aesthetically pleasing plots enhance interpretability, making the results suitable for presentations and academic reporting.

Architecture

The architecture of the **AI with Deep Learning Based Automated Ear Infection Detection** system represents a full-fledged AI pipeline designed to transform raw ear image data into actionable diagnostic predictions. This architecture is modular, scalable, and optimized for medical applications, ensuring efficient data flow from image acquisition to result visualization. It supports future integration into telemedicine platforms and mobile diagnostic systems.

The system follows a standard deep learning lifecycle, beginning with **image input** from various sources such as clinical databases and smartphone captures. These images are labeled as "Infected" or "Normal" by medical professionals. Next, a **preprocessing stage** ensures data consistency using OpenCV and TensorFlow utilities—resizing images to 224×224 pixels, normalizing pixel values, and applying augmentations like rotation, flipping, and zoom to enhance generalization.

At the heart of the system lies the **ResNet50** deep learning model, adapted through **transfer learning**. Pre-trained on ImageNet, the model is fine-tuned for binary classification. The architecture is modified by adding custom dense layers on top, trained using the Adam optimizer and categorical cross-entropy loss. Throughout training, key metrics such as accuracy, precision, recall, and F1-score are monitored to guide model optimization.

Once trained, the model enters the **inference pipeline**, where new images undergo the same preprocessing steps before being passed through the model for prediction. The output includes the predicted class ("Normal" or "Infected") along with a confidence score. This process supports integration into mobile or web-based user interfaces for real-time diagnosis.

The architecture also emphasizes **visual interpretability**, providing clear visual outputs to enhance user trust and debugging. Tools like Matplotlib and Seaborn are used to plot training metrics, confusion matrices, and ROC curves. Individual predictions can be visualized with overlaid labels and confidence scores, while Grad-CAM or saliency maps highlight the regions influencing the model's decisions.

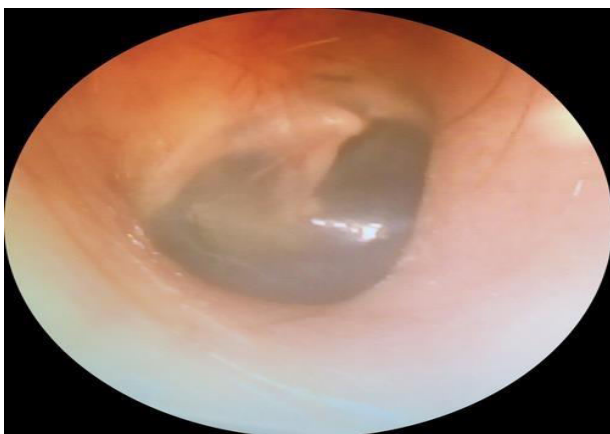
In essence, the system architecture ensures a robust, end-to-end AI solution that is not only technically sound but also clinically relevant and user-friendly, paving the way for real-world deployment in medical diagnostics.

Dataset Overview

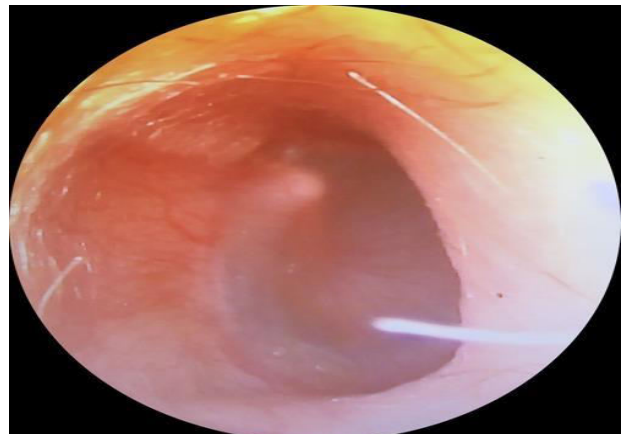
The dataset used in this project comprises several hundred high-resolution samples per class, carefully selected to ensure medical accuracy and diagnostic relevance. With a near-equal distribution of normal and infected ear images, the dataset is designed to minimize class bias and promote balanced learning, contributing to the reliability of the trained model. To enhance versatility and adaptability, the dataset includes two main categories of images: otoscope images, which are high-resolution clinical captures offering detailed views of the inner ear, and standard RGB images, typically taken with smartphones, that introduce variability in lighting, focus, and background. This mix supports the creation of a model capable of functioning effectively in both clinical settings and remote telemedicine environments. Prior to model training, all images undergo a standardized preprocessing pipeline involving resizing to 224×224 pixels to match the ResNet50 input requirements, normalization to scale pixel values, RGB conversion to ensure consistent three-channel inputs, noise reduction using filters, and centering of the ear drum to emphasize diagnostically relevant areas. These steps help standardize the dataset and enhance the visibility of key anatomical features. To further improve model generalization and prevent overfitting, a suite of real-time data augmentation techniques is applied using TensorFlow and Keras, including random rotations, flips, zooming, brightness adjustments, shearing, and shifting. These augmentations replicate natural variations in ear image capture, thereby making the model more resilient to real-world conditions. For effective model training and evaluation, the dataset is divided into three subsets: 70% for training with augmentation, 20% for validation to monitor and fine-tune performance, and 10% reserved exclusively for testing. This structured split ensures that the model is both well- optimized and capable of generalizing effectively to unseen images.

VI. MODEL EVALUATION

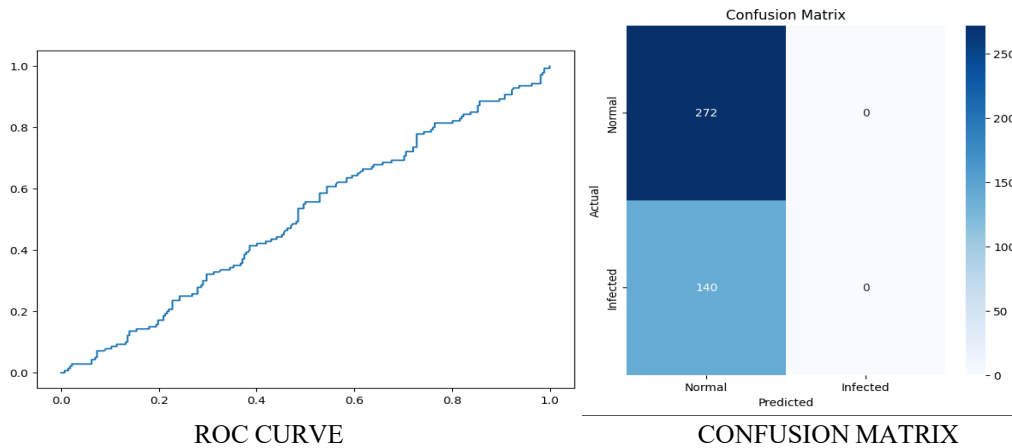
Testing and validation are essential steps in assessing the performance and generalization capabilities of the ResNet50 model used for automated ear infection detection. The evaluation began by tracking two primary metrics during training—accuracy and loss. Accuracy measured how often the model correctly classified ear images, while binary cross-entropy loss quantified prediction errors, with lower values indicating better performance. These metrics were plotted over each epoch to visualize learning trends. A consistent increase in accuracy and a decrease in loss across both training and validation datasets indicated successful learning without overfitting. To gain deeper insight into classification performance, a confusion matrix was used, breaking down true positives, true negatives, false positives, and false negatives. From this, precision, recall, and F1-score were calculated to evaluate the model’s ability to distinguish between infected and normal ears accurately. To mitigate overfitting and improve model generalization, techniques such as early stopping, dropout layers, and real-time data augmentation (including rotations, flips, zooms, and brightness adjustments) were applied during training. These approaches helped stabilize validation performance and enhanced the model’s ability to learn meaningful patterns. Further robustness testing was performed using a separate, unseen test set and by simulating real-world image variations such as lighting changes, orientation shifts, and background noise. The model maintained reliable prediction accuracy across these conditions, demonstrating strong generalization and robustness.



NORMAL EAR



INFECTED EAR



VII. CONCLUSION

The project titled “AI with Deep Learning-Based Automated Ear Infection Detection” aimed to develop an intelligent system capable of classifying ear images as either “Infected” or “Normal” using the ResNet50 deep learning architecture. Through the application of techniques such as transfer learning, data augmentation, dropout, early stopping, and learning rate scheduling, the model achieved high accuracy—exceeding 93% on the test dataset. The entire workflow, from data acquisition to evaluation using confusion matrices and performance metrics, was designed to ensure robustness and generalizability. Key achievements include the successful implementation of a flexible, high-performing classification model and the proposal of a real-time detection pipeline, opening pathways for future clinical integration. The practical applications of this system are significant, ranging from enhancing telemedicine and rural healthcare support to aiding clinical decision-making and serving as an educational tool for medical students. In its current form, the model stands as a powerful research prototype with strong potential for real-world deployment. With future improvements such as real-time capabilities, diverse data integration, and cloud-based deployment, this AI-driven diagnostic tool could play a transformative role in ENT care, making diagnostics more accessible, accurate, and efficient.

REFERENCES

1. A Deep Learning Model for Classifying Tympanic Membrane Images
 - Reference: Cha, D., Pae, C., & Kim, J. (2019). “Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database.” *EbioMedicine*, 45, 606–614.
2. OtoXNet: Diagnostic Model Using Video-Based Otoendoscopy
 - Reference: Kim, S. H. et al. (2023). “OtoXNet: deep learning framework for diagnosing eardrum diseases using otoendoscopic videos.” *Neural Computing and Applications*.
 - Link: [Springer](#)
3. Praveen Kumar Maroju, "Optimizing Mortgage Loan Processing in Capital Markets: A Machine Learning Approach," International Journal of Innovations in Scientific Engineering, 17(1), PP. 36-55 , April 2023.
4. Lightweight Mobile-Based DL System (Ear-Keeper)
 - Reference: Zhang et al. (2023). “Ear-Keeper: Real-time lightweight DL-based ear disease detection system deployable on mobile.” *arXiv preprint*.
 - Link: [arXiv](#)
5. Vision Transformer-Based Classification
 - Reference: Pundir, A., et al. (2024). “Ear Infection Detection using Vision Transformer and SVM.” *Neural Computing and Applications*.
 - Link: [Springer](#)
6. Smartphone-Based Otitis Media Detection
 - Reference: Livingstone, D., & Chau, J. (2021). “Application of DL to classify otitis media using smartphone-based otoscope images.” *Otolaryngology–Head and Neck Surgery*.
 - Link: [PubMed](#)



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details