Otitis Net: AI-Driven Scalable Paediatric Otitis Media Diagnosis from Otoscopic Images

B. Rama Mohan[1*], Elkapalli Pavani[2], Motrige Praveen[2], Dasari Narendhar[2]
[1]Professor, [2]UG Student,[1,2]Department of Computer Science and Engineering
[1,2]Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.
*Correspondence: B. Rama Mohan (ramamohan.b@kpritech.ac.in)

ABSTRACT

Otitis media represents a group of inflammatory disorders affecting the middle ear, which can lead to serious complications if not diagnosed in a timely manner. Accurate identification of these conditions remains challenging, highlighting the need for an efficient screening system supported by machine learning techniques. This work proposes an automated diagnostic framework based on deep learning for analyzing otoscopic images, aiming to address the growing prevalence of ear infections, especially among children, where such conditions are a major cause of hearing loss and frequent clinical visits. Despite the availability of digital otoscopic imaging, diagnosis is still largely dependent on manual examination by specialists, which can be subjective, time-intensive, and susceptible to variability, delays, and human error, particularly in low-resource environments. Traditional diagnosis involves distinguishing between conditions such as Acute Otitis Media, Chronic Otitis Media, Myringosclerosis, Cerumen Impaction, and normal cases, often without standardized criteria. To improve reliability and scalability, the proposed system incorporates image preprocessing, deep feature extraction using DenseNet121, and classification through multiple machine learning models including Nearest Centroid classifier (NCC), K-Nearest Neighbors (KNN), and eXtreme Gradient Boosting (XGBoost), along with a proposed ensemble approach that combines a calibrated perceptron and a dense neural network using a voting mechanism. The dataset is split into training and testing sets, and performance is assessed using metrics such as accuracy, precision, recall, F1-score, confusion matrix, ROC curves, and AUC. The experimental findings indicate that the ensemble voting model outperforms individual classifiers, offering a robust, efficient, and scalable solution for automated otitis media diagnosis and supporting clinical decision-making.

Key words: Healthcare AI, Clinical Decision Support System, Pediatric Ear Disease Detection, Automated Diagnosis System

1. INTRODUCTION

Otitis media (OM) is a widely occurring condition among children [1], commonly associated with symptoms such as fever, disrupted sleep, and acute ear infections [2]. The impact of this condition extends beyond the affected children, causing considerable distress to caregivers as well [3]. Globally, OM shows significant prevalence, reported at 9.2% in Nigeria, 10% in Egypt, 6.7% in China, 9.2% in India, 9.1% in Iran, and between 5.1% and 7.8% in Russia [4], while native Australian children exhibit an exceptionally high rate of 90% [5]. Previous studies have explored various diagnostic and treatment strategies for OM [6]. However, incorrect diagnosis can result in serious complications, including hearing impairment, cognitive developmental issues, unnecessary surgical interventions, excessive use of antibiotics, and worsening of the condition [7]. Notably, nearly 80% of patients with OM are prescribed antibiotics, which can contribute to antibiotic resistance and increased healthcare costs [8]. Hence, achieving accurate diagnosis is crucial to minimize these risks and ensure appropriate treatment.

AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(1) (2026)

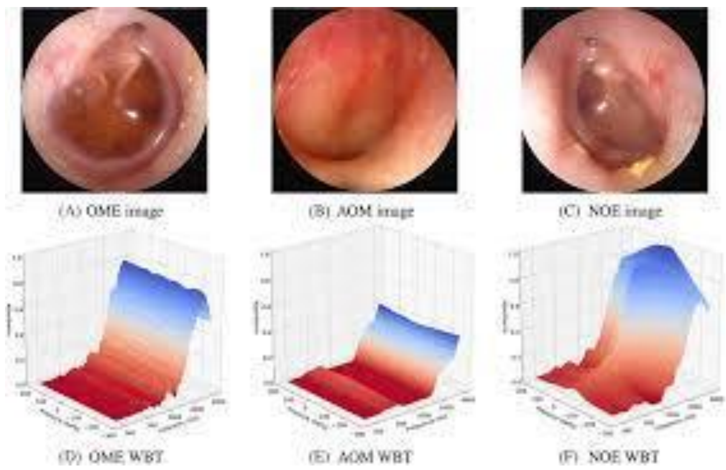


Fig. 1. Otoscopic images and WBT measurements

Diagnosing both acute and chronic middle ear infections remains a complex task. In infants, the process is particularly challenging due to narrow ear canals and the presence of cerumen, which can obstruct clear visualization during otoscopic examination [9]. Moreover, in primary healthcare and pediatric settings, diagnostic accuracy is often limited because of insufficient training and lack of familiarity with advanced techniques such as pneumatic otoscopy [10,11]. To overcome these limitations, several solutions have been proposed, including enhanced training programs for medical professionals, development of improved otoscopic tools and techniques, use of absorbance and acoustic admittance measurements, and integration of impedance-based hearing aids. Clinical studies have also compared the effectiveness of these approaches. Despite these advancements, diagnostic accuracy in primary care settings among pediatricians and otolaryngologists remains below 70%.

Otitis Media, characterized by inflammation or infection in the middle ear, is a major contributor to hearing loss across both pediatric and adult populations worldwide as shown in fig 1. It is typically identified by signs such as fluid accumulation, redness, and bulging of the tympanic membrane, observable through otoscopic imaging. However, precise diagnosis largely depends on the skill and experience of trained specialists. In rural or resource-limited regions, the shortage of such expertise often leads to delayed, inaccurate, or missed diagnoses. This creates serious public health challenges, particularly in children, where untreated OM can negatively affect speech development and learning abilities

2. LITERATURE SURVEY

Frosolini A et al. [12] integrated of AI in audiology has evolved significantly over the succeeding decades, with 87.5% of manuscripts published in the last 4 years. Most types of AI were consistently used for specific purposes, such as logistic regression and other statistical machine learning tools (e.g., support vector machine, multilayer perceptron, random forest, deep belief network, decision tree, k-nearest neighbor, or LASSO) for automated audiometry and clinical predictions; convolutional neural networks for radiological image analysis; and large language models for automatic generation of diagnostic reports. Despite the advances in AI technologies, different ethical and professional challenges are still present, underscoring the need for larger, more diverse data collection and bioethics studies in the field of audiology.

Esposito S et al. [13] proposed to improve disease detection. We showed that misdiagnosis can be dangerous or lead to relevant therapeutic mistakes. The need to improve AOM diagnosis has allowed the identification of a long list of technologies to visualize and evaluate the tympanic membrane and to assess middle-ear effusion. Most of the new instruments, including light field otoscopy, optical coherence tomography, low-coherence interferometry, and Raman spectroscopy, are far from being introduced in clinical practice. Video-otoscopy can be effective, especially when it is used in association with telemedicine, parents' cooperation, and artificial intelligence. Introduction of otologic telemedicine and use of artificial intelligence among pediatricians and ENT specialists must be strongly promoted in order to reduce mistakes in AOM diagnosis.

Binol H et al. [14] proposed framework, called SelectStitch, is useful for classifying eardrum abnormalities using a single composite image instead of the entire raw otoscope video dataset. Methods: SelectStitch consists of a convolutional neural network (CNN) based semantic segmentation approach to detect the eardrum in each frame of the otoscope video, and a stitching engine to generate a high-quality composite image from the detected eardrum regions. In this study, we utilize two separate datasets: the first one has 36 otoscope videos that were used to train a semantic segmentation model, and the second one, containing 100 videos, which was used to test the proposed method.

Byun H et al. [15] presented were to assess the performance of the Teachable machine® for diagnosing tympanic membrane lesions. A total of 3024 tympanic membrane images were used to train and validate the diagnostic performance of the network. Tympanic membrane images were labeled as normal, otitis media with effusion (OME), chronic otitis media (COM), and cholesteatoma. According to the complexity of the categorization, Level I refers to normal versus abnormal tympanic membrane; Level II was defined as normal, OME, or COM + cholesteatoma; and Level III distinguishes between all four pathologies. In addition, eighty representative test images were used to assess the performance.

Tsilivigkos C et al. [16] conducted a thorough search on MEDLINE for papers published until June 2023, utilizing the keywords ‘otorhinolaryngology’, ‘imaging’, ‘computer vision’, ‘artificial intelligence’, and ‘deep learning’, and at the same time conducted manual searching in the references section of the articles included in our manuscript. Their search culminated in the retrieval of 121 related articles, which were subsequently subdivided into the following categories: imaging in head and neck, otology, and rhinology. Thier objective is to provide a comprehensive introduction to this burgeoning field, tailored for both experienced specialists and aspiring residents in the domain of deep learning algorithms in imaging techniques in otorhinolaryngology.

Sandström J et al. [17] investigated the performance of a convolutional neural network in screening for otitis media using digital otoscopic images labelled by an expert panel. Methods: Five experienced otologists diagnosed 347 tympanic membrane images captured with a digital otoscope. Images with a majority expert diagnosis (n = 273) were categorized into three screening groups Normal, Pathological and Wax, and the same images were used for training and testing of the convolutional neural network. Expert panel diagnoses were compared to the convolutional neural network classification. Different approaches to the convolutional neural network were tested to identify the best performing model.

Song D et al. [18] integration of objective tools, like artificial intelligence (AI), could potentially improve the diagnostic process by minimizing the influence of subjective biases and variability. We systematically reviewed the AI techniques using medical imaging in otolaryngology. Relevant studies related to AI-assisted otitis media diagnosis were extracted from five databases: Google Scholar, PubMed, Medline, Embase, and IEEE Xplore, without date restrictions. Publications that did not relate to AI and otitis media diagnosis or did not utilize medical imaging were excluded. Of the 32identified studies, 26 used tympanic membrane images for classification, achieving an average diagnosis accuracy of 86% (range: 48.7–99.16%).

3. PROPOSED METHODOLOGY

The proposed system is an intelligent medical image classification framework developed to automatically diagnose Otitis Media from otoscopic images. In the initial stage, ear images are collected and systematically arranged into specific disease categories. Each image undergoes resizing and preprocessing, after which it is fed into a pretrained DenseNet121 model to obtain deep feature representations through transfer learning. These extracted features are then split into training and testing datasets to facilitate model development. The framework employs several baseline classifiers and further introduces a hybrid soft voting approach that integrates a Calibrated Perceptron with a Dense Neural Network to improve prediction accuracy and model stability. Ultimately, the trained system is deployed through a user-friendly GUI, allowing users to upload new ear images and obtain instant diagnostic predictions along with relevant performance evaluation metrics.

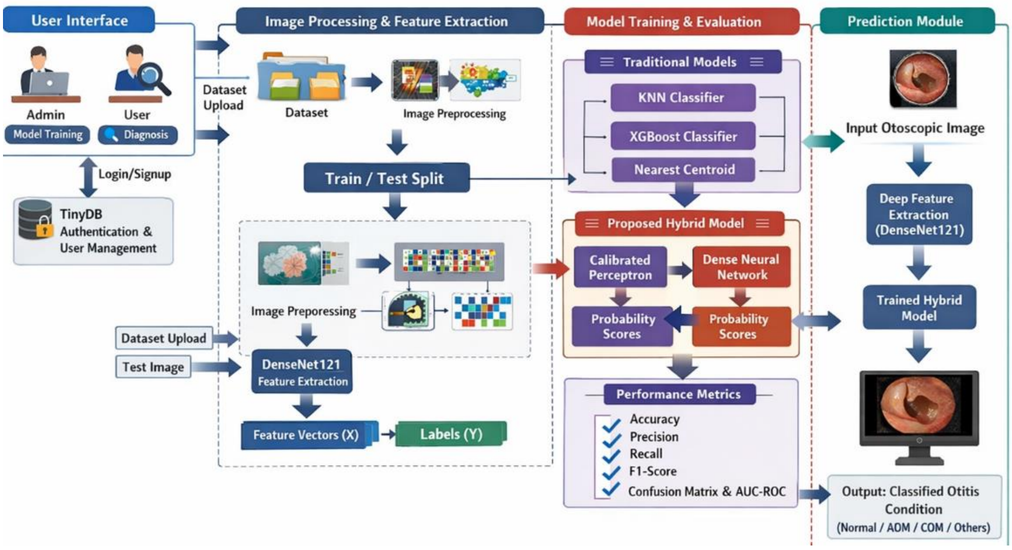


Fig. 2. Proposed System architecture of otitis diagnosis

As shown in fig 2. the system begins with collecting otoscopic ear images categorized into different disease classes such as Acute Otitis Media, Chronic Otitis Media, Cerumen Impaction, Myringosclerosis, and Normal. The dataset is organized into separate folders where each folder represents one class label. This structured format allows automatic class identification during dataset upload. Each image is resized to 128×128 pixels to maintain uniform input dimensions. The images are converted into numerical arrays and normalized using DenseNet121 preprocessing. This ensures that the pixel values are standardized and compatible with the pretrained CNN model for effective feature extraction.

A pretrained DenseNet121 model (without the top classification layer) is used to extract deep feature representations from each image. Instead of training a CNN from scratch, the system leverages transfer learning to obtain high-level discriminative features. The extracted feature vectors are stored for further classification. The extracted feature dataset is randomly shuffled and divided into training (80%) and testing (20%) sets. This ensures proper model validation and prevents overfitting. The training data is used to build models, while the testing data evaluates performance. Traditional classifiers such as Nearest Centroid, K-Nearest Neighbors (KNN), and XGBoost are trained using the extracted deep features. These models serve as baseline approaches to compare classification performance and analyze how different algorithms handle the feature space. The proposed system introduces a hybrid soft voting classifier combining a Calibrated Perceptron and a Dense Neural Network (DNN). The Perceptron captures linear decision boundaries, while the Dense Neural Network learns complex nonlinear patterns. Both models generate probability scores, which are averaged using soft voting to produce the final prediction. This improves robustness and generalization. In the final stage, a user can upload a new otoscopic image through the graphical interface. The system preprocesses the image, extracts features using DenseNet121, applies the trained hybrid model, and displays the predicted disease label along with visual output. This enables real-time automated diagnosis support.

Voting classifier with Perceptron and DNN model

A Soft Voting Hybrid Classifier is proposed to improve the accuracy and robustness of Otitis Media diagnosis. The model combines two different learning approaches: a Calibrated Perceptron (linear classifier) and a DNN (non-linear deep model). Instead of relying on a single classifier, the system averages the probability outputs of both models to make the final decision. This soft voting mechanism enhances generalization, reduces overfitting, and balances the strengths of linear and deep learning

AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(1) (2026)

approaches, resulting in improved classification performance across multiple disease categories as illustrated in fig 3.

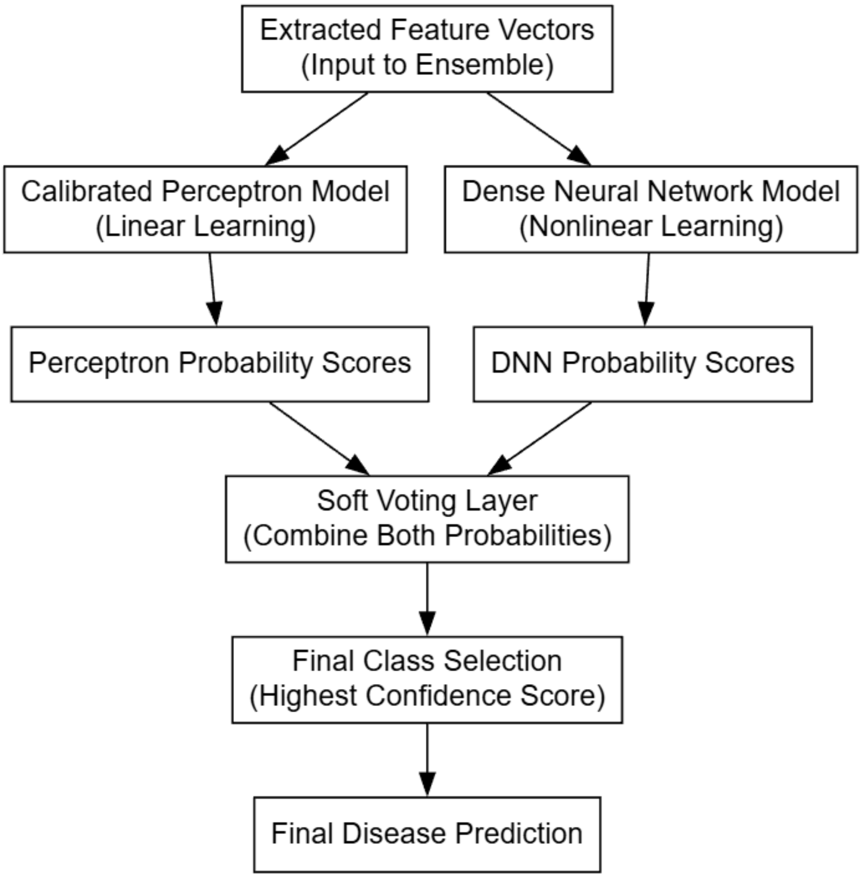


Fig 3. Internal workflow of Proposed Voting classifier

Initially, otoscopic images are passed through the pretrained DenseNet121 model to extract high-level deep features. These features represent important patterns such as texture, color variation, and structural details of the ear images. The extracted feature vectors become the input dataset for the classification models. The same extracted feature dataset is provided to both the Calibrated Perceptron and the Dense Neural Network. This ensures that both models learn from identical information while applying different learning strategies.

The Perceptron model is trained using the training dataset to learn linear decision boundaries between disease categories. Since a normal Perceptron does not generate probability outputs, it is calibrated to produce confidence scores for each class. This makes it suitable for soft voting. Simultaneously, a Dense Neural Network is trained using the same feature vectors. The network consists of multiple fully connected layers with activation functions and dropout regularization. This model captures complex and nonlinear relationships present in the extracted deep features. During testing, both trained models make predictions independently. Instead of producing only class labels, both models generate probability scores that indicate their confidence level for each disease category. After combining the probability outputs, the system selects the disease category with the highest overall confidence score. This becomes the final predicted diagnosis for the input image..

4. Result Discussion

Fig 4. depicts the GUI of the Automated Otitis Media Diagnosis system. On the right side, there is a vertical panel of buttons allowing users to perform key functions, including uploading the dataset, performing DenseNet121 feature extraction, train-test splitting, training various models like Perceptron, NearestCentroid, KNN, and uploading a test image for prediction. The left side of the interface contains a scrollable text box that displays messages and system feedback; in this screenshot, it shows that the dataset has been successfully loaded with five classes: *Acute Otitis Media, Cerumen Impaction, Chronic Otitis Media, Myringosclerosis,* and *Normal*. The background image shows a clinical setting, reinforcing the medical context of the application, while the layout combines both functionality and informative status updates in a user-friendly manner.

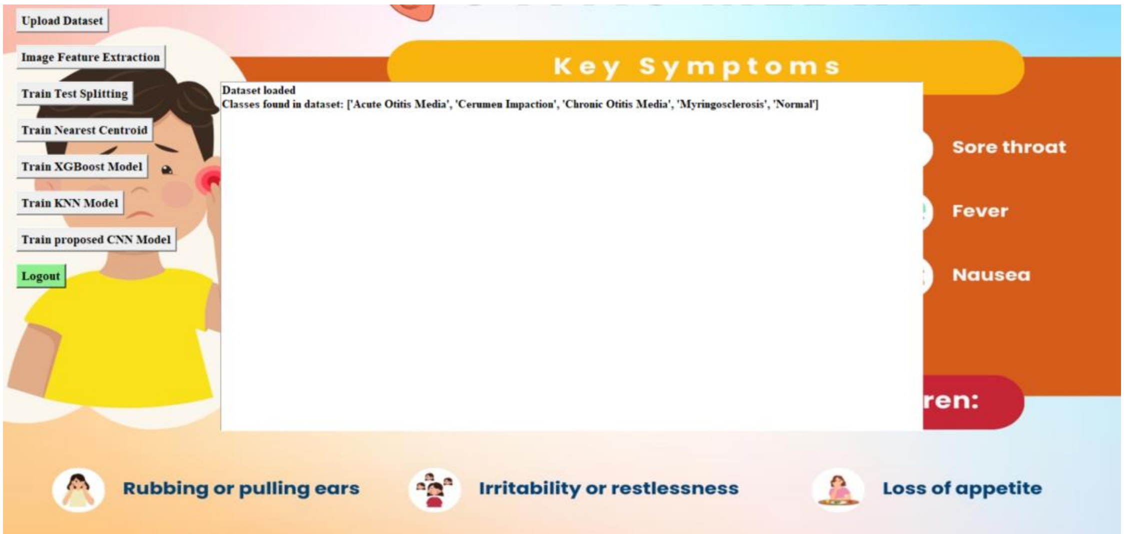


Fig 4. GUI representation and dataset uploading of Otitis media diagnosis after admin login

Fig 5 shows confusion matrix of the proposed Voting Classifier model demonstrates excellent classification performance across all five otitis categories. Most predictions lie perfectly on the diagonal, indicating correct classification for nearly all samples. Acute Otitis Media (107), Cerumen Impaction (120), and Normal (118) are classified with 100% accuracy. Chronic Otitis Media shows 133 correct predictions with only 1 misclassified as Myringosclerosis, while Myringosclerosis has 120 correct predictions with just 1 misclassified as Normal. The extremely low number of off-diagonal errors confirms that the proposed ensemble model effectively distinguishes between different ear conditions. The model achieves near-perfect accuracy with minimal misclassification, demonstrating strong reliability and superior performance compared to individual models..

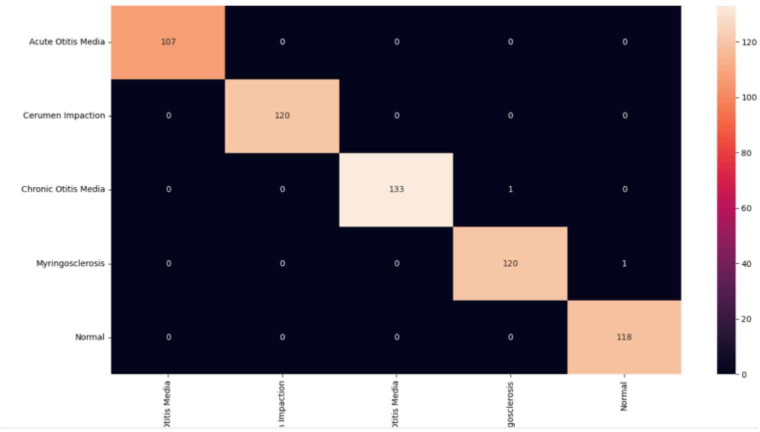


Fig 5. Confusion matrix obtained for Proposed KNN with DenseNet121 Model

Fig 6. shows ROC curve of the proposed Voting Classifier model shows an almost perfect classification performance across all five classes—Acute Otitis Media, Cerumen Impaction, Chronic Otitis Media, Myringosclerosis, and Normal. The ROC curves rise sharply toward the top-left corner and remain very close to the ideal boundary, indicating a True Positive Rate near 1.0 with an extremely low False Positive Rate. This suggests that the model has excellent discriminative capability and achieves an Area Under the Curve (AUC) value close to 1.0 for all classes. Compared to the individual models (such as KNN and XGBoost, which previously showed random performance), the proposed voting ensemble significantly improves prediction reliability and demonstrates superior multiclass otitis diagnosis performance.

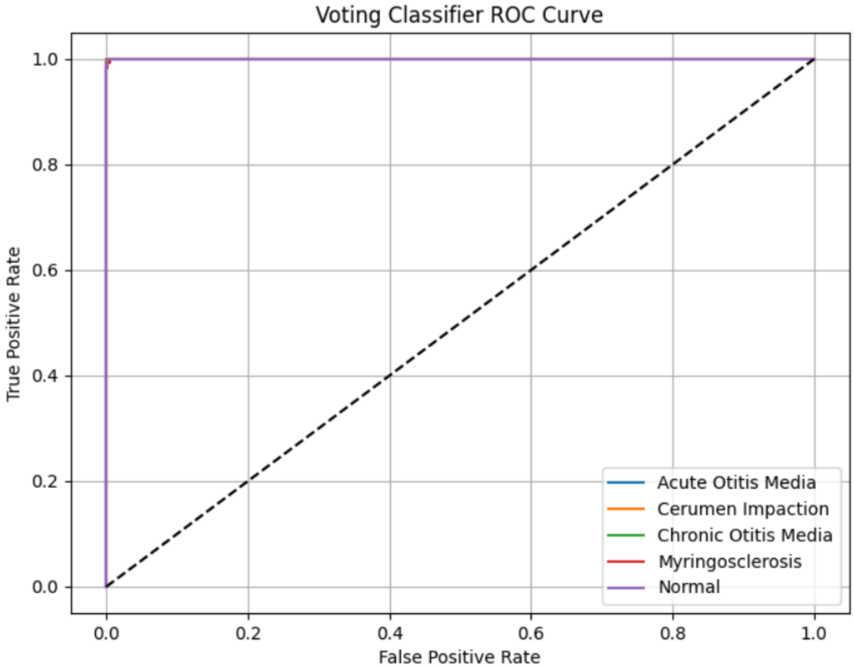
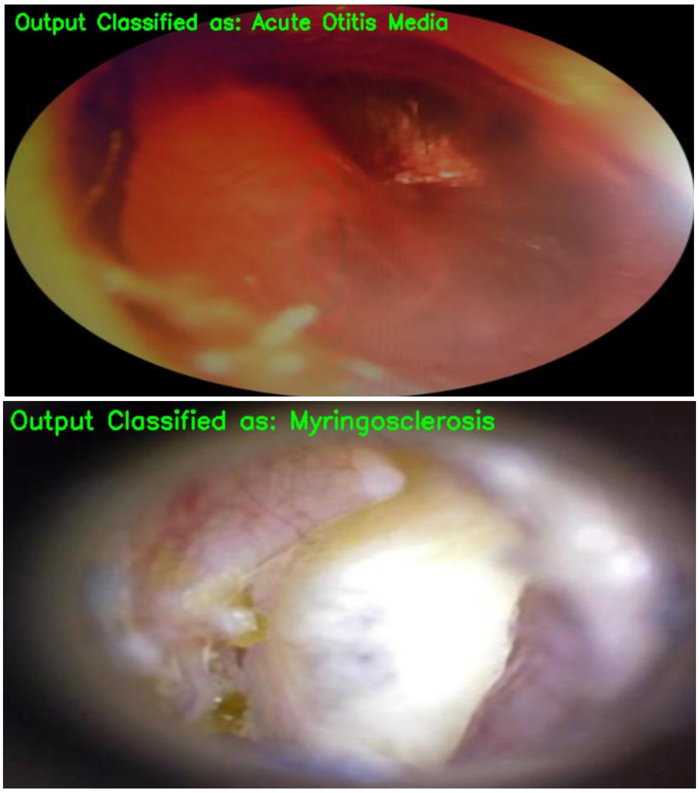


Fig 6. ROC Curve obtained for Voting Classifier

Fig 7. displays the prediction output for a test otoscopic image classified by the Proposed voting classifier. The image itself shows a tympanic membrane that is intensely red and inflamed, consistent with clinical signs of acute infection. The green text overlay at the top left confirms the model's diagnosis: "Output Classified as: Acute Otitis Media," demonstrating that the features extracted by DenseNet121 and processed by the highly accurate voting model successfully identified the characteristic pathological features of this condition.



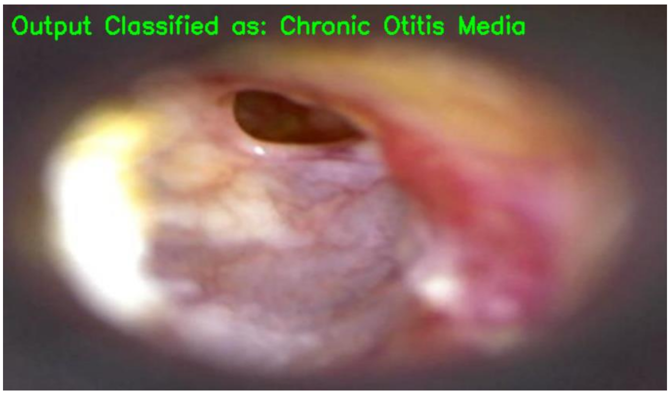


Fig 7. Prediction on Test image obtained using Voting classifier

The performance comparison presented in Table 1 clearly demonstrates significant differences among the evaluated classification models: Nearest Centroid, XGBoost, KNN, and the proposed Voting Model. The Nearest Centroid classifier achieved moderate and consistent performance, obtaining an accuracy of 82.33% along with precision, recall, and F-score values around 82%, indicating that it was able to reasonably separate the classes but lacked the capability to capture more complex patterns in the feature space. In contrast, both the XGBoost and KNN classifiers performed very poorly, each producing only 17.83% accuracy with extremely low precision (3.57%), recall (20.00%), and F-score (6.05%). This suggests that these models were unable to effectively learn discriminative patterns from the extracted features, possibly due to data distribution complexity or feature characteristics that did not suit their learning behavior. The proposed Voting Model, which combines a Calibrated Perceptron and a Dense Neural Network, significantly outperformed all baseline approaches by achieving an accuracy of 99.67%, precision of 99.67%, recall of 99.69%, and F-score of 99.68%. The near-perfect and balanced metric values indicate that the ensemble approach provided highly reliable classification across all classes, minimizing both false positives and false negatives. Overall, the table confirms that integrating linear and nonlinear learning through a voting-based ensemble greatly enhances classification stability and predictive performance compared to individual machine learning models.

Table 1: Performance comparison for the Nearest Centroid, XGBoost, KNN and proposed Voting models

Algorithms Name	Accuracy	Precision	Recall	F-score
NCC	82.33%	82.68%	82.60%	82.63%
XGBoost Classifier	17.83%	3.57%	20.00%	6.05%
KNN classifier	17.83%	3.57%	20.00%	6.05%
Voting Model	99.67%	99.6 7%	99.69%	99.68%

5. CONCLUSION

The research introduces a comprehensive and intelligent framework for the automated detection of Otitis Media, combining deep learning with classical machine learning and ensemble strategies to achieve reliable medical image classification. Instead of building a model from scratch, a pretrained DenseNet121 architecture is utilized to extract meaningful deep features from otoscopic images, enabling the system to efficiently learn complex patterns present in the data. These features serve as inputs for multiple classification models, where traditional algorithms such as NCC, KNN, and XGBoost are applied to establish baseline performance comparisons. To further enhance predictive capability, a novel ensemble approach is designed by integrating a Calibrated Perceptron with a Dense Neural Network through a soft voting mechanism, allowing the system to benefit from both linear and nonlinear decision-making strengths. This combined strategy not only improves accuracy but also ensures more stable and generalized predictions across various ear disease classes. The performance of the system is rigorously assessed using standard evaluation measures including accuracy, precision, recall, F1-score, confusion matrix, ROC analysis, and AUC values, demonstrating its effectiveness. Additionally, the system is implemented with a user-oriented graphical interface developed using Tkinter, along with secure role-based access managed through TinyDB, making it practical for real-world usage. The proposed solution offers a flexible, scalable, and efficient tool for supporting early diagnosis of ear conditions, thereby contributing to smarter and more accessible healthcare systems.

REFERENCES

[1] Boruk, M.; Paul, L.; Yelena, F.; Rosenfeld, R.M. Caregiver well-being and child quality of life. Otolaryngol. Neck Surg. 2007, 136, 159–168.

[2] Tran, T.T.; Fang, T.Y.; Pham, V.T.; Lin, C.; Wang, P.C.; Lo, M.T. Development of an automatic diagnostic algorithm for pediatric otitis media. Otol. Neurotol. 2018, 39, 1060–1065.

[3] Berman, S. Otitis media in children. N. Engl. J. Med. 1995, 332, 1560–1565.

[4] DeAntonio, R.; Yarzabal, J.P.; Cruz, J.P.; Schmidt, J.E.; Kleijnen, J. Epidemiology of otitis media in children from developing countries: A systematic review. Int. J. Pediatr. Otorhinolaryngol. 2016, 85, 65–74.

[5] Kenyon, G. Social otitis media: Ear infection and disparity in Australia. Lancet Infect. Dis. 2017, 17, 375–376.

[6] Vanneste, P.; Page, C. Otitis media with effusion in children: Pathophysiology, diagnosis, and treatment. A review. J. Otol. 2019, 14, 33–39.

[7] Crowson, M.G.; Hartnick, C.J.; Diercks, G.R.; Gallagher, T.Q.; Fracchia, M.S.; Setlur, J.; Cohen, M.S. Machine learning for accurate intraoperative pediatric middle ear effusion diagnosis. Pediatrics 2021, 147, e2020034546.

[8] Wu, Z.; Lin, Z.; Li, L.; Pan, H.; Chen, G.; Fu, Y.; Qiu, Q. Deep learning for classification of pediatric otitis media. Laryngoscope 2021, 131, E2344–E2351.

[9] Granath, A. Recurrent acute otitis media: What are the options for treatment and prevention? Curr. Otorhinolaryngol. Rep. 2017, 5, 93–100.

[10] Blomgren, K.; Pitkäranta, A. Is it possible to diagnose acute otitis media accurately in primary health care? Fam. Pract. 2003, 20, 524–527.

[11] Pichichero, M.E.; Poole, M.D. Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. Arch. Pediatr. Adolesc. Med. 2001, 155, 1137–1142.

[12] Frosolini A, Franz L, Caragli V, Genovese E, de Filippis C, Marioni G. Artificial Intelligence in Audiology: A Scoping Review of Current Applications and Future Directions. *Sensors*. 2024; 24(22):7126.

[13] Esposito S, Bianchini S, Argentiero A, Gobbi R, Vicini C, Principi N. New Approaches and Technologies to Improve Accuracy of Acute Otitis Media Diagnosis. *Diagnostics*. 2021; 11(12):2392.

[14] Binol H, Moberly AC, Niazi MKK, Essig G, Shah J, Elmaraghy C, Teknos T, Taj-Schaal N, Yu L, Gurcan MN. SelectStitch: Automated Frame Segmentation and Stitching to Create Composite Images from Otoscope Video Clips. *Applied Sciences*. 2020; 10(17):5894.

[15] Byun H, Lee SH, Kim TH, Oh J, Chung JH. Feasibility of the Machine Learning Network to Diagnose Tympanic Membrane Lesions without Coding Experience. J Pers Med. 2022 Nov 7;12(11):1855. doi: 10.3390/jpm12111855. PMID: 36579584; PMCID:

[16] Tsilivigkos C, Athanasopoulos M, Micco Rd, Giotakis A, Mastronikolis NS, Mulita F, Verras G-I, Maroulis I, Giotakis E. Deep Learning Techniques and Imaging in Otorhinolaryngology—A State-of-the-Art Review. *Journal of Clinical Medicine*. 2023; 12(22):6973.

[17] Sandström J, Myburgh H, Laurent C, Swanepoel DW, Lundberg T. A Machine Learning Approach to Screen for Otitis Media Using Digital Otoscope Images Labelled by an Expert Panel. *Diagnostics*. 2022; 12(6):1318.

[18] Song D, Kim T, Lee Y, Kim J. Image-Based Artificial Intelligence Technology for Diagnosing Middle Ear Diseases: A Systematic Review. *Journal of Clinical Medicine*. 2023; 12(18):5831.