IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 June 2024, accepted 11 July 2024, date of publication 15 July 2024, date of current version 26 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3428700

SURVEY

Artificial Intelligence-Driven Advancements in Otitis Media Diagnosis: A Systematic Review

MD. AWLAD HOSSEN RONY[1], KANIZ FATEMA[1], MOHAIMENUL AZAM KHAN RAIAAN[2], MD. MEHEDI HASSAN[3], (Member, IEEE), SAMI AZAM[4], (Member, IEEE), ASIF KARIM[4], (Member, IEEE), MIRJAM JONKMAN[4], (Member, IEEE), JEMIMA BEISSBARTH[5], FRISO DE BOER[4], SHEIKH MOHAMMED SHARIFUL ISLAM[6], AND AMANDA LEACH[5]

[1]Health Informatics Research Laboratory (HIRL), Department of Computer Science and Engineering, Daffodil International University, Dhaka 1341, Bangladesh
[2]Department of Computer Science and Engineering, United International University, Dhaka 1212, Bangladesh
[3]Computer Science and Engineering Discipline, Khulna University, Khulna 9208, Bangladesh
[4]Faculty of Science and Technology, Charles Darwin University, Casuarina, NT 0909, Australia
[5]Child Health Division, Menzies School of Health Research, Casuarina, NT 0909, Australia
[6]Institute for Physical Activity and Nutrition, Deakin University, Melbourne, VIC 3125, Australia

Corresponding authors: Sami Azam (sami.azam@cdu.edu.au) and Amanda Leach (amanda.leach@menzies.edu.au)

ABSTRACT Otitis Media (OM), predominantly affecting children, is a significant global health issue, with an estimated 360 million pediatric cases yearly worldwide. OM causes mild and moderate conductive hearing loss which can be disabling for young children, particularly during the first three years of life when brain growth is rapid, resulting in poor speech and language development, poor communication skills, and increased vulnerability on entering school. OM therefore contributes to the global burden of all-cause hearing loss. This systematic review seeks to provide a comprehensive evaluation of pre-trained Artificial Intelligence (AI) models, including both classical Machine Learning (ML) and Deep Learning (DL), in the context of OM. This review proposes six research questions, and it summarizes the body of research across multiple domains, including the diversity and quantity of source material for training and testing models, including otoscopy images, videos, and tympanometry, and the methods used to assess quality and effectiveness in real-time settings. In addition, the review aims to provide insight into the impact and potential of AI in improving OM diagnosis and cast light on the existing challenges, such as model interpretability, limited medical expert involvement, and the need for knowledge discovery and unanswered questions, including the evolving landscape of OM diagnosis within this domain. The findings of this systematic review emphasize the importance of developing more interpretable AI models that incorporate both still images of the tympanic membrane and video recordings (with multiple frames) to maximize sensitivity and specificity of the model. In addition, collaboration with consumers and medical professionals in multiple specialties (general practice, pediatricians, audiologists and ear, nose, throat (ENT) surgeons) is needed to ensure applicability and confidence of these diagnostic digital support systems in real-world healthcare settings.

INDEX TERMS Otitis media, AI, datasets, otoscopy image, ensemble model, segmentation, tympanometry.

ABBREVIATION

AI	Artificial Intelligence.
AOM	acute otitis media.
AOMwoP	acute otitis media without perforation.
AUC	area under the curve.
AM	Augmentation.
AUROC	Average area Under the Receiver Operating characteristics Curve.
CBAM	channel and spatial model.
COM	chronic otitis media.
CSOM	Chronic Suppurative OM.

The associate editor coordinating the review of this manuscript and approving it for publication was Asadullah Shaikh.

CAM	class activation mapping.
CCV	color coherence vector.
CAD	computer aided diagnosis.
CV	Computer vision.
CNN	Convolutional Neural Networks.
DL	Deep Learning.
DSC	dice similarity coefficient.
DCT	discrete cosine transform.
ENT	ear, nose, throat.
EC	Exclusion Criteria.
FNN	Feedforward Neural networks.
FCN	fully convolutional neural network.
GBM	Gradient Boosting Machines.
GCM	grid color moment.
HD	Hausdorff distance.
HOG	histograms of oriented gradient.
IP	Image Processing.
IC	Inclusion Criteria.
KNN	K-Nearest Neighbors.
LBP	local binary pattern.
LSTM	Long Short-Term Memory Networks.
ML	Machine Learning.
MAD	mean absolute distance.
MED	Middle Ear Disorders.
NLP	natural language processing.
NOE	no effusion.
OME	OM with Effusion.
OM	Otitis Media.
PPV	positive predictive value.
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
PCA	principal component analysis.
QE	quality evaluation.
RAOM	recurrent acute otitis media.
RNN	Recurrent Neural Networks.
ROI	region of interest.
RQ	Research Question.
SQ	Search Query.
Sen	sensitivity.
SEO	smartphone-enabled otoscope.
Spe	specificity.
SVM	Support Vector Machines.
TL	Transfer learning.
TM	tympanic membrane.

I. INTRODUCTION

Otitis Media (OM) is defined as any infection or inflammation in the middle ear, behind the tympanic membrane (TM) [1]. It is one of the most prevalent conditions in the world [2]. It has been estimated that 80% of children suffer from OM by the age of four years [3]. It is most common in the first year of a child's life and becomes less common in subsequent years [4]. IN Australia, Aboriginal children have a higher prevalence of middle ear infections than non-Aboriginal children, and remote areas have a greater prevalence than rural or urban areas [5]. OM is also the leading cause of conductive learning loss in adolescents [6]. Depending on its severity and duration, OM may cause transient or permanent sensorineural hearing loss. Moreover, hearing loss due to OM may result in an imbalance between sound perception of the two ears [7].

Children's brains develop rapidly during the first three years, reaching 90 percent of their adult size by age three. The growth of various brain regions is dependent on stimulation and the auditory stimuli that increase activity in the auditory cortex create the conditions necessary for development of sound processing [8].

A child who suffers from OM for an extended period during these critical years of accelerated brain development is therefore at risk of developing auditory and speech processing disorders, which may impair cognitive development [5]. OM diagnoses reflect a continuum of fluctuating states, mainly acute otitis media without perforation (AOMwoP), OM with Effusion (OME) and Chronic Suppurative OM (CSOM). AOMwoP, the most common type of OM in very young children, often occurs after a cold or the flu and is due to nasopharyngeal bacteria gaining access via the Eustachian tube to the middle ear cavity [9], [10]. It frequently occurs in children under the age of two years [11]. AOMwoP is characterized by the presence of fluid in the middle ear, bulging of the TM with possible erythema, fever and irritability. Bulging results in reduced TM mobility, reduced transmission of sound waves through the middle ear space to the inner ear and brain, [12]. It can be categorized as mild, moderate, or severe, depending on the characteristics of the TM and clinical symptoms [13].

The second most prevalent type in young children, and the most common in older children, is OME, which is characterized by accumulation of fluid in the middle ear in the absence of acute otitis media (AOM) symptoms [14]. Early symptoms of OME can be revealed by the otoscopic examination of the TM, including mucoid deposition ('glue'), the existence of an air-liquid layer, bubbles, or loss of TM light reflex and transparency [15], [16]. The least common and most severe type of OM in children, but common type in adults, is Chronic Supportive OM (CSOM) [17]. This is characterized by discharge of pus through a perforation of the TM of at least 2 weeks duration, and size of at least 2% of the pars tensa. CSOM can cause long-term harm such as degradation of the ossicles in the middle ear. CSOM is preceded by acute otitis media with perforation of the TM, AOMwiP, which is defined also as discharge through a perforation of the TM but is of duration less than 2 weeks and a perforation size less than 2%. Where the discharge resolves without repair of the TM, the condition is described as a dry perforation. Conditions of the ear canal such as foreign bodies (small beads, insects) or infection and swelling of the canal wall (otitis externa) are not related to the middle ear and do not cause conductive hearing loss [18], [19]. AOM is one of the most common reasons for the prescription of antibiotics in children [20], yet controlled trials have shown that it is generally self-limiting and can be managed with pain relief. OME affects more people than AOM. More than 360 million cases of OM in children are found around the world yearly [21]. Additionally, it is estimated that over 1.5 billion individuals, or roughly 20% of the global population, suffer from hearing loss and around 2.5 billion people are predicted to have learning loss by 2050 [22].

Tympanometry measures TM compliance, the mobility of the TM under positive or negative pneumatic pressure. When otoscopy identifies a TM as non-bulging (not AOM) but dull and non-translucent, it can be difficult to distinguish no OM from OME using otoscopy (appearance) alone [23]. More commonly, the eardrum is inspected using an otoscope to diagnose the type of otitis media (OM). However, it should be noted that these otoscopes are not digital, and therefore, images cannot be recorded or stored [24]. Increasingly, video otoscopes are being used to enable the capture and review of quality video frames to confirm the diagnosis of OM, discuss with families and patients, and compare changes over time [25]. However, inspecting a large number of video frames is time-consuming and storage is likely problematic. A comparison of video otoscopy diagnoses of AOM and OME by pediatricians and otolaryngologists revealed that the pediatricians made misdiagnoses in around 50% of all cases, whereas the otolaryngologist had a misdiagnosis rate of 27%. The average diagnostic accuracy for AOM utilizing video otoscopy is only 51%. Consequently, a novel, accurate and automated diagnostic method is required to enhance diagnosis confidence to effectively screen patients with otologic diseases based on aberrant otoscopic findings. Artificial intelligence (AI) based tools, notably diagnostic screening systems to assist clinician decision making, have influenced and enhanced traditional healthcare delivery. Although the diagnosis of OM is crucial in order to provide appropriate treatment, it has not been studied as extensively as other diseases [26], [27], [28].

Some of the studies [29], [30], [31], [32], [33] reviewed the advancement of AI in the OM diagnosis, although there are certain limitations. How often the researchers provide explainability and clinical validation has not been thoroughly discussed in state-of-the-art review papers. By addressing these limitations, this paper's primary objective is to evaluate the current literature of AI-related studies that aim to employ state-of-the-art models (classical machine learning and deep learning) to assess, predict, classify, or otherwise enhance the diagnosis of OM based on multiple diagnostic modalities (otoscopy images, videos, and tympanometry).

The main contributions of this systematic review are:

- Providing a comprehensive analysis of different pre-trained AI models, including ML and DL, pertaining to the diagnosis of OM.
- Exploring and reviewing the existing studies relevant to otoscopy images, videos, and tympanometry, and providing an overview of existing public and private OM datasets along with their limitations.
- Providing insight into the prospects of AI and its impact on diagnosis, as well as describing the open questions, challenges, and potential future developments.

The outline of this review is illustrated in Figure 1.

II. MEDICAL IMAGES OF OM

Medical images contain intricate details that are important for accurate diagnosis. Healthcare professionals rely on specific features within images to identify diseases, abnormalities, and other conditions. Understanding these characteristics helps to train AI models to recognize these patterns effectively, improving diagnostic accuracy. This is why it is important to examine the domain images from a medical perspective before creating an automated diagnostic system [34]. It helps us understand the markers that experts use for diagnosis, and this medical knowledge can be integrated with AI techniques to find the regions of interest (ROIs) of the images. For example, in TM images, there are several relevant medical features, including visibility of the malleus bone, the position and color of the TM, perforations, obstructive ear wax, middle ear fluid, and reflected light. The TM position can be neutral, retracted, or bulging (a convex tympanic membrane with a near lack of malleus bone lateral process and manubrium visibility). Middle ear fluid, the malleus bone, a cone of light (reflection of the otoscope light), and perforation may or may not be visible from the TM [35]. Since OM has several classes, the imaging features are varied. Specialists distinguish the OM classes based on these predefined features. Figure 2 illustrates the different classes. The images have been collected from reliable sources including Hawke library (provided by Dr Michael Hawke), Özel Van Akdamar hospital, and child patients during OM examination [36], [37], [38].

III. ABOUT AI TECHNIQUES

A. HARNESSING CLASSICAL MACHINE LEARNING AND DEEP LEARNING APPROACHES

In a broad sense, Artificial Intelligence refers to a machine's capacity to simulate human intelligence. To let computers perform tasks without explicit programming, classical machine learning (ML) and Deep Learning both come under the umbrella of AI [39], [40]. Mathematical model fitting is a part of the learning process. Relationships between features (or variables) are mathematically examined via multiple data points, resulting in a best-fitting model. At a basic level, ML involves providing computers with data from which to learn and make proper decisions or predictions. AI models are often used for dealing with complicated problems or uncovering hidden patterns [27]. The performance of a model depends on the quality and the quantity of the training dataset and the appropriateness of the selected algorithms. Adoption of electronic health record systems makes it easier to gather clinical data about patients. Imaging data, including otoscopy images, drug usage records, surgical and pathology records are among the types of information used. The

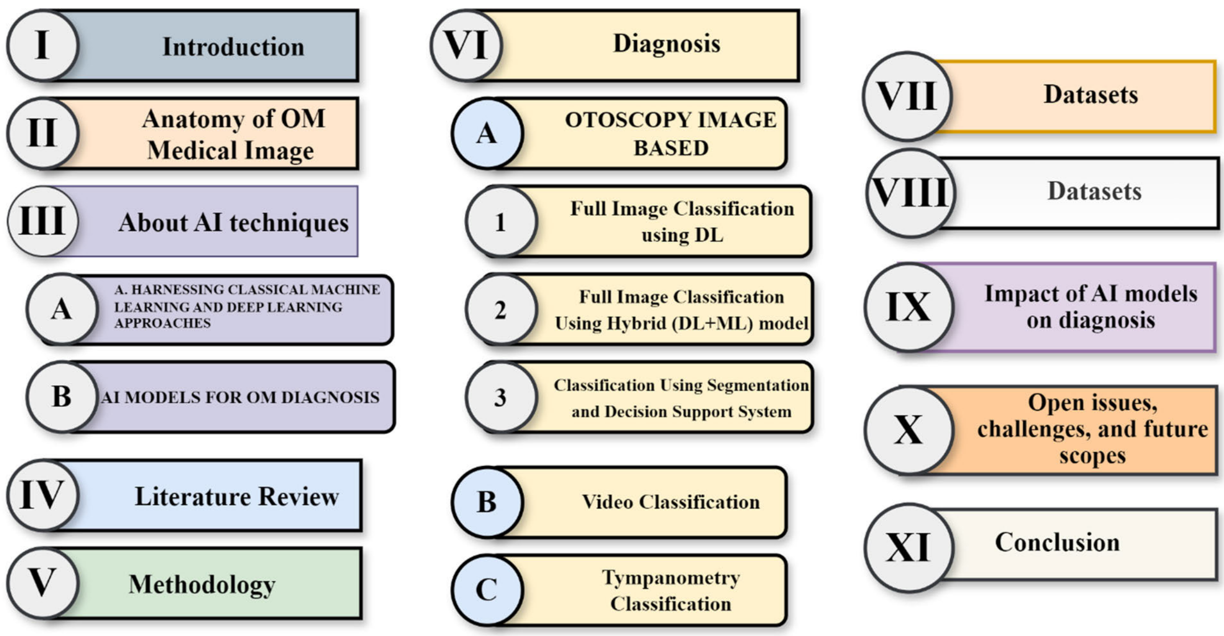

FIGURE 1. Section organization of the research.

healthcare business collects data from large-scale genetic research and consumer sources, including wearables and cellphones, which provide physiological data. In contrast to traditional hypothesis-based statistical approaches, ML algorithms may be fitted with enormous quantities of data for training purposes. ML algorithms can capture more complicated, and nonlinear connections between features and outcomes than traditional approaches [41], [42].

ML can also manage massive volumes of data and incorporate insights from various data types. DL [43], which is an extension of classical ML, works with even more complicated mathematical models than classical ML algorithms and is more detailed. It is a data processing approach that employs many processing layers with various nonlinear transformations [44]. Classical machine learning techniques, such as Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Gradient Boosting Machines (GBM), [45] have demonstrated robustness in classifying medical conditions like otitis media. Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the most frequently appearing class. It can effectively handle complex datasets with various features, making it suitable for medical diagnosis tasks [46]. SVM is a powerful classification algorithm that discovers the best hyperplane to separate classes in feature space. It is particularly effective in high-dimensional spaces, making it suitable for medical datasets with many features. By utilizing different kernel functions, SVM can handle both linear and non-linear data, aiding in the diagnosis of otitis media [47]. KNN is a simple and intuitive algorithm that classifies a data point based on the majority class of its k nearest neighbors. In diagnosing otitis media, KNN can effectively identify patterns based on similar cases in the dataset. It does not assume any underlying data distribution and is suitable for datasets with irregular or difficult-to-define decision boundaries [48]. GBM is an ensemble learning technique that sequentially combines multiple weak classifiers to build a strong classifier. It progressively improves the model's predictive performance by adding predictors that correct the mistakes of their predecessors. GBM is particularly effective when the data is imbalanced or when certain features have a higher impact on the classification task, which could be the case in diagnosing otitis media [49]. These models can accurately diagnose otitis media when trained on relevant datasets containing features related to patient symptoms, medical history, and potentially medical imaging results. However, it is crucial to ensure high-quality data and thorough validation to guarantee reliability and effectiveness in real-world scenarios. Additionally, interpreting the model outputs should always be done in consultation with medical professionals to ensure accurate diagnosis and treatment. Deep Learning models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory Networks (LSTM), excel in extracting intricate patterns from diverse data types such as medical images, time-series data, and natural language. CNNs are particularly notable for their exceptional performance in image recognition, while RNNs and LSTMs exhibit remarkable capability in sequential data analysis, making them highly valuable for tasks such as speech recognition and language translation. By integrating these models into healthcare systems, practitioners can leverage their capabilities to derive actionable insights, facilitate

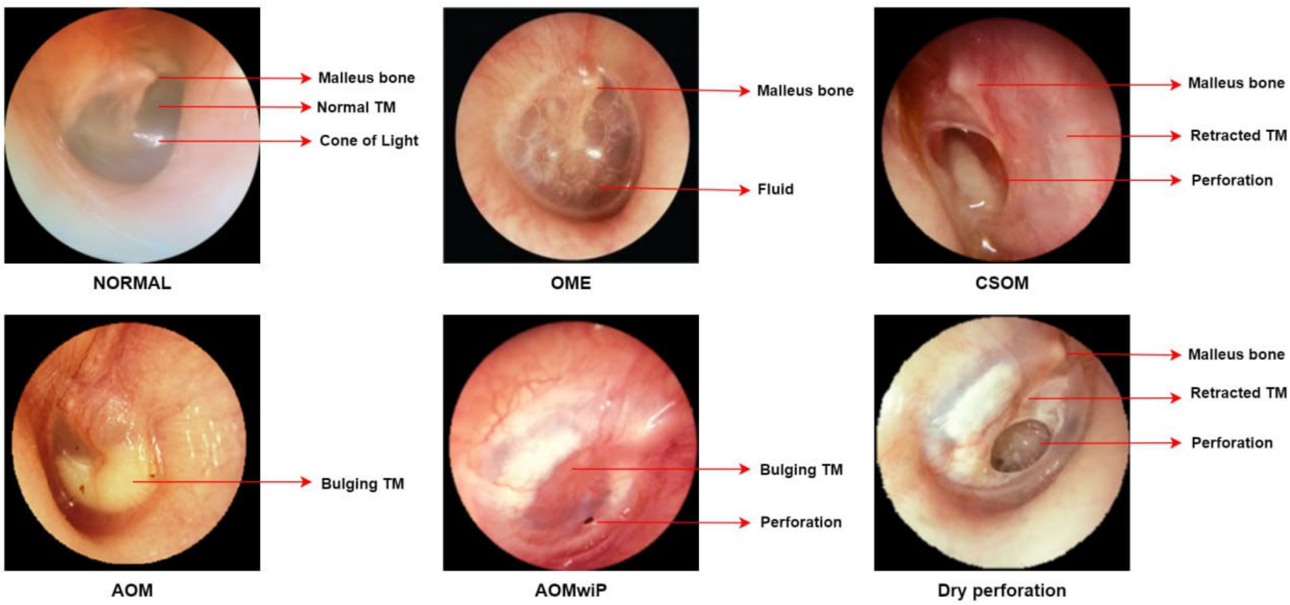

FIGURE 2. **Medical features corresponding to the OM classes.**

diagnosis, and enhance patient care. In recent years, it has produced significant advances in computer vision, speech recognition, natural language processing (NLP), bioinformatics and other fields [50]. DL tries to replicate the human brain network. The original data is abstracted layer by layer using nonlinear processing layers. Abstract characteristics are extracted from the data and applied to the unseen data for target identification, classification, or segmentation. This has the benefit of replacing manual feature selection with effective hierarchical feature extraction techniques and unsupervised or semi-supervised feature learning [51]. Despite the vast volume of available medical data, there are still several concerns, including the diversity of the type of data (maps, texts, films, etc.) and differences in data quality due to the equipment utilized [52]. Medical images form an important component of medical data which can be used with classical ML and DL. There are several types of ML and DL algorithms [53]. Tables 1 and 2 give an overview of the algorithms that have been used most extensively. Figure 3 visualizes a tree-based categorization of ML and DL algorithms used for OM detection propositions.

B. AI MODELS FOR OM DIAGNOSIS

In recent years, image classification using deep learning and its applications has been a major topic of research and development. The CNN model, in particular has become one of the most popular applications of DL for the analysis of medical images [54]. CNN models can be employed in automated tasks such as classification, detection, segmentation, and data augmentation [59]. They may also be trained with OM images to classify the different OM categories. The model is usually trained with a large dataset, and a Deep CNN model can be generated by integrating multiple CNN models. However, for the sake of simplicity, we will discuss a shallow CNN model with three convolutional layers, each followed by a max pool layer, in this section. The input image size is $224 \times 224 \times 3$. All convolutional layers are configured with 8 kernels of size 3×3, and softmax has been used as the activation function. The kernel size for all max pool layers is 3×3. When a kernel is applied to an image, the filter moves along the pixels of the image and outputs a two-dimensional array. Each kernel is intended to detect some valuable image features, but sometimes the value of these two-dimensional arrays is '0', indicating that no image features have been identified for that kernel. These can be considered as black frames in the two-dimensional array. Figure 4 depicts the architecture of the shallow CNN model for OM classification showing several black frames corresponding to the convolutional layers. In general, if the kernel is built to extract features such as horizontal lines, vertical lines, edges, ridges, corners, and more, and the image contains no vertical lines etc., the result would be that no feature is recognized, resulting in a black frame [60]. The convolution layer is comprised of a multitude of fixed-size elements that apply complex functions to the input image to extract relevant features. A sequence of convolutional, nonlinear, and subsampling stages is used to recover the high-level features [61]. The following convolutional layer receives the feature maps from the previous one [62]. The feature maps of the previous layer are used to build the feature maps of the next layer and the procedure is repeated until the deep-level characteristics are achieved. Max pool is a down sampling or pooling technique that reduces the height and width of feature maps while preserving crucial information. In addition, a dropout layer aids in keeping the deep CNN model from overfitting. Fully connected layers are used

TABLE 1. Different categories of ML algorithms.

Type of ML Algorithm	Definition
Supervised learning [41]	The model utilizes labeled data to make predictions. It is typically utilized for classification and regression. • Linear Regression • Naive Bayes • Decision Trees • Gradient Boosting Machine • Logistic Regression • Gradient Boosting • Random Forest • AdaBoost • Support Vector Machines • XGBoost • K-Nearest Neighbors • Light Gradient Boosting Machine
Unsupervised learning [45]	The training data is unlabeled, and the purpose is to discover patterns within the training data. Its primary use is to address clustering and dimension reduction problems. • K-Means Clustering • Isolation Forest • Hierarchical Clustering • Spectral Clustering • Principal Component Analysis • Mean Shift Clustering • Density-Based Spatial Clustering of Applications with Noise • t-Distributed Stochastic Neighbor
Reinforcement learning [43]	The training model discovers the environment-to-action response that maximizes cumulative returns through exploration and trial and error.
Ensemble learning [54]	Generating several models and integrating them using different approaches. This is mostly used to enhance model performance and minimize the possibility of incorrect model selection. **Bagging:** A form of ensemble learning which relies on the Decision Trees. Each decision tree can provide distinct and unrelated results. Ultimately, the decision is established by a majority-vote rule. Each decision tree is equally weighted. **Boosting:** An ensemble learning model is built via boosting. To produce models that are somewhat effective, the dataset is split into multiple subsets. This technique uses misclassified items from earlier models to create a new subset. After that, it integrates weak models with cost functions to improve performance. **Stacking:** Stacking is a meta-classifier-based ensemble approach to combine many classification models. The layers are organized in a hierarchical structure. **Voting:** Ensemble voting models include hard and soft voting. Hard voting is a majority-based voting technique used for aggregated classification, whereas soft voting aggregates several categories and predicts based on probabilities.

TABLE 2. Different categories of DL algorithms.

Type of DL Algorithm	Definition
Transfer learning (TL) [46]	The process of transferring acquired knowledge from one domain to another can result in significant time and computational resource savings during the training phase. •Residual Network •AlexNet •Extreme Inception •CapsuleNet •Inception •Visual Geometry Group •ShuffleNet •ResNeXt •MobileNet •EfficientNet •MobileNetV2 •Densely Connected Convolutional Networks •Neural Architecture Search Network •Squeeze-and-Excitation Network
Feedforward Neural networks (FNN) [55]	There are multiple layers of neurons in an FNN model, and each neuron only receives the output of the layer above it. It then sends this output to the layer below it. No feedback exists between adjacent layers.
Convolutional Neural Networks (CNN) [56]	A very popular algorithm for imaging analysis. The input image is approximately processed by the convolution layer, activation function, and pooling layer, and redundant features are removed to reduce computational complexity and prevent overfitting. The significant features are then combined via the fully connected layer to generate a prediction or classification.
Recurrent Neural Networks (RNN) [57]	RNN is focused on time series analysis and sequential data.
Transformer Network [58]	It is a renowned deep learning framework for its self-attention mechanism, parallel data processing capabilities, and state-of-the-art performance in NLP and other sequence-based tasks. Its adaptability and efficacy have made it a cornerstone of modern AI research and applications. • Vision Transformer (ViT) • Convolutional Vision Transformer • Swin Transformer • Data-efficient Image Transformer • Bidirectional Encoder representation from Image Transformers • Pyramid Vision Transformer • Tokens-to-Token Vision Transformer • Compact Convolutional Transformer • Pooling-based Vision Transformer • Cross-Attention Multi-Scale Vision Transformer

after convolutional and pooling layers to collect high-level features for classification or regression tasks. Utilizing otoscopy images, the CNN model described here performed classification in accordance with this methodology. Other deep learning architectures, such as AlexNet, GoogLeNet, ResNet (ResNet18, ResNet50, ResNet101), Inception-V3,

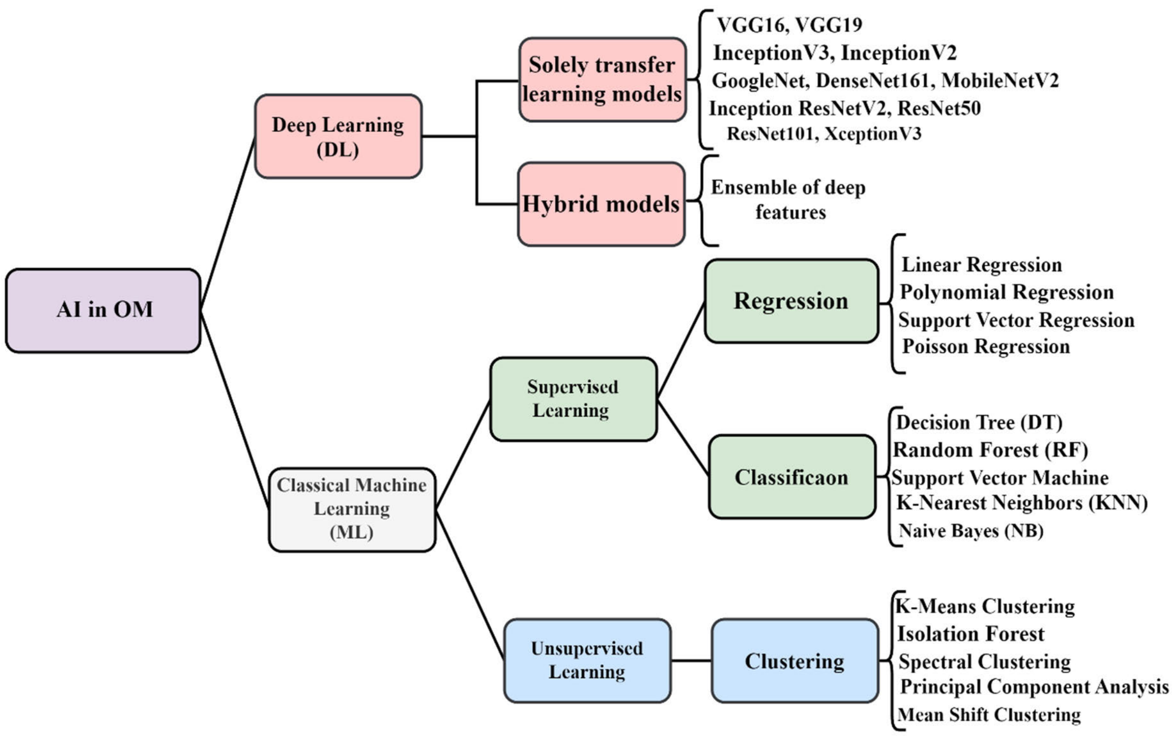

FIGURE 3. Tree-based categorization of ML and DL algorithms.

Inception-ResNet-V2, SqueezeNet, and MobileNet-V2 [55], [58], have been extensively used in medical image analysis. Each of these architectures offers unique advantages in terms of depth, computational efficiency, and feature extraction capabilities. For example, AlexNet introduced the ReLU activation function and dropout techniques to mitigate overfitting [63]. GoogLeNet [64] utilizes Inception modules for multi-scale feature extraction. ResNet's [65] deep residual learning framework effectively addresses the vanishing gradient problem, enabling the training of very deep networks. MobileNet-V2 optimizes mobile and embedded vision applications by using depthwise separable convolutions, which reduce the model size and computational cost [66].

Moreover, advanced models like Inception-ResNet-V2 [67] combine the benefits of Inception modules with residual connections to enhance performance and convergence speed. Furthermore, DenseNet [68] and EfficientNet [69] have gained traction in medical image classification. DenseNet's connectivity pattern interconnects each layer to every other layer in a feed-forward manner, improving the flow of information and gradients within the network, resulting in more accurate and efficient training. On the contrary, EfficientNet systematically scales up the network's depth, width, and resolution using a compound scaling method, striking a better balance between performance and computational efficiency.

IV. LITERATURE REVIEW

Diagnosing OM using AI has emerged as a promising and novel procedure in recent years. This technological advancement has the potential to revolutionize how clinicians identify and treat OM infections.

Some recent papers [29], [30], [31], [32] have reviewed the advances in the field of otitis media diagnosis using AI, although the number is minimal. Ngombu et al. [30] evaluated the most recent applications of AI in diagnosing OM in both pediatric and adult populations. After screening numerous articles published between January 2010 and May 2021, 25 relevant studies were identified. The most prevalent AI techniques utilized for OM diagnosis were machine learning, natural language processing, and prototyping. While some of these technologies are still in the developmental and testing stages, the review highlights their potential to improve the efficiency and accuracy of OM diagnoses. A systematic review by Song et al. [33] was conducted to examine the use of AI models in diagnosing otitis media through medical imaging. The review included a total of 26 studies that utilized tympanic membrane images, with an average accuracy rate of 86%. Additionally, three studies incorporated segmentation and classification methodologies, resulting in an average accuracy rate of 90.8%. These findings suggest that AI has the potential to significantly improve diagnostic accuracy in otitis media, which can have significant implications for telemedicine and primary care. However, this review has limitations, such as inadequate analysis of explainability and insufficient involvement of clinical experts. These limitations must be addressed to ensure patient safety and achieve optimal outcomes. Another review by Esposito et al. [31] delves into the difficulties in reliably diagnosing acute otitis media

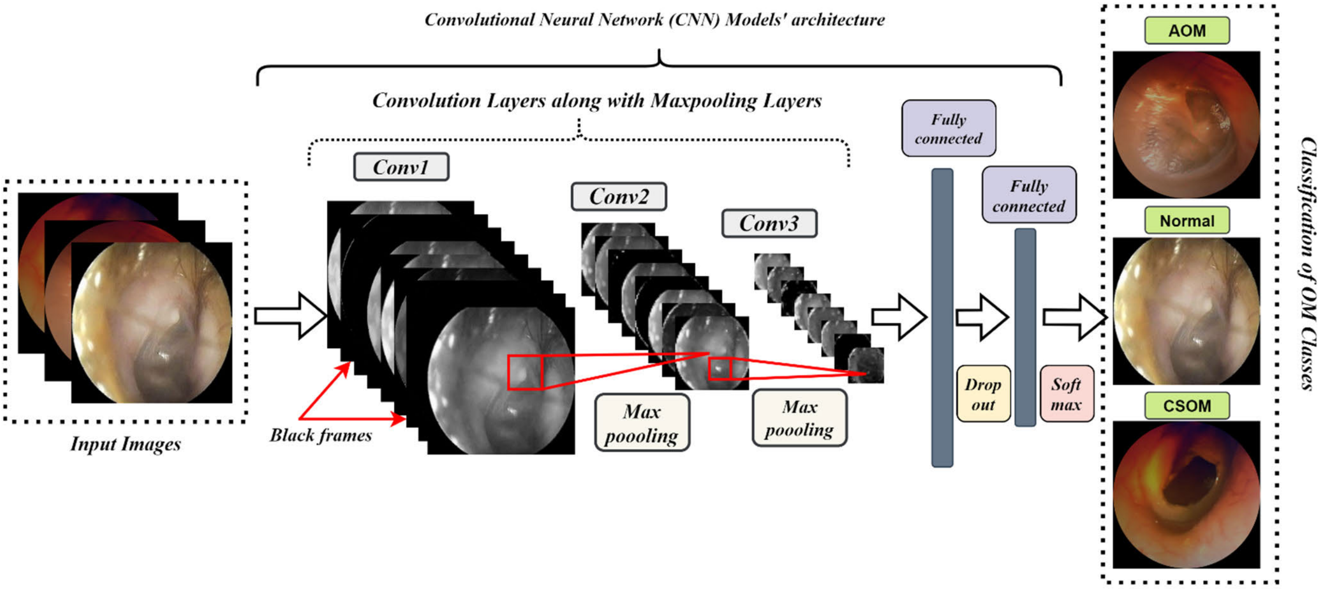

FIGURE 4. Otitis media disease detection using shallow deep learning network.

(AOM). It emphasizes the significance of comprehending AOM risk factors and implementing more effective preventive measures, especially for AOM-vulnerable children. Diagnostic uncertainty often leads to antibiotic overuse and therapeutic errors. The paper investigates novel approaches and technologies, such as light field otoscopy, optical coherence tomography, low-coherence interferometry, and Raman spectroscopy. However, the majority of these have not yet been implemented in clinical practice. The article emphasizes the potential of video-otoscopy, particularly in conjunction with telemedicine and artificial intelligence. Promoting otologic telemedicine and artificial intelligence among pediatricians and ENT specialists raises awareness of AOM diagnostic errors and improves patient care. The review by Ding et al. [32] emphasizes the use of AI in diagnosing, treating, and managing OM. AI has shown promise in healthcare, particularly in disease detection, image interpretation, and outcome prediction. The review emphasizes successful OM diagnostic applications of machine learning algorithms such as ResNet, InceptionV3, and Unet. AI's potential in OM is manifest but the application is still in its infancy. This article discusses the current use of ML and AI, key concepts, and future challenges in developing AI-assisted OM technologies However, this study has not discussed the explainability of the AI-based models and clinician involvement. In another study, a systematic evaluation of sixteen ML models for classifying Middle Ear Disorders (MED) using Tympanic Membrane (TM) images is detailed by Cao et al. [29]. The study was completed per the PRISMA guidelines, where sensitivity, specificity, and area under the curve (AUC) were considered when extracting results. The evaluation of 20,254 TM images across 25 ML approaches yielded an accuracy range of 76.00% to 98.26%, with 68.8% of studies exhibiting a low risk of bias. The suggested conclusion indicates that ML can effectively differentiate between normal ears and MED, highlighting the importance of establishing a standardized protocol for acquiring and annotating TM images.

In addition, the classification of emotions through the recording of EEG signals has proven to be highly beneficial for diagnosing Otitis Media. Haapala et al. [70] conducted a study to examine the impact of recurrent acute otitis media (RAOM) on involuntary auditory attention in toddlers. They found that children with a history of RAOM exhibited reduced lP3a amplitude and delayed LN latency, suggesting potential long-term effects on the neural mechanisms of attention. These findings indicate that RAOM may impede the development of involuntary attention control in young children. In another study, Jotic et al. [71] aimed to assess the severity of symptoms of depression, anxiety, and stress in patients with chronic otitis media (COM), as well as explore the influence of patient demographics and COM characteristics on these symptoms. The results revealed that a significant proportion of patients experienced anxiety (70.57%), stress (49.37%), and depression (13.29%). Moreover, more severe COM symptoms were found to be positively correlated with higher levels of these mental health issues. Factors such as hearing loss and symptoms like drainage, hearing problems, and tinnitus were significant predictors of increased anxiety, stress, and depression. Ahmed et al. [72] introduced a novel method called InvBase for baseline removal in EEG-based emotion classification. This method demonstrated significant improvements in the accuracy of classifying valence and arousal compared to traditional subtractive and no-baseline correction methods.

As discussed, the number of survey papers for AI-based diagnosis of OM is limited, although a few studies [73], [74]

TABLE 3. Research questions.

Research Question (RQ)	Description
RQ1. AI Models	How often ML and DL models improve diagnosing otitis media?
RQ2. Preprocessing	How much of the AI-based health care informatics paper focuses on image/video preprocessing?
RQ3. Model Explainability Features	Do AI, ML, and DL models incorporate explainability features?
RQ4. Medical Expert Involvement	How can increased collaboration with medical experts improve model accuracy and relevance in otitis media diagnosis?
RQ5. Data Quality	What strategies ensure the aural images used for training AI models in otitis media diagnosis are precise and diverse?
RQ6. Web Application	How often has AI-based paper introduced the presence of web applications?

TABLE 4. Identification of sources for relevant research papers.

Electronic Database	Type	URL
Scopus	Search Engine	https://www.scopus.com/ (accessed on 5 October 2023)
Web of science	Search Engine	https://www.webofscience.com/ (accessed on 5 October 2023)
Springer	Digital Library	https://www.springer.com/gp (accessed on 5 October 2023)
Google Scholar	Search Engine	https://scholar.google.com.au (accessed on 5 October 2023)
Science Direct—Elsevier	Digital Library	https://www.sciencedirect.com (accessed on 5 October 2023)
MDPI	Digital Library	https://www.mdpi.com (accessed on 5 October 2023)
ResearchGate	Social Networking Site	https://www.researchgate.net (accessed on 5 October 2023)
IEEE Xplore	Digital Library	https://ieeexplore.ieee.org/Xplore/home.jsp (accessed on 5 October 2023)

reviewed OM diagnosis. Based on the preceding analysis, none of the examined studies have adequately elucidated the precise contributions of AI to the field of OM diagnosis. The crucial issue is whether AI can actually aid clinicians in their diagnostic processes and whether these healthcare professionals comprehend the results of AI-generated predictions. It is essential to recognize that the intersection of AI and OM diagnosis remains a challenging, dynamic, and complex domain. However, these obstacles are thoroughly addressed in our review. Through our analysis, we seek to shed light on the current state of AI applications in OM diagnosis, examining its potential benefits and the obstacles that must be overcome to realize its maximum potential.

V. METHODOLOGY

In our review of AI-based Otitis media diagnosis, we adhere to a systematic review approach to ensure the highest quality of research synthesis [56], [57]. In this section, we first conduct a systematic review approach of diagnosing OM utilizing AI, and then analyze the quality of the papers in a more in-depth overview.

A. RESEARCH QUESTIONS AND OBJECTIVES

The research questions presented in this review paper focus on diagnosing otitis media using AI and are illustrated in Table 3.

The objective is to investigate the advancement of AI technologies and their practical implementation in OM diagnosis. The research questions illustrated in Table 3 aim to evaluate the role of AI, ML, and DL models in diagnosing otitis media. This evaluation includes assessing improvements in diagnostic accuracy, the significance of image/video preprocessing in healthcare informatics, integrating explainability features in the models, and studying the impact of collaboration with medical experts. Additionally, the objective is to investigate strategies for obtaining high-quality and diverse aural images for training models and examine the prevalence of web applications in AI-based studies. This approach will enhance the effectiveness of AI in diagnosing otitis media.

B. SYSTEMATIC REVIEW OF EXISTING PROPOSALS TO DIAGNOSIS OM UTILIZING AI

Kitchenham [75] state that a systematic review is an effective method for compiling existing studies and identifying research gaps that may lead to new areas of inquiry. As such, a comprehensive overview of the present state of AI-based diagnosis in the context of OM was generated through a systematic review. The review protocol is divided into two phases: (1) the initial search and (2) the formulation of inclusion and exclusion criteria.

The search phase includes defining academic databases, digital libraries, and search engines that can be used to search for eligible studies. The most recent access to these reputable electronic databases, such as Scopus, web of science, and Google Scholar, was recorded on 5 October 2023 as part of our research, see Table 4. We used a predefined search strategy that included specific keywords and inclusion criteria to identify relevant articles pertaining to AI-based Otitis media diagnosis. The selection procedure involved evaluating the relevance of titles and abstracts, followed by a thorough review of the full texts of potentially eligible studies.

An extensive literature search has been compiled to find most recent research articles. Table 5 presents a list of keyword searches for retrieving relevant articles and studies at the intersection of AI and ML technologies, including concepts such as deep learning, convolutional neural networks, data mining, computer-assisted diagnosis, computer-assisted surgery, and computer vision. Simultaneously, it aims to investigate medical subjects related to OM, including terms like "ear," "eardrum," "tympanic membrane," "ear disease," and various abbreviations for OM. Using these keywords, we effectively identified and gained access to

TABLE 5. Literature search keywords.

	Non- Medical Subject Headings Terms	Medical Subject Headings Terms
Search Query (SQ)	("artificial intelligence" OR "machine learning" OR "deep learning" OR "convolutional neural networks" OR "data mining" OR "computer-assisted diagnosis" OR "computer-assisted surgery" OR "computer vision")	("eardrum" OR "tympanic membrane" OR "ear disease" OR "otitis" OR "otitis media" OR "Otitis media with effusion " OR "acute otitis media" OR "chronic otitis media" OR "chronic suppurative otitis media" OR "cholesteatoma ")

TABLE 6. Inclusion and exclusion overview.

List of Inclusion and Exclusion Criteria	
Inclusion Criteria (IC)	
IC1	Must contain at least one of the specified keywords
IC2	Must be present in one of the designated databases
IC3	Published within the preceding decade (2014–2023)
IC4	Requires publication in a reputable journal or conference
IC5	The examined research should possess congruent title, abstract, and full text
Exclusion Criteria (EC)	
EC1	Eliminate duplicates
EC2	Inability to access the entire paper
EC3	Irrelevant to AI-Based OM Diagnosis
EC4	Documents not in English

scientific publications and research findings pertaining to the application of AI and ML techniques within the context of OM, thereby facilitating a thorough investigation.

The subsequent phase entails establishing inclusion criteria (IC) and exclusion criteria (EC) based on previously established standards. These criteria were carefully formulated to improve the accuracy of our research paper selection procedure. Table 6 provides an exhaustive list of IC and EC for this study. Studies that fell under the EC category were immediately deemed ineligible. To obtain a more refined set of search results, we screened the titles, abstracts, and full texts of the studies more thoroughly.

(a) Title: Excluded were papers that did not match at least one of the keywords enumerated in Table 5.

(b) Abstract: Only papers meeting at least 40 percent of the ICs were retained for evaluation.

(c) Full text: Papers were required to discuss proposals that addressed AI-based diagnosis of OM.

Figure 5 shows the visualization of the PRISMA diagram, representing the systematic review methodology used in this study. Five hundred thirty papers were initially identified in the identification stage. During the Screening Stage, the abstracts and titles of these papers were carefully examined, resulting in 252 papers. In the Eligibility Stage, the full text of the papers was thoroughly evaluated, resulting in the selection of 131 papers that met our inclusion criteria. Finally, in the last inclusion stage, only 32 papers met our criteria and were included in our analysis.

C. QUALITY ASSESSMENT OF SYSTEMATIC REVIEWS

In this review, we employ evaluation techniques to ensure that the papers selected for inclusion in this systematic review satisfy stringent criteria for credibility and suitability. Utilizing quality evaluation (QE) queries from a prior systematic literature review [76], [77]. Each QE is scored as "No" (0), "Partial" (1), or "Yes" (2), with a total of four questions per exam. The cumulative score for each query ranges from 0 to 6, giving each publication a maximum possible score of 24. All reviewers unanimously concurred that papers with a cumulative score between 0 and 12 should be deemed insufficient and excluded from the review. In contrast, papers with a cumulative score between 13 and 24 are eligible for inclusion in the review. The outcomes of the quality assessments conducted on the selected papers for this systematic review are summarized in Table 7. In addition, the rigorous quality assessment process ensures that only papers meeting the highest standards of dependability are included in the review, thereby enhancing the integrity of the overall findings. This meticulous evaluation is essential for maintaining the credibility and validity of the conclusions of the systematic review.

According to the results, all 32 papers met the required criteria and are therefore included in the review.

VI. DIAGNOSIS

As previously stated, OM classes are diagnosed using otoscopy images, videos, and tympanometry. Consequently, different researchers employed different AI methods to analyze OM diseases. Recent progress in the diagnosis based on AI is described below. This section is divided into three sections: otoscopy images with three subsections (classification of OM images using transfer learning, classification using hybrid methods, segmentation and decision system based), otoscopy videos, and tympanometry.

A. OTOSCOPY IMAGE BASED

The main goal of this part is to distinguish OM classes using AI methods based on otoscopy images. This section contains three subsections: classification of OM images using solid deep learning methods, classification using hybrid methods of deep learning and machine learning, segmentation, and decision support system methods.

1) IMAGE CLASSIFICATION USING DEEP LEARNING MODELS

Basaran et al. [78] developed a computational model using the Faster R-CNN and pre-trained the CNN model to differentiate normal from abnormal TMs, using 1692 augmented otoscope images (raw image 282). The Faster R-CNN finds the TMs in otoscope images and produces a patch encompassing the TM only. Pre-trained CNNs were then re-trained with these patches, utilizing a transfer learning strategy. The VGG16 model provided a classification accuracy of 90.48%. In another study, Raiaan et al. [79] utilized about 655 distinct

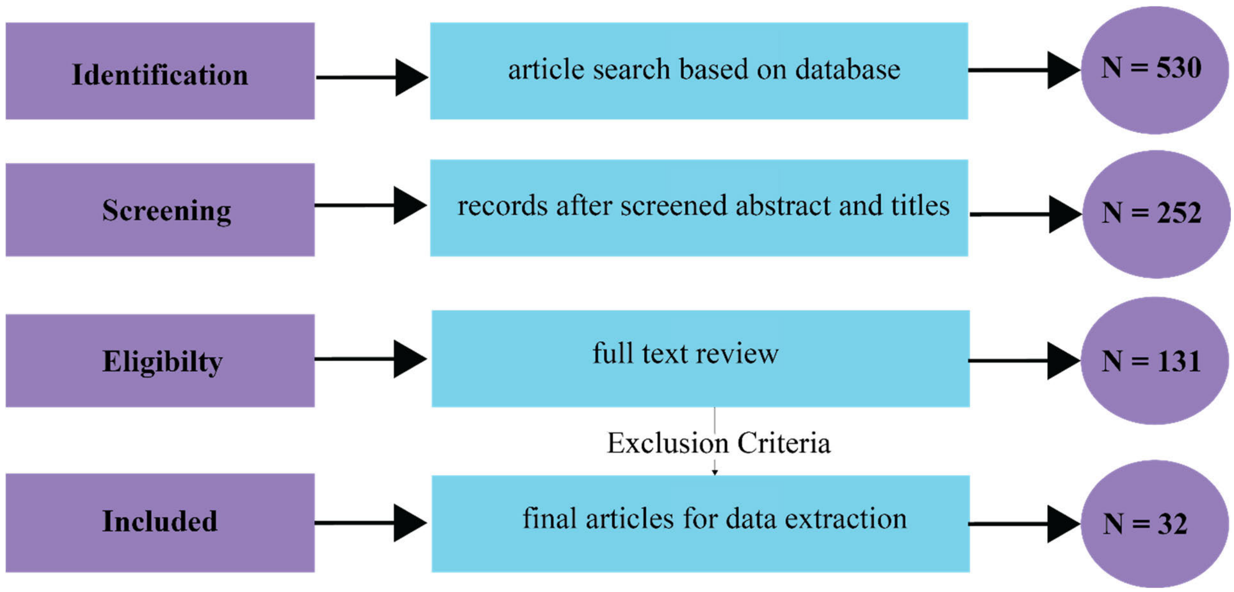

FIGURE 5. Systematic review flow diagram.

otoscopic images of 63 cerumen impactions, 120 images of tympanostomy, 346 images of normal TM, 50 images of TM with myringosclerosis, 44 images of otitis externa, and 32 images of TMs with perforations. They proposed a CNN model using three convolutional layers, batch normalization and dropout layers. Their model had a 84.4% accuracy in classification.

Khan et al. [80] developed a novel automated detection CNN model (DenseNet) to diagnose TM and middle ear infections. They utilized around 2,484 otoscopic images, including normal, OME, and chronic OM (COM) with perforation. DenseNet161 outperformed other models with a high accuracy of 94.9%, precision of 95.2%, a recall of 95%, and an F1-score of 95.1% in classification. Furthermore, their approach obtained 0.99 in terms of the Average area Under the Receiver Operating characteristics Curve (AUROC) and used class activation mapping (CAM) to ensure proper detection of the region of interest (ROI). Sundgaard et al. [81] proposed an approach for classification employing deep metric learning to the 1336 otoscopy images of three classes including OME (533 image), AOM (145 images) and no effusion (NOE) (658 images). They utilized circular hough transform methods for cropping images and performed down sampling and horizontal flipping techniques. Several loss function techniques including constructive loss, triplet loss and multi-class N-pair loss were employed where the triplet loss function obtained the highest accuracy of 85% in classification. The deep metric methods provide useful insight into the decision making of a neural network.

A study by Tsutsumi et al. [82] introduced a multiclass-classifier network and website for classifying TM pathologies based on otoscopic images. They utilized about 400 normal (196) and abnormal (204) images, including AOM (116), otitis external (44), CSOM (23), and cerumen impaction (21). They employed several deep learning models: ResNet50, InceotionV3, InceptionResNetV2, and MobileNetV2. The MobileNetV2 model had the best relative performance with high AUC-ROC (0.902) and accuracy ranging between 73% and 77% in binary classification. The macro-AUC-ROC and the accuracy for multiclass classification were respectively 0.91 and 66%. They employed this MobileNetV2 model as a proof-of concept publicly accessible website for real time OM predictions. In another study, researchers [83] represented a novel computer aided diagnosis (CAD) support model based on a CNN network. For improving the generalized ability of the proposed model, a combination of the channel and spatial model (CBAM), residual blocks and hyper column technique were embedded into the proposed model. They used a total of 956 raw images of three classes: AOM (119), chronic supportive otitis media (CSOM) (63), and Earwax (140). Their proposed model achieved a 98.26% accuracy in classification.

In another study [84] researchers focused on developing the "ResNet18+Suffle" network to diagnose TM images into four categories: normal (1180), OME (400), COM (627), or cholesteatoma (165). They utilized cropping and resizing to eliminate unusual border lines from images and also employed flipping and rotation augmentation methods to balance the dataset. Their proposed method resulted in an accuracy of 97.18% while a five k-fold cross validation demonstrated that the network accurately diagnosed OM with an accuracy greater than 93%. They also employed the Grad-CAM method to identify the essential ROI from images. Sandstrom et al. [85] utilized two different data sets: dataset1 consisting of normal (183), pathological (44), and wax (46) and dataset2, also with normal (123), pathological (206),

TABLE 7. A summary of scores used for determining the quality of papers.

Authors [Reference]	QE 1: Is the publication related to the diagnosis of OM using AI?	QE 2: Is the suggested solution well-articulated?	QE 3: Are those clearly stating the challenges addressed by their proposed solution?	QE 4: Did the publication outline the limitations of the proposed solutions?	Summary Points
Basaran et al. [78]	6	6	5	5	22
Devon et al. [79]	6	6	6	5	23
Khan et al. [80]	6	5	5	5	21
Sundgaard et al., [81]	6	6	5	5	22
Kotaro et al., [82]	6	6	5	5	22
Alhudhaif et al., [83]	5	5	5	4	19
Byun et al., [84]	6	5	5	5	21
Sandstrom et al. [85]	6	6	5	4	21
Caliskan et al. [19]	5	5	5	4	19
Christopher et al. [86]	6	5	5	4	20
Viscaino et al. [87]	6	5	4	4	19
Yeonjoo et al. [88]	6	6	5	5	22
Zeng et al. [89]	6	5	6	4	21
Habib et al. [90]	6	5	5	4	20
Hermanus et al. [35]	6	6	5	5	22
Cha et al. [91]	5	5	5	4	19
Comert et al. [92]	6	6	5	5	22
Viscaino et al. [93]	6	5	5	4	20
Xinyu et al. [94]	5	5	4	4	18
Shie et al. [95]	6	6	5	4	21
Pham et al. [96]	0	6	6	6	18
Pham et al. [97]	0	6	6	6	18
Chen et al. [98]	0	6	6	6	18
Kim et al. [99]	6	6	5	4	21
Myburgh et al. [100]	6	6	5	4	21
Tozar et al. [102]	6	6	6	5	23
Chan et al. [103]	6	5	5	4	20
Moshtaghi et al. [104]	5	5	5	4	19
Sundgaard et al. [25]	6	5	5	5	21
Grais et al., [106]	6	6	5	5	22
Sundgaard et al. [107]	6	5	5	4	20
Merchant et al. [108]	5	5	5	5	20

and wax (60) images. They employed an image cropping method to remove the black areas from the image and randomly removed images from classes with a higher number of images. In addition, they generated lowest resolution images from a high-resolution image to increase the dataset. The GoogleNet transfer learning model was employed to classify the images. The overall accuracy of the convolutional neural network was above 90% in all except one approach. Sensitivity to finding ears with wax or pathology was above 93% in all cases and specificity was 100%.

Caliskan et al. [19] performed the diagnostic task in two stages. In the first stage, a CNN model was applied to images obtained using the otoscope device and deep features were obtained. To separate normal and abnormal images the VGG16 model was used. In the second stage, the activation maps of the fc6 and fc7 layers consisting of 4,096 features and fc8 layer consisting of 1,000 features of the VGG16 model were obtained. The deep features obtained from all activation maps were then merged and a new feature set was created and fed to the support factor machine (SVM) classifier. They applied a total of 956 middle ear images of two classes: normal (535 images), and abnormal (421 images). The highest accuracy of 82.17% was obtained for the fc6.

A study by Tseng et al. [86] used 834 otoscopic images including cholesteatoma (197), abnormal non-cholesteatoma (457) and normal (180). Their main focus was to detect cholesteatoma. They applied eight retrained CNN models: VGG19, MobileNet2, DenseNet201, InceptionV3, ResNet152V2, Xception, InceotionResNetV2, and NASNetLarge. Their final trained CNN's model demonstrated a strong performance, achieving accuracies of 83.8%-98.5%, 75.6%-90.1%, 87.0%-90.4% respectively for differentiating cholesteatona from normal, cholesteatoma from abnormal non-cholesteatoma, and cholesteatoma from non-cholesteatoma (normal+abnormal non-cholesteatoma). Viscaino et al. [87] explored color wavelengths dependence in a model that diagnoses four types of middle and external ear diseases including normal, COM, OME and ear wax plug. They extracted around 22,000 key frames from 200 videos (195 patients) from the otolaryngology department of the clinical hospital of the university of Chile (HCUCH). They employed several methods including Kullback-Leibler divergence score, variance of the laplacian method, analyzing principal component analysis (PCA) and K- means clustering for extracting key frames from videos. They also employed multiple augmentation methods including rotations, zoom-in, zoom-out, horizontal and vertical flips to overcome the over fitting. Their proposed CNN model acquired a highest accuracy of 92%, a sensitivity of 85%, a specificity of 95%, a precision of 86% and an F1-score of 85% in diagnose.

Choi et al. [88] proposed a CNN model customizing the EfficiemrNet-B4 architecture to predict the primary class: OME (1,630 images), COM (1,534 images), and none (3,466 images) without OME and COM and secondary classes: attic cholesteatoma (893 images), meningitis (1,083 images), ventilating tube (1,676 images) and otomycosis (181 images). In addition, the model accurately predicted the primary class with a dice similarity coefficient (DSC) of 95.19%; in secondary classes, the diagnosis of cholesteatoma and meningitis received a DSC of 88.37% and 88.28%, respectively. Zeng et al. [89] developed and validated a DL model to identify atelectasis and attic retraction pocket in cases of OME using multi-center otoscopic images. Threefold random cross-validation has been utilized with 6393 OME otoscopic images. The DL model exhibited a detection accuracy of 79% for atelectasis and 89% for attic retraction pocket, with corresponding AUC values of 0.87 and 0.87. The DL algorithm demonstrated promise as a valuable instrument for precise diagnosis and identification of atelectasis and attic retraction pocket in OME otoscopic images, as evidenced by Class Activation Mapping.

Habib et al. [90] evaluate the applicability of DL-based AI algorithms in detecting OM based on otoscopic images. A comprehensive collection of 1842 otoscopic images was gathered from three distinct sources. These images were classified into diagnostic categories of normal or abnormal. The internal performance of AI-otoscopy algorithms exhibited notable accuracy (mean AUC: 0.95). However, when evaluated on external otoscopic images that were not utilized in the training process, a decline in performance was detected (mean AUC: 0.76). In contrast to the algorithm's internal performance, its external performance was noticeably inferior. Results emphasize the necessity for additional investigation into data augmentation and pre-processing methodologies, which can improve the algorithm's applicability in clinical settings. Table 8 illustrates all the details about these studies.

2) IMAGE CLASSIFICATION USING ENSEMBLE DEEP LEARNING AND CLASSICAL MACHINE LEARNING MODELS

In this section, we try to gain insight in classifying OM images using deep learning and classical machine learning techniques, reviewing papers where deep learning and classical machine learning techniques were used to classify OM images. Hermanus et al. [35] introduced an automated approach using image processing methods to classify OM classes into five categories: normal (n-TM), AOM, OME, CSOM with perforation, and obstructing wax or foreign bodies (O/W). After assessment of ear images by experts, the final dataset contained around 489 images, including O/W (120), n-TM (123), AOM (80), OME (80), and CSOM with perforation (86). The image was provided as an input image to the system after image processing and feature extraction were performed to classify unknown images. After extracting features, the decision -tree (DT) classified the feature vector associated with the input image. The output, which consists of the extracted features and the final diagnosis, was then presented to the user. An accuracy of 80.6% was achieved for images taken with commercial video-otoscopes. In comparison, an accuracy of 78.7% was acquired for images captured on-site with a low-cost custom-made video otoscope.

A study by, Cha et al. [91] utilized approximately 10,544 otoscopy images for training nine state-of-the-art models, Squeezanet, AlexNet, ResNet18, MobileNet-v2, GoogleNet, ResNet50, ResNet101, inceptionTrsnet-v2, for classifying six classes of ear diseases (normal, attic retraction, tympanic perforation, otitis external, and tumor. After assessing deep learning models, they integrated the InceptionV3 and ResNet101 models based on their performance. Before that, they employed several data augmentation methods (axis translation, rotation, scaling, and flips) to balance the dataset. The ensemble model obtained a mean accuracy of 93.67% and a sensitivity and specificity for normal and abnormal classes of 93.69% and 96.82%, respectively. The accuracy, sensitivity, and specificity for the attic retraction or adhesive OM are respectively 85.78%, 93.69%, and 98.25%. Similarly, the accuracy, specification, and sensitivity for otitis externa are 77.91%, 99.02%, and 89.33%.

In another study, Zafer et al. [92] combined fine-tuned deep features along with classical machine learning models to obtain optimal results in classification. They focused on classifying 956 images of normal, AOM, CSOM, and ear wax. Initially, they trained several deep learning models with images and generated a hybrid model consisting of a combination of fused fine-tuned deep features provided by DCNNs. The weights of the last few layers of the models were updated, the results of the models evaluated and compared to each other, and the updated final fully connected layers of the DCNNs were concatenated. The new feature set was used as input to the machine learning models and a support vector machine (SVM) resulting in 99.47% accuracy, 99.35% sensitivity, and 99.77% specificity in classification.

Viscaino et al. [93] also proposed an automated CAD based on classical machine learning models and image processing techniques for the diagnosis of OM images. They used 880 images 220 in each of four classes: normal, ear wax, myringoesclerosis, and COM. Several image processing methods, including the Laplacian method's variability for evaluating the images' blurriness and a Circular Hough transform method, were employed to identify the ROI from images. Three different feature extraction methods: filter bank, discrete cosine transform (DCT), and color coherence vector (CCV) were employed to extract the essential features

TABLE 8. A literature demonstration of image classification using solely deep learning models.

Authors [Reference]	Dataset collection	Class name & total raw image	Image Processing and Augmentation Methods	Deep Learning Models used with performance	Model Explainability Features	Presence of web Application (Yes/No)	Limitations
Basaran et al. [78]	**Name:** N/A **Hospital name:** Özel Van Akdamar Hospital in Turkey between 10/2018 and 1/2019 **Patient no: 282**	1. Normal (154), 2. Abnormal (128) Total image: 282 After augmentation: Normal (924), Abnormal (768) **Total**: 1692	**IP:** N/A **AM:** 1.Flipping 2.Rotation	Model: VGG16 Accuracy (90.48%)	✗	✗	1.Dataset has a lack of OM images 2. Lack of model fine tuning 3. Lack of proper image preprocessing 4. Lack of model fine tuning 5. Segmentation is missing 6. Missing comparison with other datasets 7. Low accuracy in classification
Devon et al. [79]	**Name:**N/A **Hospital name:** N/A **Patient no:** N/A	1.Normal (346) 2.Cerumen impactions (63), 3. Tympanostomy (120) 4. TM with myringosclerosis (50) 5. Otitis externa (44) 6. TM with perforations (32) Total image: 655	**IP:** N/A **AM:** 1.Rotation 2.Mirroring	Model: CNN Accuracy (88.4%)	✗	✗	1.Dataset has a lack of OM images 2. Lack of image preprocessing 3. Lack of model fine tuning 4. Segmentation is missing 5. Missing comparison with other datasets 6. Low accuracy in classification
Khan et al. [80]	**Name:**Oto-endoscopic images (OEIs) **Hospital Name: N/A** **Patient no:** 1,427	1. Normal (897) 2. OME (1241) 3. COM with perforation (346) Total image:2484	**IP:** Normalization **AM:** 1.Random Resized Crop 2.Horizontal flip 3.Rotation	Model: DenseNet161 Accuracy (94.9%), precision (95.2%), recall (95%) and F1-score (95.1%)	✓	✓	1. Segmentation is missing 2. Missing of comparison with other datasets
Sundgaard et al. [81]	**Name:**N/A **Hospital name:** Kamide ENT clinic, Shizouka,Japan **Patient no:**519	1.No effusion (NOE) (658 images), 2.OME (533 images) 3. AOM (145 images) Total image: 1336	**IP:** Cropping method (Circular Hough Transform) **AM:** 1.Down sampling Horizontal flipping	Triplet loss function (85%)	✗	✗	1.Dataset has a lack of OM images 2. Lack of model fine tuning 3. Lack of proper image preprocessing 4. Segmentation is missing 5. Missing comparison with other datasets 6. Low accuracy in classification
Kotaro et al [82]	**Name:** N\A **Hospital name:** Van Akdamar Hospital eardrum database **Patient no:282**	1. Normal (196 images) 2. Abnormal: (AOM+OE+CSOM+ Cerumen impaction) Total image: 400	**IP:**1. Resize **AM:** 1.Rotation range 180, 2.Sheering range 0.3, 3.Zoom range 0.6, 4.Random brightness change between ranges of 0.2–2.7, 5. Random horizontal flips 6.Vertical flips	Model: MobileNetV2 Binary classification: AUC-ROC (0.902) Accuracy range between (0.73-0.77) Multi-class classification: AUC-ROC (0.91) Accuracy (66%)	✗	✓	1. Limited Sample size. 2. Lack of image preprocessing 3. Lack of model fine tuning 4. Segmentation is missing 5. Low accuracy in classification 6. Missing comparison with other datasets
Alhudhaif et al., [83]	**Name: N/A** **Hospital name:** Özel Van Akdamar Hospital between 10/2018 and 06/2019 **Patient no:N/A**	1.Normal (535), 2. AOM (119), 3. CSOM (63), 4. Ear wax (140), 5. others (99) Total image 956 images	**IP:** N/A **AM:** 1.Horizontal scroll ratio, 2.Vertical scroll ratio, 3.Zoom, 4.Rotation, 5.Sharpness value, etc.	Model: Proposed CNN Accuracy (98.26%), Sensitivity (97.68%), Specificity (99.30%)	✗	✗	1. Limited sample size. 2. Lack of proper image preprocessing 3. Lack of model fine tuning 4. Segmentation is missing 5. Missing comparison with other datasets
Byun et al., [84]	Name: N/A Source name: Otolaryngology department for ear problems from January 2015 to December 2018 Patient no: N/A	1.Normal (1180), 2.OME (400), 3.COM (627) 4. Cholesteatoma (165). Total image: 2272	**IP:** 1.Cropping 2. Resize **AM:** 1.Flip, 2.Flop, 3. Rotation	Model: ResNet18 + Shuffle Accuracy (97.18%)	✗	✗	1. Missing comparison with other datasets 2. Lack of model fine tuning 3. Segmentation is missing
Sandstrom et al. [85]	Dataset1 name: N/A Source name: (Welch Allyn Digital Macroview Otoscope, Welch Allyn Inc., Skaneateles Falls, New York, NY, USA) between March and May 2016. Dataset2 name: N/A (Public)	Dataset1: 1.Normal (183) 2.Pathological (44) 3. Wax (46) Total images:273 Dataset2: 1.Normal (123) 2.Pathological (206) 3. Wax (60) Total images:389	**IP:** 1.Cropping **AM:** Generating low resolution images from a high-resolution image	Model: GoogleNet / InceptionV1 1.Overall accuracy above 0.9 in all except one approach. 2.Sensitivity to finding ears with wax or pathology was above 93% in all cases and specificity was 100%.	✗	✗	1.Dataset has a lack of Pathological and Wax images. 2. Lack of proper image preprocessing 3. Lack of model fine tuning 4. Segmentation is missing

TABLE 8. ***(Continued.)*** **A literature demonstration of image classification using solely deep learning models.**

Caliskan et al. [19]	**Name:**N/A **Hospital name:** Van Akdamar Hospital inTurkey **Patient no: N/A**	1. Normal (535), 2. Abnormal (421) Total image: 956	**IP:** N/A **AM:** N/A	Model: VGG16 model with fc6 layer accuracy (82.17%), Sensitivity (71.43%) Specificity (90.62%) and F1-score (77.92%)	✗	✗	1.Dataset has a limited number of OM images 2. Lack of proper image preprocessing 3. Lack of model fine tuning 4. Segmentation is missing 5. Missing comparison with other datasets 6. Low accuracy in classification
Christopher et al. [86]	**Name:**N/A **Hospital name:**N/A **Patient no: N/A**	1.Normal (180), 2.Cholesteatoma (197), 3.Abnormal non-cholesteatoma (457), Total image: 834	**IP:** Employed but not mentioned **AM:** 1.Rotation 2.Horizontal reflection 3.Vertical reflection	Their final trained CNN model acquired. 83.8%–98.5% (cholesteatoma Vs normal), 75.6%–90.1% (cholesteatoma Vs abnormal non-cholesteatoma), and 87.0%–90.4% (cholesteatoma Vs non-cholesteatoma) (abnormal non-cholesteatoma + normal)	✓	✗	1. Limited Sample size. 2. Segmentation is missing 3. Missing comparison with other datasets
Viscaino et al. [87]	**Name:** N/A **Hospital name:** Otolaryngology department of the Clinical Hospital of the University of Chile (HCUCH) between 2018 and 2019. Patient no: 195	1.Normal (5500) 2. COM (5500) 3. OME (5500) 4. Ear wax plug (5500) Total image (keyframes): 22,000 (from 200 videos)	**IP:** 1. Kullback-Leibler divergence score 2. Applied variance of the laplacian method) 3. principal component analysis (PCA) and. K-means clustering) **AM:** 1.Rotations 2.Zoom-in 3.Zoom-out 4.Horizontal flips 5.Vertical flips	Model: A single green channel CNN model acquired accuracy (92%), sensitivity (85%), specificity (95%), precision (86%), and F1-score (85%)	✓	✗	1. Lack of model fine tuning 2. Segmentation is missing 3. The accuracy does not reached to the state-of-art level. 4. Evaluation conducted on a single dataset only
Yeonjoo et al. [88]	**Name:** N/A **Hospital name:** The otologic clinic in Asan Medical Center from Jan 2018 to Dec 2020. **Patient no:** N/A	Category-1: 1.OME (1,630 images) 2. COM (1,534 images), 3. none (3,466 images) without OME and COM Total image: 6,630 **Category-2:** 1. attic cholesteatoma (893 images) 2. Meningitis (1,083 images), 3. Ventilating tube (1,676 images) 4. Otomycosis (181 images). Total image:3,833	**IP:** Circular Cropping **AM:** 1. Rotation (−90° to 90°) 2. Translation shift (0–20% of image size in horizontal and vertical axes), 3. Zoom (0–20%), 4. Horizontal flip, 5. Brightness change (0–20%) 6. downscale (0–50%)	Model: Modified EfficientNet-B4 model Dice Coefficient (DSC) First class: 95.19% - Second class: 88.37% and 88.27% respectively for attic cholesteatoma and meningitis	✓	✗	1. Lack of model fine tuning 2. No preprocessing technique to improve the quality of the images. 3. Segmentation is missing 4. The accuracy does not reach the state-of-art level.
Zeng et al. [89]	Name: N/A Hospital name: Sun Yat-sen Memorial Hospital of Sun Yat-sen University, the Third Afliated Hospital of Sun Yat-sen University, and Zhujiang Hospital of Southern Medical University Patient no: N/A	OME otoscopic images: For detecting 1.Atelectasis 2. attic retraction pocket Total image: 6393	IP: N/A AM: N/A	Model: DL model Attic retraction pocket: Threefold cross-validation Accuracy (89%) AUC (0.89) Sensitivity (0.93) Specificity (0.62) Atelectasis: Threefold cross-validation Accuracy (79%) AUC (0.87) Sensitivity (0.71) Specificity (0.84)	✗	✗	1.Dataset has a lack of OM images 2. Lack of model fine tuning 3. Lack of proper image preprocessing 4. Segmentation is missing 5. Missing comparison with other datasets 6. Low accuracy in classification
Habib et al. [90]	Name: N/A Hospital name: (a) Van, Turkey, (b) Santiago, Chile, and (c) Ohio, USA Patient no: N/A	First source (a): Normal, AOM), and COM (848) Second source (b): normal or COM (540) Third source (c): normal or OME (454) Total image: 1842	**IP: N/A** **AM: N/A**	AI-otoscopy algorithms exhibited notable accuracy (mean AUC: 0.95).	✗	✗	1. Lack of model fine tuning 2. No preprocessing technique to improve the quality of the images. 3. Segmentation is missing 4. Low accuracy in classification

from images and then feed the classical machine learning models to classify images. The SVM model obtained the highest accuracy of 93.9%, a sensitivity of 87.8%, a specificity of 95.9%, and a positive predictive value (PPV) of 87.7% in classification.

Zeng et al. [94] utilized 20,542 images of eight classes, including normal eardrum (4217), cholestestoma of the middle ear (818), CSOM (3169), external auditory cana bleeding (694), impacted cerumen (5453), otomycosis external (2256), secretory otitis media (2448), and tympanic membrane classification (1037). Image cropping and scaling were performed to process the data and they also employed augmentation methods: flipping vertically and horizontally and rotation (90 and 180 degrees) to overcome the overfitting issue.

Afterward, they deployed DenseNet-BC1615 and DenseNet-BC169 models along with their ensemble classifier. The ensemble model consists of integrating different essential features of the models. So, the classification accuracy was obtained on DensNet-BC1615 and DensNet-BC169, and the ensemble classifier reached 94.94%, 95.08%, and 95.59%, respectively. Table 9 shows all the essential information corresponding to the literature.

3) IMAGE CLASSIFICATION USING SEGMENTATION AND DECISION-MAKING PROCESS

We also reviewed several papers where researchers focused on segmentation methods to extract ROIs from images and classify them based on accuracy while developing a decision support system. Shie et al. [95] introduced the OM image segmentation method in their work to classify OM. They focused on diagnosing four classes, normal, AOM, OME, and COM, using private datasets (865 images). They did not apply any image processing and augmentation techniques in their work. The TM area was segmented with an ''active contours'' method, which minimizes an energy function to evolve the active contour. The grid color moment (GCM), histograms of oriented gradient (HOG), local binary pattern (LBP), and gabor features were extracted from the segmented images and classification was done using several machine learning models. The AdaBoost model obtained an accuracy of 88.06%, a sensitivity of 91.57%, a specificity of 79.87%, and an F1-score of 0.914.

Pham et al. [96] proposed a segmentation method named EAR-UNet by integrating EfficientNet for the encoder, an attention gate for the skip connection path, and the decoder part is constructed based on the residual blocks from the Reset model. They also proposed a new loss function term for the neural networks to perform segmentation tasks. They used around 1012 images of normal and OM (AOM, OME, and COM) classes. To prevent overfitting data augmentation was done using rotation, and vertical and horizontal flipping. Their proposed EAR-UNet model obtained an accuracy of 95.8%, a dice similarity coefficient (DSC) of 929%, a Jaccard coefficient (Jac) of 868%, a sensitivity (Sen) of 92%, a specificity of (Spe) 976%, Hausdorff distance (HD) of 9.290 and a mean absolute distance (MAD) of 2.984 in classification.

In another study [97] TMs were automatically segmented from otoscopy images with a deep learning method. A hybrid loss function, integrating the dice loss and active contour loss, was applied with a fully convolutional neural network (FCN). During learning, the model considers the Dice similarity and the required boundary contour information, including the contour length and areas within and outside the contour. The proposed loss function was then used with the fully convolutional network to segment the tympanic membrane. They employed this approach to 1139 images of normal and OM. Their method resulted in a mean Hausdroff distance (AHD) of 19.189, DSC of 0.895, and HD of 19.189.

Chen et al. [98] collected 2820 images of ten classes, normal, AOM, CSOM, OME, TM perforation, acute myringitis, cerumen impaction, ventilation tube, TM retraction, and otomycosis. After image processing (canny edge detection), augmentation (vertical and horizontal flipping and color transformations), and splitting, images were fed to the CNN model for training purposes. The best-performing models were identified and integrated into a small CNN model which was converted into a mobile phone-based program. CAM was utilized for identifying key features for the CNN model and their method performed with 98.0% accuracy in classification.

A study by, Kim et al. [99] proposed a ResNet152 UNet++ segmentation method by applying the ResNet152 layer structure to the encoders in the UNet++ model to accurately identify the TM location and affected area. This segmentation method was applied to 9792 augmented images (from 1632 real images). Their method performed well compared with other segmentation methods and the DenseNet161 model had an accuracy of 91.4% and a recall of 90%. In contrast, without performing segmentation, the same model classified OM classes with an accuracy of 90.5% and recall of 87.5%.

Myburgh et al. [100] utilized a total amount of 389 images, including normal (123), OME (69), CSOM (86), and W/O (60) in their work. Image processing was focused on cropping the unusual areas from images and detecting the blurry images. After that, several features that are relevant to medical structures (color detection, blob detection, and position detection) were extracted from the images. They proposed a neural network as a classifier and compared it to a decision tree. The decision three achieved an 81.58% accuracy while the neural network resulted in an 86.84% accuracy in diagnosis.

Table 10 gives an overview of all the literature discussed in this section.

After reviewing all the papers, key and major limitations are pointed out, such as

1. Most researchers employed private datasets with a limited number of images and then used several augmentation methods to extend their dataset, see Table 9-10.

2. Since ear images include unusual backgrounds, reflection of lights, and other artifacts, it is usually necessary to generate denoised and artifact-free images. However, there is an absence of using proper image processing methods.

3. Although the proposed CNN models seemed to yield satisfactory results, they were not evaluated with multiple new large datasets. Table 9 shows that several studies proposed an ensemble deep learning model by integrating various deep learning model features.

4. In addition, some researchers extracted handcrafted features (filter bank, DCT, CCV, etc.) from processed images to classify them into various OM categories.

As a large data hub for OM is lacking, they employed several augmentation techniques to increase the number of images. However, a dataset with a large number of original images would be preferable. As deep features often do not provide essential medical information [101] it is worthwhile

TABLE 9. Literature demonstration of image classification using hybrid (deep learning and machine learning) process.

Authors [Reference]	Dataset collection	Class name & total raw image	Image Processing (IP) and Augmentation Methods (AM)	Feature extraction methods	Ensemble Models' name with performance	Model Explainability Features	Presence of web Application (Yes/No)	Limitations
Hermanus et al. [35]	**Name:** N/A **Hospital name:** N/A **Patient no: 282**	1. O/W (120), 2. n-TM (123), 3. AOM (80), 4. OME (80), 5. CSOM with perforation (86). **Total image: 489**	**IP:** N/A **AM:** N/A	N/A	Model: Decision Tree Accuracy (80.61%) Specificity: AOM (0.92) O/W (0.97) positive predictive value (PPV): AOM (0.75) O/W (0.89) Negative predictive values (NPV): range between 0.94 and 0.95	✗	✗	1. It has a lack of Normal, AOM, and OME images. 2. Lack of image preprocessing 3. Segmentation is missing 4. Missing of comparison with other datasets 5. Low accuracy in classification
Cha et al. [91]	**Name:** N/A **Hospital name:** outpatient clinic in Severance Hospital otorhinolaryngology department from the year 2013 to 2017 **Patient no:** N/A	1.Normal (4,342) 2.Tympanic perforation (3,370) 3.Attic retraction (1,122) 4.AOM or Myringitis or externa (430) 5.OME (787) 6.Middle ear or EAC tumor or cerumen impaction (493) Total image: 10,544	**IP:** N/A **AM:** 1. X and Y translation of input images from –45 to 45 pixels, 2. Rotations from –30 to 30 degrees 3. Scales between 0.8 and 1.2 4. left/right flips to render translation, rotation, scale and left/right	Ensemble models deep features	Model: ensemble model (InceptionV3 and ResNet101) Mean Accuracy (93.67%) 1.Normal, OME, tympanic perforation, and tumors: Accuracy (90%) Sensitivity (93.69%), Specificity (96.82%) 2.Attic retraction or adhesive otitis media: Accuracy (85·78%) Sensitivity (93.69%) Specificity (98.25%) 3.Otitis externa or myringitis: Accuracy (77.91%) Specificity (99·02%) Sensitivity (89·33%)	✗	✗	1. Lack of model fine tuning 2. No preprocessing technique to improve the quality of the images. 3. Segmentation is missing 4. The accuracy does not reach the state-of-the-art level. 5. Evaluation conducted on a single dataset only
Comert et al. [92]	**Name:** N/A **Hospital name:** Özel Van Akdamar Hospital in Turkey between 10/2018 and 06/2019. **Patient no:** N/A	1.Normal (535) 2.AOM (119) 3.CSOM (63) 4.Earwax TM (140) 5. Others (99) Total image: 956	**IP:** N/A **AM:** N/A	Deep CNNs features (AlexNet, VGG16, VGG19, GoogleNet, ResNet18, RestNet50, ResNet101)	After fused fusion: Integrating of the fused fine-tuned deep features and SVM model Accuracy (99.47%) Sensitivity (99.35%) Specificity (99.77%)	✗	✗	1.Dataset has a lack of Normal, AOM, and CSOM images. 2. Lack of image preprocessing 3. Lack of model fine tuning 4. Segmentation is missing 5. Lack of handcrafted features 6. Missing of comparison with other datasets
Viscaino et al. [93]	Name: N/A Hospital name: Otolaryngology Department of the Clinical Hospital from Universidad de Chile. Patient no:180	1. Normal (220) 2. Earwax cases (220) 3.Myringoesclerosis (220) 4. COM (220) Total image: 880	**IP:** 1. Evaluate blurriness using the variance of the Laplacian method, 2. Applied a circular Hough transform finding the ROIs **AM:** N/A	1.Filter bank 2.Discrete cosine transform (DCT) 3.Color coherence vector (CCV)	Model: SVM Accuracy (93.9%) Sensitivity (87.8%) Specificity (95.9%) PPV (87.7%)	✗	✗	1. Limited Sample size. 2. The accuracy does not reach the state-of-the-art level. 3. Missing of comparison with other datasets
Xinyu et al. [94]	Name: N/A Hospital name: Department of otolaryngology in the people's hospital of Shenzhen Baoan District, from July 2016 to August 2019 Patient no: 41,056	1. Normal eardrum (4217) 2. Cholestestoma of the middle ear (818) 3. CSOM (3169) 4. External auditory cana bleeding (694) 5. Impacted cerumen (5453) 6. Otomycosis external (2256) 7. Secretory otitis media (2448) 8. Tympanic membrane classification (1037) Total image: 20,542	**IP:** 1. Cropped 2. Scale **AM:** 1. Flipped horizontally 2. Flipped vertically and 3. Rotate (90 degree) 4. Rotate (180 degree)	Deep features After integrating multiple models	Model: 1.DensNet-BC169 Accuracy (95.08%), 2.DensNet-BC1615 Accuracy (94.94%) 3.Ensemble model Accuracy (95.59%)	✗	✓	**1.** Missing of comparison with other datasets

to identify the medical features first. For example, a normal class image may include malleus bones, cones of light, flaccid, and umbo but classification using deep neural networks does not provide insight in the biomarkers of the images. Segmentation methods play a crucial role in identifying the ROIs from images and discarding an image's unwanted background or areas. Segmenting essential markers and classifying them using a decision tree method can aid the specialist in diagnosing cases without increasing the model's complexity.

B. OTOSCOPY VIDEOS

Although video-based classification has not been widely utilized in diagnosing OM to date, its potential needs to be explored. Despite the limited adoption of video technology at present, its incorporation could result in significant positive changes. Although video classification integrating AI has not yet supplanted current methods of diagnosing OM, some research has been conducted utilizing video only. To assess vestibular impairment in children with OME and vertigo, Tozar et al. [102] used the Video Head Impulse Test (vHIT) and compared the results to those of healthy children. Thirty pediatric OME patients with vertigo and thirty healthy children aged 4 to 15 participated in the study. Mean vHIT gains were comparable between OME patients with dizziness and healthy children, indicating that significant vestibular impairment is not typically present in these OME cases. However, covert saccades were observed in some OME patients, indicating mild vestibular impairment.

Another research study conducted in pediatric emergency departments by Chan et al. [103] compared the prescription rates of antimicrobials by clinicians using a smartphone otoscope versus those using a conventional otoscope to diagnose Acute Otitis Media (AOM) in children. There was only a minor difference in the likelihood of prescribing antibiotics between the smartphone group (18.8%) and the conventional group (18.0%). However, most clinicians in the smartphone group (73%) preferred the smartphone otoscope over the conventional one.

The objective of the research of Moshtaghi et al. [104] was to compare the efficacy of a smartphone-enabled otoscope (SEO) to microscopic otoscopy in detecting and evaluating tympanic membrane (TM) pathology in an otology/neurotology practice. The SEO had an accuracy of 96% in identifying normal TMs and 100% at detecting pathology. Patients were receptive to the technology, with 93% feeling comfortable with its use and 88% finding the acquired images helpful in comprehending their condition, indicating its potential as a useful screening tool, particularly in telemedicine settings.

None of the studies mentioned in this section used video and AI in their diagnostic procedures. However, incorporating AI into these studies can potentially accelerate diagnosis and improve diagnostic precision. AI has demonstrated extraordinary speed and accuracy in analyzing medical images such as TM images. By incorporating AI algorithms that can rapidly and accurately identify pathological patterns in these images, diagnosing conditions such as OM could be expedited significantly, reducing the time and resources required for comprehensive evaluations.

C. TYMPANOMETRY

Tympanometry is a diagnostic tool used to evaluate the acoustic characteristics of the ear canal by employing an acoustic probe at 226 Hz or 1 kHz and a microphone to measure sound [25]. This analysis provides quantitative information regarding tympanic membrane (TM) and middle ear health. In addition, it helps detect the presence or absence of fluid. Tympanometry is important for the evaluation and diagnosis of OM. The accumulation of fluids can result in effusion and a flat tympanogram when negative pressure persists in the middle ear for an extended period of time [25], [105]. As a result, aberrant pressure readings can serve as an early indicator of impending middle ear infection, allowing healthcare professionals to diagnose and treat the condition promptly [106].

Several studies [25], [106], [107], [108] have used tympanometry and Wideband Tympanometry (WBT) measurements to diagnose OM. Analyzing WBT measurements, Sundgaard et al. [25] presented an automatic diagnostic algorithm that used a convolutional neural network to detect OM. The method detected OM with a high overall accuracy of 92.6% but did not differentiate between specific classes. However, their study did show the potential of deep learning for automatic OM diagnosis and demonstrated the superiority of wideband tympanograms over conventional techniques. In another study, Grais et al. [106] aimed to utilize ML techniques to automatically diagnose middle ear conditions, specifically otitis media with effusion (OME), by analyzing Wideband Absorbance Immittance (WAI) data. They collected 672 sets of WAI data from normal middle ears and ears with OME and found significant differences in absorbance values, allowing classifiers to attain an automated diagnostic accuracy of approximately 80%. The research also identified specific frequency-pressure regions, ranging from 1090Hz to 2310Hz and -40 to $+90$ daPa, as crucial for interpreting WAI data, and demonstrated the potential of ML tools to improve the diagnosis of OM.

A study by, Sundgaard et al. [107] intended to determine the inter-rater reliability of diagnosing OME, AOM, and no effusion cases using otoscopy images and, in some instances, WBT measurements. Four physicians of the ear, nose, and throat (ENTs) independently evaluated the otoscopy images and WBT results of 1409 cases. The results revealed an overall diagnostic agreement of 57% among the four ENTs, with the highest agreement and certainty observed in cases of AOM (77.0% and 90%, respectively) and the lowest in cases with no effusion (34.0% and 58%, respectively). The combination of WBT measurements with otoscopy images improved diagnostic certainty and agreement, demonstrating the value of incorporating WBT for more precise diagnoses without invasive procedures.

In another study, Merchant et al. [108] examined the effect of middle-ear effusion volume on wideband acoustic

TABLE 10. Literature demonstration of image classification using segmentation and decision-making process.

Authors [Reference]	Dataset collection	Class name & raw image	Image Processing (IP) and Augmentation Methods (AM)	Segmentation methods	Hand crafted/ medical feature extraction	Ensemble Models' name with performance	Model Explainability Features	Presence of web Application	Limitations
Shie et al. [95]	Name: N/A Hospital name: N/A Patient no:	1.Normal 2.AOM 3.OME 4.COM Total image: 865	IP: N/A AM: N/A	1."Active contour" (Segment TM area from whole image)	Features name: Mid-level features (inputting the low-level features to the test data): GCM, HOG, LBP, and Gabor)	Model: AdaBoost **Before segmentation:** Accuracy (83.75%) Sensitivity(90.63%) Specificity (67.84%) F1-score (0.874) **After segmentation:** Accuracy (88.06%) Sensitivity (91.57%) Specificity (79.87%) F1-score (0.914)	✗	✗	1. Limited Sample size. 2. Lack of performance metrics comparison. 3. The methodology description is not briefed in detail. 4. The accuracy does not reach the state-of-art level.
Pham et al. [96]	Name: N/A Source name: Collected by otologists from Karl Storz, Tullingen, Germany. Patient no: N/A	1.Normal(505) 2.OM(507): AOM(100), OME(111) COM (296). Total image: 1012	IP: N/A AM: 1.Vertical flips, 2. Horizontal flips 3. Rotation	Proposed UNet model (EAR-UNet)	N/A	Model: EAR-UNet DSC (0.929), Jac (0.868) Acc (0.958) Sen (0.920) Spe (0.976) HD (9.290) MAD (2.984)	✗	✗	1.Limited sample size. 2.No proper preprocessing steps for denoising the image. 3.Evaluation conducted on a single dataset only.
Pham et al. [97]	Name: N/A Source name: Collected by otologists from Karl Storz, Tullingen, Germany. Patient no: N/A	1.Normal(509) 2.OM(630): Total image: 1139	IP: Resize AM: 1.Rotation 2.Vertical flips, 3. Horizontal flips	Proposed deep learning based segmentation method	N/A	Model:Fully convolutional neural networks (FCNs) DSC (0.895), HD (19.189) APD (6.429)	✗	✗	1. No preprocessing technique to improve the quality of the images. 2. No proper comparison with existing imaging techniques.
Chen et al. [98]	Name: N/A Source name: Taipei Veterans General Hospital in Taiwan between Jany 1, 2011 to Dec 31, 2019. Patient no: N/A	1.Normal 2.AOM 3.CSOM 4. OME 5.TM perforation 6. Acute myringitis 7. Cerumen impaction 8. Ventilation tube 9. TM retraction 10. Otomycosis. Total image: 2820	IP: 1.Flips (vertical, horizontal and both) 2. Colour transformations (histogram normalisation, gamma correction, and Gaussian blurring). AM: 1.Flips (vertical, horizontal and both) 2. Colour transformations (histogram normalisation, gamma correction, and Gaussian blurring).	Proposed CNN model	N/A	Model: Proposed CNN Accuracy (98.0%)	✓	✓	1. The quality of the images may vary due to the acquisition from different sources and use of a smartphone camera. 2. No proper validation technique to assess the robustness of the model. 3. They built a smartphone-based edge AI system so the security issue related to dataset and training should be emphasized.
Kim et al. [99]	Name:N/A Hospital name: Korea University Ansan Hospital Patient: N/A	Raw image: 1632 After augmentation 1. Normal (2782) 2. Perforation (3626) 3. Retraction (1,866) 4. Cholesteatoma (954) Total image: 9,792	IP: N/A AM: 1. Five augmented images were generated by rotating the original image five times by 60◦	ResNet152 UNet++	N/A	Model: DenseNet161 Before segmentation: Accuracy (90.5%) Recall (87.5%) loss (0.447) After segmentation: Accuracy (91.4%) Recall (90.0%) loss (0.536)	✓	✗	1.The proposed model generates many parameters which require more computational cost. 2.The segmentation is limited to TM region only. 3.Proper image analysis and broader region segmentation is missing.
Myburgh et al. [100]	Name:N/A Hospital name: Korea University Ansan Hospital Patient: N/A	1. Normal (1370) 2. OM disease (1227): (AOM, SOM, MOM, COM w/o P, COM w P, traumatic TM, Sclerosis TM, Tube, Chole) Total image:2597	IP: N/A AM: 1.Random horizontal flip	Mask R-CNN model	Five substructures (malleus, umbo, cone of light, pars flaccida, and annulus) of a normal TM image	Model: Proposed deep learning model (with combinations of five substructures) AUC (ranging from 0.905 to 0.932	✗	✗	1.No proper preprocessing and augmentation technique is employed. 2.Privacy and security concern since it integrates cloud server technique. 3.Do not highlight the acceptability and generalizability of the paper.

immittance in children with effusion-associated OM. It was discovered that absorbance, a specific measure of acoustic immittance, decreased systematically as effusion volume increased, especially in the 1–5 kHz frequency range. A multivariate logistic regression approach demonstrated high accuracy in classifying OM based on effusion presence and volume, with absorbance serving as a reliable indicator of middle-ear effusion volume in children with effusion-positive OM.

These studies indicate that tympanometry, specifically WBT in conjunction with AI and ML, can potentially improve the accuracy and efficacy of OM diagnosis. These technologies enable automated OM detection with high accuracy and the capability to differentiate between distinct OM classes. In addition, the combination of WBT measurements and otoscopy images enhances diagnostic certainty and agreement, thereby reducing the need for invasive procedures. AI tools based on tympanometry have the potential to provide valuable assistance to healthcare personnel, which may result in more precise and non-invasive OM diagnoses, thereby enhancing patient care if these AI tools' black box nature can be made interpretable and explainable.

VII. DATASETS

Table 11 presents a list of datasets used in various studies related to OM image classification. Table 11 shows that most datasets are private, and that images were collected from different hospitals and resources.

For example, dataset 16 is a private dataset that contains only 2820 images of ten classes (nine disease classes and a normal class). Dataset 13 contains 20542 images of eight classes. This dataset seems quite satisfactory compared to others in terms of diversity. In contrast, dataset 5 contains 22000 images (frames extracted from 200 auto endoscopy videos) but only four classes. So, there is a lack of variation in this dataset. In other cases, the dataset consists of a small number of images. As the third dataset is a public dataset, most researchers have used it for classification tasks [78], although the number of images is limited.

VIII. DISCUSSION

After conducting a comprehensive analysis of the 32 papers, several crucial insights and answers to our research questions.

A. RESEARCH QUESTION 1: EFFICACY OF AI MODELS IN DIAGNOSING OTITIS MEDIA

Our review suggests that ML and DL models, specifically Hybrid CNN and Transfer Learning techniques, are frequently utilized to classify OM. A smaller portion of the studies employ UNet architectures for segmentation purposes. The emphasis on these advanced methodologies highlights their potential to enhance the accuracy of diagnoses. In particular, Transfer Learning utilized pre-trained models on extensive datasets, enabling improved generalization on OM datasets even with limited data availability. These models have exhibited promising results in enhancing diagnostic accuracy, mitigating the inherent subjectivity associated with manual diagnoses, and facilitating early detection and treatment.

B. RESEARCH QUESTION 2: SIGNIFICANCE IN AI-BASED HEALTHCARE INFORMATICS

The role of image and video preprocessing is crucial for the successful implementation of AI models in healthcare informatics. Of the 32 papers reviewed, 19 (59.37%) included preprocessing, while 40.62% did not. The studies that incorporated preprocessing generally reported better performance metrics than those that did not. Preprocessing techniques, such as noise reduction, normalization, and augmentation, were found to improve the quality of the input data, thereby enhancing the robustness and accuracy of the models. This finding emphasizes the importance of integrating preprocessing steps in the analysis of medical images to ensure the development of reliable and effective diagnostic tools.

C. RESEARCH QUESTION 3: MODEL EXPLAINABILITY FEATURES

Enhancing the explainability of AI models is crucial for increasing trust and ensuring their suitability in clinical environments. However, only 7 out of 32 papers (21.87%) included explainability features such as Class Activation Maps (CAM) and feature map extraction. This low adoption rate emphasizes a significant gap in the current research landscape. For AI models to be implemented in real-world clinical practice, researchers must provide transparent and interpretable outcomes that medical professionals can understand and rely on. Therefore, future research efforts should prioritize integrating explainability features to mitigate this gap and facilitate broader acceptance and implementation of AI-powered diagnostic tools in healthcare.

D. RESEARCH QUESTION 4: MEDICAL EXPERT INVOLVEMENT

The involvement of medical experts in developing and validating AI models for OM diagnosis is limited, as seen from our literature analysis. The lack of medical expertise in many studies creates a disconnect between AI research and its practical application in clinical settings. Increased collaboration with medical professionals can enhance the relevance and accuracy of these models by incorporating domain-specific knowledge, validating model outputs, and ensuring that the developed tools meet clinical requirements and standards. This collaboration is essential for translating AI advancements into reliable, effective, widely adopted diagnostic instruments.

E. RESEARCH QUESTION 5: DATA QUALITY- ENSURING ACCURACY AND DIVERSITY

The quality and diversity of ear images utilized in training AI models are crucial for their effectiveness. Public datasets for Otitis Media diagnosis are limited, while private datasets are more common. These datasets include various modalities, such as images, videos, and tympanometry, which can enhance the training process. Ensuring data quality involves strict annotation protocols, multimodal data integration, and

TABLE 11. Dataset descriptions.

Dataset no	Source	Private / Public	Name of the diseases	No of images	Reference
1	Oto-endoscopic images (OEIs)	Private	normal, COM with TM perforation, and OME	2,871	[80]
2	Kamide ENT clinic, Shizouka, Japan	Private	acute otitis media, otitis media with effusion, and no effusion	1336	[81]
3	Özel Van Akdamar Hospital in Turkey between 10/2018 and 1/2019	Public	Normal, Abnormal (AOM, Earwax, Miringoskleroz, Tympanostomy tubes, CSOM, Otitis externa)	282	[78]
			Normal, Abnormal (AOM, CSOM, Earwax, Otitis externa, Tympanosclerosis, Ear ventilation tube, Pseudo-membranes, Foreign bodies in the ear)	956	[19]
			Normal, Abnormal (acute otitis media, otitis externa, chronic suppurative otitis media, cerumen impaction)	400	[82]
			Normal, AOM, CSOM, Ear wax, others	956	[83]
			AOM, CSOM, Earwax, Normal	857	[92]
4	Department of Otolaryngology – Head and Neck Surgery, Rutgers New Jersey Medical School, Newark, New Jersey, USA	Private	Normal, Cholesteatoma, Abnormal non-cholesteatoma	834	[86]
5	Otolaryngology department of the Clinical Hospital of the University of Chile (HCUCH) between 2018 and 2019.	Private	Normal, COM, OME, Ear wax plug	22000(from 200 videos)	[87]
			Normal, Earwax cases, Myringoesclerosis, COM	880	[93]
6	The otologic clinic in Asan Medical Center from Jan 2018 to Dec 2020.	Private	Category-1(OME, COM, none without OME and COM), Category-2(attic cholesteatoma, Meningitis, Ventilating tube, Otomycosis)	6630, 3833	[88]
7	Otolaryngology department for ear problems from January 2015 to December 2018	Private	Normal, OME, COM, Cholesteatoma	2372	[84]
8	(Welch Allyn Digital Macroview Otoscope, Welch Allyn Inc., Skaneateles Falls, New York, NY, USA) between March and May 2016.	Private	Normal, Pathological, Wax	273	[85]
9 10	Dr. Björn Åberg, Oslo, Norway and Dr. Thorbjörn Lundberg, Umeå, Sweden for providing us with high quality photographs from their respective TM image collections.	Private	Normal, Pathological, Wax	389	[85]
			O/W, n-TM, AOM, OME, CSOM with perforation	489	[35]
			Normal, OME, CSOM, W/O	389	[100]
11	Outpatient clinic in Severance Hospital otorhinolaryngology department from the year 2013 to 2017	Private	Normal, Tympanic perforation, Attic retraction, AOM or Myringitis or externa, OME, Middle ear or EAC tumor or cerumen impaction	10544	[91]
11	Department of otolaryngology in the people's hospital of Shenzhen Baoan District, from July 2016 to August 2019	Private	Normal eardrum,Cholestestoma of the middle ear, CSOM, External auditory cana bleeding, Impacted cerumen, Otomycosis external, Secretory otitis media, Tympanic membrane classification	20542	[94]
12	Seven ear-nose-throat (ENT) physicians of Cathay General Hospital.	Private	Normal, AOM, OME, COM	865	[95]
13 14	Collected by otologists from Karl Storz, Tullingen, Germany.	Private	Normal, OM (AOM, OME, COM)	1012	[96]
			Normal, OM	1139	[97]
15	Department of Otolaryngology at Taipei Veterans General Hospital(TVGH) in Taiwan from January 1st, 2011 to December 31st, 2019	Private	Normal, AOM, CSOM, OME, TM perforation, Acute myringitis, Cerumen impaction, Ventilation tube,TM retraction, Otomycosis	2820	[98]
15	Korea University Ansan Hospital	Private	Normal, Perforation, Retraction, Cholesteatoma	1632	[99]
16	Department of Otorhinolaryngology, Wonju Severance Christian Hospital, from 2015 to 2020	Private	Normal,OM disease(AOM, SOM, MOM, COM w/o P, COM w P, traumatic TM, Sclerosis TM,Tube, Chole)	2597	[100]

robust preprocessing techniques. High-quality and diverse datasets are essential for developing accurate and generalizable AI models that can be applied in various clinical scenarios.

F. RESEARCH QUESTION 6: WEB APPLICATION-INTEGRATION IN AI-BASED RESEARCH

Despite their potential advantages, only 12.5% of the analyzed papers incorporated web applications. Integrating web applications can improve the accessibility and usability of AI diagnostic tools, facilitating real-time analysis and broader implementation in clinical practice. Web-based platforms offer user-friendly interfaces for healthcare professionals, streamlining the diagnostic process and enabling remote consultations. Future research should prioritize the development of web applications to maximize the impact of AI-driven diagnostic tools, aligning with the need for increased involvement of medical experts and practical applicability.

The analysis of these 32 papers provides crucial insights into the current state and future directions of AI-based diagnostic tools for Otitis Media. While significant progress has been made in utilizing advanced ML and DL models, several gaps persist in preprocessing practices, model explainability, medical expert involvement, data quality, and web application integration. Addressing these gaps through focused research and development efforts is essential for effectively translating AI innovations into practical, reliable, and widely adopted healthcare tools.

IX. IMPACT OF AI MODELS ON DIAGNOSIS

AI models have significantly impacted medical diagnostics. Medical imaging is a key feature of this invention that has become integral to modern healthcare because it facilitates more precise diagnoses and informed treatment decisions. Some key effects of AI models on medical diagnostics:

- The integration of AI with medical imaging has ushered in a new era of precision and efficiency in diagnosis. DL algorithms in particular have emerged as a powerful tool for streamlining and enhancing image processing, transforming the landscape of medical diagnosis.
- AI algorithms are very good at identifying patterns, which is a necessary ability for the intricate interpretation of medical imaging. Through training on large datasets of annotated medical images, these algorithms may be trained to identify abnormalities and crucial patterns that would be imperceptible to the human eye.
- AI-powered image fermentation techniques facilitate the identification of specific structures or ROIs, which is highly advantageous for medical treatment and surgical planning.
- In massive datasets, AI algorithms can find biomarkers and genetic markers that impact susceptibility to illness and response to treatment. Clinicians can use this to develop more successful individualized treatment plans with fewer adverse effects. Improved AI-based image interpretation accuracy leads to a decrease in diagnostic errors, enabling more precise treatment and potentially lowering the need for antibiotic prescriptions.

Although AI has great potential to revolutionize healthcare, combining AI's innovative potential with healthcare's ethical standards is challenging, and it will need the combined efforts of engineers, healthcare professionals, lawmakers, and patients. When a doctor tries to diagnose the OM disease manually, there is a high chance of human error that could, directly and indirectly, affect the patients (child or adult), especially in medical prescriptions. AI could aid them in assessing otoscopy videos by converting them into frames and selecting clear TM frames using an automated system. In addition, AI can improve their efficiency and accuracy by giving medical professionals data-driven insights to help them make decisions and use their clinical knowledge to understand AI recommendations. This reduces the time and effort required and aids the patients and specialists in diagnosing OM diseases accurately. Thus, the advancement of AI technology can lead to improved OM diagnosis.

X. OPEN ISSUES AND FUTURE SCOPE FOR RESEARCH

A. OPEN ISSUES AND DISCUSSIONS

After extensive review of the papers related to otitis media, we have identified several unresolved questions and challenges in this research domain.

1) MODEL INTERPRETABILITY AND LACK OF MEDICAL EXPERT INVOLVEMENT

In the context of OM diagnosis using AI, we conducted a comprehensive review of 32 papers, revealing a pervasive difficulty in applying ML and DL models within the medical domain to diagnose ear infections. This challenge centers on model interpretability, as these models often operate as inscrutable "black boxes." This dearth of transparency is a major barrier for healthcare professionals who are seeking insight into the decision-making process underlying the models. Given the critical implications of healthcare decisions on patients' well-being, this raises legitimate concerns about the reliability and safety of these AI models for the real-life diagnosis process. Interpretable models are crucial because they equip doctors with the knowledge to use and effectively employ these technologies when diagnosing OM. In this context, informed decision-making requires a comprehensive understanding of how these models generate their recommendations. In addition, interpretability plays a crucial role in ethical oversight, ensuring that irrelevant criteria or biases do not exert an inordinate influence on ML and DL algorithmic decisions. Notably, within the scope of OM diagnosis, the absence of explainable models in the 32 reviewed papers raises the question of whether AI is readily implementable in real life since the studies focus on ensuring accuracy rather than discovering the explain ability of the proposed model. As discussed in this review paper, significant progress has been made in the domain of AI to diagnose OM. However, the models developed by data scientists frequently lack explainability, as discussed, leading to challenges in collaboration between medical experts and the AI community. Because the results produced by these advanced models often lack clear understanding, posing a challenge for healthcare practitioners to understand the fundamental mechanisms that decide the outcomes. The lack of transparency in this model hinders efficient collaboration among computer scientists, data scientists, and medical professionals, thereby impeding the progress and implementation of AI models in the healthcare sector. Medical professionals are concerned about the absence of transparency, which causes them to be hesitant to place trust in the outcomes generated by these models. As a result, the incorporation of artificial intelligence into medical practice and research has been constrained, emphasizing the critical requirement for AI models that are far more transparent and interpretable to promote collaboration and establish confidence among healthcare professionals. The limited participation of medical experts in the research process, perpetuates the dominance of applied research within the IT context and emphasizes the critical need for more transparent and collaborative approaches in this domain for knowledge discovery.

2) LACK OF KNOWLEDGE DISCOVERY

The lack of interpretability and the limited involvement of medical experts have created a significant barrier to knowledge discovery. This deficiency is especially apparent when contemplating the potential insights that could be obtained through collaboration between clinicians and AI specialists. For instance, if clinicians actively participated and shared their knowledge on distinguishing the position, degree of bulging, and other clinically relevant information related to eardrum abnormalities in medical images, AI experts could use this valuable information to develop marker-based approaches and facilitate knowledge discovery. Such collaborations could result in the development of more robust decision support systems that incorporate a deeper understanding of the medical nuances, which is presently absent and remains an open issue in this field.

B. FUTURE WORK

The future work outlined here holds significant promise in advancing the application of ML and DL in the diagnosis of OM, with potential benefits spanning various aspects of healthcare:

1) IMPROVED MODEL INTERPRETABILITY

Enhancing Improved Model Interpretability: Enhancing the interpretability of ML and DL models can have profound benefits. By incorporating explainable AI techniques and developing visualization tools, we can demystify the decision-making process of complex algorithms. For instance, to address middle ear issues using temporal bone computed tomography, an automated segmentation technique can be employed to segment the significant ROI from ear images. This involves implementing a 3D convolutional neural network to extract the 3D ROI and diagnose pathological ears and disease classes. Additionally, gradient-weighted class activation mapping techniques can be utilized to generate heatmaps that emphasize the crucial ROI of the middle ear, aiding in decision-making processes. This will help clinicians to trust these models and empower them to understand the reasoning behind the AI-generated recommendations. With improved model interpretability, clinicians can confidently integrate AI suggestions into their clinical decision-making processes, such as on which basis the models give predictions. Additionally, it will lead to the real-time AI assessment of an image, ultimately leading to more accurate and efficient diagnoses and treatment plans.

2) INVOLVEMENT OF MEDICAL EXPERTS

Collaborating closely with medical experts is crucial to improving model performance and accuracy. Medical professionals possess valuable domain knowledge that can refine data annotation and region of interest (ROI) segmentation, ensuring that training data aligns closely with real-world clinical scenarios. This partnership enhances the models' ability to detect diseases accurately, reduce false positives, and minimize biases. Ultimately, it results in AI systems that are more reliable and clinically relevant, instilling trust among healthcare providers and facilitating seamless integration into medical practice.

3) ENHANCING DATA QUALITY

Critical to the performance of ML and DL models is the quality of the data used for training. Hence, it is important to ascertain that the aural images are precise, comprehensive, and reflective of practical clinical situations. In the case of OM, this may entail gathering data from diverse populations and ensuring that the data encompasses a diversity of OM diagnoses, informed by national OM guidelines. Currently very few studies utilize more than four or five-class labeling.

4) HYPERPARAMETER TUNING

Hyperparameters are pre-training parameters that determine aspects of a model, including the learning rate and the number of layers. By fine-tuning these hyperparameters, the complexity of the model can be reduced. It is evident that most of the studies did not involve extensive hyperparameter tuning. Additionally, studies have not included hyperparameter optimization algorithms to improve their model. This can be achieved using techniques such as grid or random search and will eventually produce a more optimal model for OM diagnosis utilizing AI.

Implementing these recommendations has the potential to revolutionize the medical field by bridging the gap between cutting-edge ML and DL technologies and their practical application in healthcare. Improved model interpretability, enhanced medical expert involvement, and interdisciplinary collaboration are essential in developing AI systems that optimize OM diagnosis and contribute to better patient outcomes, reduced healthcare costs, and more efficient healthcare delivery. Ultimately, these advances will benefit both patients and healthcare providers by providing them with powerful and trustworthy tools to improve healthcare quality and accessibility.

XI. CONCLUSION

Otitis Media is a global public health issue with a notably high prevalence, tragically resulting in serious health concerns in multiple age groups. Studies employing various ML and DL methodologies in the field of OM diagnosis have shown that the development of AI may facilitate disease detection. This review is based on extensive literature research related to the use of AI techniques to diagnose OM. The review examines the applications of AI-based models for various modalities, including otoscopy images, videos, and tympanometry. In addition, it casts light on existing OM datasets by revealing their limitations and constraints.

It can be concluded that previous research employing AI techniques has primarily focused on image-based applications. Despite the potential importance for real-time OM detection, there has been limited exploration of video data. Interpretability of models needs to be further explored and collaboration with medical experts is crucial in these research endeavors.

To improve the applicability of future research, it is necessary to develop more interpretable model disease detection systems that operate on both images and video recordings, as well as otoscopy and tympanometry, and involve medical professionals in knowledge discovery. Such an approach has the potential to significantly enhance the applicability of this research to actual healthcare settings.

REFERENCES

[1] A. J. Leach and P. S. Morris, "Antibiotics for the prevention of acute and chronic suppurative otitis media in children," *Cochrane Database Systematic Rev.*, pp. 1–16, Oct. 2006.

[2] H. Joe and Y. J. Seo, "A newly designed tympanostomy stent with TiO_2 coating to reduce pseudomonas aeruginosa biofilm formation," *J. Biomaterials Appl.*, vol. 33, no. 4, pp. 599–605, Oct. 2018.

[3] A. Danishyar and J. V. Ashurst, *Acute Otitis Media*. USA: StatPearls Publishing, 2017.

[4] R. M. Rosenfeld, "Evidence-based otitis media," PMPH, Shelton, CT, USA, Tech. Rep., 2003.

[5] C. J. Williams and A. M. Jacobs, "The impact of otitis media on cognitive and educational outcomes," *Med. J. Aust.*, vol. 191, no. S9, pp. S69–S72, Nov. 2009.

[6] O. Hong, M. J. Kerr, G. L. Poling, and S. Dhar, "Understanding and preventing noise-induced hearing loss," *Disease-a-Month*, vol. 59, no. 4, pp. 110–118, Apr. 2013.

[7] J. Roberts, L. Hunter, J. Gravel, R. Rosenfeld, S. Berman, M. Haggard, J. Hall, C. Lannon, D. Moore, L. Vernon-Feagans, and I. Wallace, "Otitis media, hearing loss, and language learning: Controversies and current research," *J. Develop. Behav. Pediatrics*, vol. 25, no. 2, pp. 110–122, Apr. 2004.

[8] M. W. Casby, "Otitis media and language development: A meta-analysis," *Amer. J. Speech-Language Pathol.*, vol. 10, no. 1, pp. 65–80, Feb. 2001.

[9] N. H. Davidoss, Y. K. Varsak, and P. L. S. Maria, "Animal models of acute otitis media—A review with practical implications for laboratory research," *Eur. Ann. Otorhinolaryngology, Head Neck Diseases*, vol. 135, no. 3, pp. 183–190, Jun. 2018.

[10] J. Pitaro, S. Waissbluth, M.-C. Quintal, A. Abela, and A. Lapointe, "Characteristics of children with refractory acute otitis media treated at the pediatric emergency department," *Int. J. Pediatric Otorhinolaryngol.*, vol. 116, pp. 173–176, Jan. 2019.

[11] E. Roy, K. Z. Hasan, F. Haque, A. Siddique, and R. B. Sack, "Acute otitis media during the first two years of life in a rural community in Bangladesh: A prospective cohort study," *J. Health, Population, Nutrition*, vol. 25, no. 4, p. 414, 2007.

[12] G. D. Castillo-Aguas, C. García-Vera, J. Urkin, M. Moretto, M. V. Spreitzer, P. Keronen, A. Werner, L. Reali, K. Geitmann, P. Poloskey, B. Kartousova, W. Sauseng, M. Schumacher, S. Reingold, and C. Sánchez-Pina, "Pediatricians? Attitudes in management of acute otitis media and ear pain in Turkey," *Int. J. Pediatric Otorhinolaryngol.*, vol. 107, pp. 14–20, Apr. 2018.

[13] K. Kitamura, Y. Iino, Y. Kamide, F. Kudo, T. Nakayama, K. Suzuki, H. Taiji, H. Takahashi, N. Yamanaka, and Y. Uno, "Clinical practice guidelines for the diagnosis and management of acute otitis media (AOM) in children in Japan-2013 update," *Auris Nasus Larynx*, vol. 42, no. 2, pp. 99–106, 2013.

[14] S. Shah-Becker and M. M. Carr, "Current management and referral patterns of pediatricians for acute otitis media," *Int. J. Pediatric Otorhinolaryngol.*, vol. 113, pp. 19–21, Oct. 2018.

[15] R. Karli, A. Karli, A. Aksoy, and S. Açikgöz, "Orta kulak efüzyonlarýnda timpanogram ile otoskopik bulgularýn karşılaştırılması," *Dicle Tıp Dergisi*, vol. 40, no. 1, pp. 54–56, 2013.

[16] T. Els and I. P. Olwoch, "The prevalence and impact of otitis media with effusion in children admitted for adeno-tonsillectomy at Dr George Mukhari Academic Hospital, Pretoria, South Africa," *Int. J. Pediatric Otorhinolaryngol.*, vol. 110, pp. 76–80, Jul. 2018.

[17] N. S. Tsilis, P. V. Vlastarakos, V. F. Chalkiadakis, D. S. Kotzampasakis, and T. P. Nikolopoulos, "Chronic otitis media in children: An evidence-based guide for diagnosis and management," *Clin. Pediatrics*, vol. 52, no. 9, pp. 795–802, Sep. 2013.

[18] P. S. Morris, A. J. Leach, P. Silberberg, G. Mellon, C. Wilson, E. Hamilton, and J. Beissbarth, "Otitis media in young aboriginal children from remote communities in Northern and Central Australia: A cross-sectional survey," *BMC Pediatrics*, vol. 5, no. 1, pp. 1–10, Dec. 2005.

[19] A. Çalişkan, "Classification of tympanic membrane images based on VGG16 model," *Kocaeli J. Sci. Eng.*, vol. 5, no. 1, pp. 105–111, May 2022.

[20] M. E. Pichichero, "Diagnostic accuracy of otitis media and tympanocentesis skills assessment among pediatricians," *Eur. J. Clin. Microbiol. Infectious Diseases*, vol. 22, no. 9, pp. 519–524, Sep. 2003.

[21] L. Monasta, L. Ronfani, F. Marchetti, M. Montico, L. V. Brumatti, A. Bavcar, D. Grasso, C. Barbiero, and G. Tamburlini, "Burden of disease caused by otitis media: Systematic review and global estimates," *PLoS ONE*, vol. 7, no. 4, Apr. 2012, Art. no. e36226.

[22] S. Functions. *World Report on Hearing*. World Health Org. Accessed: Oct. 9, 2023. [Online]. Available: https://www.who.int/publications/i/item/world-report-on-hearing

[23] A. Yar Muhammad, K. Anwar, and S. Khan, "The hearing impaired child-pattern of tympanograms in otitis media with effusion," *J. Saidu Med. College, Swat*, vol. 13, no. 2, pp. 34–39, Jun. 2023.

[24] C.-H. Tseng, Y.-S. Sung, and S. Fuh, "An eardrum image capture guidance program for the otoscope," in *Proc. IPPR Conf. Comput. Vis., Graph., Image Process.*, Tainan, Taiwan, 2018, pp. 1–8, Paper 19.

[25] J. V. Sundgaard, P. Bray, S. Laugesen, J. Harte, Y. Kamide, C. Tanaka, A. N. Christensen, and R. R. Paulsen, "A deep learning approach for detecting otitis media from wideband tympanometry measurements," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 7, pp. 2974–2982, Jul. 2022.

[26] N. Shirai and D. Preciado, "Otitis media: What is new?" *Current Opinion Otolaryngol. Head Neck Surg.*, vol. 27, no. 6, pp. 495–498, 2019.

[27] A. M. Bur, M. Shew, and J. New, "Artificial intelligence for the otolaryngologist: A state of the art review," *Otolaryngol.-Head Neck Surg.*, vol. 160, no. 4, pp. 603–611, Apr. 2019.

[28] E. You, V. Lin, T. Mijovic, A. Eskander, and M. G. Crowson, "Artificial intelligence applications in otology: A state of the art review," *Otolaryngol.-Head Neck Surg.*, vol. 163, no. 6, pp. 1123–1133, Dec. 2020.

[29] Z. Cao, F. Chen, E. M. Grais, F. Yue, Y. Cai, D. W. Swanepoel, and F. Zhao, "Machine learning in diagnosing middle ear disorders using tympanic membrane images: A meta-analysis," *Laryngoscope*, vol. 133, no. 4, pp. 732–741, Apr. 2023.

[30] S. Ngombu, H. Binol, M. N. Gurcan, and A. C. Moberly, "Advances in artificial intelligence to diagnose otitis media: State of the art review," *Otolaryngol.-Head Neck Surg.*, vol. 168, no. 4, pp. 635–642, Apr. 2023.

[31] S. Esposito, S. Bianchini, A. Argentiero, R. Gobbi, C. Vicini, and N. Principi, "New approaches and technologies to improve accuracy of acute otitis media diagnosis," *Diagnostics*, vol. 11, no. 12, p. 2392, Dec. 2021.

[32] X. Ding, Y. Huang, X. Tian, Y. Zhao, G. Feng, and Z. Gao, "Diagnosis, treatment, and management of otitis media with artificial intelligence," *Diagnostics*, vol. 13, no. 13, p. 2309, Jul. 2023.

[33] D. Song, T. Kim, Y. Lee, and J. Kim, "Image-based artificial intelligence technology for diagnosing middle ear diseases: A systematic review," *J. Clin. Med.*, vol. 12, no. 18, p. 5831, Sep. 2023.

[34] M. A. K. Raiaan, K. Fatema, I. U. Khan, S. Azam, M. R. U. Rashid, M. S. H. Mukta, M. Jonkman, and F. De Boer, "A lightweight robust deep learning model gained high accuracy in classifying a wide range of diabetic retinopathy images," *IEEE Access*, vol. 11, pp. 42361–42388, 2023.

[35] H. C. Myburgh, W. H. van Zijl, D. Swanepoel, S. Hellström, and C. Laurent, "Otitis media diagnosis for developing countries using tympanic membrane image-analysis," *EBioMedicine*, vol. 5, pp. 156–160, Mar. 2016.

[36] *Ear and Hearing Problems*. Accessed: Oct. 9, 2023. [Online]. Available: https://remotephcmanuals.org.au/document/35732.html
[37] O. Duggineni. *OtoscopeData*. Kaggle. Accessed: Oct. 9, 2023. [Online]. Available: https://www.kaggle.com/datasets/omduggineni/otoscopedata
[38] *Air-Fluid Levels in Ear*. Accessed: Oct. 9, 2023. [Online]. Available: https://www.mdedge.com/familymedicine/article/77660/air-fluid-levels-ear
[39] G. S. Handelman, H. K. Kok, R. V. Chandra, A. H. Razavi, M. J. Lee, and H. Asadi, ''EDoctor: Machine learning and the future of medicine,'' *J. Internal Med.*, vol. 284, no. 6, pp. 603–619, Dec. 2018.
[40] D. Chicco, ''Ten quick tips for machine learning in computational biology,'' *BioData Mining*, vol. 10, no. 1, p. 35, Dec. 2017.
[41] A. Rajkomar, J. Dean, and I. Kohane, ''Machine learning in medicine,'' *New England J. Med.*, vol. 380, no. 14, pp. 1347–1358, 2019.
[42] R. C. Deo, ''Machine learning in medicine,'' *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.
[43] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, ''High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images,'' *Comput. Biol. Med.*, vol. 155, Mar. 2023, Art. no. 106646.
[44] E. Akleman, ''Deep learning,'' *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[45] S. Azam, S. Montaha, M. A. K. Raiaan, A. K. M. R. H. Rafid, S. H. Mukta, and M. Jonkman, ''An automated decision support system to analyze malignancy patterns of breast masses employing medically relevant features of ultrasound images,'' *J. Imag. Informat. Med.*, vol. 37, no. 1, pp. 45–59, Jan. 2024.
[46] M. A. K. Raiaan, N. M. Fahad, M. S. H. Mukta, and S. Shatabda, ''Mammo-light: A lightweight convolutional neural network for diagnosing breast cancer from mammography images,'' *Biomed. Signal Process. Control*, vol. 94, Aug. 2024, Art. no. 106279, doi: 10.1016/j.bspc.2024.106279.
[47] A. Roy and S. Chakraborty, ''Support vector machine in structural reliability analysis: A review,'' *Rel. Eng. Syst. Saf.*, vol. 233, May 2023, Art. no. 109126.
[48] A. X. Wang, S. S. Chukova, and B. P. Nguyen, ''Ensemble k-nearest neighbors based on centroid displacement,'' *Inf. Sci.*, vol. 629, pp. 313–323, Jun. 2023.
[49] J. Guo, S. Yun, Y. Meng, N. He, D. Ye, Z. Zhao, L. Jia, and L. Yang, ''Prediction of heating and cooling loads based on light gradient boosting machine algorithms,'' *Building Environ.*, vol. 236, May 2023, Art. no. 110252.
[50] L. Deng and D. Yu, ''Deep learning: Methods and applications,'' *Found. Trends Signal Process.*, vol. 7, nos. 3–4, pp. 197–387, 2014.
[51] H. A. Song and S.-Y. Lee, ''Hierarchical representation using NMF,'' in *Proc. 20th Int. Conf. Neural Inf. Process. (ICONIP)*, Daegu, Korea. Cham, Switzerland: Springer, Nov. 2013, pp. 466–473.
[52] Q. Zhang, D. Zhao, and X. Chi, ''XXXX (Review for deep learning based on medical imaging diagnosis),'' *XXXX*, vol. 44, no. Z11, pp. 1–7, 2017.
[53] A. M. Fayyaz, M. I. Sharif, S. Azam, A. Karim, and J. El-Den, ''Analysis of diabetic retinopathy (DR) based on the deep learning,'' *Information*, vol. 14, no. 1, p. 30, Jan. 2023.
[54] P. Ghosh, S. Azam, R. Quadir, A. Karim, F. M. J. M. Shamrat, S. K. Bhowmik, M. Jonkman, K. M. Hasib, and K. Ahmed, ''SkinNet-16: A deep learning approach to identify benign and malignant skin lesions,'' *Frontiers Oncol.*, vol. 12, Aug. 2022, Art. no. 931141.
[55] N. M. Fahad, S. Sakib, M. A. K. Raiaan, and M. S. H. Mukta, ''SkinNet-8: An efficient CNN architecture for classifying skin cancer on an imbalanced dataset,'' in *Proc. Int. Conf. Electr., Comput. Commun. Eng. (ECCE)*, Feb. 2023, pp. 1–6.
[56] M. A. K. Raiaan, S. Sakib, N. M. Fahad, A. A. Mamun, M. A. Rahman, S. Shatabda, and M. S. H. Mukta, ''A systematic review of hyperparameter optimization techniques in convolutional neural networks,'' *Decis. Analytics J.*, vol. 11, Jun. 2024, Art. no. 100470.
[57] M. A. K. Raiaan, M. S. H. Mukta, K. Fatema, N. M. Fahad, S. Sakib, M. M. J. Mim, J. Ahmad, M. E. Ali, and S. Azam, ''A review on large language models: Architectures, applications, taxonomies, open issues and challenges,'' *IEEE Access*, vol. 12, pp. 26839–26874, 2024.
[58] I. U. Khan, M. A. K. Raiaan, K. Fatema, S. Azam, R. U. Rashid, S. H. Mukta, M. Jonkman, and F. De Boer, ''A computer-aided diagnostic system to identify diabetic retinopathy, utilizing a modified compact convolutional transformer and low-resolution images to reduce computation time,'' *Biomedicines*, vol. 11, no. 6, p. 1566, May 2023.
[59] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, ''Rethinking the inception architecture for computer vision,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2818–2826.
[60] S. Azam, S. Montaha, K. U. Fahim, A. K. M. R. H. Rafid, M. S. H. Mukta, and M. Jonkman, ''Using feature maps to unpack the CNN 'black box' theory with two medical datasets of different modality,'' *Intell. Syst. with Appl.*, vol. 18, May 2023, Art. no. 200233.
[61] A. Aimar, H. Mostafa, E. Calabrese, A. Rios-Navarro, R. Tapiador-Morales, I.-A. Lungu, M. B. Milde, F. Corradi, A. Linares-Barranco, S.-C. Liu, and T. Delbruck, ''NullHop: A flexible convolutional neural network accelerator based on sparse representations of feature maps,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 3, pp. 644–656, Mar. 2019.
[62] M. Sarıgül, B. M. Ozyildirim, and M. Avci, ''Differential convolutional neural network,'' *Neural Netw.*, vol. 116, pp. 279–287, Aug. 2019.
[63] M. A. K. Raiaan, N. M. Fahad, S. Chowdhury, D. Sutradhar, S. S. Mihad, and M. M. Islam, ''IoT-based object-detection system to safeguard endangered animals and bolster agricultural farm security,'' *Future Internet*, vol. 15, no. 12, p. 372, Nov. 2023.
[64] M. Al-Qizwini, I. Barjasteh, H. Al-Qassab, and H. Radha, ''Deep learning algorithm for autonomous driving using GoogLeNet,'' in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2017, pp. 89–96.
[65] K. Fatema, M. A. H. Rony, S. Azam, M. S. H. Mukta, A. Karim, M. Z. Hasan, and M. Jonkman, ''Development of an automated optimal distance feature-based decision system for diagnosing knee osteoarthritis using segmented X-ray images,'' *Heliyon*, vol. 9, no. 11, Nov. 2023, Art. no. e21703.
[66] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, and J. J. Kang, ''Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM,'' *Sensors*, vol. 21, no. 8, p. 2852, Apr. 2021.
[67] C. A. Ferreira, T. Melo, P. Sousa, M. I. Meyer, E. Shakibapour, P. Costa, and A. Campilho, ''Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2,'' in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer, May 2018, pp. 763–770.
[68] Y. Zhu and S. Newsam, ''DenseNet for dense flow,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 790–794.
[69] M. Z. Hasan, S. Montaha, I. U. Khan, M. M. Hassan, A. A. Mahmud, A. K. M. R. H. Rafid, S. Azam, A. Karim, S. Prountzos, E. Alexopoulou, U. B. Ashraf, and S. M. S. Islam, ''Fast and efficient lung abnormality identification with explainable AI: A comprehensive framework for chest CT scan and X-ray images,'' *IEEE Access*, vol. 12, pp. 31117–31135, 2024.
[70] S. Haapala, E. Niemitalo-Haapola, A. Raappana, T. Kujala, K. Suominen, E. Jansson-Verkasalo, and T. Kujala, ''Long-term influence of recurrent acute otitis media on neural involuntary attention switching in 2-year-old children,'' *Behav. Brain Functions*, vol. 12, no. 1, pp. 1–8, Dec. 2015.
[71] A. D. Jotic, A. M. Opankovic, Z. Z. Radin, L. Cvorovic, K. R. S. Vujovic, S. B. Krejovic-Trivic, B. M. Bukurov, B. R. Milicic, and J. D. Stojanovic, ''Symptoms of depression, anxiety and stress in patients with chronic otitis media,'' *PLoS ONE*, vol. 17, no. 7, Jul. 2022, Art. no. e0270793.
[72] M. Z. I. Ahmed, N. Sinha, E. Ghaderpour, S. Phadikar, and R. Ghosh, ''A novel baseline removal paradigm for subject-independent features in emotion classification using EEG,'' *Bioengineering*, vol. 10, no. 1, p. 54, Jan. 2023.
[73] R. Ikeda, H. Hidaka, M. Ito, Y. Kamide, H. Kuroki, A. Nakano, H. Yoshida, H. Takahashi, Y. Iino, Y. Harabuchi, and H. Kobayashi, ''Pharmacotherapy focusing on for the management of otitis media with effusion in children: Systematic review and meta-analysis,'' *Auris Nasus Larynx*, vol. 49, no. 5, pp. 748–754, Oct. 2022.
[74] T. Marom, J. Pitaro, U. K. Shah, S. Torretta, P. Marchisio, A. T. Kumar, P. C. Barth, and S. O. Tamir, ''Otitis media practice during the COVID-19 pandemic,'' *Frontiers Cellular Infection Microbiol.*, vol. 11, Jan. 2022, Art. no. 749911.
[75] B. Kitchenham, ''Procedures for performing systematic reviews,'' *Keele, U.K., Keele Univ.*, vol. 33, pp. 1–26, Jul. 2004.
[76] D. Elangovan, C. S. Long, F. S. Bakrin, C. S. Tan, K. W. Goh, S. F. Yeoh, M. J. Loy, Z. Hussain, K. S. Lee, A. C. Idris, and L. C. Ming, ''The use of blockchain technology in the health care sector: Systematic review,'' *JMIR Med. Informat.*, vol. 10, no. 1, Jan. 2022, Art. no. e17278.

[77] I. C. A. Pilares, S. Azam, S. Akbulut, M. Jonkman, and B. Shanmugam, "Addressing the challenges of electronic health records using blockchain and IPFS," *Sensors*, vol. 22, no. 11, p. 4032, May 2022.

[78] E. Başaran, Z. Cömert, and Y. Çelik, "Convolutional neural network approach for automatic tympanic membrane detection and classification," *Biomed. Signal Process. Control*, vol. 56, Feb. 2020, Art. no. 101734.

[79] D. Livingstone, A. S. Talai, J. Chau, and N. D. Forkert, "Building an otoscopic screening prototype tool using deep learning," *J. Otolaryngol. Head Neck Surg.*, vol. 48, no. 1, pp. 1–5, Dec. 2019.

[80] M. A. Khan, S. Kwon, J. Choo, S. M. Hong, S. H. Kang, I.-H. Park, S. K. Kim, and S. J. Hong, "Automatic detection of tympanic membrane and middle ear infection from oto-endoscopic images via convolutional neural networks," *Neural Netw.*, vol. 126, pp. 384–394, Jun. 2020.

[81] J. V. Sundgaard, J. Harte, P. Bray, S. Laugesen, Y. Kamide, C. Tanaka, R. R. Paulsen, and A. N. Christensen, "Deep metric learning for otitis media classification," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102034.

[82] K. Tsutsumi, K. Goshtasbi, A. Risbud, P. Khosravi, J. C. Pang, H. W. Lin, H. R. Djalilian, and M. Abouzari, "A web-based deep learning model for automated diagnosis of otoscopic images," *Otology Neurotol.*, vol. 42, no. 9, pp. e1382–e1388, 2021.

[83] A. Alhudhaif, Z. Cömert, and K. Polat, "Otitis media detection using tympanic membrane images with a novel multi-class machine learning algorithm," *PeerJ Comput. Sci.*, vol. 7, p. e405, Feb. 2021.

[84] H. Byun, S. Yu, J. Oh, J. Bae, M. S. Yoon, S. H. Lee, J. H. Chung, and T. H. Kim, "An assistive role of a machine learning network in diagnosis of middle ear diseases," *J. Clin. Med.*, vol. 10, no. 15, p. 3198, Jul. 2021.

[85] J. Sandström, H. Myburgh, C. Laurent, D. W. Swanepoel, and T. Lundberg, "A machine learning approach to screen for otitis media using digital otoscope images labelled by an expert panel," *Diagnostics*, vol. 12, no. 6, p. 1318, May 2022.

[86] C. C. Tseng, V. Lim, and R. W. Jyung, "Use of artificial intelligence for the diagnosis of cholesteatoma," *Laryngoscope Investigative Otolaryngol.*, vol. 8, no. 1, pp. 201–211, Feb. 2023.

[87] M. Viscaino, M. Talamilla, J. C. Maass, P. Henríquez, P. H. Délano, C. Auat Cheein, and F. Auat Cheein, "Color dependence analysis in a CNN-based computer-aided diagnosis system for middle and external ear diseases," *Diagnostics*, vol. 12, no. 4, p. 917, Apr. 2022.

[88] Y. Choi, J. Chae, K. Park, J. Hur, J. Kweon, and J. H. Ahn, "Automated multi-class classification for prediction of tympanic membrane changes with deep learning models," *PLoS ONE*, vol. 17, no. 10, Oct. 2022, Art. no. e0275846.

[89] J. Zeng, W. Deng, J. Yu, L. Xiao, S. Chen, X. Zhang, L. Zeng, D. Chen, P. Li, Y. Chen, H. Zhang, F. Shu, M. Wu, Y. Su, Y. Li, Y. Cai, and Y. Zheng, "A deep learning approach to the diagnosis of atelectasis and attic retraction pocket in otitis media with effusion using otoscopic images," *Eur. Arch. Oto-Rhino-Laryngol.*, vol. 280, no. 4, pp. 1621–1627, Apr. 2023.

[90] A.-R. Habib, Y. Xu, K. Bock, S. Mohanty, T. Sederholm, W. B. Weeks, R. Dodhia, J. L. Ferres, C. Perry, R. Sacks, and N. Singh, "Evaluating the generalizability of deep learning image classification algorithms to detect middle ear disease using otoscopy," *Sci. Rep.*, vol. 13, no. 1, p. 5368, Apr. 2023.

[91] D. Cha, C. Pae, S.-B. Seong, J. Y. Choi, and H.-J. Park, "Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database," *EBioMedicine*, vol. 45, pp. 606–614, Jul. 2019.

[92] C. Zafer, "Fusing fine-tuned deep features for recognizing different tympanic membranes," *Biocybernetics Biomed. Eng.*, vol. 40, no. 1, pp. 40–51, Jan. 2020.

[93] M. Viscaino, J. C. Maass, P. H. Delano, M. Torrente, C. Stott, and F. A. Cheein, "Computer-aided diagnosis of external and middle ear conditions: A machine learning approach," *PLoS ONE*, vol. 15, no. 3, Mar. 2020, Art. no. e0229226.

[94] X. Zeng, Z. Jiang, W. Luo, H. Li, H. Li, G. Li, J. Shi, K. Wu, T. Liu, X. Lin, F. Wang, and Z. Li, "Efficient and accurate identification of ear diseases using an ensemble deep learning model," *Sci. Rep.*, vol. 11, no. 1, p. 10839, May 2021.

[95] C.-K. Shie, H.-T. Chang, F.-C. Fan, C.-J. Chen, T.-Y. Fang, and P.-C. Wang, "A hybrid feature-based segmentation and classification system for the computer aided self-diagnosis of otitis media," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2014, pp. 4655–4658.

[96] V.-T. Pham, T.-T. Tran, P.-C. Wang, P.-Y. Chen, and M.-T. Lo, "EAR-UNet: A deep learning-based approach for segmentation of tympanic membranes from otoscopic images," *Artif. Intell. Med.*, vol. 115, May 2021, Art. no. 102065.

[97] V.-T. Pham, T.-T. Tran, P.-C. Wang, and M.-T. Lo, "Tympanic membrane segmentation in otoscopic images based on fully convolutional network with active contour loss," *Signal, Image Video Process.*, vol. 15, no. 3, pp. 519–527, Apr. 2021.

[98] Y.-C. Chen, Y.-C. Chu, C.-Y. Huang, Y.-T. Lee, W.-Y. Lee, C.-Y. Hsu, A. C. Yang, W.-H. Liao, and Y.-F. Cheng, "Smartphone-based artificial intelligence using a transfer learning algorithm for the detection and diagnosis of middle ear diseases: A retrospective deep learning study," *eClinicalMedicine*, vol. 51, Sep. 2022, Art. no. 101543.

[99] T. Kim, K. Oh, J. Kim, Y. Lee, and J. Choi, "Development of ResNet152 UNet++-based segmentation algorithm for the tympanic membrane and affected areas," *IEEE Access*, vol. 11, pp. 56225–56234, 2023.

[100] H. C. Myburgh, S. Jose, D. W. Swanepoel, and C. Laurent, "Towards low cost automated smartphone- and cloud-based otitis media diagnosis," *Biomed. Signal Process. Control*, vol. 39, pp. 34–52, Jan. 2018.

[101] H.-P. Chan, R. Samala, L. Hadjiiski, and C. Zhou, "Deep learning in medical image analysis," *Adv. Exp. Med. Biol.*, vol. 1213, pp. 3–21, Feb. 2020, doi: 10.1007/978-3-030-33128-3_1.

[102] M. Tozar, E. Cömert, Z. Şencan, G. Şimşek, N. B. Muluk, and R. Kılıç, "Video head impulse test in children with otitis media with effusion and dizziness," *Int. J. Pediatric Otorhinolaryngol.*, vol. 129, Feb. 2020, Art. no. 109783.

[103] K. N. Chan, A. Silverstein, L. N. Bryan, C. E. McCracken, W. K. Little, and A. L. Shane, "Comparison of a smartphone otoscope and conventional otoscope in the diagnosis and management of acute otitis media," *Clin. Pediatrics*, vol. 58, no. 3, pp. 302–306, Mar. 2019.

[104] O. Moshtaghi, R. Sahyouni, Y. M. Haidar, M. Huang, A. Moshtaghi, Y. Ghavami, H. W. Lin, and H. R. Djalilian, "Smartphone-enabled otoscopy in neurotology/otology," *Otolaryngol.-Head Neck Surg.*, vol. 156, no. 3, pp. 554–558, 2017.

[105] E. M. Grais, X. Wang, J. Wang, F. Zhao, W. Jiang, Y. Cai, L. Zhang, W. Lin, and H. Yang, "The use of machine learning tools to analyse wideband absorbance immittance in normal and ears with otitis media with effusion," *SSRN Electron. J.*, pp. 1–12, Nov. 2020.

[106] E. M. Grais, X. Wang, J. Wang, F. Zhao, W. Jiang, Y. Cai, L. Zhang, Q. Lin, and H. Yang, "Analysing wideband absorbance immittance in normal and ears with otitis media with effusion using machine learning," *Sci. Rep.*, vol. 11, no. 1, p. 10643, May 2021.

[107] J. V. Sundgaard, M. Värendh, F. Nordström, Y. Kamide, C. Tanaka, J. Harte, R. R. Paulsen, A. N. Christensen, P. Bray, and S. Laugesen, "Inter-rater reliability of the diagnosis of otitis media based on otoscopic images and wideband tympanometry measurements," *Int. J. Pediatric Otorhinolaryngol.*, vol. 153, Feb. 2022, Art. no. 111034.

[108] G. R. Merchant, S. Al-Salim, R. M. Tempero, D. Fitzpatrick, and S. T. Neely, "Improving the differential diagnosis of otitis media with effusion using wideband acoustic immittance," *Ear Hearing*, vol. 42, no. 5, pp. 1183–1194, Sep. 2021.