RESEARCH ARTICLE

Artificial Intelligence With Deep Learning Based Automated Ear Infection Detection

IBRAHIM M. MEHEDI[1], (Senior Member, IEEE), MUHAMMAD SHEHZAD HANIF[1,2],
MUHAMMAD BILAL[1,2], MAHENDIRAN T. VELLINGIRI[1],
AND THANGAM PALANISWAMY[1], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering (ECE), King Abdulaziz University, Jeddah 21589, Saudi Arabia
[2]Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: Ibrahim M. Mehedi (imehedim@gmail.com)

ABSTRACT Artificial intelligence (AI) related to intelligent control in healthcare denotes using AI techniques to enhance the management and control of healthcare processes and systems. Damage to the inner and middle ear caused by accidents and diseases even causes hearing impairment in the ear that has been harmed or injured. Traditional otoscopy devices were utilized to check the tympanic membrane (TM) to identify OM in medical practice, and a conclusion is drawn depending on the outcomes of the examination. While developing a computer-aided method to support the OM diagnosis, it is possible to focus on methods like feature extraction, image pre-processing, classification, and image segmentation. The existing methodology of detecting the ear infection experiences a reduction of accuracy due to the influence of the noise in the input ear image. This presence of noise affects the feature extraction process, directly influences the accuracy in detection process. To overcome this issue, in this manuscript, a Deep learning (DL) is utilized to find biomedical ear infections by examining images of the eardrum and ear canal. The process includes training a DL method with a large dataset of ear images, where the images were labeled as either not infected or infected. With this motivation, this article emphasizes the design of Bayesian optimization with a deep learning-based automated ear infection detection and classification (BODL-AEIDC) model. The BODL-AEIDC technique exploits the DL model with a metaheuristic optimization algorithm for the ear infection classification process. The BODL-AEIDC technique employs a Wiener filtering (WF) based noise removal process to eliminate the noise data. In addition, the BODL-AEIDC technique exploits W-Net-based segmentation and the EfficientNet model for feature extraction purposes. Moreover, the BODL-AEIDC technique employs a fuzzy Restricted Boltzmann machine (FRBM) model for ear infection detection. Furthermore, the BO algorithm is utilized to adjust the FRBM technique's hyperparameter values effectively. The BODL-AEIDC technique's experimental outcomes occur using the medical dataset. The comprehensive comparative study stated the enhanced performance of the BODL-AEIDC approach over other existing methods.

I. INTRODUCTION

Otitis media (OM), or ear infection, refers to an illness usually found in children below the age of three years [1]. In the

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar[ID].

U.S., the medical expenses of ear infections are predicted to be $ 3 to 5 billion annually. Three kinds of OM exist: chronic otitis media with effusion, acute otitis media (AOM), and otitis media with effusion (OME) [2]. Due to bacterial infection, AOM may occur in the middle of the ear, bringing about the build-up of fluid. OME leads to fluid build-up in

the middle of the ear because of inflammation, which is much more severe than AOM [3]. The standard diagnosis depends on pneumatic otoscopy, the benchmark for distinguishing AOM from OME [4]. A pneumatic otoscope can measure the deflection of TM under pressure by constituting an airtight seal. On top of the complex procedure of constituting an airtight seal, the pneumatic otoscope neither finds bacteria nor differentiates bacterial and viral infections [5]. Inefficacy in distinguishing AOM from OME or bacterial and viral infections may result in the prescription of antibiotics, leading to drug-resistant microorganisms. Thus, precisely identifying bacteria accountable for OM was a medically unfulfilled demand [6].

An otoscope is a commonly utilized tool to detect the issues connected with ear pain [7]. The primary variable inspected in the middle of the ear is the presence of liquid in the TN, transparency, and coloration. In many cases, victims are recommended topical or oral antibiotics; though being adequate for pain, infection persists because of the development of antimicrobial-resistant bacteria [8]. The frequent usage of bacterial biofilms and antibiotics can assist in the rise of antimicrobial resistance. Currently, with the developments in image processing technologies, it is possible to identify a wider variety of diseases than was previously conceivable. Above all, while evolving a computer-aided method to help in the OM diagnosis, it will likely focus on methods like feature extraction, picture pre-processing, classification, and image segmentation.

AI has started to affect medicine. However, some applications are relevant to hearing, and AI is absent from hearing healthcare [9]. Therefore, the chances are described here to use prevailing technologies to develop medical applications that model the auditory system to allow fundamental advancements in hearing research. Biological replication is unnecessary [10]: many significant medical difficulties in hearing are solved through methods with no relation to the auditory system or methods that mimic only some aspects of its function (like DNNs for sound source separation).

This article emphasizes the design of Bayesian optimization with a deep learning-based automated ear infection detection and classification (BODL-AEIDC) model. The BODL-AEIDC technique employs a Wiener filtering (WF) based noise removal process to eliminate the noise data. Besides, the BODL-AEIDC technique exploits W-Net-based segmentation and the EfficientNet model for feature extraction. In addition, the BODL-AEIDC technique employs a fuzzy Restricted Boltzmann machine (FRBM) model for ear infection detection. Finally, the BO algorithm is utilized to adjust the FRBM method's hyperparameter values effectively. The experimental outcomes of the BODL-AEIDC method take place using the medical dataset.

II. LITERATURE REVIEW
Huang et al. [11] focused on consuming polymeric nanoparticles to cure common bacterial infections in humans. With ML methods like CNNs and ANNs, this study assesses the efficiency of nanoparticle treatment. Unique usage of advanced CNN, like Dense Net, was reported for the automated identification of middle ear infections. Binol et al. [12] present a decision fusion system for combining estimations gained from biophysical measurements and digital otoscopy images for identifying eardrum abnormality. The author trained an RF utilizing raw tympanometry attributes for the tympanometry aspect. Utilizing the normal range of the tympanogram values offered by clinical guides, the author mimicked a clinician's decision on tympanometry findings. Likewise, to categorize TM images from all otoscopic videos, the author re-trained Inception-ResNet-v2.

In [13], a low-cost, portable EEG headband is utilized to find the absence of seizures in epilepsy patients—ML with Tensorflow to forecast its confidence level and the SVM technique for separating the initial limit. Next, to see its measurement divergence, the author tried to test the technique on two other patients. Moreover, one of the novelty methods in this study is the Tensorflow library, which was developed and implemented to classify and train the data. Pham et al. [14] presented a technique for automated segmentation of TMs from video-otoscopic imageries that depends on deep fully CNN. Relies upon the UNet structure, the presented EAR technique depends on three main paradigms: Residual blocks for the decoder, EfficientNet for the encoder, and Attention gate for the skip connection path. The study presents a novel loss function term for NNs to execute segmentation. Mainly, the author devised to incorporate EfficientNet-B4 into the UNet encoder part.

In [15], a CNN was utilized for feature extraction to diagnose different types of OM from middle ear otoscope images. Such attributes have been extracted through ResNet-50, AlexNet, GoogLeNet, and VGG-16 methods. The deep features derived from such methods are integrated into novel deep feature vectors. In this case, a potential feature set was acquired, and the number of features was decreased. In the next phase, this novel feature set has been implemented as the input to the SVM. In [16], a new CAD support technique depends on the CNN was advanced. To enhance the generalized capability of the presented technique, integration of channel and spatial model (CBAM), hypercolumn, and residual blocks method is entrenched into the presented technique. Bapi et al. [17] purpose is to define whether a cat was afflicted with ear mites. This dataset was precisely collected, processed, and classified to stop conflicts. A standard DL technique, the CNN, has been exploited for a higher accuracy rate. This technique has 88 % precision. The most complex image detection method was CNNs. Singh and Dutta [18] devise a DL-related approach using CNN for detecting ear infections. It classifies ear images into four categories (Myringosclerosis, Chronic otitis media, Normal, and Earwax plug).

III. THE PROPOSED MODEL
In this study, we have concentrated on developing the BODL-AEIDC technique for intelligent control in the

healthcare sector. The BODL-AEIDC technique exploits the DL model with a metaheuristic optimization algorithm for the ear infection classification process. To accomplish this, the BODL-AEIDC technique comprises WF-based noise removal, W-Net segmentation, EfficientNet feature extraction, FRBM-based classification, and BO-based hyperparameter tuning. Fig. 1 demonstrates the overall flow of the BODL-AEIDC algorithm.

A. IMAGE PRE-PROCESSING

The WF technique is applied in this study for the noise removal process. WF is a signal processing method that decreases signal noise [19]. A linear filter depends on the statistical properties of the noise and signal. WF aims to predict the signal from its noisy observation by considering the noise and signal features. The WF is designed between the noise and signal's power spectral density (PSD) and the cross-power spectral density (CPSD). The filter was devised in the frequency domain and implemented to the noisy signal to acquire a prediction of the original signal. The WF possess two dimensions and is analogy to the equation 1.

$$W(u, v) = F(u, v) * B(u, v) \quad (1)$$

where, the W(u,v) is the wiener filtering function which is analogy to the ideal fourier function F(u,v) and the blurring function b(u,v). In this particular scenario, B takes the form of a sinc function: if three pixels within a line harbor information originating from a common point on an image, the resultant digital image will appear to have undergone convolution with a three-point boxcar in the temporal domain. In an ideal situation, it would be possible to deduce a Fest, or F estimate, provided that the values of W and B are already known. The fourier function F(u,v) shall be computed using the relation between the Wiener and the blurring function as defined in equation 2.

$$F(u, v) = \frac{|B(u, v)|^{2.W(u,v)}}{|B(u, v)|^{2.B(u,v)}} + K(u, v) \quad (2)$$

where K(u,v) is the constant function chosen for the purpose of optimizing the estimated fourier function.

The WF operates by implementing a weighting function to noisy signals in the frequency domain, where the weighting function is devised to reduce the mean squared error between the original signal and its prediction. The weighting function was a function of the CPSD and PSD of noise and signal and can be calculated through the Wiener-Hopf equation.

B. W-NET BASED SEGMENTATION

The W-Net model is applied to segment the images. Afterward, pre-processing and W-net-based segmentation were performed to obtain a map [20]. This system retains localization and content data utilizing the decoded and encoded paths. Additionally, edge data was preserved to maintain consistency and sharpen the image in segmentation. This network was planned as a progression of the U-Net. Next, one AE was executed by linking two U-Net topologies. In all the

u-net, an expansive path (decoder) and a contracting (encoder) based architecture were performed.

The initial W-net model is a contracting path encompassing several blocks. The BN and three-stage convolution layers interspersed with ReLU were the key components. To construct a single convolution block, this essential element was considered twice. By applying 2×2 layers of max-pooling, the block was joined. The crucial target data could be maintained, and the number of variables could be reduced using max pooling. In the expanding path, the convolution layer count was 8, increasing from 8 to 128 in the contracting path.

The decoded part was a second-wide path. Convolution layer and upsampling made up its structures. The input was upscaled four times in the decoder's portion, and in the encoder part, it was downscaled once. To recover lost information at the time of the convolution and max-pooling procedures, the feature map from the corresponding feature map from the decoder module has a concatenated encoder path. The second part is corresponding to the first; however, in the first U-Net section, the outcomes of the unit were positioned at similar levels, and the outcomes of the topmost pooling layer were merged.

After upsampling the decoder's final component and the last amalgamation of the encoder's primary unit, there is a further block similar to all others. In that block, an 11 convolutional layer, the final layer, and a softmax activation function were exploited to match feature maps and the desired number of classes. This technique integrated cross-entropy loss (CEL) and total-variation loss (CT-loss).

$$Loss = L_{cr-etrp} + L_{total-var} \quad (3)$$

$$L_{cr-etrp} = L\left(sr'_n, pc_n\right) = -\sum_i^K pc_i \log\left(sr'_i\right) \quad (4)$$

$$L_{total-var} = L(\{sr\})'_n$$
$$= \sum_{\xi}^{W-1} \sum_{\eta}^{H-1} \|sr'_{\xi+1,\eta} - sr'_{\xi,\eta}\| \left\| sr'_{\xi+1,\eta} - sr'_{\xi,\eta} \right\| 1 \quad (5)$$

Here, H and W denote the height and width of input images. The pixel value at that position in standard segmentation map $\{PC\}$ signifies the pseudo segmentation mask created by index that maximized the measured segmentation maps value; $\{sr'_n\}$ specifies the sample n's normal segmentation maps. Owing to the features of total-variation loss, the CT loss support in reducing time and utilizing memory, the segmentation mask is compressed, negating the necessity for post processing.

C. EFFICIENTNet FEATURE EXTRACTION

In this stage, the EfficientNet model is utilized to derive feature vectors. EfficientNet is a set of CNN methods that attain accuracy on image classification tasks [21] while also being computationally scalable and efficient. EfficientNet methods utilize an integration of methods to attain their efficacy. Compound scaling includes scaling up the width,
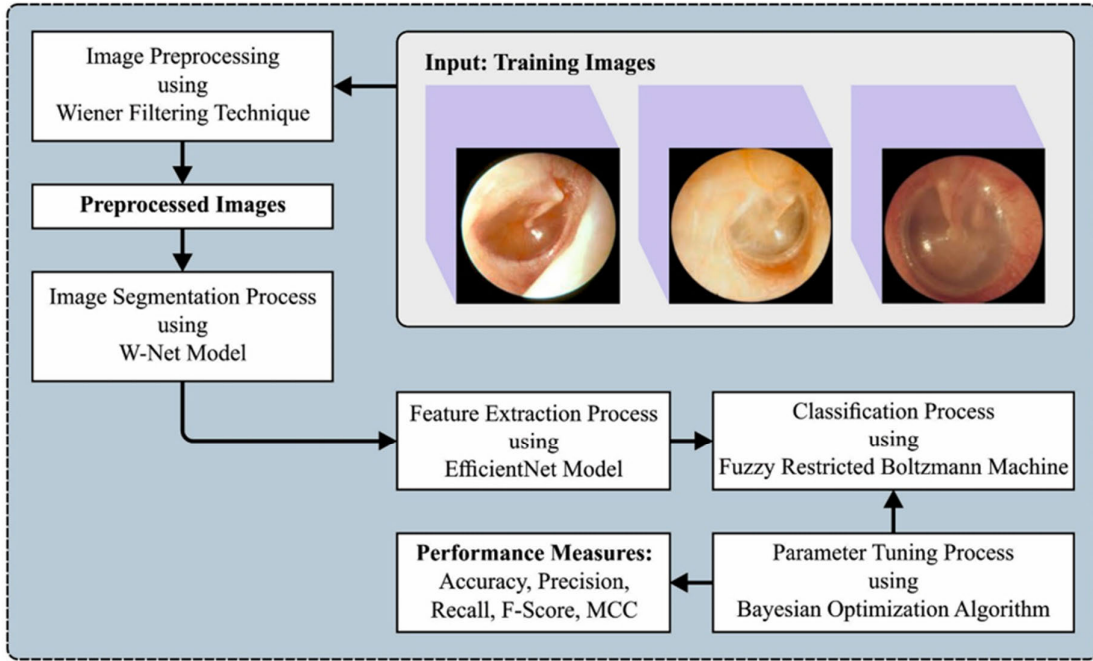
FIGURE 1. Overall flow of BODL-AEIDC algorithm.

depth, and resolution of the network concurrently while preserving a balance between the three. It allows the network to be more powerful and more efficient. It uses an amalgamation of mobile inverted bottleneck (MBConv) blocks, convolutional layers, and squeeze-and-excitation (SE) blocks for reducing computations and requires several parameters. It is developed through an automated ML (AutoML) technique, which searches an ample space of possible architectures and hyperparameters to find the optimum method combination. Such techniques have attained accuracy on a wide range of image classifier tasks while being highly efficient regarding memory usage and computing resources.

D. EAR INFECTION DETECTION USING THE FRBM MODEL

For the ear infection detection process, the FRBM model is used. FRBM is beneficial and related to the typical RBM. FRBM proposes higher extraction feature abilities than classic RBM, as the set-up parameters are fuzzy [22]. If the projected with noisy data, FRBM's resilience can be enhanced. The visible layer (VL) and hidden layer (HL) exist in FRBM. No connections exist among neurons in similar layers, and every connection occurs among neurons in other layers. Input data can be transferred to 'm' visible units($v_1, v_2, \ldots, VP$), and input data features are extracted utilizing n' HLs ($h_1, h_2, \ldots, HQ$).

The fuzzy energy function, $E(v, h, \theta)$, is demonstrated by Eq. (4).

$$\tilde{E}\left(v, h, \tilde{\theta}\right) = -\sum_{k=1}^{p} \tilde{a}_k v_k - \sum_{l=1}^{q} \tilde{b}_l h_l - \sum_{k=1}^{p}\sum_{l=1}^{q} h_l \tilde{w}_{k,l} v_k \tag{6}$$

whereas $\tilde{\theta} = \{\tilde{a}, \tilde{b}, \tilde{w}_{k,l}\}$ denotes the fuzzy parameters, $\tilde{a}, \tilde{b}$, and $W_{k,l}$ implies fuzzy numbers. $\tilde{a}$ and $\tilde{b}$ signify the biases of VL and HL; correspondingly, $\tilde{w}_{k,l}$ represents the connection weighted between the kth visible and l^{th} hidden units. The fuzzy free energy function can be represented as Eq. (7).

$$\tilde{G}(v) = -ln \sum_{\tilde{h}} e^{-E\left(v, h, \tilde{\theta}\right)} = -\sum_{k=1}^{p} \tilde{a}v_k$$
$$- \sum_{l=1}^{q} \ln\left(1 + e^{\left(\tilde{b} + \Sigma_{k=1}^{p} \tilde{w}_{k,l} v_k\right)}\right) \tag{7}$$

In which $\tilde{G}(v)$ represents the fuzzy energy function.

Utilizing membership function (MF) can be needed, taking superior sensitivity and resolution because of smaller parameters and the required short fluctuation interval. The fuzzy number $\tilde{a}$ is defined by Eq. (8).

$$g_{\tilde{a}}(v) = \begin{cases} 0, & v \leq a_k^L \\ \dfrac{v - a_k^L}{a_k^M - a_k^L}, & a_k^L < v < a_k^M \\ \dfrac{a_k^R - v}{a_k^R - a_k^M}, & a_k^M < v \leq a_k^R \\ 0, & v \geq a_k^R \end{cases} \tag{8}$$

whereas a_{kR}, a_{kL}, and a_{kM} signifies right bound, left bound, and center of connection weight correspondingly. $\tilde{b}$ and $\tilde{w}_k$ are even attained by the same approaches. Fig. 2 shows the structure of FRBM.

FRBMs contain VL and HL, which act together optimally if the parameter can be optimally tuned. The primary function

develops non-linear as the optimal solution can be turned into maximal probability questions about fuzzy numbers. When the primary function is defuzzified, the problem can be decreased to a predictable maximal probability method. Some defuzzification systems were obtainable: area bisector, maximal membership, the center of centroid, and crisp possibilistic mean value (CPMV). Eq. (9) determines the defuzzified free energy function once the MF is symmetric.

$$
\begin{aligned}
\tilde{G}(v) &= \int_0^1 \alpha(G^L(\alpha) + G^R(\alpha))d\alpha \\
&\approx -\sum_{k=1}^p \frac{a_k^L + a_k^U}{2} v_k \\
&\quad -\sum_{l=1}^q \frac{\ln(1 + e^{b_l^L + \Sigma_{k=1}^p W_{kl}^L v_k})}{2} \quad k = 1 l = 1 \\
&= \frac{1}{2}[-\sum_{k=1}^p a_k^L v_k - \sum_{l=1}^q \ln(1 + e^{(b_R^L + \Sigma_{k=1}^p W_{kl}^L v_k)}) \\
&\quad -\sum_{k=1}^p a_k^R v_k - \sum_{l=1}^q \ln(1 + e^{(b_R^R + \Sigma_{k=1}^p W_{kl}^R v_k)}) \\
&= \frac{1}{2}\left[G\left(v, W^L, a^L, b^L\right) + G\left(v, W^R, a^R, b^R\right)\right] \quad (9)
\end{aligned}
$$

whereas $G(V)$ denotes the defuzzified $\tilde{G}(v)$, $G^L(\alpha)$, and $G^R(\alpha)$ implies left and right boundaries of interval $[G^L(\alpha), G^R(\alpha)]$ that written the α-cut of fuzzy number $G(\alpha)V$.

The probability function of free energy, $(v|\tilde{\theta})$ is offered in Eq. (10).

$$
P\left(v \mid \tilde{\theta}\right) = \frac{e^{-\tilde{G}_s(v)}}{\Sigma_v e^{-\tilde{G}_s(v)}} \quad (10)
$$

The purpose of negative log probability was determined by Eq. (9).

$$
\ln N(a, \tilde{\theta}) = -\ln \prod_{k=1}^{qr} P(v^k|\tilde{\theta}) = -\sum_{k=1}^{qr} \ln P\left(\left(v^k \mid \tilde{\theta}\right)\right) \quad (11)
$$

To develop the optimum solution for parameters based on Eq. (10), the stochastic gradient descent (SGD) method can be executed in Eq. (9).

$$
m_{\tilde{\theta}}^i n(-\sum_{v \in r} \ln P\left(\left(v \mid \tilde{\theta}\right)\right)) \quad (12)
$$

The partial derivative of $\ln((v|\tilde{\theta})$ can be calculated. The contrastive divergence can be utilized for approximating partial derivatives of the maximal log probability gradient to shorten the computation complexity.

E. HYPERPARAMETER TUNING USING BO ALGORITHM
Finally, the BO approach is utilized to adjust the hyperparameter values of the FRBM method. The BO technique is a hyperparameter optimization technique [23]. Compared with the grid and random search process, the BO algorithm finds

the fittest model parameter more rapidly. All the parameters are in a specific interval, constituting hyperparameter space. The predictive accuracy of the model is different for the various hyperparameter combinations. The accuracy of the model gets increased by the optimizer technique, and the estimated outcome is nearer to the actual value:

$$
f(x) = \sqrt{\frac{\Sigma_{i=1}^{num}(y(x_i) - \hat{y}_i)^2}{num}} \quad (13)
$$

In Eq. (13), *num* indicates the length of the input time series. $\hat{y}_i$ denotes the true value, and $y(x)$ shows the forecasted value. The BO algorithm attains the optimum parameter, making the target reach a minimal value.

The objective function of the BO algorithm is given by:

$$
x^* = arg \min_{x \in X} f(x) \quad (14)
$$

In Eq. (14), χ shows the set of hyperparameter combinations. X denotes the parameter combination, and x^* refers to the better parameter.

The BO algorithm finds new assessment points by increasing the acquisition for weighing the dispersal of evaluation points and the expansion of predictive accuracy; then, it re-imports them as input in the method for obtaining novel output such that it continuously updates and finds model parameters. Then, the Gaussian function is selected as a distribution hypothesis, and later, the acquisition function is used for selecting the next point in the posterior function. The Gaussian process (GP) was an expansion of multi-dimensional Gaussian distribution determined by the mean and covariance. The covariance function of GP was its kernel function $k(x, x')$. The kernel function evaluates distance between, x and x'.

$$
f(x) \sim GP\left(\mu(x), k\left(x, x'\right)\right) \quad (15)
$$
$$
\mu(x) = E[f(x)] \quad (16)
$$
$$
k\left(x, x'\right) = E\left[(f(x) - \mu(x))\left(f(x') - \mu(x')\right)\right] \quad (17)
$$

Generally, the mean function is fixed to 0; after, the following equation formulates the abovementioned Gaussian method:

$$
K_n = \begin{pmatrix} k(x_1, x_1) & \ldots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \ldots & k(x_{n'}x_n) \end{pmatrix} \begin{pmatrix} k & k(xx) \\ & \vdots \\ k(xx) & \end{pmatrix} \quad (18)
$$

The covariance matrix equation can be updated once the novel set of assessment samples is added to each assessment point as:

$$
\begin{cases} K_{n+1} = \begin{pmatrix} k_n & k_n^T \\ k_n & k_{(x_{n+1}, x_{n+1})} \end{pmatrix} \\ k_n = [k(x_{n+1}, x_1), k(x_{n+1}, x_2), \ldots, k(x_{n+1'}x_n)] \end{cases} \quad (19)
$$

By applying the updated covariance matrix, the posterior probability can be obtained:

$$
P\left(f_{n+1} \mid D_{n+1'}x_{n+1}\right) \sim N(\mu_{n+1}(x), \sigma_{n+1}^2(x)) \quad (20)
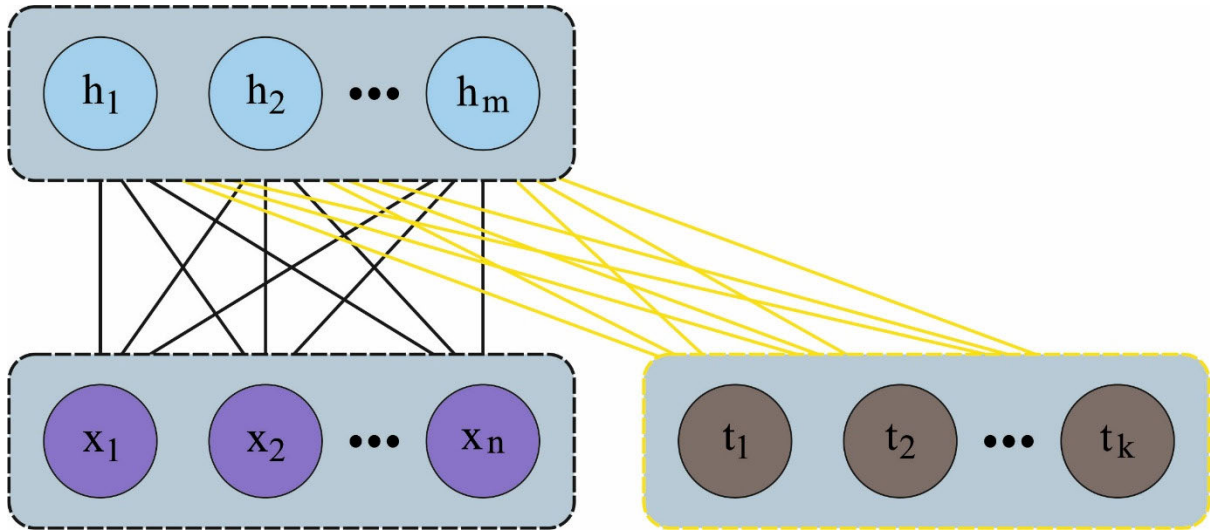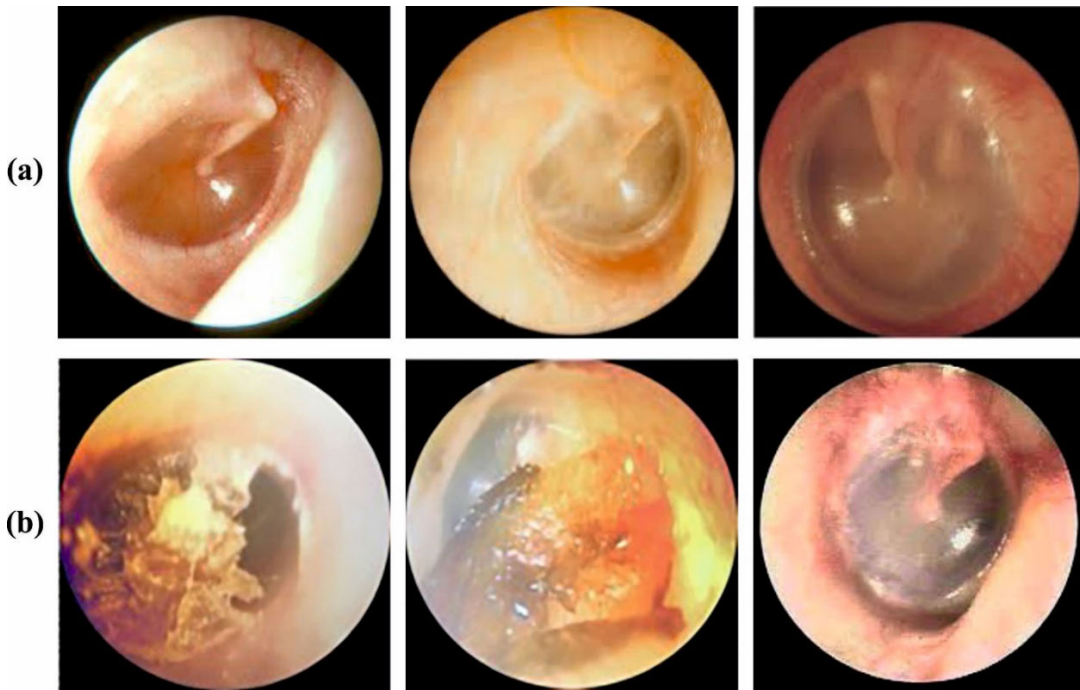$$

FIGURE 2. Architecture of FRBM.

FIGURE 3. a) Normal images b) Earwax images.

In Eq. (20), $\mu_{i+1}(x)$ and $\sigma_{i+1}^2(x)$ represent the variance and mean of $f(x)$ at $i+1$ steps, and D denotes the observation data.

The function value is sampled from joint posterior distribution via estimating the covariance matrix and mean value. The sampling function determines the next point that needs to be estimated to reduce resource consumption and find the optimum value faster. The UCB function was selected as a sampling function, and the formula is given below:

$$x_{i+1} = \text{argmax} H\left(x \mid D_i\right) = \text{argmax} \mu\left(x\right) + \zeta_{i+1}^{\frac{1}{2}} \sigma_i\left(x\right) \quad (21)$$

In Eq. (21), $+1$, $\zeta_{i+1}^{\frac{1}{2}}$ represent the constant, x_{i+1} denotes the selected hyperparameter of step i, and $H(XD)$ shows the UCB acquisition function.

IV. RESULTS AND DISCUSSION
In this section, the ear infection detection and classification results of the BODL-AEIDC technique are examined in detail. Fig. 3 represents the average and earwax images.

Table 1 and Fig. 4 provide the PSNR analysis of the WF with other pre-processing techniques. The results imply that the WF reaches increasing PSNR values over other

TABLE 1. PSNR analysis of WF with other pre-processing approaches.

Samples	(%)	Wt. Median Filter	Adpt. Median Filter	HE-ACWM	Wiener Filter
	30	21.17	20.83	22.49	24.21
IMG-1	40	17.82	22.08	21.89	23.21
	50	19.58	22.78	22.09	24.00
	30	21.86	24.09	27.01	28.90
IMG-2	40	25.58	23.49	27.33	28.90
	50	23.66	25.37	24.04	26.95
	30	22.46	23.16	27.10	28.31
IMG-3	40	27.13	25.48	26.10	28.74
	50	24.87	25.05	24.47	26.94
	30	21.08	23.84	24.45	26.23
IMG-4	40	24.91	29.09	28.52	30.54
	50	23.70	23.40	26.28	27.70
	30	24.79	26.16	27.07	28.93
IMG-5	40	24.97	27.67	28.30	29.78
	50	29.25	28.25	29.68	31.49
	30	25.27	25.70	27.20	28.71
IMG-6	40	23.34	29.81	27.69	31.63
	50	29.87	30.30	28.93	31.80

TABLE 2. RI analysis of the W-NET model with other segmentation approaches.

Rand Index (RI)				
No. of Images	ACM	GC	DWT	W-Net
IMG-1	63.36	65.21	71.86	74.79
IMG-2	81.55	87.13	93.02	94.84
IMG-3	48.32	42.29	58.18	60.53
IMG-4	71.61	78.29	82.02	84.19
IMG-5	91.13	90.98	97.66	98.53
IMG-6	79.65	87.39	98.09	98.86
IMG-7	14.02	12.36	26.01	28.60
IMG-8	21.60	24.10	40.52	43.44
IMG-9	77.17	71.99	87.69	90.33
IMG-10	31.96	48.44	50.02	52.07

models. For instance, on IMG-1 with 30% of noise, the WF obtains an increasing PSNR of 24.21dB while the WMF, AMF, and HE-ACWM models offer decreasing PSNR of 21.17dB, 20.83dB, and 22.49dB, respectively. Concurrently, on IMG-3 with 30% of noise, the WF gains an increasing PSNR of 28.31dB while the WMF, AMF, and HE-ACWM methods offer to decrease PSNR of 22.46dB, 23.16dB, and 27.10dB correspondingly. Simultaneously, on IMG-6 with 30% of noise, the WF acquires an increasing PSNR of 28.71dB while the WMF, AMF, and HE-ACWM approaches offer decreasing PSNR of 25.27dB, 25.70dB, and 27.20dB correspondingly.

The rand index (RI) investigation of the W-Net method with other segmentation approaches is given in Table 2 and
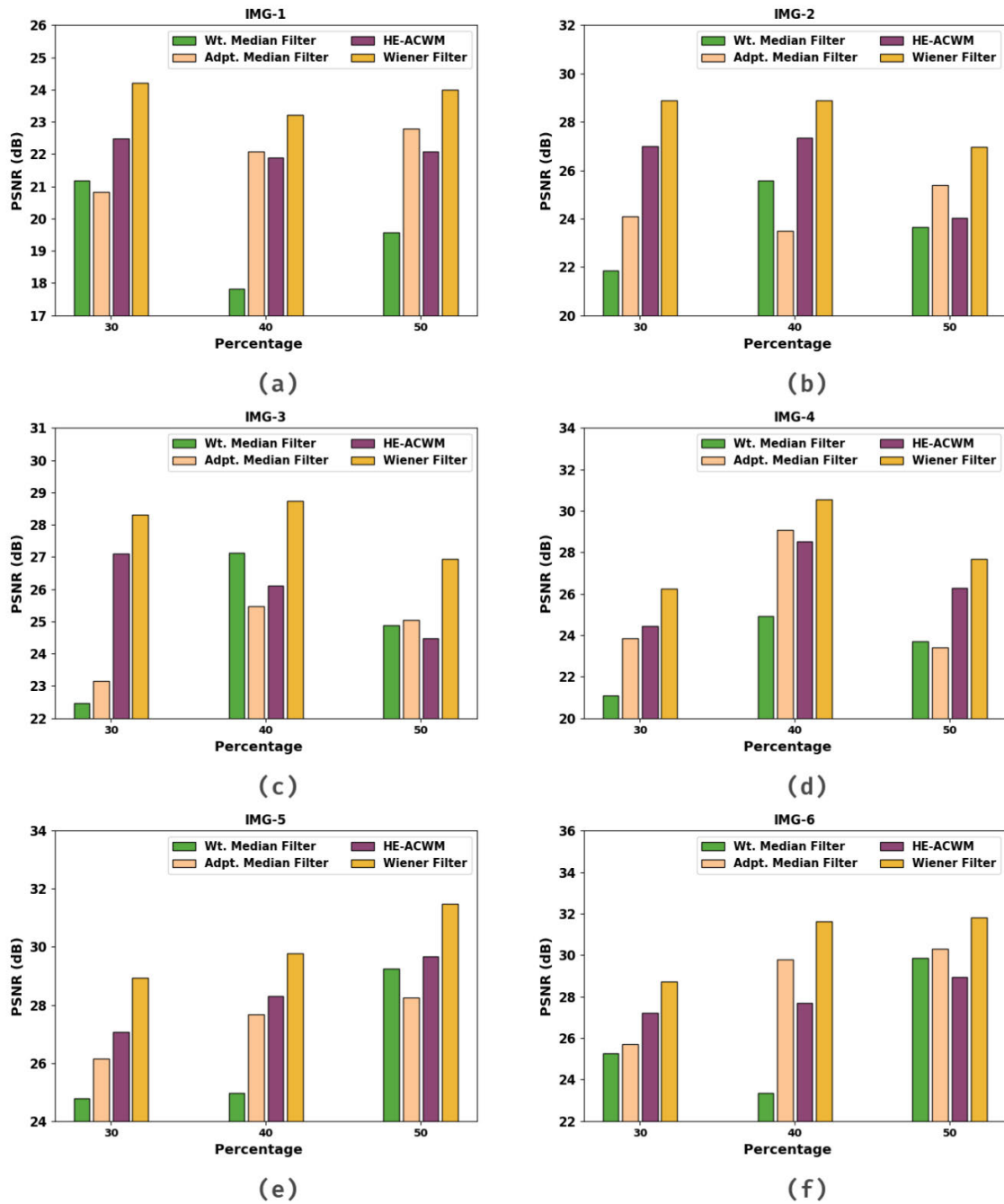
FIGURE 4. PSNR analysis of WF with other pre-processing approaches (a-f) IMG 1-6.

Fig. 5. The results demonstrated the effectual performance of the W-Net model with increasing RI values. For instance, on IMG-1, the W-Net model offers a higher RI of 74.79. Meanwhile, on IMG-5, the W-Net model offers a higher RI of 98.53. Additionally, on IMG-7, the W-Net model offers a higher RI of 28.60. At last, on IMG-10, the W-Net model offers a higher RI of 52.07

The segmentation results of the W-Net approach with existing methods are compared in terms of global consistency error (GCE) in Table 3 and Fig. 6. The figure highlights the improved performance of the W-Net approach over other segmentation methods. For example, on IMG-1, the W-NET method obtains a decreased GCE of 43.76%, whereas the ACM, GC, and DWT models attain increased GCE of 84.67%, 61.61%, and 44.90%, respectively. Followed by IMG-5, the W-NET approach obtains decreased GCE of 65.38%, whereas the ACM, GC, and DWT models attain increased GCE of 81.48%, 84.86%, and 66.27%, respectively.

TABLE 3. GCE analysis of the W-NET model with other existing approaches.

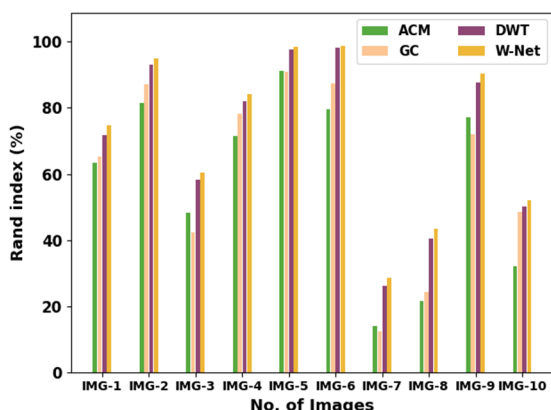| Global Consistency Error (%)				
No. of Images	ACM	GC	DWT	W-Net
IMG-1	84.67	61.61	44.90	43.76
IMG-2	52.54	51.09	40.56	39.78
IMG-3	7.95	10.76	5.65	5.02
IMG-4	72.66	64.23	50.45	49.61
IMG-5	81.48	84.86	66.27	65.38
IMG-6	75.61	55.39	34.49	33.75
IMG-7	38.29	29.77	11.38	10.74
IMG-8	30.94	51.04	40.34	29.80
IMG-9	11.37	30.68	21.31	10.84
IMG-10	39.36	21.44	13.07	12.01

FIGURE 5. RI analysis of the W-Net model with other segmentation approaches.
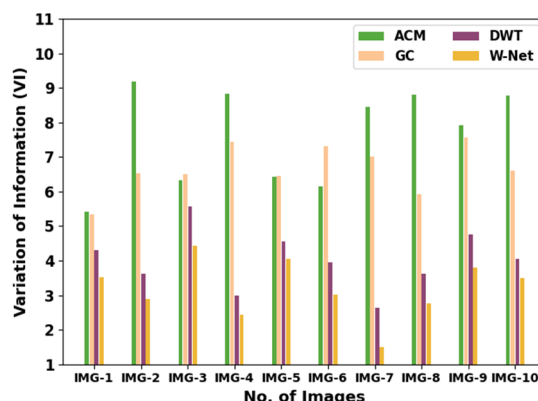

FIGURE 7. VI analysis of the W-Net model with other existing approaches.
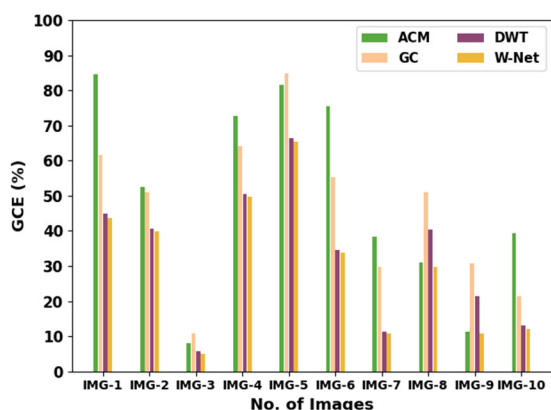

FIGURE 6. GCE analysis of the W-Net model with other existing approaches.

In addition, on IMG-7, the W-NET method obtains a decreased GCE of 10.74%, whereas the ACM, GC, and DWT approaches attain increased GCE of 38.29%, 29.77%, and 11.38%. Lastly, on IMG-10, the W-NET approach obtains a decreased GCE of 12.01%, whereas the ACM, GC, and DWT methods attain increased GCE of 39.36%, 21.44%, and 13.07%.

The segmentation results of the W-Net model with existing techniques are compared in terms of variation of information (VI) in Table 4 and Fig. 7.

The figure showcased the improved performance of the W-Net model over other segmentation approaches. For instance, on IMG-1, the W-NET model obtains a decreased VI of 3.5245, whereas the ACM, GC, and DWT models attain increased VI of 5.4044, 5.3364, and 4.2965, respectively. Followed by, on IMG-5, the W-NET approach gains decreased VI of 4.0581 whereas the ACM, GC, and DWT techniques attain increased VI of 6.4182, 6.4491, and 4.5611 correspondingly. In addition, on IMG-7, the W-NET technique obtains a decreased VI of 1.5022, whereas the ACM,

TABLE 4. VI analysis of the W-NET model with other existing approaches.

Variation of Information (VI)				
No. of Images	ACM	GC	DWT	W-Net
IMG-1	5.4044	5.3364	4.2965	3.5245
IMG-2	9.1727	6.5313	3.6129	2.8859
IMG-3	6.3235	6.5067	5.5734	4.4244
IMG-4	8.8174	7.4397	2.9853	2.4263
IMG-5	6.4182	6.4491	4.5611	4.0581
IMG-6	6.1507	7.3133	3.9414	3.0164
IMG-7	8.4553	6.9978	2.6412	1.5022
IMG-8	8.7956	5.9202	3.6194	2.7584
IMG-9	7.9241	7.5572	4.7665	3.8075
IMG-10	8.7836	6.5992	4.0505	3.4885

TABLE 5. Details of the dataset.

Class	No. of Instances
Normal (Class-0)	100
AOM (Class-1)	100
Total Number of Instances	200

TABLE 6. Classifier outcome of BODL-AEIDC Approach on 70:30 of TRP/TSP.

Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	MCC	SSIM	Jaccard Coefficient	Kappa Coefficient
Training Phase (70%)								
Class-0	88.89	84.85	88.89	86.82	75.65	79.21	33.21	34.24
Class-1	87.01	90.54	87.01	88.74	75.65	78.25	34.02	33.28
Average	87.95	87.69	87.95	87.78	75.65	78.73	33.61	33.76
Testing Phase (30%)								
Class-0	97.30	100.00	97.30	98.63	96.56	83.25	34.74	34.26
Class-1	100.00	95.83	100.00	97.87	96.56	91.76	33.48	32.85
Average	98.65	97.92	98.65	98.25	96.56	87.50	35.69	34.85

GC, and DWT models attain increased VI of 8.4553, 6.9978, and 2.6412, respectively. Finally, on IMG-10, the W-NET algorithm gains decreased VI of 3.4885, whereas the ACM, GC, and DWT methods attain increased VI of 8.7836, 6.5992, and 4.0505 correspondingly. The classification outcomes of the BODL-AEIDC technique are determined with the dataset, comprising 200 samples and two classes, as represented in Table 5.

In Fig. 8, the confusion matrices of the BODL-AEIDC technique is illustrated. On 70% of TRP, the BODL-AEIDC technique recognizes 56 instances into class-0 and 67 instances into class-1. Meanwhile, on 30% of TSP, the BODL-AEIDC method recognizes 36 instances into class-0 and 23 instances into class-1. Moreover, on 80% of TRP, the BODL-AEIDC algorithm recognizes 73 instances into class-0 and 79 instances into class-1.
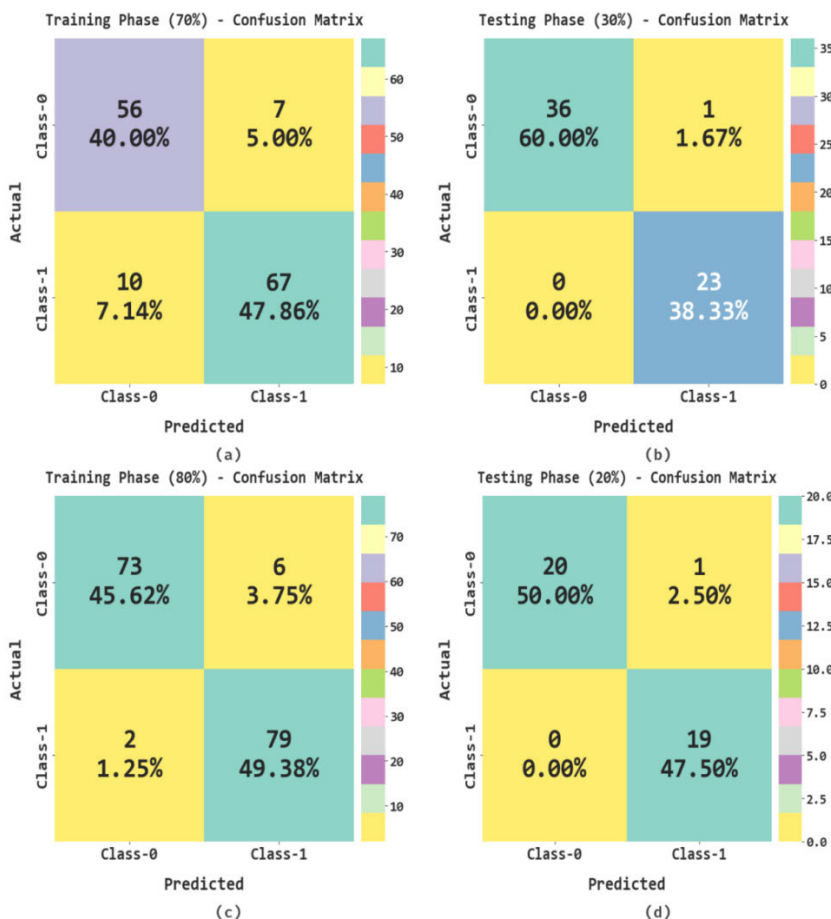
FIGURE 8. Confusion matrices of BODL-AEIDC approach (a-b) 70:30 of TRP/TSP and (c-d) 80:20 of TRP/TSP.

TABLE 7. Classifier outcome of BODL-AEIDC approach on 80:20 of TRP/TSP.

Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	MCC	SSIM	Jaccard Coefficient	Kappa Coefficient
			Training Phase (80%)					
Class-0	92.41	97.33	92.41	94.81	90.11	88.68	34.69	36.48
Class-1	97.53	92.94	97.53	95.18	90.11	89.59	35.95	35.74
Average	94.97	95.14	94.97	94.99	90.11	89.13	35.32	36.11
			Testing Phase (20%)					
Class-0	95.24	100.00	95.24	97.56	95.12	90.58	36.44	39.48
Class-1	100.00	95.00	100.00	97.44	95.12	91.65	37.29	36.75
Average	97.62	97.50	97.62	97.50	95.12	91.11	36.86	38.11

In Table 6, the overall classifier results of the BODL-AEIDC technique are shown. The results indicate that the BODL-AEIDC technique classified two classes efficiently. For instance, on 70% of TRP, the BODL-AEIDC technique gains average $accu_y$ of 87.95%, $prec_n$ of 87.69%, $reca_l$ of 87.95%, F_{score} of 87.78%, and MCC of 75.65%.

Concurrently, on 30% of TSP, the BODL-AEIDC method gains average $accu_y$ of 98.65%, $prec_n$ of 97.92%, $reca_l$ of 98.65%, F_{score} of 98.25%, and MCC of 96.56%.

Table 7 shows the overall classifier results of the BODL-AEIDC technique. The results indicate that the BODL-AEIDC method classified two classes efficiently. For
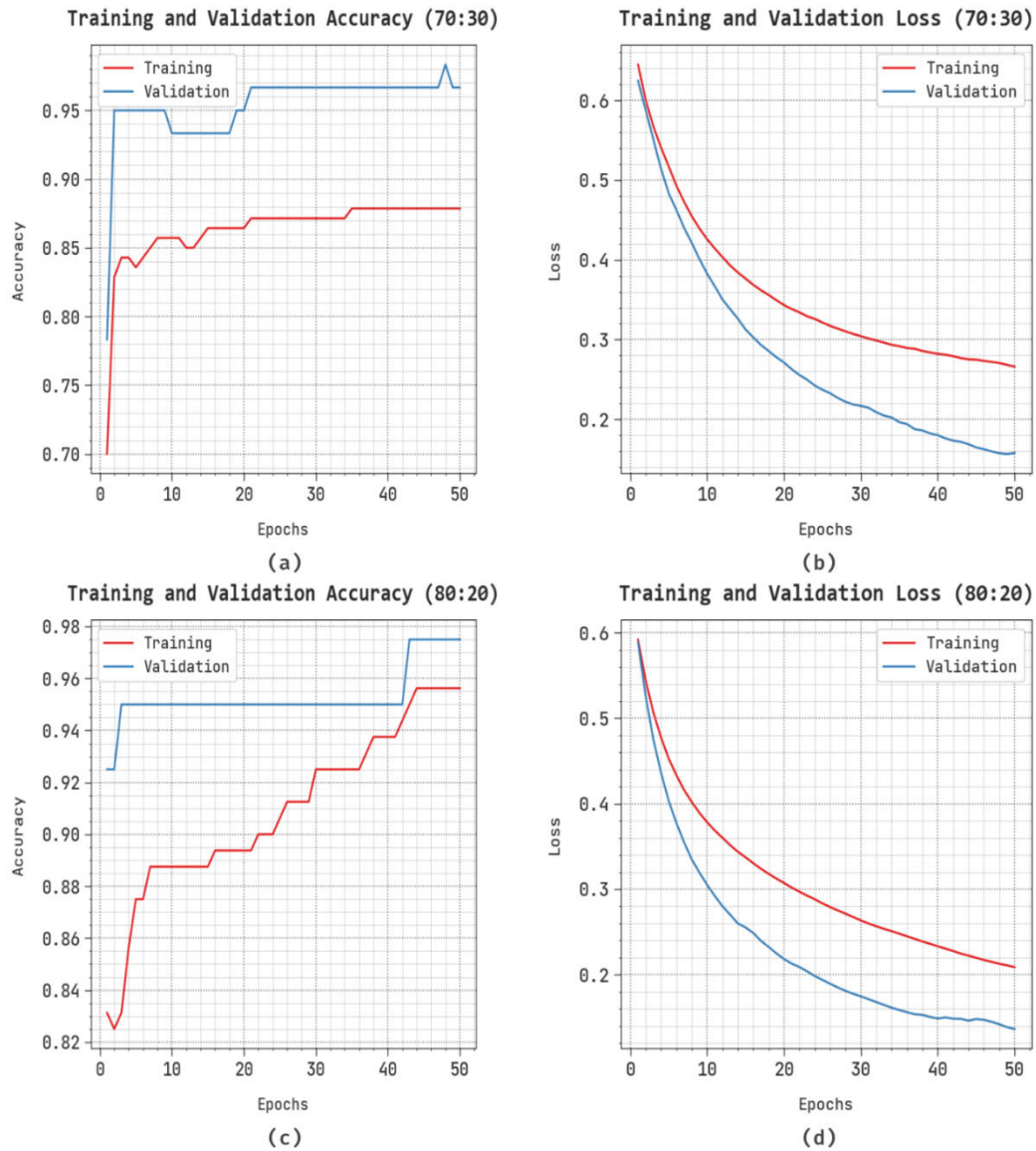
FIGURE 9. (a–c) Accuracy curve on 70:30 and 80:20 and (b–d) Loss curve on 70:30 and 80:20.

TABLE 8. Comparative outcome of BODL-AEIDC approach with other existing techniques.

Methods	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}
BODL-AEIDC	98.65	97.92	98.65	98.25
SVM	96.75	95.40	96.77	95.60
XGBoost	96.14	96.19	96.35	95.47
ELM	95.90	95.60	95.88	96.49
AlexNet	96.05	96.32	96.76	95.69
KNN	96.74	96.97	95.91	96.29

instance, on 80% of TRP, the BODL-AEIDC approach gains average $accu_y$ of 94.97%, $prec_n$ of 95.14%, $reca_l$ of 94.97%, F_{score} of 94.99%, and MCC of 90.11%. In parallel, on 20% of TSP, the BODL-AEIDC approach acquires average $accu_y$ of 97.62%, $prec_n$ of 97.50%, $reca_l$ of 97.62%, F_{score} of 97.50%, and MCC of 95.12%.
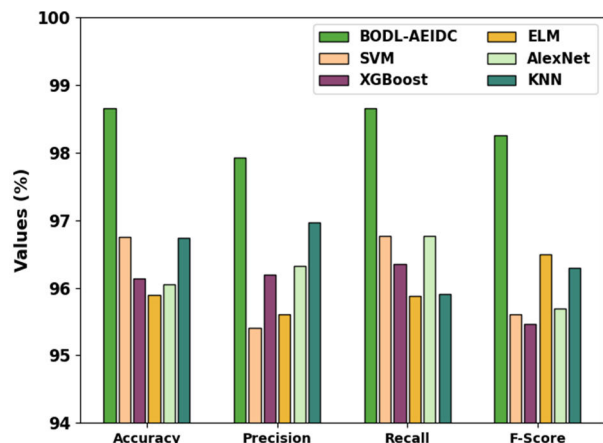
FIGURE 10. Comparative outcome of BODL-AEIDC approach with other existing techniques.

Fig. 9 reveals the classifier results of the method on 70:30 and 80:20. Figs. 9a-9c reveals the accuracy analysis of the BODL-AEIDC technique under 70:30 and 80:20. The figure shows that the BODL-AEIDC approach reaches increasing accuracy values over increasing epochs. Also, the increasing validation accuracy over training accuracy displays that the BODL-AEIDC approach learns efficiently on the test dataset. Lastly, Figs. 9b-9d exemplifies the loss study of the BODL-AEIDC method under 70:30 and 80:20. The results specify that the BODL-AEIDC method reaches closer training and validation loss values. The BODL-AEIDC approach learns efficiently on the test dataset.

A comparative result analysis of the BODL-AEIDC method with other existing models is given in Table 8 and Fig. 10 [24]. The figure indicates that the ELM approach reaches lower performance over other models. At the same time, the SVM, XGBoost, AlexNet, and KNN models obtain slightly increasing outcomes. However, the BODL-AEIDC technique outperforms the existing ones with maximum accu_y of 98.65%, prec_n of 97.92%, reca_l of 98.65%, and F_score of 98.25%. These results stated the improved performance of the BODL-AEIDC technique.

V. CONCLUSION

In this article, we have concentrated on developing the BODL-AEIDC technique for intelligent control in the healthcare sector. The presented BODL-AEIDC technique exploits the DL model with a metaheuristic optimization algorithm for the ear infection classification process. To accomplish this, the BODL-AEIDC technique comprises WF-based noise removal, W-Net segmentation, EfficientNet feature extraction, FRBM-based classification, and BO-based hyperparameter tuning. The BO algorithm is utilized to effectively adjust the hyperparameter values of the FRBM method, which in turn improves the classification results. The experimental outcomes of the BODL-AEIDC algorithm take place using the medical dataset. The comprehensive comparison study stated the enhanced performance of the BODL-AEIDC technique over other methods. The performance of the proposed

work yields a better level of accuracy with 98.65%, 97.92% of precision, 98.65% of recall and 98.25% of F score. In the upcoming years, the performance of the BODL-AEIDC algorithm can be enhanced with the feature fusion-based DL models. In addition, the performance of the proposed work shall further be enhanced with the inclusion of fuzzy logic or Artificial Intelligence algorithms for classification with better level of accuracy.

REFERENCES

[1] M. G. Crowson, D. W. Bates, K. Suresh, M. S. Cohen, and C. J. Hartnick, "'Human vs machine' validation of a deep learning algorithm for pediatric middle ear infection diagnosis," *Otolaryngol.-Head Neck Surg.*, vol. 169, no. 1, pp. 41–46, Jul. 2023.

[2] R. G. Kashani, M. C. Młyńczak, D. Zarabanda, P. Solis-Pazmino, D. M. Huland, I. N. Ahmad, S. P. Singh, and T. A. Valdez, "Shortwave infrared otoscopy for diagnosis of middle ear effusions: A machine-learning-based approach," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, Jun. 2021.

[3] Y. S. Park, J. H. Jeon, T. H. Kong, T. Y. Chung, and Y. J. Seo, "Deep learning techniques for ear diseases based on segmentation of the normal tympanic membrane," *Clin. Experim. Otorhinolaryngol.*, vol. 16, no. 1, pp. 28–36, Feb. 2023.

[4] J. E. N. Pedersen, "Digital otoscopy with AI diagnostic support: Making diagnosis of ear disease more accessible," *ENT Audiol. News*, Jun. 2020. [Online]. Available: https://www.entandaudiologynews.com/contributors/person/jenny-elise-nesgaard-pedersen

[5] M. G. Crowson, C. J. Hartnick, G. R. Diercks, T. Q. Gallagher, M. S. Fracchia, J. Setlur, and M. S. Cohen, "Machine learning for accurate intraoperative pediatric middle ear effusion diagnosis," *Pediatrics*, vol. 147, no. 4, 2021, Art. no. e2020034546.

[6] H. Al-Sadr, M. Popescu, and J. M. Keller, "Early sepsis recognition based on ear localization using infrared thermography," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Madrid, Spain, Dec. 2018, pp. 823–830.

[7] J. Won, G. L. Monroy, R. I. Dsouza, D. R. Spillman, J. McJunkin, R. G. Porter, J. Shi, E. Aksamitiene, M. Sherwood, L. Stiger, and S. A. Boppart, "Handheld briefcase optical coherence tomography with real-time machine learning classifier for middle ear infections," *Biosensors*, vol. 11, no. 5, p. 143, May 2021.

[8] J. Chan, S. Raju, R. Nandakumar, R. Bly, and S. Gollakota, "Detecting middle ear fluid using smartphones," *Sci. Transl. Med.*, vol. 11, no. 492, pp. –2, 2019, Art. no. eaav110.

[9] G. P. Rajamanickam, "A multispectral imaging method and device to detect and quantify the presence of fluid in the middle ear to facilitate the diagnosis and triage of ear infections," Ph.D. dissertation, Dept. Eng. Manag., Massachusetts Inst. Technol., Cambridge, MA, USA, 2020.

[10] A. R. Habib, M. Kajbafzadeh, Z. Hasan, E. Wong, H. Gunasekera, C. Perry, R. Sacks, A. Kumar, and N. Singh, "Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis," *Clin. Otolaryngol.*, vol. 47, no. 3, pp. 401–413, 2022.

[11] Z. Huang, S. Chen, H. E. Ali, D. H. Elkamchouchi, J. Hu, E. Ali, J. Zhang, and Y. Huang, "Application of CNN and ANN in assessment the effect of chemical components of biological nanomaterials in treatment of infection of inner ear and environmental sustainability," *Chemosphere*, vol. 331, Aug. 2023, Art. no. 138458.

[12] H. Binol, A. C. Moberly, M. K. K. Niazi, G. Essig, J. Shah, C. Elmaraghy, T. Teknos, N. Taj-Schaal, L. Yu, and M. N. Gurcan, "Decision fusion on image analysis and tympanometry to detect eardrum abnormalities," *Proc. SPIE*, vol. 11314, pp. 375–382, Mar. 2020.

[13] E. Sutanto, T. W. Purwanto, F. Fahmi, M. Yazid, W. Shalannanda, and M. Aziz, "Implementation of closing eyes detection with ear sensor of muse EEG headband using support vector machine learning," *Int. J. Intell. Eng. Syst.*, vol. 16, no. 1, pp. 460–473, 2023.

[14] 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.

[15] E. Başaran, Z. Cömert, and Y. Çelik, "Neighbourhood component analysis and deep feature-based diagnosis model for middle ear otoscope images," *Neural Comput. Appl.*, vol. 34, no. 8, pp. 6027–6038, Apr. 2022.

[16] 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.

[17] A. M. Bapi, N. Nabi, S. A. R. Shuvo, P. Chakraborty, and S. A. Khushbu, "CatEarMites: An approach of detecting ear mites of cat using convolutional neural network," in *Proc. 13th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Kharagpur, India, Oct. 2022, pp. 1–6.

[18] A. Singh and M. K. Dutta, "Diagnosis of ear conditions using deep learning approach," in *Proc. Int. Conf. Commun., Control Inf. Sci. (ICCISc)*, Idukki, India, Jun. 2021, pp. 1–5.

[19] T. Wang, H. Guo, X. Yan, and Z. Yang, "Speech signal processing on graphs: The graph frequency analysis and an improved graph Wiener filtering method," *Speech Commun.*, vol. 127, pp. 82–91, Mar. 2021.

[20] X. Xia and B. Kulis, "W-Net: A deep model for fully unsupervised image segmentation," 2017, *arXiv:1711.08506*.

[21] N. Siddique, S. Paheding, M. Z. Alom, and V. Devabhaktuni, "Recurrent residual U-Net with EfficientNet encoder for medical image segmentation," *Proc. SPIE*, vol. 11735, pp. 134–142, Apr. 2021.

[22] S. Sorguli and H. Rjoub, "A novel energy accounting model using fuzzy restricted Boltzmann machine—Recurrent neural network," *Energies*, vol. 16, no. 6, p. 2844, Mar. 2023.

[23] Y.-T. Bai, W. Jia, X.-B. Jin, T.-L. Su, J.-L. Kong, and Z.-G. Shi, "Nonstationary time series prediction based on deep echo state network tuned by Bayesian optimization," *Mathematics*, vol. 11, no. 6, p. 1503, Mar. 2023.

[24] M. J. Abdulaal, I. M. Mehedi, A. J. Aljohani, A. H. Milyani, M. Mahmoud, M. K. Sahu, A. M. Abusorrah, and R. J. Meem, "Intelligent control techniques for the detection of biomedical ear infections," *Comput. Intell. Neurosci.*, vol. 2022, Sep. 2022, Art. no. 9653513, doi: 10.1155/2022/9653513.

MUHAMMAD BILAL was a Postdoctoral Researcher with KAIST, South Korea. He is currently an Educator, a Researcher, and a Maker. He is also an Associate Professor with the Department of Electrical and Computer Engineering, King Abdulaziz University. His research interests include digital image/signal processing, machine learning/AI, digital/analog circuit design, embedded systems, and robotics.

MAHENDIRAN T. VELLINGIRI received the B.E. degree in electrical and electronics engineering from the Maharaja Engineering College, Avinashi, affiliated to Bharathiyar University, Coimbatore, Tamil Nadu, India, in 2000, the M.E. degree in power electronics and drives from the K. S. Rangasamy College of Technology, Tiruchengode, affiliated to Anna University, Chennai, Tamil Nadu, in 2006, and the Ph.D. degree in electrical engineering from Anna University, in 2014. He is currently an Assistant Professor with the Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include soft computing applications in control of power electronics drives, control systems, electrical machines, solar energy, and power systems.

IBRAHIM M. MEHEDI (Senior Member, IEEE) received the Ph.D. degree from The University of Tokyo, Japan. He is a Distinguished Expert in control systems, renewable energy, biomedical engineering, AI, aerospace engineering, and biosensors. He has taught a diverse array of courses at King Abdulaziz University (KAU), Saudi Arabia, and the King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia, for over a decade. He has published over 100 peer-reviewed journal articles, holds five U.S. patents, and more than 1300 citations. He was recognized as a top 2% scientist worldwide according to the prestigious listing generated by Stanford University and published by Elsevier. In addition, he serves on the review panel for the U.S. National Defense Science and Engineering Graduate (NDSEG) Fellowship Program.

THANGAM PALANISWAMY (Senior Member, IEEE) received the B.E. degree in computer hardware and software engineering from Avinashilingam University, India, in 2001, and the M.E. degree in computer science and engineering and the Ph.D. degree in information and communication engineering from Anna University, India, in 2007 and 2013, respectively. She has a total teaching experience of 15 years in various reputed engineering colleges in Tamil Nadu. She is currently an Associate Professor with the Department of Electrical and Computer Engineering, King Abdulaziz University, Saudi Arabia. Her research interests include databases, data processing and mining, medical image analysis, image processing, cryptography, embedded systems, and the Internet of Things. Her contributions in professional societies include IEEE, the International Association of Engineers, Indian Society for Technical Education, and the International Association of Computer Science and Information Technology.

MUHAMMAD SHEHZAD HANIF received the B.Sc. degree in electrical engineering from the University of Engineering and Technology, Lahore, Pakistan, in 2001, and the M.S. degree in engineering sciences and the Ph.D. degree in computer engineering from Sorbonne University, Paris, France, in 2006 and 2009, respectively. He is currently an Associate Professor with the Department of Electrical and Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia. His research interests include machine learning, image analysis, and information fusion.

• • •