Journal of Digital Imaging (2022) 35:947–961
https://doi.org/10.1007/s10278-022-00617-8

ORIGINAL PAPER

Classification of Ear Imagery Database using Bayesian Optimization based on CNN-LSTM Architecture

Kamel K. Mohammed[1,4] · Aboul Ella Hassanien[2,4] · Heba M. Afify[3,4]

Received: 15 November 2021 / Revised: 25 February 2022 / Accepted: 27 February 2022 / Published online: 16 March 2022

Abstract

The external and middle ear conditions are diagnosed using a digital otoscope. The clinical diagnosis of ear conditions is suffered from restricted accuracy due to the increased dependency on otolaryngologist expertise, patient complaint, blurring of the otoscopic images, and complexity of lesions definition. There is a high requirement for improved diagnosis algorithms based on otoscopic image processing. This paper presented an ear diagnosis approach based on a convolutional neural network (CNN) as feature extraction and long short-term memory (LSTM) as a classifier algorithm. However, the suggested LSTM model accuracy may be decreased by the omission of a hyperparameter tuning process. Therefore, Bayesian optimization is used for selecting the hyperparameters to improve the results of the LSTM network to obtain a good classification. This study is based on an ear imagery database that consists of four categories: normal, myringosclerosis, earwax plug, and chronic otitis media (COM). This study used 880 otoscopic images divided into 792 training images and 88 testing images to evaluate the approach performance. In this paper, the evaluation metrics of ear condition classification are based on a percentage of accuracy, sensitivity, specificity, and positive predictive value (PPV). The findings yielded a classification accuracy of 100%, a sensitivity of 100%, a specificity of 100%, and a PPV of 100% for the testing database. Finally, the proposed approach shows how to find the best hyperparameters concerning the Bayesian optimization for reliable diagnosis of ear conditions under the consideration of LSTM architecture. This approach demonstrates that CNN-LSTM has higher performance and lower training time than CNN, which has not been used in previous studies for classifying ear diseases. Consequently, the usefulness and reliability of the proposed approach will create an automatic tool for improving the classification and prediction of various ear pathologies.

Keywords Ear imagery database · Convolutional neural networks (CNN) · Hyperparameters · Bayesian Optimization · Long short-term memory (LSTM)

Introduction

The appropriate early detection of ear conditions is very significant to avoid hearing impairment [1]. The clinical examination of the ear is based on an otoscope and tuning fork to measure the ear function [2].

The otoscope is used as a diagnostic device to screen for ear canal and tympanic membrane (TM) that provide otoscopic images. The exact diagnosis of the ear conditions provides good control of the otoscopic examination for magnification and illumination [3].

The ear conditions are classified according to clinical symptoms, otoscopic images, age, duration, frequency, and complications [4].

The pathological conditions for the middle ear, such as otitis media and TM perforation, and the external ear,

✉ Heba M. Afify
hebaaffify@yahoo.com

Kamel K. Mohammed
tawfickamel@gmail.com

Aboul Ella Hassanien
aboitcairo@gmail.com

[1] Center for Virus Research and Studies, Al Azhar University, Cairo, Egypt

[2] Faculty of Computers and Information, Cairo University, Giza, Egypt

[3] Systems and Biomedical Engineering Department, Higher Institute of Engineering in Shorouk Academy, Al Shorouk City, Cairo, Egypt

[4] Scientific Research Group in Egypt (SRGE), Cairo, Egypt

such as the external auditory canal (EAC), have unique characteristics that affected the auditory system [5]. These otologic diseases may be difficult to discover by diagnostic imaging, especially for ear tumors [6].

However, the otoscope's diagnostic accuracy ranges from 72 to 82% with otolaryngologists [7]. It means that the diagnosis by otoscope suffered from low accuracy due to the unavailability of specialists in otologic disorders and difficulties in determining ear diseases [8]. It has been noted that pediatricians achieved the diagnostic accuracy of 50% for middle ear pathology and 62% for acute otitis media (AOM) with pneumatic otoscopy [9]. At the same time, the video otoscopy showed a diagnostic accuracy of 51% for AOM and 46% for serous otitis media (SOM) [10].

Also, the limitation of ear conditions is based on the real-time recognition of otitis media [11].

Most former otologic diagnoses using feature extraction have poor generalizability for complex cases of ear images [12].

The goal of this approach is to develop a deep learning-based system that integrates the CNN and LSTM networks to automatically classify the four ear disorder types. CNN is utilized to extract features in the current proposal, and LSTM is used to classify ear disorders based on those features. The LSTM network has a built-in memory that allows it to learn from long-term imperative events. Layers are completely connected in fully connected networks, but nodes among layers are connectionless and process only one input. The nodes in an LSTM are linked from a directed graph along a temporal sequence that is considered an input with a defined order [13]. The results in [13] conceded that the combining of CNN and LSTM is very useful for enhancing the outcomes of categorization. The values of model parameters, such as weights and biases, must be learned throughout the training phase to obtain an efficient deep learning LSTM model. The weight and biases are controlled by hyperparameters. These hyperparameters cannot be directly acquired because they need to be set suitably. Therefore, hyperparameter adjustment is needed, and it will take several iterations to find the best parameter combination. Evolutionary optimization, random search, and grid search are some of the optimization strategies that have been suggested for the hyperparameter tuning procedure. Several researchers have suggested that Bayesian optimization is a preferred technique for locating a global optimum [14]. Because it does not have any underlying space assumptions, it additionally was designed for black-box functions, and it takes no derivatives.

In this paper, the proposed CNN-LSTM and CNN architectures are applied both on the ear imagery dataset and compare the performance of two proposed models with previous work on the same database.

Related works

The main drawbacks of related works for otologic conditions were unavailable data in the public sources and small samples for training and testing procedures. This means that there is still an open area for scientists to produce an accurate diagnosis approach for otologic conditions.

Accordingly, the shortage of ear diagnostic strategies demands a new approach. Deep learning algorithms may play a major role in otoscopic image diagnostic to remove misinterpretation and guarantee reliable treatment [15]. Classification of otoscopic images, including AOM and otitis media with effusion (OME), is applied with accuracies ranging from 73.11% to 91.41% [16].

Moreover, machine learning algorithms are used to distinguish among normal ear, AOM, OME, and COM with an accuracy of 88.06% [17]. Also, there are two accuracy values for each method in machine learning algorithms in [18] for diagnosing otitis media. It achieved 81%·and 58% by the decision tree, and 86%·and 84% by the neural network method [18].

Recently, classification of TM lesions is presented using CNN with an accuracy of 97.9% for identifying the TM side and 91.0% for identifying the perforation presence [19] with discharge and cholesteatoma, which are hard to discover than direct perforation.

Also, Zafer [20] proposed a transfer learning approach to distinguish among normal AOM, chronic suppurative otitis media (CSOM), and earwax TM. However, the problem of this research is based on an insufficient ear dataset. Viscaino et al. [21] employed a machine learning approach and image processing methods for classifying between four ear conditions, including normal, myringosclerosis, earwax plug, and COM, with an accuracy of 93.9%.

The previous computational works for the ear conditions diagnosis are summarized in Table 1. The most dataset for ear conditions is private data [12, 16–19, 22–24], while the public data is available in two previous works only [20, 21]. However, the ear imagery dataset [21] is larger than the dataset [20] for four classes. The previous works for ear conditions [12, 16–18, 21–23] performed feature extraction and machine learning algorithms with an accuracy ranging from 73 to 89% on the different databases. The recent works for ear conditions [19, 20, 24] were performed on CNN architecture with higher classification accuracy than machine learning algorithms. Using CNN architecture is based on its need for large training time and a large dataset to produce good results. The large dataset consisting of 10,544 samples for six classes [24] is applied to CNN architecture with an accuracy of 93.67%. Also, the time cost for ear conditions diagnosis using the feature extraction is represented in some previous works

Table 1 A comparison of the related works for ear condition diagnosis

Ref	Dataset	Number of images	Number of classes	Data description	Implemented techniques	Accuracy
[16]	Private	186	2	Normal and otitis media	Image features, machine learning algorithms such as k-nearest neighbors, decision tree, linear discriminant analysis, Naïve Bayes, multilayer neural networks, support vector machine	73.11% to 91.41%
[22]	Private	100	3	Normal, otitis media, other pathologies	Color image distribution and Bayesian decision rule	62.5% to 74%
[23]	Private	181	3	AOM, OME, no effusion	Vocabulary, grammar, and decision tree	89.9%
[17]	Private	865	4	Normal, AOM, OME, COM	Feature-based segmentation, local binary pattern, the histogram of oriented gradients, and AdaBoost	88.06%
[12]	Private	486	5	Obstructing wax or foreign bodies in the external ear canal, normal, AOM, OME, CSOM	Image processing, visual features, and decision tree	80.6% for images taken with commercial video-otoscopes, and 78.7% for images captured on-site with a low-cost custom-made video-otoscope
[18]	Private	389	5	Normal, obstructing wax or foreign bodies in the external ear canal, AOM, OME, CSOM	Feature-based description, decision tree, and neural networks	81%·and 58% by decision tree, and 86%·and 84% by neural network method
[19]	Private	1818	2	TM left side and TM right side	Class activation map, and CNNs	97.9% for identifying the TM side and 91.0% for identifying the perforation presence
[24]	Private	10,544	6	Normal, tympanic perforation, attic retraction or atelectasis, myringitis or acute otitis media/external, OME, middle ear or EAC tumor or cerumen impaction	Nine public CNN models, Inception- V3, ResNet-101, Ensemble classifiers, fivefold cross-validation	93.·67%
[20]	Public	857	4	AOM, CSOM, Earwax, normal	Pre-trained DCNNs, deep features, and support vector machine algorithms such as artificial neural network, k-nearest neighbor, decision tree, and support vector machine	93.05% for DCNNs by VGG-16 and 99.47% by the combination of the fused fine-tuned deep features and support vector machine model
[21]	Public	880	4	Myringosclerosis, earwax plug, COM, normal	Feature extraction, machine learning algorithms such as support vector machine, k-nearest neighbors, and decision trees	93.9%

[16, 22]. The feature-based segmentation is applied for ear conditions diagnosis in some previous works [12, 17, 21]. The recent public datasets [20, 21] supported research works concerning comparing their algorithms to others. However, this public database [20] is not balanced to classify the different classes of ear diseases.

The main objective of the proposed approach is to apply a public dataset supplying a large number of images used in [21] on the CNN networks for differentiating four ear classes. Also, the proposed approach based on optimized LSTM architecture guaranteed a high accuracy and low training time than the previous works.

Materials and Methods

Structure of The Proposed Approach

In this paper, we proposed an approach based on CNN and bidirectional long short-term (BiLSTM) [25] for identifying four external and middle ear conditions. The CNN is utilized to extract features. Afterward, the CNN output is fed into the BiLSTM model, which is used to identify ear disorders based on those features. The Bayesian optimization [26] is used for tuning the hyperparameters of an LSTM architecture. The benefit of this optimizer is to find the best hyperparameters, which achieve the best accuracy on the testing dataset. The structure of the proposed approach is illustrated in Fig. 1. The main steps for the proposed approach are summarized in the following:

1. CNN is used to extract features.
2. Bayesian optimization is used to choose hyperparameter values.
3. LSTM architecture is used as a classifier that contains a BiLSTM layer, 2000 hidden units, and a dropout layer afterward followed by a fully connected layer. The dropout layer is used for solving the overfitting challenge [27]
4. The evaluation process is used to calculate the best accuracy from 30 iterations. Then, calculate sensitivity, specificity, and PPV [28].

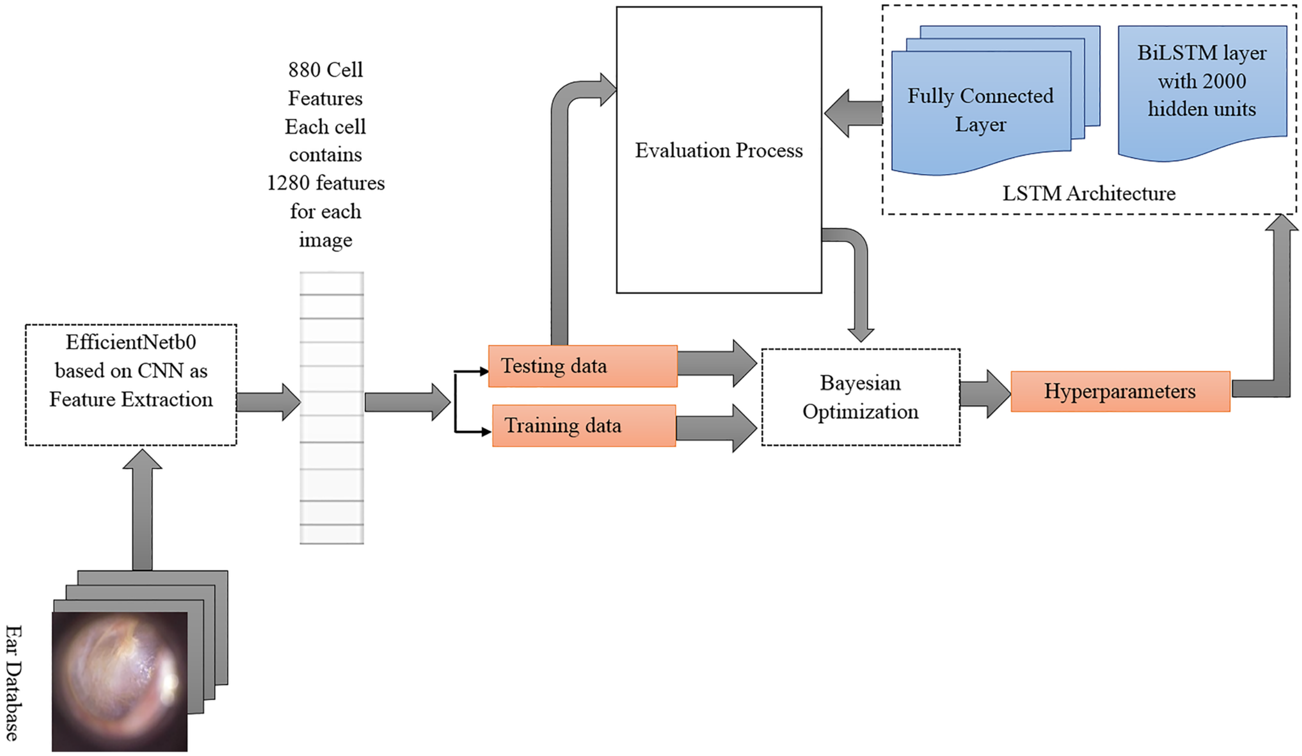

880 Cell Features Each cell contains 1280 features for each image
Evaluation Process
Fully Connected Layer
BiLSTM layer with 2000 hidden units
LSTM Architecture
EfficientNetb0 based on CNN as Feature Extraction
Testing data
Training data
Bayesian Optimization
Hyperparameters
Ear Database

Fig. 1 Structure of the proposed approach based on Bayesian optimization with the LSTM network architecture for diagnosing ear imagery dataset

5. The feedback structure is based on repeating all the above steps for 30 iterations to select the best results.

Description of Ear Imagery Database

The ear imagery database [29] is publicly available for 880 otoscopic images extracted from otolaryngologists in the Clinical Hospital from Universidad de Chile. The otoscopic images are collected from the patients with right and left ears using a digital otoscope DE500. All samples in the database were stored as RGB images with 420 × 380 pixels resolution, as shown in Fig. 2. This ear database holds four different classes, including normal, myringosclerosis, earwax plug, and COM, for which each class has 220 images.

Convolutional Neural Network

CNNs [26] were created precisely for dealing with image classification to acquire temporal and spatial dependencies and select suitable features. However, CNNs are suffered from a lack of the ability to learn sequential correlations and high time-consuming [25]. A kernel or filter in a CNN is a small square matrix that is used to acquire a specific feature from the input image. To create an output feature map, each filter is convolved with the input feature map. Finally, the outcomes are added together to provide one value in the output. Convolving the input feature map f with the kernel k (x, y) yields an output feature map. The three kinds of layers in a CNN, namely convolutional, pooling, and fully connected layers, are described as the following:

1. CNN's convolution layer has the majority of the computation and acts as a filter to collect essential features from an input picture.
2. The pooling layer is utilized to downsample in a feature map.
3. A fully linked layer means that every neuron in this layer is fully linked to the preceding layer, and it also collects positional and rotational invariant features from an input feature map.

The layer functions similarly to a traditional perception in that it integrates all of the input to generate the output categories. The inputs are the resolution (R) that refers to the size of a feature map, width (W) that refers to the number of channels/filters, depth (D) that refers to the number of layers, and F that refers to the size of the filter used in a layer. Those variables have a substantial impact on performance [30].

In this paper, two layers in CNN were considered for extracting the features during candidate training including the convolutional and pooling layers. Also, the LSTM network is used to solve sequential modeling problems and computational complexity in CNNs [27].

EfficientNetb0

This study used the EfficientNetb0 [31] which is a novel type of CNN network with incredible parameter efficiency and speed. To considerably increase the model's efficiency and accuracy, the EfficientNetb0 utilizes all dimensions of the recombination constant unified scaling model. Rather than freely expanding network dimensions such as D, R, and W, EfficientNetb0 utilizes simple and efficient recombination constant to enlarge the CNN in a further structured approach. The EfficientNetb0 model is more accurate and efficient than the previous CNN model, by utilizing this novel scaling technique and automated machine learning) AutoML(innovation. Additionally, the EfficientNetb0 parameter amount and floating-point operations per second (FLOPs) are both decreased by an order of magnitude [32].

Feature Extraction

The ear imagery database utilized in this approach is in RGB color space with 420 × 380 pixels resolution. In this study, the EfficientNetb0 pre-trained deep learning model was utilized to extract deep features from those images as shown in Fig. 1. The feature size that the pre-trained model extracts on each image in the fully connected layer was cell array 1280 × 1. Furthermore, EfficientNetb0 architectures

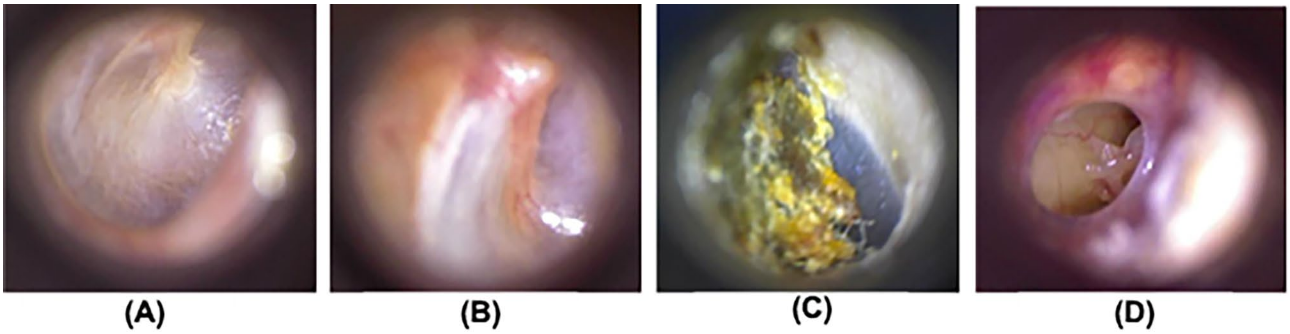


Fig. 2 Samples for Ear imagery database: (**A**) Normal ear, (**B**) Myringosclerosis, (**C**) Earwax plug, (**D**) Chronic otitis media (COM)

extracted 1280 features for each image. Finally, we obtained 880 cells, and these features were arranged in cell array to obtain 1280 × 1 cells for each image.

Bayesian Optimization

The Bayesian optimization [33] is a powerful tool for optimizing hyperparameters to avoid computational costs and leverage robust CNN architecture. This optimizer keeps former findings to select the best hyperparameters for the estimation process. This optimizer is successfully used for better classification when the data are complex [34]. Generally, hyperparameters are a group of parameters implemented for the learning procedure and consist of an integer or variable, which ranges from the lower to upper bound values. The superior hyperparameters should have a low loss function and high accuracy of the algorithm, considering the training time. The choice of hyperparameters is varied according to the algorithm objective.

The optimization process is dependent on Baye's theorem [35], which contains prior information of the objective function and updates posterior information to minimize loss and increase the classification accuracy. The posterior distribution is based on the Gaussian process [36] to update the former results of the objective function. Also, the acquisition function [26] is employed to define a balance between exploring new areas in the objective space and exploiting areas already known to have adequate values. Baye's theorem is based on a model Z and observation Y as the following equation.

$$P(Z|Y) = (P(Y|Z)P(Z))/P(Y) \tag{1}$$

where P(Z|Y) is the posterior probability of Z given Y, P(Y|Z) is the likelihood of Y given Z, and P(Y) is the prior probability of Y.

The Bayesian hyperparameter optimization is formulated by Eq. (2) as in previous work [35].

$$y^* = argminf(y) \tag{2}$$

Y is a set of hyperparameters in the domain, and f(y) is an objective score to reduce the error ratio during the learning procedure. The Bayesian optimization is implemented to find the minimum function f(y) on a bounded set Y.

Hyperparameters

In this paper, the selected hyperparameters are learning rate, momentum, regularization, and max epoch [37]. Previously, those four hyperparameters achieved good optimization results for the same database [38]. Therefore, hyperparameters were employed to make a real comparison with another study. The learning rate (α) is used to classify the comprehensive patterns in images based on the gradient loss function error. If the learning rate is low, important patterns may be unintentionally excluded. The momentum (δ) is used to detect the whole image without losing important elements based on updating the previous gradients. The objective of momentum value is to facilitate the gradient descent proceedings that reduce vertical oscillations, redirecting a good path to local optima with fewer iterations than the random gradient. The regularization (λ) allows good generalizability without overfitting in the database to obtain better predictions. The regularization is based on minimizing the weight by a small factor called weight decay to avoid the model complexity. In this paper, ridge regularization is used by computing the squares of all the feature weights based on loss function (L). Equations (3)–(5) for the hyperparameters are shown as the following:

$$\theta_{new} = \theta_{old} - \alpha\left(\frac{\delta}{\delta\theta_{old}}\right) * gradient \tag{3}$$

$$V_t = V_t^{\delta} + 1 + \alpha\Delta wL(W, Z, y) \tag{4}$$

$$W = W - V_t \tag{5}$$

where δ is the momentum ranging from 0 to 1, Δw is the gradient change, α is the learning rate, V_t is the variation in momentum according to the weight. L (W, Z, y) is the loss function of weight (W), model (Z), and hyperparameters (y). The gradient is used to find the new feature based on the old feature and momentum. Momentum with the stochastic gradient takes the exponentially weighted averages of the gradients. The objective of momentum is to smooth out the gradient descent phases, which reduces vertical oscillations, forwarding a better path to local optima with a lesser number of iterations than the stochastic gradient.

The loss function L is used to indicate regularization (λ), as shown in Eq. (6). If regularization has zero value, there is no effect on L. If regularization has a high value, there is underfitting. The zero value of regularization produced its best value.

$$L(y, z) = \sum_{i=1}^{n} (zi - yi)^2 + \lambda \sum_{i-1}^{n} \theta i^2 \tag{6}$$

where θi is the feature vector and y, z are the input values for the ith iteration.

LSTM Architecture

The proposed LSTM architecture for deep learning predictive network includes four layers. Figure 3 displays the proposed LSTM model structure. The first layer acts as the input of LSTM architecture that consists of features extracted from the CNN network. The second layer of the LSTM architecture is BiLSTM, which consists of 2000 hidden units. The third layer is the dropout layer that defined as 0.5 to prevent overfitting as well as it is commonly used and received satisfying performance. The final layer is a fully connected layer with an output size corresponding to the number of classes and a softmax layer as shown in Fig. 3.

The LSTM [30] structure is a unit of recurrent neural networks (RNNs); recently, it performs better classifications than deep neural network systems for the classification of signal and sound. RNNs are neural networks in addition to memories that are capable of recalling all data recorded in the previous element in sequential order. In different words, RNNs are an effective method to utilize data from somewhat lengthy series, because they carry out identical tasks for each element in the series, with output based on all preceding calculations. A feed-forward neural network with an extra cyclic loop is known as an RNN. The information is carried from one-time step to the next by this cyclic loop. Cyclic loops are a type of short-term memory that stores and recovers past data over time scales. The current time-step is estimated using the previous state and the current state by a recurrent neural network, which learns temporal patterns. When recurrent neural networks must learn long-term dependencies in time steps, the problem of vanishing gradients occurs. Therefore, when propagating across several layers of an RNN to learn long-term dependencies in time steps, the gradient vector either rises or decays exponentially. LSTM addresses the vanishing gradient problem. To solve the vanishing and exploding gradient problem, LSTM suggests memory blocks rather than traditional RNN units [30]. The key difference between it and RNNs is that it adds a cell state to save long-term states. An LSTM network can recall and link earlier data to data collected in the present [31].

Through this study, the bidirectional LSTM (BiLSTM) [25] is used to classify the four types from the ear imagery database based on data division into 90% for the training and 10% for testing.

Bayesian Optimization for LSTM Architecture

Many works [39, 40] presented to define the extent of hyperparameters' influence on CNN architecture. Some of the hyperparameters are more important than others. It is noted that the time cost and classification error are increased when using inappropriate hyperparameters.

The performance of LSTM architecture depended on a good setting for hyperparameters, including learning rate, momentum, regularization, and maximum epoch. It is noted that the proposed architecture cannot be implemented to tune the number of layers and the filter size. The reason is that the architecture and connectivity of structure are considered as a sequential decision problem. Using LSTM architecture only is based on a long time for tuning hyperparameters and limited accuracy.

The function of the optimizer is to find the best hyperparameters with less time through training samples. Compared with traditional optimization methods [14, 41] and manual tuning, these methods are based on the trial-and-error concept, which is ineffective for choosing hyperparameters. Hence, Bayesian optimization is suitable for tuning hyperparameters.

In this paper, the Bayesian optimization algorithm is applied to tune hyperparameters for LSTM architecture [42, 43], including a BiLSTM layer and a fully connected layer. The Bayesian optimization created the objective function for the Bayesian optimizer using the training and testing database as inputs. The objective function trained an LSTM architecture and returned the classification error on the testing set.

For implementation, the input of LSTM architecture consists of four hyperparameters extracted from the Bayesian optimization and training dataset. The output of LSTM architecture is combined with the testing dataset for evaluating the classification process as shown in Fig. 4.

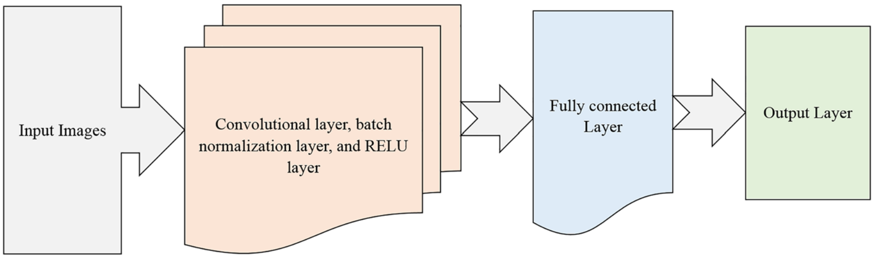

Input Images
Convolutional layer, batch normalization layer, and RELU layer
Fully connected Layer
Output Layer

Fig. 3 Structure of LSTM Network

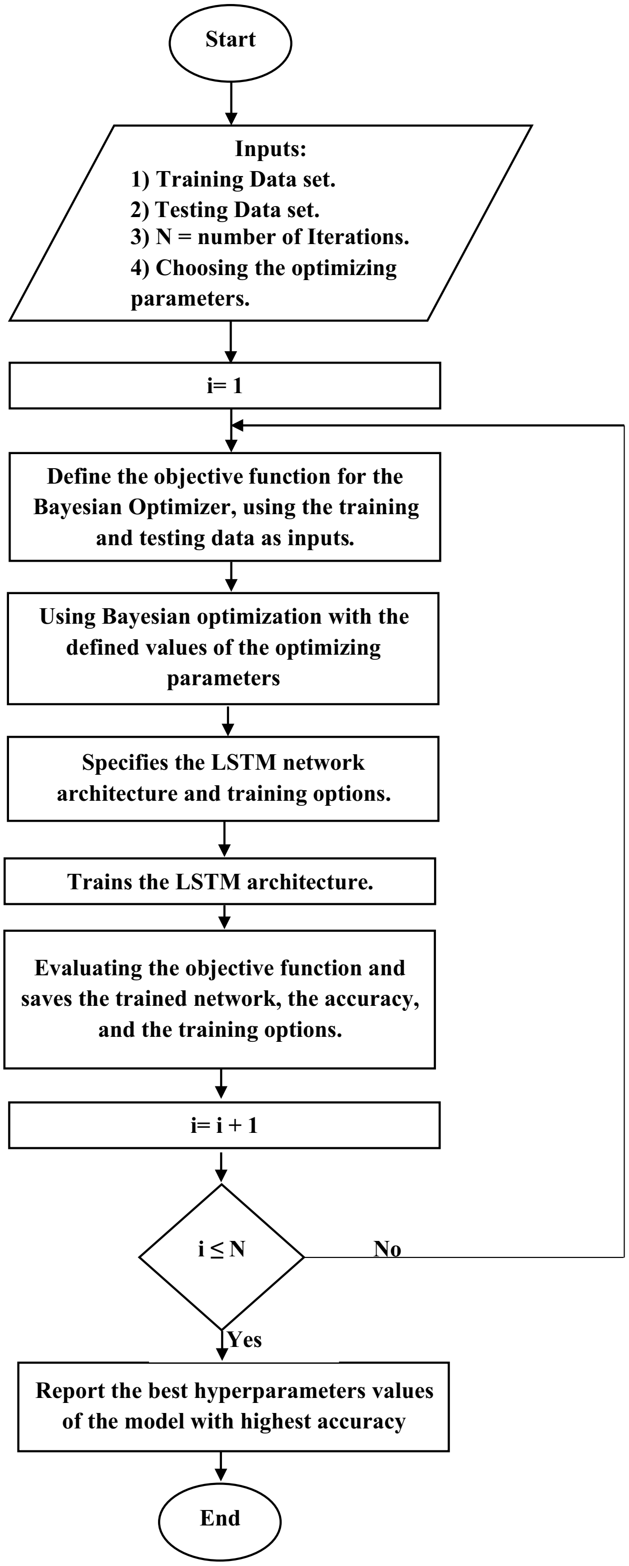

Start
Inputs:
1) Training Data set.
2) Testing Data set.
3) N = number of Iterations.
4) Choosing the optimizing parameters.
i= 1
Define the objective function for the Bayesian Optimizer, using the training and testing data as inputs.
Using Bayesian optimization with the defined values of the optimizing parameters
Specifies the LSTM network architecture and training options.
Trains the LSTM architecture.
Evaluating the objective function and saves the trained network, the accuracy, and the training options.
i= i + 1
i ≤ N
No
Yes
Report the best hyperparameters values of the model with highest accuracy
End

Fig. 4 Flowchart of the Bayesian optimization with LSTM architecture

The Bayesian optimization about LSTM architecture is explained in the following steps:

1. Select the initial optimizing hyperparameters from images.
2. Evaluate the objective function by using an acquisition function [26].
3. Run for 30 iterations
4. Select the best-optimized values
5. Use the optimized hyperparameters in the validate database.

Optimization Criterion

There are some factors for classifier's performance that has been related to endeavors aimed to achieve the best results.

In this paper, the evaluation process for LSTM architecture [13] is based on four metrics: accuracy, sensitivity, specificity, and PPV, as used in previous work [21]. Accuracy is the overall effectiveness of a classifier. Sensitivity is used to avoid false negative samples, while specificity is used to identify the samples without diseases correctly. PPV is used as a precision value to calculate the accurately classified samples according to the sum of classified samples. Equations (7)–(10) described the four metrics.

$$\text{Accuracy} = \frac{\sum_{i=1}^{c} \frac{TP_i+TN_i}{TP_i+FN_i+FP_i+TN_i}}{c} \tag{7}$$

$$\text{Sensitivity} = \frac{\sum_{i=1}^{c} \frac{TP_i}{TP_i+FN_i}}{c} \tag{8}$$

$$\text{Specificity} = \frac{\sum_{i=1}^{c} \frac{TN_i}{TN_i+FP_i}}{c} \tag{9}$$

$$\text{PPV} = \frac{\sum_{i=1}^{c} \frac{TP_i}{TP_i+FP_i}}{c} \tag{10}$$

Table 2 The selected hyperparameters for the proposed approach

Hyperparameters	Initial value	Final value	Type
Max Epoch	20	100	int
Learning rate (α)	1×10^{-4}	1	log
Momentum (δ)	0.8	0.98	log
Regularization (λ)	1×10^{-10}	0.01	log

Table 3 Hyperparameter values for five iterations during the tuning process

Iterations	Learning rate	Momentum	Regularization	Max Epoch	Accuracy
1	0.000224	0.97692	0.0031402	91	100%
2	0.89076	0.88748	1.6897×10^{-9}	93	77.3%
3	0.29861	0.95071	0.002767	49	86.4%
4	0.001644	0.91862	1.741887×10^{-7}	24	96.6%
5	0.073819	0.88923	9.8598×10^{-9}	87	98.8%

where *c* is the number of categories, TP is true positive, which is a correctly classified category, TN is true negative, which is correctly classified as not relating category, FP is false positive, which is incorrectly classified category, and FN is a false negative, which is incorrectly classified as not relating category.

Experimental Results

The proposed approach is executed to classify four categories: normal, myringosclerosis, earwax plug, and COM obtained from the ear imagery database under 5 GB RAM GPU and MATLAB 2020 software.

In the training and testing of the proposed approach, the CNN architecture adopted Bayesian optimization through four hyperparameters, including learning rate, momentum, regularization, and maximum epoch, for automatic diagnosis of four ear conditions. The hyperparameters are optimized via the layers, and extensive optimization is carried out through the testing dataset.

According to the previous work [25], the LSTM network is more effective on classification accuracy than the CNN. In this paper, the LSTM architecture is used to classify the four groups extracted CNN-EfficientNetb0 architecture from the ear imagery dataset [29] which each image has a size 420 × 380. The total ear database occupied 880 images that contain 220 for each type. The proposed LSTM architecture based on a deep learning network mainly includes four layers as in Fig. 3. Different from a previous study [21], we combined CNN and LSTM [33] to obtain a more precise estimation under the ear imagery database.

The selected hyperparameters for Bayesian optimization are shown in Table 2. The optimal hyperparameters for five iterations only using Bayesian optimization are shown in Table 3. Table 3 shows that the four selected hyperparameters affect the LSTM accuracy. The best feasible points for the hyperparameters used for testing data using 30 iterations are shown in Table 4. The best values for the hyperparameters are employed for iterations to minimize the generalization error and increase the classification accuracy. The obtained results for the proposed approach are reported accuracy of 100%, sensitivity of 100%, specificity of 100%, and PPV of 100%, using the testing set as shown in the confusion matrix in Fig. 5.

Table 4 Best observed feasible values for 30 iterations

Iterations	Learning rate	Momentum	Regularization	Network depth
30 iterations	0.000224	0.97692	0.0031402	91

For every iteration, the images extracted from the acquisition function are estimated over the objective function. The images are added to the data to update its posterior through a feedback structure.

The objective function is employed as a Gaussian process [37] to motivate a posterior distribution. The relationship between the minimum objective obtained over the 30 iterations and the number of function evaluations is illustrated in Fig. 6.

The training progress of the different optimized LSTMs is shown in Fig. 7. According to the comparison between Table 3 and Fig. 7, the first iteration for learning rate of 0.000224, momentum of 0.97692, regularization of 0.0031402, max epoch of 91, and accuracy of 100% is displayed in Fig. 7a. The fourth iteration for learning rate of 0.001644, momentum of 0.91862, regularization of 1.741887×10^{-7}, max epoch of 24, and accuracy of 96.6% is displayed in Fig. 7b. The third iteration for learning rate of 0.29861, momentum of 0.95071, regularization of 0.002767, max epoch of 49, and accuracy of 86.4% is displayed in Fig. 7c.

The training progress of the CNN-EfficientNetb0 [31] and combined CNN-LSTM [33] is shown in Fig. 8a, b, respectively. The receiver operating characteristic (ROC) curve for CNN and combined CNN-LSTM is shown in Fig. 9a, b, respectively.

For one iteration, the training time on CNN and CNN-LSTM networks is 637 min and 25 min, respectively. This means that the proposed approach (CNN-LSTM) with Bayesian optimization achieved a lower training time rather than the CNN classifier only.

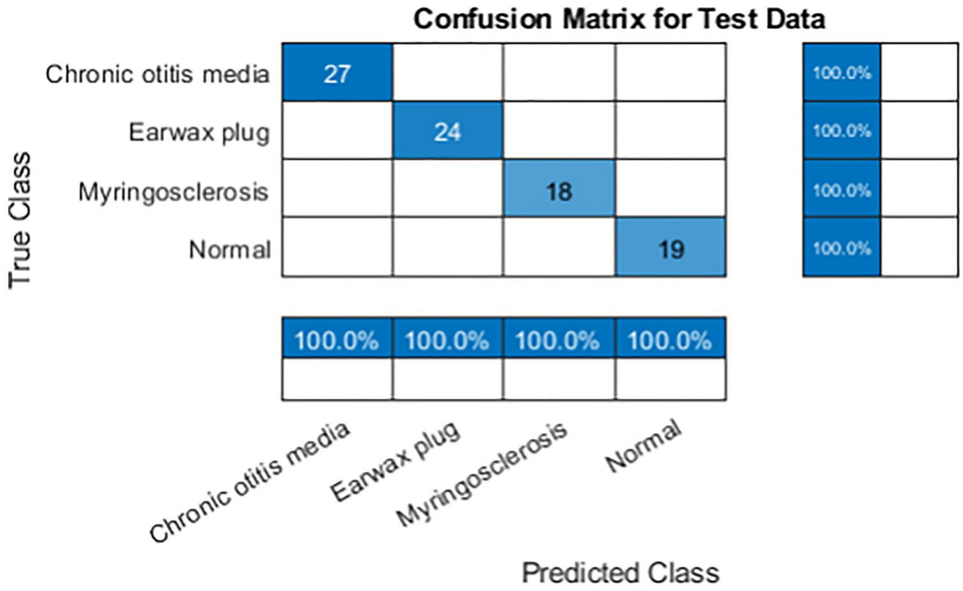


Fig. 5 The confusion matrix for four ear classes on testing images

Comparison with Other Previous Works

The proposed approach based on CNN-LSTM [33] and CNN classifier [31] was built on the ear imagery dataset [29]. The performance of these two approaches was compared by calculating accuracy, sensitivity, specificity, and PPV, and applying them to 90% training data with 10% test data, 80% training data with 20% test data, and 70% training data with 30% test data. For 90% training data and 10% test data, the CNN-LSTM performed best among the two approaches, the accuracy, sensitivity, specificity, and PPV on testing dataset reaching 100% as in Table 5. As the CNN classifier was used, the accuracy reached 86.3% as in Fig. 8a, which is the lowest among the previous work [21]. A previous work using machine learning algorithms to classify the ear imagery dataset yielded an accuracy of 93.9%.

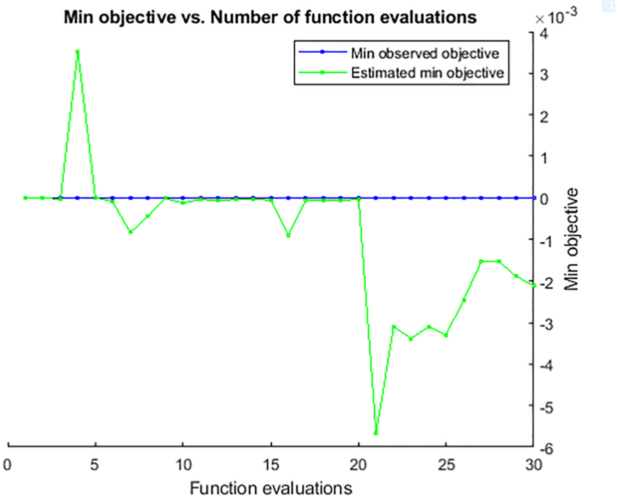


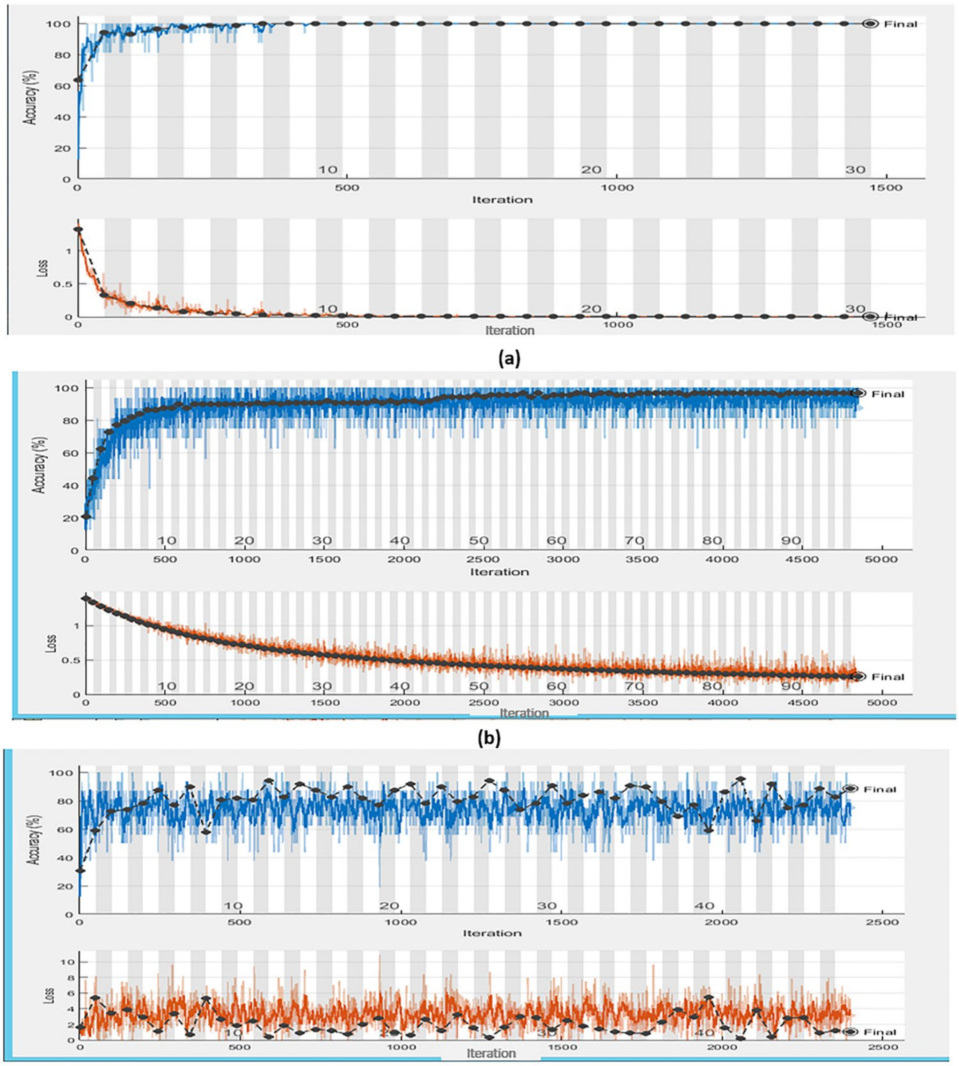
Fig. 6 Minimum objective and number of function evaluations for 30 iterations

As the CNN-LSTM and Bayesian optimization were used on the ear imagery dataset [29], there were increases in the four metrics of the testing dataset. For 80% of training data and 20% of test data, the CNN-LSTM achieved accuracy, sensitivity, specificity, and PPV on a test dataset of 100%. For 70% of training data and 30% of test data, the CNN-LSTM achieved accuracy, sensitivity, specificity, and PPV on a test dataset of 99.62%, 99.65%, 99.88%, and 99.62%, respectively.

Discussion

Generally, the image-assisted ear detection approach is essential to support otolaryngologists during the disease diagnosis task [44]. In some cases, otolaryngologists do not have adequate training and experience to assign the right decisions [45]. Therefore, alternative solutions are needed to assist otolaryngologists in improving diagnostic accuracy.

The researchers discussed the private and public databases of image-assisted detection procedures for ear conditions, as shown in Table 1. It was noted that there is no possibility for making a good comparison between previous works of ear conditions diagnosis due to the differences in the samples, classes, and implemented techniques. The contribution of this work relates to how images of ear diseases including normal, myringosclerosis, earwax plug, and COM are classified with high accuracy and less computational time.



(a)

(b)

(c)

Fig. 7 Training progress for the different optimized LSTMs according to different iterations. (**a**) First iteration with an accuracy of 100%, (**b**) fourth iteration with an accuracy of 96.6%, and (**c**) third iteration with an accuracy of 86.4%

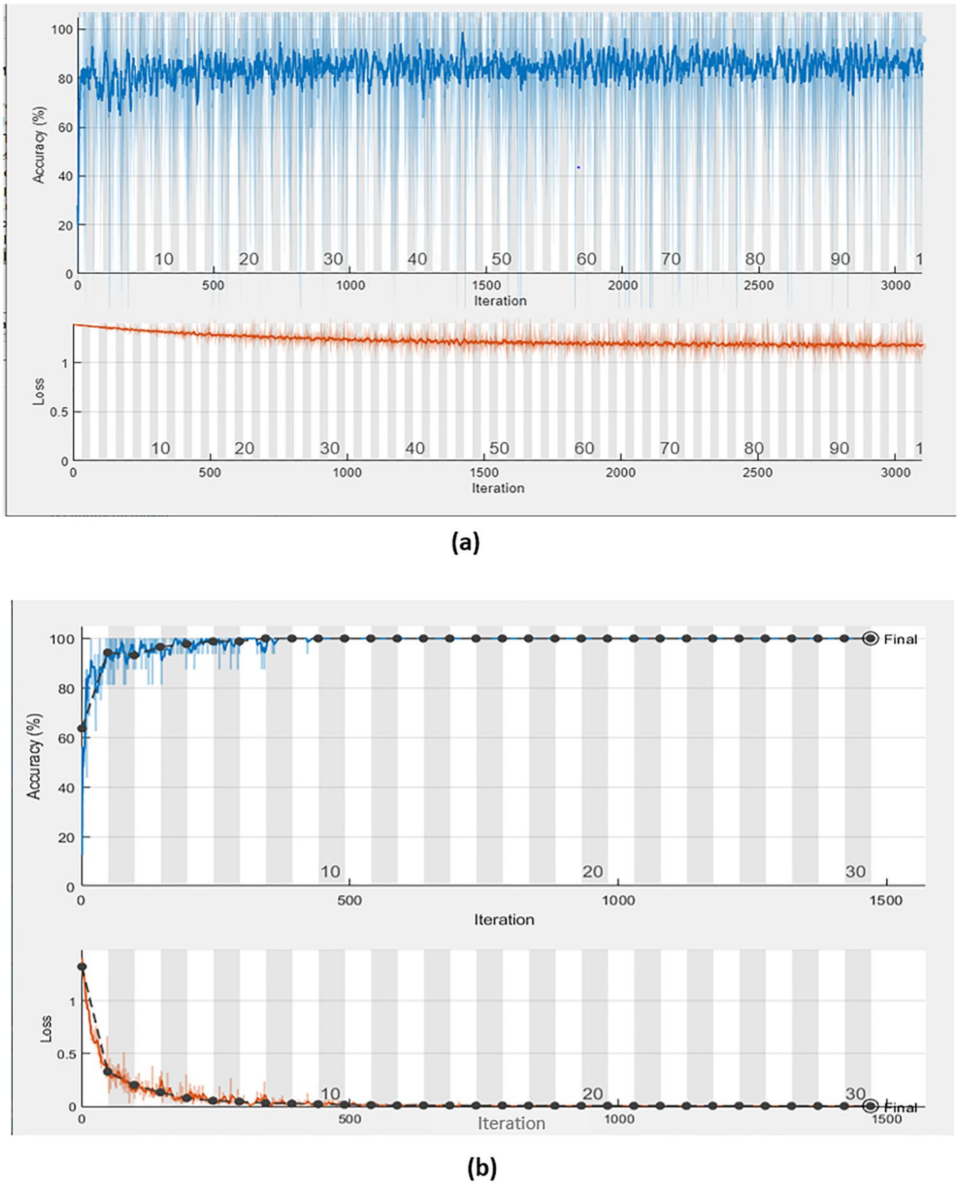
The proposed approach used the recent public database, namely the ear imagery dataset containing 880 images [29] that previously applied to machine learning algorithms with an accuracy of 93.9% [21]. On the other hand, the private dataset containing 389 images applied to machine learning algorithms with an accuracy of 86.8% [18]. It means that the ear dataset with large samples is more effective for improving the classification process. Large samples are not available in the otolaryngology field, which created many restrictions for CNN architecture.

In the proposed approach, we demonstrated a model based on CNN and BiLSTM for identifying four external and middle ear conditions. Our method is implemented to differentiate ear categories in the ear imagery dataset and attained an accuracy of 100%. The test error was 0%, which achieved a low rate during testing the proposed approach. These findings outperform what is reported in the literature [21]. The Bayesian optimization extended the iterations automatically and stopped when it reached the maximum optimized values.

The Bayesian optimization for LSTM architecture generated 90% training images to produce the best hyperparameters that help in the multi-class classification process. The multi-class classification process generated 10% testing images to record the approach performance by the evaluated four metrics.

The confusion matrix referred to the classifier's accuracy value, which is based on the relation between the correct predicted samples in the matrix diagonal and the misclassified samples on the outside of the matrix diagonal.

Fig. 8 The training progress of (**a**) the CNN based on EfficientNetb0 and (**b**) combined CNN-LSTM



The comparison between the proposed approach and previous works concerning four metrics is shown in Table 5. From experimental results, the incorporation of CNN-LSTM with Bayesian optimization is very beneficial in the ear imagery diagnosis where the CNN is insufficient to build a more effective classification model.

In this paper, a novel approach based on CNN-LSTM with Bayesian optimization is suggested that has potential application in diseases classification. The limitation of this work is based on small-scale studies on the ear imagery database, and the performance of the proposed approach has not been compared with different databases because there are not enough public databases for ear diseases. Therefore, this work focused on comparison with two algorithms and different data divisions for the ear imagery database.

Furthermore, the proposed approach is qualified for diagnosing otoscopic suspicion cases rather than the optical examination of the ear by a manual otoscope. In the future, the proposed approach will expand to validate the final results according to the opinion of otolaryngologists.

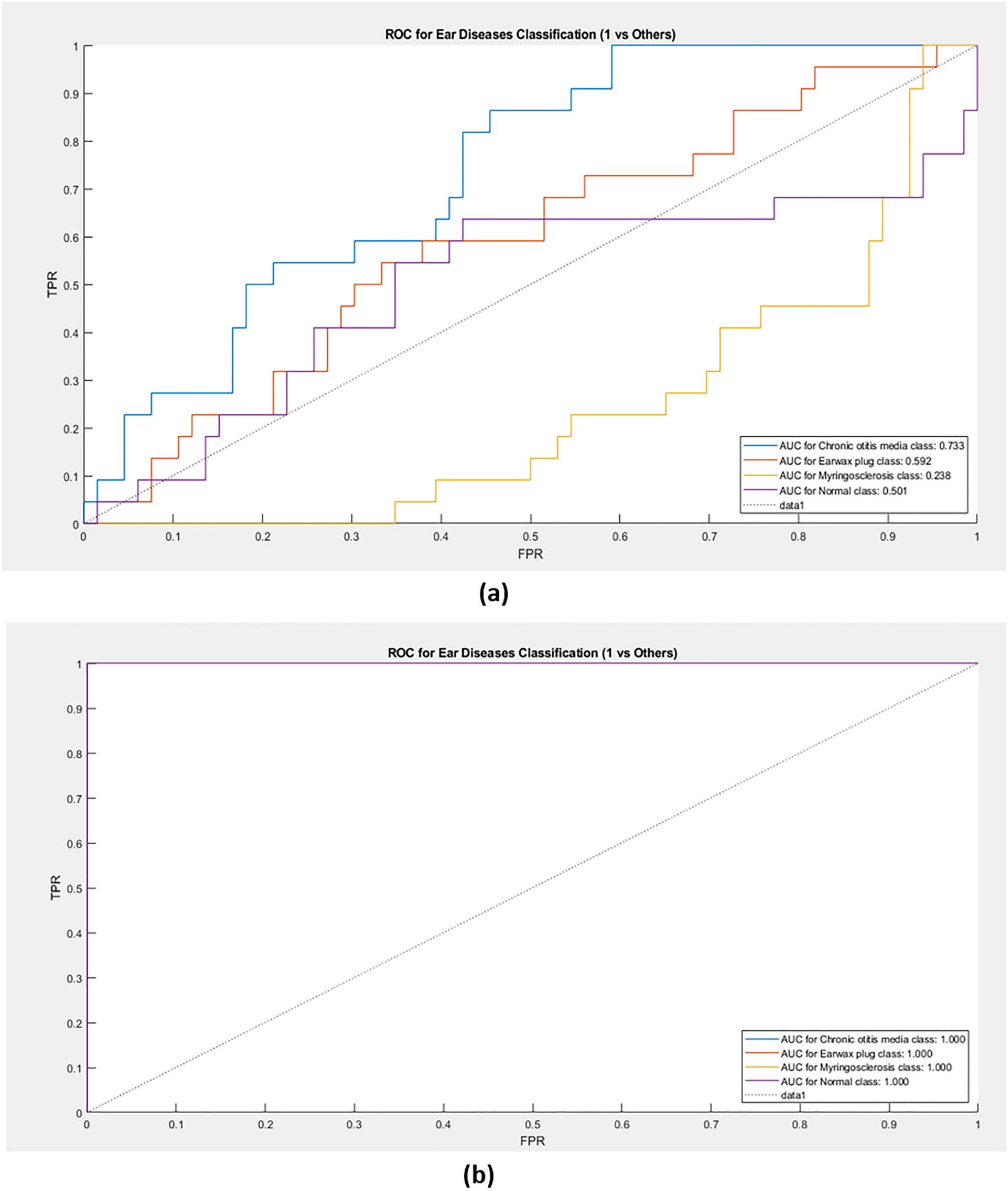


Fig. 9 ROC curve for (**a**) CNN only and (**b**) combined CNN-LSTM

Table 5 Performance of the proposed approach and other previous works in terms of four metrics

	Accuracy %	Sensitivity %	Specificity %	PPV %
Proposed approach (CNN-LSTM) with Bayesian optimization	**100**	**100**	**100**	**100**
CNN classifier without LSTM and Bayesian optimization	86.3	86.3	95.4	86.2
Previous work [21] based on machine learning	93.9	87.8	95.9	87.7

Conclusion

The otologic conditions are extremely subspecialized, causing diagnosis challenges for an otolaryngologist, potentially leading to detrimental patient results. The problems in otologic diagnoses are focused on lack of specialists, self-decisions from general practitioners, expensive diagnostic devices, and limited databases. This paper could overcome these problems by the CNN-LSTM approach based on Bayesian hyperparameter optimization for the automatic diagnosis of ear conditions. It was noted that the proposed approach obtained the best findings by using ear conditions that have been estimated before, such as myringosclerosis, earwax plug, normal, and COM in previous works. Finally, otologic image processing will enhance patient care in otolaryngology, and future efforts will advance the perspective of ear disease prediction models.

References

1. Block, S.L.; Mandel, E.; Mclinn, S.; Pichichero, M.E.; Bernstein, S.; Kimball, S.; Kozikowski, J. Spectral gradient acoustic reflectometry for the detection of middle ear effusion by pediatricians and parents. Pediatr. Infect. Dis. J. 1998, 17, 560–564.
2. Wang X, Valdez TA, Bi J. Detecting tympanostomy tubes from otoscopic images via offline and online training. Comput Biol Med. 2015; 61:107–18.
3. Lieberthal, A.S.; Carroll, A.E.; Chonmaitree, T.; Ganiats, T.G.; Hoberman, A.; Jackson, M.A.; Jo_e, M.D.; Miller, D.T.; Rosenfeld, R.M.; Sevilla, X.D. The diagnosis and management of acute otitis media. Pediatrics 2013, 131, e964–e999
4. Harnsberger HR (1995) The temporal bone: external, middle and inner ear segments. In: Gay SM (ed) Handbook of head and neck imaging Mosby, St. Louis 426–458
5. Agnieszka Trojanowska & Andrzej Drop & Piotr Trojanowski & Katarzyna Rosińska-Bogusiewicz & Janusz Klatka & Barbara Bobek-Billewicz, External and middle ear diseases: radiological diagnosis based on clinical signs and symptoms, Insights Imaging (2012) 3:33–48
6. Devaney KO, Boschman CR, Willard SC, Ferlito A, Rinaldo A (2005) Tumours of the external ear and temporal bone. Lancet Oncol 6:411–420.
7. Moberly, A.C.; Zhang, M.; Yu, L.; Gurcan, M.; Senaras, C.; Teknos, T.N.; Elmaraghy, C.A.; Taj-Schaal, N.; Essig, G.F. Digital otoscopy versus microscopy: How correct and confident are ear experts in their diagnoses? J. Telemed. Telecare 2018, 24, 453–459.
8. Pichichero ME, Poole MD. Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. Arch Pediatr Adolesc Med. 2001; 155(10):1137–42
9. Asher E, Leibovitz E, Press J, Greenberg D, Bilenko N, Reuveni H. Accuracy of acute otitis media diagnosis in community and hospital settings. Acta Paediatr. 2005; 94(4):423–8.
10. Pichichero ME, Poole MD. Comparison of performance by otolaryngologists, pediatricians, and general practioners on an otoendoscopic diagnostic video examination. Int J Pediatr Otorhinolaryngol. 2005; 69(3):361–6.
11. Davies, J.; Djelic, L.; Campisi, P.; Forte, V.; Chiodo, A. Otoscopy simulation training in a classroom setting: A novel approach to teaching otoscopy to medical students. Laryngoscope 2014, 124, 2594–2597.
12. Myburgh HC, van Zijl WH, Swanepoel D, Hellstrom S, Laurent C. Otitis media diagnosis for developing countries using tympanic membrane image analysis. EBioMedicine. 2016; 5:156–60.
13. Wang Y, Wu Q, Dey N, Fong S, Ashour AS. Deep back propagation–long short-term memory network based upper-limb sEMG signal classification for automated rehabilitation. Biocybern Biomed Eng 2020; 40:987–1001
14. J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, Curran Associates, Inc., 2012, pp. 2951–2959.
15. Tran, T.-T.; Fang, T.-Y.; Pham, V.-T.; Lin, C.; Wang, P.-C.; Lo, M.-T. Development of an Automatic Diagnostic Algorithm for Pediatric Otitis Media. Otol. Neurotol. 2018, 39, 1060–1065.
16. Mironica, I.; Vertan, C.; Gheorghe, D.C. Automatic pediatric otitis detection by classification of global image features. In Proceedings of the E-Health and Bioengineering Conference (EHB), Iasi, Romania, 24–26 November 2011; pp. 1–4.
17. Shie, C.-K.; Chang, H.-T.; Fan, F.-C.; Chen, C.-J.; Fang, T.-Y.;Wang, P.-C. A hybrid feature-based segmentation and classification system for the computer aided self-diagnosis of otitis media. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 4655–4658.
18. Myburgh HC, Jose S, Swanepoel DW, Laurent C. Towards low cost automated smartphone- and cloud-based otitis media diagnosis. Biomed Sig Process Control 2018; 39:34–52.
19. Je Yeon Lee 1827 Seung-Ho Choi, Jong Woo Chung, Automated Classification of the Tympanic Membrane Using a Convolutional Neural Network Appl. Sci. 2019 9
20. Zafer C. Fusing fine-tuned deep features for recognizing different tympanic membranes. Biocybernetics and Biomedical Engineering. 2020 January 1; 40(1):40–51.
21. Michelle Viscaino, Juan C. Maass, Paul H. Delano, Mariela Torrente, Carlos Stott, Fernando Auat Cheein, Computer-aided diagnosis of external and middle ear conditions: A machine learning approach, PLOS ONE, 2020.
22. Vertan C, Gheorghe DC, Ionescu B. Eardrum color content analysis in video-otoscopy images for the diagnosis support of pediatric otitis. ISSCS 2011 - Int Symp Signals, Circuits Syst Proc. 2011. pp. 129–32.
23. Kuruvilla A, Shaikh N, Hoberman A, Kovacevic´ J. Automated diagnosis of otitis media: vocabulary and grammar. J Biomed Imaging 2013; 2013:27.

24. Cha D, Pae C, Seong S-B, Choi JY, Park H-J. Automated diagnosis of ear disease using ensemble deep learning with a big otoendoscopy image database. EBioMedicine 2019; 45:606–14.
25. Muhammet Fatih Aslan, Muhammed Fahri Unlersen, Kadir Sabanci, Akif Durdu, CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection, Appl Soft Comput. 2021 Jan; 98: 106912.
26. Joy TT, Rana S, Gupta S, Venkatesh S. Hyperparameter tuning for big data using Bayesian optimisation. 23rd International Conference on Pattern Recognition (ICPR) Cancún Center, Cancún, México, pp.2575- 2580, December 4–8, 2016.
27. Hamad Naeem· Ali Abdulqader Bin-Salem, A CNN-LSTM network with multi-level feature extraction-based approach for automated detection of coronavirus from CT scan and X-ray images, Appl Soft Comput . 2021, 113, Part A,113:107918.
28. Swaminathan S, Qirko K, Smith T, Corcoran E, Wysham NG, Bazaz G, et al. A machine learning approach to triaging patients with chronic obstructive pulmonary disease. PloS one. 2017; 12(11): e0188532.
29. http://www.ctganalysis.com/Category/otitis-media
30. L.T. Duong, P.T. Nguyen, C. Di Sipio, D. Di Ruscio, Automated fruit recognition using efficientnet and mixnet, Comput. Electron. Agric. 171 (2020) 105326.
31. Yin, Xuqiang, Wu, Dihua, Shang, Yuying, Jiang, Bo, Song, Huaibo, 2020. Using an EfficientNet-LSTM for the recognition of single Cow's motion behaviours in a complicated environment. Comput. Electron. Agric. 177, 105707.
32. R. Zhu, X. Tu, and J. Xiangji Huang, Chapter seven - deep learning on information retrieval and its applications, in Deep Learning for Data Analytics (H. Das, C. Pradhan, and N. Dey, eds.), pp. 125 – 153, Academic Press, 2020.
33. Islam MZ, Islam MM, Asraf A, A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Informatics in Medicine Unlocked 20 (2020) 100412.
34. Shahriari, Bobak, Swersky, Kevin, Wang, Ziyu, Adams, Ryan P., and de Freitas, Nando. Taking the human out of the loop: A review of bayesian optimization. Technical report, Universities of Harvard, Oxford, Toronto, and Google DeepMind, Proceedings of the IEEE 104.1 (2015): 148–175.
35. Kochanski G, Golovin D, Karro J, Solnik B, Moitra S, Sculley D. Bayesian optimization for a better dessert. In: 31st conference on neural information processing systems (NIPS) Long Beach, CA, USA, pp.1–10, 2017.
36. Kramer O, Ciaurri DE, Koziel S (2011) Derivative-free optimization. In: Computational optimization, methods and algorithms. Springer, pp. 61–83
37. Rasmussen, C. E. and Williams, C. K. I. Gaussian Processes for Machine Learning. In summer school on machine learning, Springer, Berlin, Heidelberg, pp. 63-71, 2006.
38. A. Helen Victoria, G. Maragatham, Automatic tuning of hyperparameters using Bayesian optimization, Evolving Systems, pp.1–7, 2020.
39. A. Koutsoukas, K.J. Monaghan, X. Li, Jun Huan, Deep learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data, J. Cheminformat. 9 (42) (2017).
40. Y. Yoo, Hyperparameter optimization of deep neural network using univariate dynamic encoding algorithm for searches, Knowl.-Based Syst. 178 (2019) 74–83.
41. S.R. Young, D.C. Rose, T.P. Karnowski, S. Lim, R.M. Patton, Optimizing deep learning hyper-parameters through an evolutionary algorithm, in Proceedings of the Workshop on Machine Learning in High-Performance Computing Environments (MLHPC 2015), ACM, Austin, Texas, 2015, pp. 1–5.
42. J. Li, P. Li, D. Guo, X. Li, Z. Chen, Advanced prediction of tunnel boring machine performance based on big data, Geosci. Front. 12 (1) (2020) 331–338.
43. S. Gonçalves, P. Cortez, S. Moro, A deep learning classifier for sentence classification in biomedical and computer science abstracts, Neural Comput. & Applic. 32 (11) (2020) 6793–6807.
44. Senaras C, Moberly AC, Teknos T, Essig G, Elmaraghy C, Taj-Schaal N, et al. detection of eardrum abnormalities using ensemble deep learning approaches. Proceeding in medical imaging 2018: Computer- Aided Diagnosis. 2018 February 27; Houston USA; 10575, pp.105751A.
45. Huang YK, Huang CP. A depth-first search algorithm based otoscope application for real-time otitis media image interpretation. Parallel Distrib Comput Appl Technol PDCAT Proc 2018; 2017(Decem):170–5.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.