Republic of Iraq
Ministry of Higher Education and Scientific Research
Tikrit University
College of Computer Science and Mathematics
Department of Computer Science

Analysis of Tympanomertic Data Based on ABC-LSTM Hybrid Model for Classification of Otitis Media

A Thesis submitted

By

Mohammed Mahmoud Hussein

To

The Council of the College of Computer Science and Mathematics, University of Tikrit

in partial fulfillment of the Requirements for obtaining the Master's Degree in Computer Science

Supervised By

Asst. Prof. Salwa Khalid Abdulateef

2025 A.D. **1447 A.H.**

بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

﴿ يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ ﴾

صدق الله العظيم

[سورة المجادلة: الآية 11]

Dedication

This thesis is dedicated to my beloved mother, whose endless love and sacrifices have shaped my path;

To the memory of my martyred father, whose spirit remains a source of pride and inspiration;

To my precious daughter, who gives meaning and hope to my life;

And to my brothers, whose constant support and encouragement have been my strength.

With deep gratitude and sincere appreciation, I also dedicate this work to my esteemed supervisor,

Asst. Prof. Salwa Khalid Abdulateef, for her invaluable guidance and unwavering encouragement.

Acknowledgment

First foremost, my grateful thanks are due to Allah, Almighty, for granting me the power and energy to conduct this humble work. I would like to express my deepest gratitude to my supervisor Asst. Prof. Salwa Khalid Abdulateef, for her efforts, encouragement, and valuable comments and suggestions. Without her support and careful guidance, this thesis would not have been possible. Special thanks go to all my professors in the Computer Science Department for their invaluable input during my graduate studies. In addition, I would like to express my heartfelt gratitude to my family who supported and encouraged me to complete my study. In addition, I would like to thanks Dr. Mona Jassim Mohammed (The First Center in Mosul) and Dr. Qaed Jaafar (Tikrit Teaching Hospital), for their hleps. Finally, yet importantly, I am very grateful to all those who assisted and supported me to complete this study.

Abstract

Otitis media (OM) is typified by inflammation and accumulation of fluid in the region beyond the eardrum and contains the little bones that are involved in hearing. OM is a leading cause of hearing impairment globally. Accurate diagnosis of OM is a critical in medical field, but its diagnosis in primary care is hindered by limited equipment and specialist expertise, variations in clinicians' interpretations, and misdiagnoses. In the recent times, deep learning techniques offer a feasible approach for automating the classification of diseases in OM. This thesis focuses on the classification method for OM diseases based on tympanometry data. Tympanometry data consist of pressure and compliance curves representing middle ear function.

The proposed method consists of four phases: collecting a new dataset from five clinical centers for 892 patients and data preprocessing, dataset splitting, model creation and training, after that the evaluation phase. This thesis introduces a new model that combines Artificial Bee Colony algorithm with Long Short-Term Memory (ABC-LSTM) to optimize hyperparameters for classifying five classes (A, B, C, Ad, As) of OM diseases. The proposed model can deal with the temporal data because of LSTM and optimize LSTM's hyperparameter because of the ABC algorithm, which makes the proposed model very suitable to classify OM diseases using tympanometry data.

The proposed model (ABC-LSTM) demonstrated robust convergence during training and achieved an accuracy of 95.96%, precision of 96.11%, recall of 96.32%, and F-score of 96.16% on the test dataset. In addition, it significantly outperformed all other models in the conducted experiments

on a variety of metrics for the newly dataset, as well as the previously published works. These results indicate a viable categorization for OM diseases based on tympanometry data to aid classification OM, contributing to medical and computational disciplines, and show comparable agreement to subspecialist doctors in diagnosis OM in early stages.

Supervisor's Certification

I certify that thesis entitled by **"Tympanometry Based on Deep Learning Techniques for Accurate Diagnosis of Otitis"** was prepared under my supervision at the Department of Computer Science / College of Computer Science and Mathematical / University of Tikrit as a Fulfillment of the requirement for the degree of Master in Mathematics.

Signature:
Asst. Prof. Dr. Salwa Khalid Abdulateef
Date: / / 2025

Report of Director of Postgraduate Studies Committee

According to the recommendations presented by the supervisor and the linguistic evaluator of this thesis, I nominate this thesis to be forwarded to discussion.

Signature:

Date: / / 2025

Report of the Head of the Department of Computer Science

According to the recommendations presented by the supervisor and the linguistic evaluator of this thesis and the director of postgraduate studies committee, I nominate this thesis to be forwarded to discussion.

Signature:

Date: / / 2025

EXAMINATION COMMITTEE CERTIFICATION

We certify that we have read the thesis entitled " **Tympanometry Based on Deep Learning Techniques for Accurate Diagnosis of Otitis "** and examining committee examined the student " **Mohammed Mahmoud Hussein** " in its contents and what is related with it in / / 2025 and that in our opinion it is adequate with ( ) standing as a thesis for the degree of Master of science in Computer Science.

Signature:

Date: / / 2025

Member

Signature:

Date: / / 2025

Member

Signature:

Date: / / 2025

Member

Signature:

Date: / / 2025

Member and supervisor

Final Approval by the Dean of the College of Computer Science and Mathem

Date: / / 2025

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Stands for
ABC	Artificial Bee Colony
ABC-CNN	Artificial Bee Colony with Convolutional Neural Network
ABC-LSTM	Artificial Bee Colony with Long Short-Term Memory
ABC-MLP	Artificial Bee Colony with Multi-Layer Perceptron
Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AOM	Acute Otitis Media
AUC	Area Under the Curve
BGD	Batch Gradient Descent
CNN	Convolutional Neural Network
COM	Chronic Otitis Media
CSOM	Chronic Suppurative Otitis Media
CT	Computed Tomography
daPa	decaPascal
DL	Deep Learning
DSC	Dice Similarity Coefficient
ECV	Ear Canal Volume
ETD	Eustachian Tube Dysfunction
FC	Fully Connected
FN	False Negatives
FP	False Positives
FRBM	Fuzzy Restricted Boltzmann Machine
GAP	Global Average Pooling
HD	Hausdorff Distance
LSTM	Long Short-Term Memory
MEC	Middle Ear Cholesteatoma
MED	Middle ear disease
ML	Machine Learning
MLP	Multi-Layer Perceptron

OM	Otitis Media
OME	Otitis Media with Effusion
OSU	Ohio State University
PA	Peak Admittance
PP	Peak Pressure
ReLU	Rectified Linear Unit
RMSprop	Root Mean Square Propagation
RNN	Recurrent Neural Network
ROI	Region Of Interest
SGD	Stochastic Gradient Descent
SI	Swarm Intelligence
SMOTE	Synthetic Minority Over-sampling Technique
SPL	Sound Pressure Level
SVM	Support Vector Machine
tanh	Hyperbolic Tangent
TM	Tympanic Membrane
TMR	Tympanic Membrane Retraction
TN	True Negatives
TP	True Positives

Chapter One

Introduction and Research Background

1.1 Introduction

This chapter provides an overview of the research direction, specifies the problem statement, and describes the rationale for undertaking this study to attain its goals. Section (1.2) presents a brief background on the research topic, offering context and highlighting key developments. Section (1.3) defines the research problem, emphasizing the existing challenges and gaps in knowledge. Section (1.4) outlines the research objectives, detailing the specific goals this study aims to accomplish. Section (1.5) highlights the scope of this study. Finally, Section (1.6) shows the organization of the thesis. Together, these sections establish a clear foundation for the thesis.

1.2 Background of the Study

Hearing is one of our most important senses. It is fundamental to building relationships and for humans to communicate verbally with friends, families, and peers [1]. Children learn to speak through hearing sounds. Therefore, hearing impairments can interfere with a child's speech and language development and increase their risk of additional disabilities, which is any condition that increases the difficulty of participating in certain activities or effectively interacting with the world around them. Prompt detection, diagnosis, and intervention are essential for the effective treatment of hearing loss to prevent adverse impacts on sleep, psychosocial well-being, interpersonal communications, school readiness, and speech and language development in children [2].

The human ear is divided into three parts: the outer ear, the middle ear, and the inner ear, as illustrated in Figure (1.1). The eardrum (tympanic membrane - TM) is a thin, cone-shaped membrane that divides the exterior and middle ears [3]. The middle ear anatomy is made up of three little bones (malleus, incus, and stapes) that transfer sound waves to the inner ear. The middle ear also has a Eustachian canal, which connects it to the nose. This tube aids in the equalization of air pressure in the middle ear, which is required for optimal sound transmission.

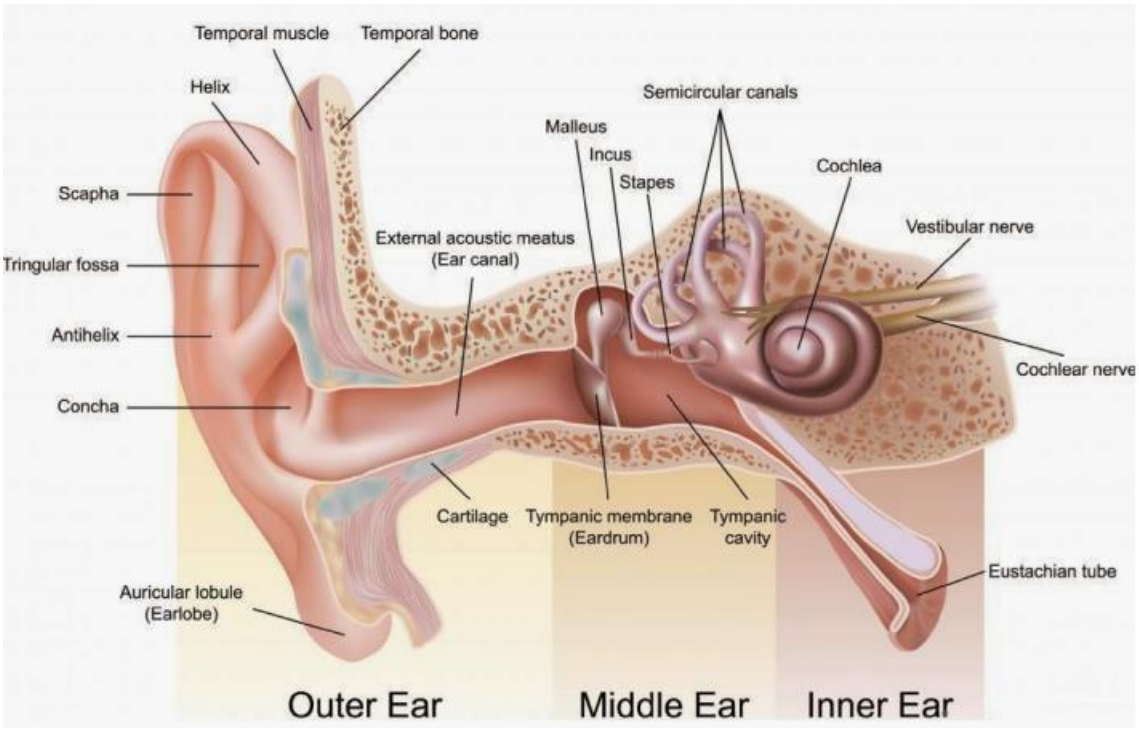


Figure 1.1: The outer ear, middle ear, and inner ear of the human ear [4].

The outer ear comprises the auricle (pinna), which collects sound waves and directs them into the ear canal for amplification. The sound waves subsequently strike the eardrum, inducing vibrations. The vibrations in the middle ear are amplified and conveyed by three small ossicles—the malleus, incus, and stapes—through the oval window to the inner ear.

Amplification in the middle ear primarily arises from two mechanisms. The primary contribution arises from the disparity in size between the area of the eardrum and the stapes footplate that connects to the oval window. The eardrum encompasses an area of roughly 55 mm^2, whereas the surface area of the stapes footplate is approximately 3.2 mm^2 [5]. External sound waves in the ear canal apply force to the larger surface area of the eardrum, subsequently concentrating this energy onto the smaller surface area of the stapes footplate, so generating significantly higher pressure at that location. The secondary amplification mechanism arises from the lever-like function of the ossicular chain.

As the malleus is longer than the incus, it traverses a greater distance; nonetheless, the incus exerts higher force, so amplifying the pressure conveyed by the stapes faceplate to the oval window. The Eustachian tube regulates air pressure equilibrium by linking the middle ear to the pharyngeal cavity [6]. In the inner ear, vibrations traverse the fluid-filled cochlea, stimulating nerve cells that transmute the vibrations into electrical messages. The impulses are subsequently transmitted to the brain for auditory processing. The inner ear houses the vestibular system, which is essential for balance maintenance.

Middle ear disease (MED) refers to a group of conditions that affect the middle ear, such as trauma or inflammation, leading to disturbances of normal middle ear function [7]. It encompasses a wide range of disease-causing conditions, the most common being otitis media (OM), arising from bacterial or viral causes [8]. Other MEDs include eustachian tube dysfunction (ETD), cholesteatoma, and fungal infections.

OM is a common childhood condition, where the Global Burden of Disease (GBD) study indicates that the incidence of Otitis Media (OM) among children rose by 40,940,535 from 1990 to 2021. This underscores the significance of addressing OM in children as a pivotal strategy for alleviating the burden of noncommunicable diseases within this demographic [9]. with a bimodal prevalence, where the first and most prominent peak occurs in children around two years old and the second peak around five years old [10]. OM can be associated with colds, as otopathogens that typically reside in the nasopharynx can ascend the eustachian tube to the middle ear, manifesting into OM.

The clinician may use an otoscope and a tympanometry test to diagnose otitis media. An otoscope is a medical equipment that commonly used to inspect the auditory canal for conditions such as cerumen impaction and acute otitis media [11]. Tympanometry is an acoustic test that evaluates eardrum vibration in response to varying air pressures within the ear canal [12], as shown in Figure (1.2). It is widely used to detect middle ear effusion with high accuracy but can be challenging to perform on children because it requires them to stay still. During the test, a tympanometer equipped with a microphone alters the air pressure while emitting a low-pitched tone—typically 226 Hz, though other frequencies may be used when needed. The device records the eardrum's movement, and the results are displayed on a tympanogram, reflecting the ear's pressure–mobility relationship. Artificial intelligence (AI) is revolutionizing healthcare by improving diagnostic accuracy, treatment planning, and patient outcomes [13]. Deep learning (DL) models, in particular, have shown significant

potential in analyzing complex medical data, enabling faster and more precise disease detection [14].

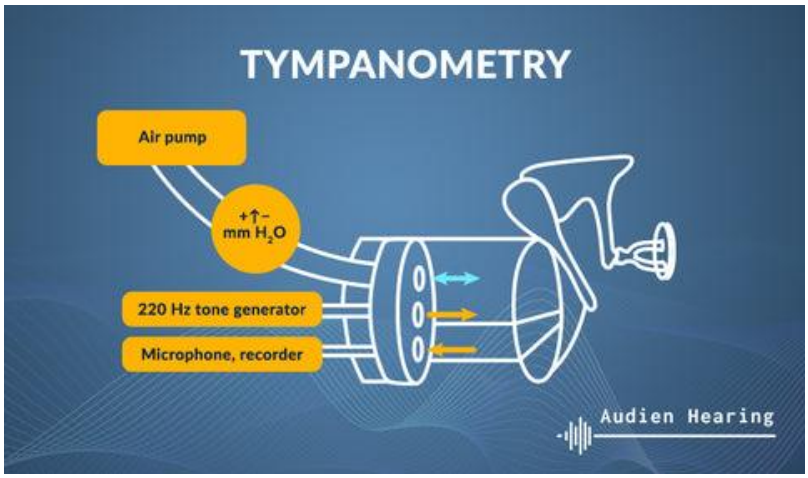


Figure 1.2: Tympanometry [15].

In the case of Otitis, DL-based classification systems can assist in distinguishing different types of the condition with high accuracy, reducing misdiagnosis and improving treatment strategies. By leveraging advanced AI techniques, healthcare professionals can make more informed decisions, leading to better patient care [16]. Therefore, this thesis introduces a method capable of classifying Otitis media problems, and to reduce human errors in diagnosis, relying on ABC-LSTM model.

1.3 Problem Statement

Otitis Media, a prevalent ear condition, requires precise identification of its different types to facilitate appropriate treatment and management. An effective diagnosis of middle-ear diseases has the potential to prevent hearing loss and antibiotic resistance. In low- and middle-income nations, at least half of otitis media cases cause hearing loss if left untreated, middle-ear diseases can lead to a variety of consequences, including balance issues, meningitis, and brain abscess [17]. Middle-ear problems are diagnosed with

tympanometry and otscopy testing. Tympanometry is the better devices for accurate diagnosis for OM [18]. However, the diagnose remains difficult for many audiologists because of the lack of experience of medical staff on how to use the device properly, this made to misdiagnosis if results are not interpreted correctly [19]. In addition, this device may be unavailable in nursing homes for the elderly [20], where AI can provide portable, smartphone-based tympanometry systems that autonomously evaluate eardrum responses, enabling precise remote assessments in nursing homes. In addition, the children cannot respond reliably to standard hearing tests due to excessive movement or fear. Therefore, accurate diagnosis of OM remains a real challenge due to the reliance on physicians' expertise and the difficulty of using current equipment. Hence, the need to develop an intelligent system capable of analyzing medical data or images to improve the accuracy and speed of diagnosis and reduce the possibility of error.

1.4 Research Objectives

The primary objective of this thesis is to classifying OM diseases based on tympanometry data to help the doctors in diagnosis. The sub-objectives of this thesis are:

- To create a new tympanometry dataset (tymp-OM) collected from various medical centers in Iraq, which contains five categories: one healthy and four diseases. The dataset have two formats: tympanograph image and electronic APX.
- To optimize the performance of an LSTM-based deep learning model for tympanometry classification using Artificial Bee Colony (ABC) algorithm for hyperparameter tuning.

- To assess the effectiveness of the proposed model in OM diagnosis using standard classification measures.

1.5 Research Scope

The research scope in this thesis revolves around several points, they are as follow:

1. This thesis classify OM diseases (one healthy and four diseases) based on Tympanometry data.
2. The dataset is private which obtained from various medical clinic in Iraq (Tikrit, Musol, and Baghdad).
3. Using LSTM + ABC Algorithm.

1.6 Outline of the Thesis

This thesis is organized, including this chapter, as outlined below:

Chapter Two: contains the main OM disease, artificial intelligence, deep learning, the techniques utilized in this thesis will be discussed in this chapter. Also, reviews existing studies on Otitis Media diagnosis using artificial intelligence methods, and identifies research gaps addressed in this thesis.

Chapter Three: details the methodology adopted in the research, including dataset collection, data preprocessing, model development, and evaluation. It explains the design and optimization of the Long Short-Term Memory (LSTM) model using the Artificial Bee Colony (ABC) algorithm.

Chapter Four: presents the experimental results obtained from training and evaluating the proposed model. It analyzes performance metrics such as

accuracy, precision, recall, and F1-score, discussing the findings in comparison with existing methods.

Chapter Five: presents the benefit and summarizes conclusion of the thesis. In addition, concludes suggestion directions for future investigations in AI-based tympanometry classification and OM diagnosis.

Chapter Two
Literature Review

2.1 Introduction

This chapter delineates the theoretical underpinnings essential for comprehending the suggested intelligent diagnostic methodology for Otitis Media (OM). It commences by elucidating the physiological and pathological principles pertaining to the middle ear, encompassing prevalent conditions that influence auditory systems. The following sections delineate tympanometry as a diagnostic method, detailing its concepts, measuring parameters, and clinical relevance. The chapter then presents the fundamental ideas of AI, Machine Learning (ML), and Deep Learning (DL), highlighting their significance in medical diagnosis and data-informed decision-making. This chapter concentrates on advanced neural architectures, including Convolutional Neural Networks (CNN), Multi-Layer Perceptrons (MLP), and Long Short-Term Memory (LSTM) models. The chapter finishes with a summary of optimization algorithms, evaluation metrics, and software tools, providing a thorough theoretical foundation for the forthcoming research methodology and experimental design.

2.2 Overview of Otitis Media

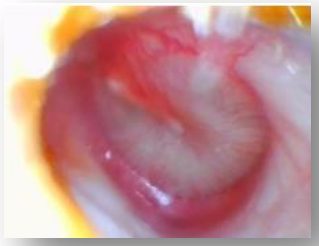
Otitis Media (OM) diseases is a type of infectious disease caused by viruses and/or bacteria in the middle ear cavity. OM is a common and potentially serious condition characterized by inflammation or infection of the middle ear [21]. It is particularly prevalent among children and represents a major cause of hearing impairment in developing countries.

The condition manifests in various forms, such as Acute Otitis Media (AOM), Otitis Media with Effusion (OME), and Chronic Otitis Media (COM), each with distinct pathological features and clinical implications [22].

- Acute otitis media (AOM) is a condition in which fluid and mucus accumulate inside the middle ear, resulting in ear pain, fever, or temporary hearing loss. It grows swiftly, resulting in swelling and redness [21].
- Otitis medium with effusion (OME): Fluid and mucus continue to accumulate after the beginning of AOM.
- Chronic otitis media (COM) is characterized by the presence of fluids in the middle ear for extended periods of time without an infection. While this condition usually not cause serious disease, it may create complications if new ear infections occur.
- Eustachian tube dysfunction: Eustachian tube dysfunction (ETD) refers to the failure of the eustachian tube to perform any of its functions. ETD usually presents symptoms of pressure (aural fullness), “popping sensations”, “underwater sensations”, crackling, ringing, muffled hearing, or own voice sounding louder (autophony).
- Perforated TM: Perforations of the TM refer to a hole or a tear in the eardrum. It can be caused by trauma or as a complication of AOM or Chronic Suppurative Otitis Media (CSOM). CSOM is identified as long-standing inflammation of the middle ear and mastoid mucosa with a perforated TM and persistent ear discharge [8].
- Tympanic membrane retraction (TMR) is a syndrome characterized by the inward displacement of a portion of the tympanic membrane

into the middle ear cavity, sometimes referred to as a retraction pocket, and is often observed in juvenile otorhinolaryngology. The incidence of TMR in children is documented to be between 8% and 10% [23]. Figure (2.1) shows middle ear conditions for visual representations [24].

Normal



Acute otitis media (AOM)

Otitis media with effusion (OME)

Perforated tympanic membrane

Retracted tympanic membrane

Figure 2.1: Middle ear conditions.

An infection of the upper respiratory tract brought on by a virus or bacteria causes inflammation of the nasopharynx and the Eustachian tube, which prevents the latter and keeps fluid in the middle ear. This increases bacterial adhesion and colonization. Additionally, Eustachian tube dysfunction results in negative middle ear pressure, which permits nasopharyngeal germs and/or viruses to enter the middle ear and cause inflammation and infection. Mistakes or delays in diagnosing otitis media may have detrimental effects such as persistent inflammation, loss of hearing, and harm to the eardrum [25]. Most serious consequences may be mastoiditis, meningitis, or even brain abscesses [26]. In addition, over-prescribed antibiotics because of misdiagnosis can lead to resistance, which is difficult to treat [27]. Therefore, prompt and precise diagnosis is critical in avoiding complications and employing appropriate strategies.

2.3 Tympanometry Technique

The Tympanometry is a non-invasive diagnostic technique and it’s a recent development that measures the mobility of the tympanic membrane and middle ear structures in response to changes in air pressure. Tympanometry provides both qualitative and quantitative data, enabling clinicians to detect abnormalities such as fluid accumulation, negative pressure, or tympanic membrane perforations. This technique uses a series of both positive and negative pressure offsets to acoustically define the ear canal. Conclusions on middle ear health and eardrum movement may then be made [28]. Figure (2.2) depicts how the determination of absorbance tympanometry works. An acoustic probe that forms an airtight seal in the ear canal is used.

The probe's microphone contains a sound-capturing device that emits sound at 226 Hz or 1 kHz at a level of around 85dB SPL. The level of sound is termed as sound pressure level (SPL). Adults generally utilize 226 Hz while 1 kHz is utilized in pediatric tympanometry, but the frequency range depends on the patient. The resulting sound pressure level in the ear canal is determined by the variations in the energy of sound that is absorbed and reflected.

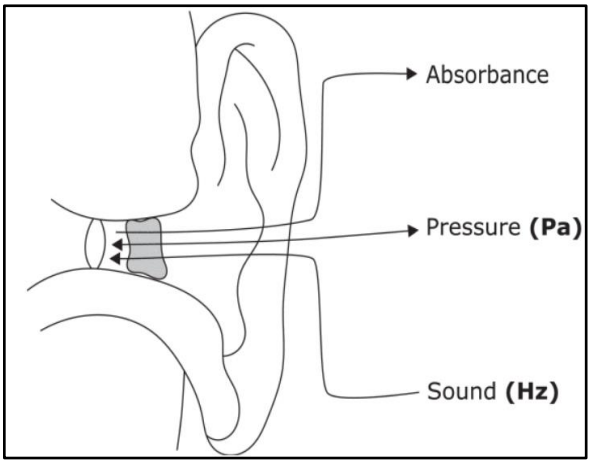


Figure 2.2: Standard absorbance tympanometry using an acoustic probe with an airtight seal in the ear canal [29].

During the measurement, the device changes the ear's pressure from +200 to -400 decaPascal (daPa). As the middle ear structures and eardrum tension fluctuate, the absorbed energy also changes with the pressure change. A tympanogram is then used to show these effects [29], which displays the admittance spectrum in contrast to the pressure showing the greatest level of diagnostic understanding. Quantitative information can be produced by tympanometry, e.g. numerical and graphical data of pressures that are both positive and negative which are generated and the amount of sound energy absorbed by the middle ear system together with ear canal volume [30]. Among these measurements:

- Peak Pressure (PP): As the pressure delivered to the ear canal declines from 200 daPa, the acoustic energy of the probe tone continues. The sound of the probe tone and the energy transmitted to the middle ear reach a max level when the pressures from both parties of the TM are balanced. Although pressures in the range of –50 to +50 daPa can be regarded as reasonable in adults, thoroughly monitored circumstances the 95% range in healthy subjects is –20 to +20 daPa. Moreover, pressures as low as –100 daPa may not be of clinical significance [31].
- Compliance: It measures the flexibility or mobility of the eardrums and middle ear system in reaction to sound or pressure [32]. It is computed from the tympanogram by analyzing the peak of the acoustic admittance curve, which represents the maximum mobility of the eardrums and middle ear system. Reduced compliance suggests stiffness or fluid in the middle ear, while increased compliance may indicate hypermobility or tympanic membrane damage.
- Ear Canal Volume (ECV): ECV can be measured and displayed on the tympanogram printout, and the normal range is age-dependent. Moreover, across all age groups, females have a lower ECV than men. The most important utilization of the ECV measurement is to make a distinction between normal and perforated TMs or between tympanostomy tubes that are obstructed and those that are working [33]. Adult ECV ranges from 0.6 to 1.5 cm3, whereas children's ECV ranges from 0.4 to 1.0 cm3.
- Tympanometric Width (Gradient): Gradient is the pressure interval including one half peak of the admittance in the plane of the TM. In

the plane of the TM, gradient has a low association with Peak Admittance (PA) and a narrow normal distribution that is independent of pump speed, making it the preferable measure, bringing complementary instead of redundant knowledge about the middle-ear transmission system [34]. A value of less than 200 daPa may be considered typical for children aged one to seven if the gradient is used as a description.

Tympanometry data are typically depicted as a tympanogram, categorized into several categories as shown in Figure (2.3) according to the patterns established by Liden and Jerger [35].

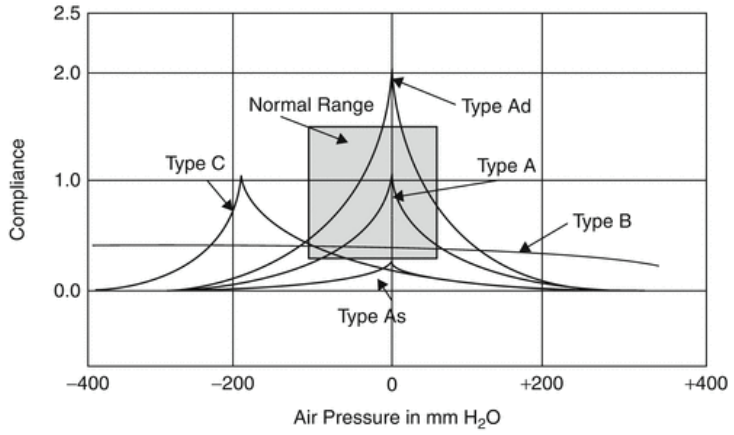


Figure 2.3: Five types of data tympanograms [36].

- Type A denotes typical middle ear function, marked by normal pressure and compliance.
- Type As (shallow) signifies reduced compliance, often associated with inflexible middle ear components, as seen in otosclerosis.
- Type Ad (deep) configuration signifies markedly increased compliance, which may be attributable to ossicular discontinuity or a relaxed tympanic membrane.

- Type B is characterized by a flat trace, usually indicating fluid in the middle ear, a perforated tympanic membrane, or impacted cerumen.
- Type C displays a peak with negative pressure, signifying Eustachian tube dysfunction, which is commonly associated with otitis media with effusion (OME).

This classification system aids in diagnosing various middle ear problems and guides appropriate treatment techniques.

2.4 Artificial Intelligence

The cognitive abilities of the human mind can be replicated by robots using the artificial intelligence (AI) branch of computer science. AI frameworks attempt to address puzzles that these algorithms are incapable of explaining in the classical sense [37]. Artificial intelligence, with all its advantages, is now transforming human society. AI has also been accepted in the educational sphere toward betterment of guiding students learning processes and other educational activities [38]. In the context of enhancing human comfort and technology advancement, researchers and AI scientists are using AI tools and methodologies [39].

Intelligent machines capable of human-like learning and reasoning are the products of AI technologies. It has found successful application in several industrial domains Computer vision, speech recognition, autonomous vehicles, and picture categorization AI has set numerous milestones. AI applies a range of logic, probabilistic and economic techniques, as well as search and mathematical optimization algorithms [40]. The domain of artificial intelligence encompasses several domains, including Computer science, mathematics, psychology, linguistics,

philosophy, neuroscience, and artificial psychology [41]. Implementation of AI will facilitate development of design procedures based on data-driven security protocols.

Determining methods for enabling a machine to recognize and formulate medical abstractions and clinical concepts and solve complex healthcare problems is a key focus of AI research. Advances in processing power, the development of sophisticated algorithms, and the continuous accumulation of large-scale medical data have led to a rapid expansion of AI applications across almost every area [42]. Whether in diagnostics, personalized treatment, or administrative tasks, Artificial Intelligence (AI) is increasingly playing a transformative role. In the context of medicine, AI can be broadly defined as the ability of machines to simulate human clinical reasoning, learning from data to support or even enhance decision-making in patient care and medical research, where the common attribute of AI is the ability to simulate reasoning processes of human beings in learning and resolving tasks or challenges [43]. AI research covers a range of topics including Machine learning (ML), Deep learning (DL), and other related fields.

2.5 Machine Learning

Recognition of patterns in data and making decisions with minimal human help is possible through the utilization of machine learning, a subdivision of artificial intelligence. Based on their learning strategies and approaches to problem-solving, machine learning algorithms can be roughly divided into three main categories: reinforcement learning, unsupervised learning, and supervised learning. The combination of the objectives of

these forms of data-driven machine learning gives rise to almost all applications of machine learning [44].

2.6 Neural Networks

Neural networks are a category of algorithms engineered to identify patterns by mimicking the functioning of the human brain. They comprise interconnected layers of nodes (or neurons) that interpret input data via weighted connections, allowing the model to learn intricate functions. The neural network inspired by biological neural systems, designed for tasks such as pattern recognition, data classification, and prediction [45]. NNs exist in numerous forms and types, including shallow and deep neural networks. The terms "shallow" and "deep" refer to the quantity of layers in a neural network. Shallow neural networks possess a limited number of layers, generally only a single hidden layer, whereas deep neural networks encompass multiple hidden layers. A neural network is considered a deep neural network when it includes two or more hidden layers between the input and output layers.

2.7 Deep Learning in Medical Diagnosis

ML focuses on algorithms that learn patterns from data using manually engineered features, while DL is a subset of ML that uses multi-layer neural networks to automatically extract features and learn complex representations directly from raw data without manual intervention. DL is regarded as an advanced kind of AI that enables computers to learn and make predictions without being explicitly programmed. It allows computers to automatically extract, analyze, and comprehend meaningful information from raw data [46]. The DL architecture outperforms classical approaches

in current scenarios, involving complicated challenges like computer vision and human language understanding. DL can tackle complicated issues using multilayer structures, making the problem-solving process faster and the outcomes more accurate. Multilayer is a subsampling technique used in the DL architecture. This makes DL extremely effective in tackling complicated issues [47]. DL is a technology developed from Artificial Neural Networks (ANN) and is considered one of the most significant advancements in intelligent operation. Well-known DL techniques include Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNN or ConvNet), and recurrent neural networks (long short-term memory).

2.7.1 Multi-Layer Perceptron (MLP)

In deep learning, a multilayer perceptron (MLP) is a name for a feedforward neural network consisting of fully connected neurons with nonlinear activation functions, organized in layers, notable for being able to distinguish data that is not linearly separable.

This primary architecture of the multi-layered neural networks will be used for carrying out the necessary DL tasks. MLP extracts simple features in the first layer and feeds those simple features into the subsequent layers to extract more complex concepts.

DL models have been shown to learn useful representations of raw data and perform well in dealing with complex engineering problems related to healthcare management [48].

Figure (2.4) presents the multi-layered feed-forward neural network (MLP-NN), which consists of many layers. Layer 1 includes an input layer that matches the feature space and is followed by multiple nonlinearity

layers. The final layer includes an output layer that matches the output space [49].

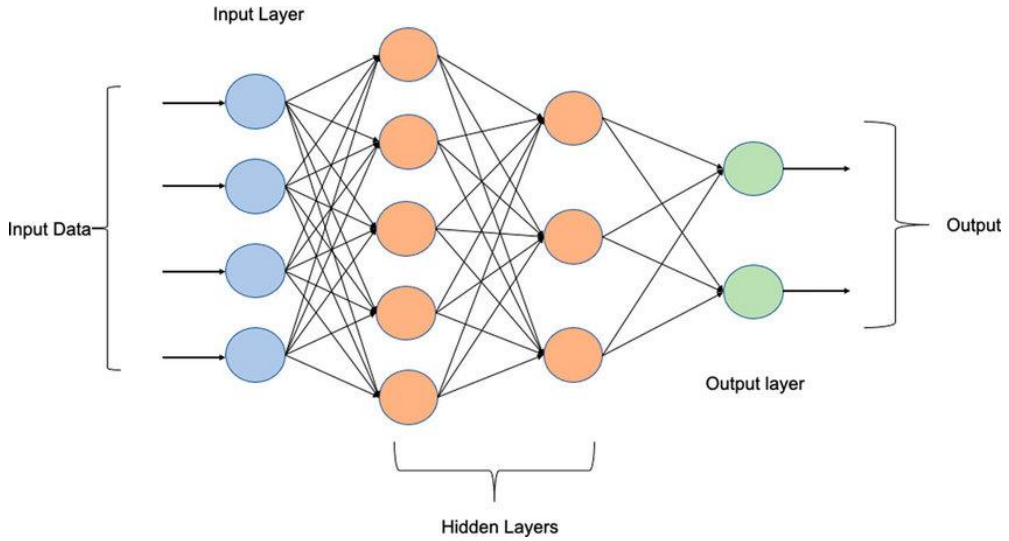


Figure 2.4: Multi-layer perceptron (MLP-NN) [49].

2.7.2 Convolutional Neural Networks (CNN)

In the field of DL, the CNN is the most famous and commonly employed algorithm. The main benefit of CNN compared to its predecessors is that it automatically identifies the relevant features without any human supervision [50]. CNNs have been extensively applied in a range of different fields, including computer vision [51], speech processing [52], Face Recognition [53], etc.

The structure of CNNs was inspired by neurons in human and animal brains, similar to a conventional neural network. More specifically, in a cat's brain, a complex sequence of cells forms the visual cortex; this sequence is simulated by the CNN. Goodfellow et al. [54] identified three key benefits of the CNN: equivalent representations, sparse interactions, and parameter sharing.

Unlike conventional fully connected (FC) networks, shared weights and local connections in the CNN are employed to make full use of 2D input-data structures like image signals.

This operation utilizes an extremely small number of parameters, which both simplifies the training process and speeds up the network. This is the same as in the visual cortex cells. Notably, only small regions of a scene are sensed by these cells rather than the whole scene (i.e., these cells spatially extract the local correlation available in the input, like local filters over the input). The benefit of convolution is that it efficiently detects and learns spatial patterns such as edges, textures, and shapes in data by using small filters that scan across the input.

A commonly used type of CNN, which is similar to the multi-layer perceptron (MLP), consists of numerous convolution layers preceding sub-sampling (pooling) layers, while the ending layers are FC layers. An example of CNN architecture for image classification is illustrated in Figure (2.5).

The CNN architecture processes images through multiple computational stages. First, the input images are fed into the network, where convolutional layers extract spatial features such as edges, textures, and color variations using learnable filters. Each convolution is followed by a max-pooling layer, which reduces the spatial dimensions and retains the most significant features, improving computational efficiency and robustness to small shifts in the image.

After feature extraction, the resulting feature maps are flattened and passed into fully connected layers, which combine the extracted features to

form high-level representations. Dropout is applied to prevent overfitting by randomly deactivating some neurons during training. Finally, an output layer classify probabilities, identifying the input image as one of the classes.

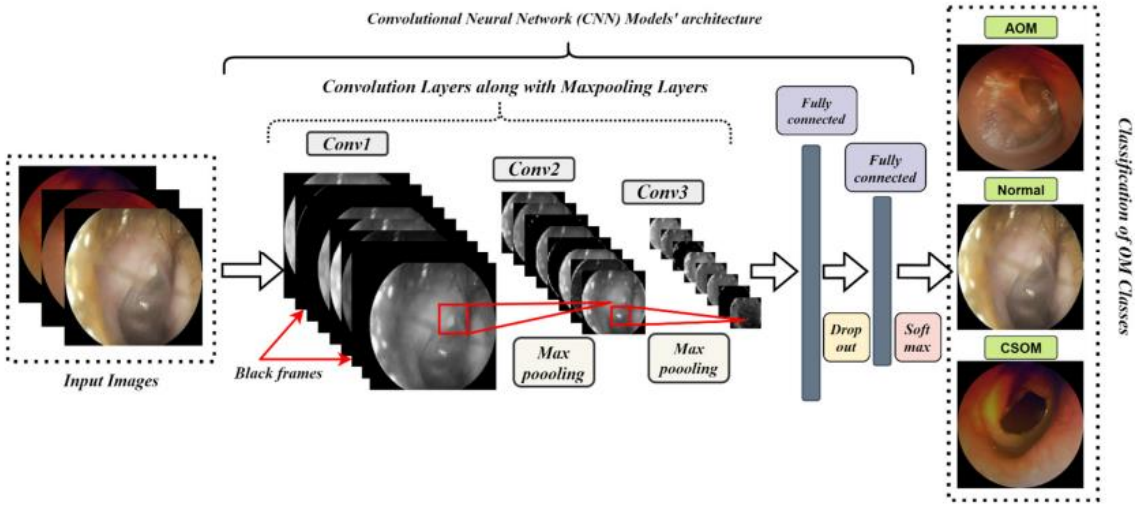


Figure 2.5: An example of CNN architecture for image classification [55].

2.7.2.1 CNN Layers

Convolutional Neural Networks (CNNs) are composed of multiple specialized layers that work together to automatically extract hierarchical features from input data. Each layer type serves a distinct purpose, contributing to the network's ability to learn complex patterns efficiently. The foundational building block is the convolutional layer, which applies learnable filters to local regions of the input, detecting spatial patterns such as edges, textures, or shapes through parameter-sharing, significantly reducing computational complexity. Following this, pooling layers (e.g., max or average pooling) down sample feature maps by summarizing local regions, enhancing translational invariance and reducing spatial dimensions while retaining critical information. Non-linear activation layers, such as ReLU (Rectified Linear Unit), introduce sparsity and non-linearity to the model, enabling it to capture intricate relationships in the data. Deeper

architectures often incorporate normalization layers (e.g., batch normalization) to stabilize training by standardizing inputs to subsequent layers, mitigating internal covariate shift. For high-level tasks like classification, fully connected layers aggregate learned features into global representations, mapping them to output classes [56].

The convolutional layer is the cornerstone of CNNs, responsible for detecting local patterns (e.g., edges, textures) through learned filters (kernels) [57].

- Operation:
  - A kernel (e.g., 3×3, 5×5) slides over the input image or feature map, computing the dot product between the kernel weights and the corresponding input region. For example, consider a simple (3 x 3) kernel applied to a (5 x 5) grayscale image. The kernel slides over the image and, at each position, multiplies its values element-wise with the corresponding (3 x 3) region of the image. The resulting nine products are then summed to produce a single output value, representing one pixel in the output feature map. This process repeats across the entire image, allowing the convolutional layer to detect specific local features such as edges or corners.
  - Mathematically, the output feature map h^k for the k-th kernel is computed as:

$$h^k = f(W^k * x + b^k) \tag{2.1}$$

Where W^k is the kernel weight matrix, x is the input, b^k is the bias, and f is the activation function.

- Key Concepts:
    - **Stride**: The step size of the kernel sliding (e.g., stride=1 shifts the kernel one pixel at a time). Larger strides reduce spatial dimensions. Using an appropriate stride helps control the output size and computational efficiency, allowing the network to balance feature detail with processing speed.
    - **Padding**: Adding zeros around the input to preserve spatial resolution (e.g., "same" padding retains input size). Padding preserves edge information by maintaining the original spatial dimensions, ensuring that features near image borders are not lost during convolution.
    - **Sparse Connectivity**: Each neuron connects only to a local region, reducing parameters and computational cost.
    - **Weight Sharing**: Kernels reuse the same weights across the entire input, enhancing efficiency and translation invariance.
- Benefits:
    - Captures spatial hierarchies by learning low-level to high-level features.
    - Reduces parameter count compared to fully connected layers.

The pooling layer down samples feature maps to reduce spatial dimensions and computational complexity while retaining critical information [58].

- Types:
    - **Max Pooling:** Selects the maximum value in a window (e.g., 2×2), emphasizing the most salient features. Max pooling highlights strong activations that indicate key features.

- **Average Pooling:** Computes the average value in a window, smoothing features. Average pooling preserves overall contextual information by considering all pixel contributions equally.
- **Global Average Pooling (GAP):** Reduces each feature map to a single value by averaging all elements, often used before classification layers.

- Purpose:
  - Invariance to small translations and distortions.
  - Reduces overfitting by lowering parameter count.

Activation functions introduce non-linearity, enabling the network to model complex relationships [59]. Activation functions determine how a neuron's input is transformed into output, enabling the network to capture relationships in the data. Without activation functions, a neural network would behave like a linear model, limiting it to only linear mappings regardless of depth. Non-linear activation functions, on the other hand, allow the network to learn complex patterns and interactions, making it capable of modeling intricate, real-world relationships. Common functions include:

- ReLU (Rectified Linear Unit):

$$f(x) = max(0, x) \tag{2.2}$$

Where:

$f(x)$: The output of the ReLU function.

x: The input value to the function

$max(0, x)$: The function that returns the maximum value between 0 and x.

- Advantages: Computationally efficient, mitigates vanishing gradient issues.
- Variants:
  - **Leaky ReLU**: Allows small negative values to prevent "dead neurons":

$$f(x) = \begin{cases} x, if\ x > 0 \\ mx, otherwise \end{cases} \quad (2.3)$$

Where:

$f(x)$: The output of the Leaky ReLU function.

x: The input value to the function.

m: A small, constant slope for negative inputs

- Sigmoid and Tanh:
  - Historically used but prone to vanishing gradients in deep networks.

The fully connected layer aggregates high-level features for final predictions (e.g., classification) [60].

- Structure:
  - Neurons connect to all activations from the previous layer.
  - Input is flattened into a vector (e.g., from a 3D feature map).
  - Outputs class probabilities using **softmax** (for classification) or continuous values (for regression).
- Role:

- Translates extracted features into interpretable outputs (e.g., class labels).

2.7.2.2 Loss Functions

Loss functions quantify prediction errors to guide parameter updates during training [61]:

- **Cross-Entropy Loss** (for classification):

$$H(p, y) = \sum_i y_i \log(p_i) \quad (2.4)$$

Where:

y_i: The true label for class i.

p_i: The predicted probability from the model that the input belongs to class i.

- Measures divergence between predicted probabilities pi and true labels yi.

- **Euclidean Loss** (for regression):

$$H(p, y) = \frac{1}{2N}\sum_{i=1}^{N}(p_i - y_i)^2 \quad (2.5)$$

Where:

N: The total number of data points in the batch or dataset.

y_i: The true label for class i.

p_i: The predicted probability from the model that the input belongs to class i.

- Computes mean squared error between predictions and targets.

2.7.2.3 Regularization

To prevent overfitting, CNNs employ regularization methods:

- **Dropout:** Randomly deactivates neurons during training to force redundancy in feature learning. Dropout ensures the network does not rely too heavily on any single neuron, promoting more robust feature learning.
- **Batch Normalization:** Normalizes layer outputs to zero mean and unit variance, stabilizing training and reducing dependency on initialization [62]. It improves training stability and accelerates convergence by reducing internal covariate shift and making the network less sensitive to weight initialization.

$$Normalized\ Output = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}} \quad (2.6)$$

Where:

μ: The mean of the values.

σ^2: The variance of the values.

ϵ: (epsilon): A very small constant added for numerical stability to prevent division by zero.

CNNs process data hierarchically:

1. **Early Layers**: Detect edges, corners, and textures.
2. **Middle Layers**: Capture complex patterns (e.g., shapes).
3. **Late Layers**: Recognize high-level semantic features (e.g., object parts).
4. **FC Layers**: Synthesize features for task-specific predictions.

Together, these layers enable CNNs to excel in tasks ranging from image recognition to medical diagnosis.

2.7.2.4 Optimizers

Optimizers play a critical role in minimizing the loss function by iteratively updating network parameters (e.g., weights, biases) through gradient-based learning. The learning rate, a hyper-parameter defining the step size for parameter updates, must be carefully chosen to balance convergence speed and stability.

Below, is the outline prominent gradient-based optimization algorithms and their enhancements.

Batch Gradient Descent (BGD): BGD updates parameters once per epoch after computing the gradient over the entire training dataset. It is stable and produces consistent convergence for small datasets. However, it requires substantial computational resources and may converge slowly or to local optima for large datasets (non-convex problems) [63].

Stochastic Gradient Descent (SGD): SGD updates parameters for each training sample, making it memory-efficient and faster for large datasets. However, frequent updates introduce noisy gradients, leading to unstable convergence behavior [64].

Mini-batch Gradient Descent: This approach divides the dataset into mini-batches and updates parameters after processing each batch. It combines the stability of BGD with the efficiency of SGD, offering steady convergence, reduced memory usage, and computational practicality [65].

Adam Optimizer: Adam adapts learning rates for each parameter by combining momentum and RMSprop benefits [66]. It uses moving averages of gradients ($E[\delta]^t$) and squared gradients ($E[\delta^2]^t$) to scale updates:

$$w_{ijt} = w_{ijt-1} - \frac{\eta}{\sqrt{E[\delta^2]^t + \epsilon}} \cdot E[\delta]^t \quad (2.7)$$

Where:

w_{ijt}: The value of the weight parameter connecting neuron j to neuron i at the current timestep t.

w_{ijt-1}: The value of the same weight parameter at the previous timestep t-1.

η (eta): The global learning rate, a hyperparameter that controls the overall step size of the update.

$E[\delta]^t$: The bias-corrected estimate of the first moment (the mean) of the gradients at timestep t. It represents the moving average of past gradients.

$E[\delta^2]^t$: The bias-corrected estimate of the second moment (the uncentered variance) of the gradients at timestep t. It represents the moving average of the squares of past gradients.

ϵ: (epsilon): A very small constant added for numerical stability to prevent division by zero.

Adam is computationally efficient, robust to noisy gradients, and widely used for deep networks. Practical Considerations:

- Learning Rate (η): Critical for balancing convergence and stability. Too high a rate causes oscillations; too low slows training.
- Local vs. Global Minima: Momentum and adaptive methods like Adam mitigate local minima traps in non-convex optimization.
- Resource Efficiency: Mini-batch GD and Adam optimize memory and computation, making them suitable for large-scale datasets.

In summary, optimizer selection depends on dataset size, computational resources, and the problem's convexity. Mini-batch GD with enhancements like Adam often provides an effective balance of speed, stability, and accuracy in CNN training.

2.7.3 Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) are a type of neural network designed to process sequential data by retaining a recollection of previous inputs through recurrent connections. An RNN must possess a minimum of three hidden levels. The fundamental design of RNNs comprises input units, output units, and hidden units, with the hidden units executing all computations through weight adjustments to generate the outputs [67]. The RNN model features a unidirectional flow of information from the input units to the hidden units, along with a directional loop that evaluates the error of the current hidden layer against that of the preceding hidden layer, subsequently adjusting the weights between the hidden levels.

The vanishing gradient problem may occur in RNNs when gradient-based learning techniques are employed for weight updates. Weights are adjusted based on the updated ratio of the partial derivative of the error function throughout each training iteration. In certain instances, the gradient may be exceedingly minimal. These erroneous signals may either escalate or dissipate, so inhibiting the alteration of the weight's value. The disappearance of erroneous signals may lead to fluctuations in the weights. With an elusive error, the learning process either requires an excessive duration or fails entirely [68].

2.7.4 Long Short-Term Memory (LSTM)

The Long Short-Term Memory (LSTM) represent a class of deep learning models and a type of RNN specifically designed to handle sequential and time-dependent data. Unlike traditional feedforward neural networks, LSTMs incorporate cyclical connections that enable information persistence across different time steps, allowing them to model dynamic temporal behavior. This capability makes LSTMs particularly suitable for applications such as speech recognition, language modeling, and time-series analysis.

In the context of OM diseases classification, LSTMs can be applied to analyze temporal patterns in tympanometry data, where pressure and compliance readings vary over time during the measurement process. The standard RNNs face challenges such as vanishing and exploding gradients during training, limiting their ability to capture long-term dependencies. On the other hand, LSTMs designed to model sequential data by mitigating vanishing gradient issues through gated memory cells. LSTMs maintain and update an internal cell state via input, output, and forget gates, enabling them to capture long-range dependencies in time-series data [69].

In the context of tympanometric time-series, LSTMs can model dynamic pressure-compliance curves over the duration of the test, offering a powerful approach to real-time OM diagnosis. A popular choice for forecasting are Long Short-Term Memory (LSTM) cells.

Since LSTMs incorporate memory units that explicitly allow the network to learn when to "forget" previous hidden states and when to update hidden states given new information, they have been utilized

effectively for sequences or temporally based data [70]. The LSTM architecture is shown in Figure (2.6). The C_t is unit memory, X_t is the input, and P_t is the output. The F_t, i_t, and o_t are the weight matrices.

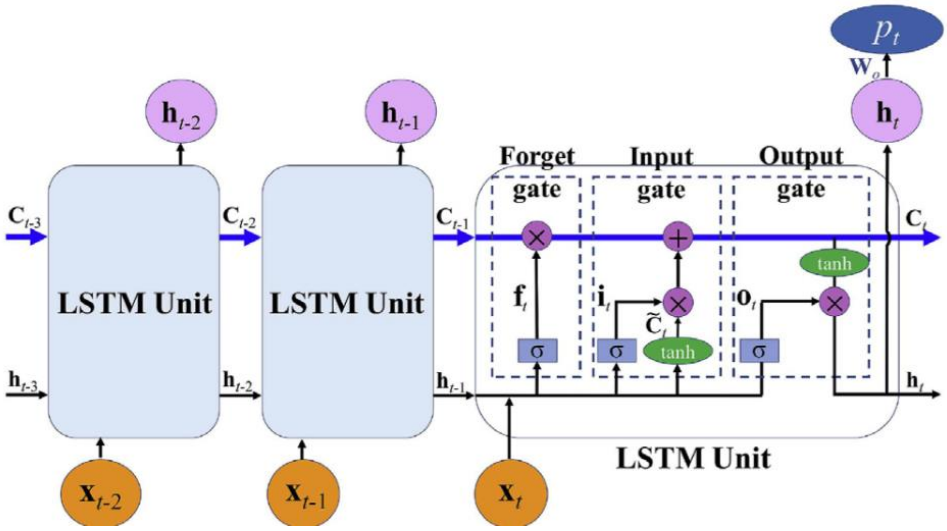


Figure 2.6: Long Short-term Memory Neural Network [70].

The LSTM unit is composed of a cell state and three regulatory gates [71]:

1. **Forget Gate**: Decides which information to discard from the cell state by outputting values between 0 (forget) and 1 (retain).

$$f^{(t)} = \sigma\left(W_f\, x^{(t)} + R_f\, y^{(t-1)} + p_f \cdot c^{(t-1)} + b_f\right) \quad (2.8)$$

Where:

The symbol (·) is the point-wise multiplication of two vectors.

$x^{(t)}$: The current input.

$y^{(t-1)}$: The output of the LSTM in the last iteration.

$c^{(t-1)}$: The cell value of the LSTM in the last iteration.

W_f ,R_f, p_f: The weights associated with $x^{(t)}$, $y^{(t-1)}$, $c^{(t-1)}$.

b_f: The bias vector.

2. **Input Gate**: Determines which new information to store in the cell state. It uses a sigmoid activation function to filter inputs and a hyperbolic tangent (tanh) function to transform values into a range of [-1, 1].

$$i^{(t)} = \sigma\left(W_i\, x^{(t)} + R_i\, y^{(t-1)} + p_i \cdot c^{(t-1)} + b_i\right) \quad (2.9)$$

Where:

The symbol (·) is the point-wise multiplication of two vectors.

$x^{(t)}$: The current input.

$y^{(t-1)}$: The output of the LSTM in the last iteration.

$c^{(t-1)}$: The cell value of the LSTM in the last iteration.

W_i ,R_i, p_i: The weights associated with $x^{(t)}$, $y^{(t-1)}$, $c^{(t-1)}$.

b_i: The bias vector.

3. **Output Gate**: Controls which parts of the cell state are exposed as the hidden state for the next time step, again using sigmoid and Tanh functions.

$$o^{(t)} = \sigma\left(W_o\, x^{(t)} + R_o\, y^{(t-1)} + p_o \cdot c^{(t-1)} + b_o\right) \quad (2.10)$$

Where:

The symbol (·) is the point-wise multiplication of two vectors.

$x^{(t)}$: The current input.

$y^{(t-1)}$: The output of the LSTM in the last iteration.

$c^{(t-1)}$: The cell value of the LSTM in the last iteration.

W_o ,R_o, p_o: The weights associated with $x^{(t)}$, $y^{(t-1)}$, $c^{(t-1)}$.

b_o: The bias vector.

For example, assume the following inputs: the current input $x^{(t)} = [0.6]$, the previous output $y^{(t-1)} = [0.4]$, the previous cell value $c^{(t-1)} = [0.5]$, and the weights $W_f = [0.8]$, $R_f = [0.4]$, $p_f = [0.2]$, $b_f = [0.1]$, $W_i = [0.7]$, $R_i = [0.3]$, $p_i = [0.1]$, $b_i = [0.05]$, $W_o = [0.5]$, $R_o = [0.4]$, $p_o = [0.2]$, $b_o = [0.1]$.

The forget gate determines which portion of the previous cell value should be retained: $f^{(t)} = \sigma\,(0.8 \times 0.6 + 0.4 \times 0.4 + 0.2 \times 0.5 + 0.1) \approx 0.73$. Thus, approximately 73% of the previous memory is preserved. The input gate regulates how much new information enters the cell value: $i^{(t)} = \sigma\,(0.7 \times 0.6 + 0.3 \times 0.4 + 0.1 \times 0.5 + 0.05) \approx 0.69$. The output gate determines the hidden state for this time step: $o^{(t)} = \sigma\,(0.5 \times 0.6 + 0.4 \times 0.4 + 0.2 \times 0.5 + 0.1) \approx 0.68$. These computed gate activations demonstrate how the LSTM selectively controls the flow of information at each time step. By balancing the retention of relevant past information through the forget gate, the integration of new input through the input gate, and the generation of the current hidden state through the output gate, the LSTM effectively maintains temporal dependencies within sequential data.

2.8 Artificial Intelligence in Healthcare

AI can enable healthcare systems to achieve their quadruple aim by democratising and standardising a future of connected and AI augmented care, precision diagnostics, precision therapeutics and, ultimately, precision medicine [72]. Research in the application of AI healthcare continues to accelerate rapidly, with potential use cases being demonstrated across the healthcare sector (both physical and mental health) including drug discovery, virtual clinical consultation, disease diagnosis, prognosis, medication management and health monitoring.

AI today (and in the near future): Currently, AI systems are not reasoning engines, which means they cannot reason the same way as human physicians, who can draw upon common sense or clinical intuition and experience [73]. Instead, AI resembles a signal translator, translating patterns from datasets. AI systems today are beginning to be adopted by healthcare organisations to automate time consuming, high volume repetitive tasks. Moreover, there is considerable progress in demonstrating the use of AI in precision diagnostics (diabetic retinopathy and radiotherapy planning).

AI in the medium term (the next 5–10 years): In the medium term, AI will be significant progress in the development of powerful algorithms that are efficient (require less data to train), able to use unlabelled data, and can combine disparate structured and unstructured data including imaging, electronic health data, multi-omic, behavioural and pharmacological data. In addition, healthcare organisations and medical practices will evolve from being adopters of AI platforms, to becoming co-innovators with technology partners in the development of novel AI systems for precision therapeutics.

AI in the long term (>10 years): In the long term, AI systems will become more intelligent, enabling AI healthcare systems achieve a state of precision medicine through AI-augmented healthcare and connected care. Healthcare will shift from the traditional one-size-fits-all form of medicine to a preventative, personalized, data-driven disease management model that achieves improved patient outcomes (improved patient and clinical experiences of care) in a more costeffective delivery system.

2.9 Swarm Intelligence (SI)

The swarm intelligence (SI) algorithm is a simulation technique designed to replicate biological collective intelligence. The inherent parallelism and distributed nature of SI algorithms facilitate the resolution of intricate nonlinear problems, exhibiting advanced attributes of self-adaptability, resilience, and search efficacy. To date, numerous optimization techniques inspired by swarm intelligence exist, including conventional particle swarm optimization (PSO) and ant colony optimization (ACO). In recent years, other advancements have emerged, including the artificial bee colony (ABC), bacterial foraging algorithm (BFO), and butterfly optimization algorithm (BOA) [74].

SI algorithms seek the optimal solution with heuristic information. It is applicable to a diverse range of optimization challenges, including dynamic optimization issues, multi-objective optimization problems, and NP problems. The continuous advancement of IoT demonstrates significant potential for SI in IoT-related applications. The SI algorithms were developed to examine how basic individuals can generate sophisticated and intricate swarm optimization behaviors via cooperation, organization, information exchange, and learning within a swarm [75].

2.10 Artificial Bee Colony (ABC)

Swarm Intelligence (SI) denotes a collective computational methodology inspired by the decentralized, self-organizing behaviors observed in natural systems. In SI, basic agents engage locally with each other and their surroundings, resulting in the formation of sophisticated, intelligent global behavior without centralized oversight. This paradigm has been extensively utilized in optimization problems, where collaboration among agents facilitates effective exploration and exploitation of the solution space. Optimization algorithms use a technique to find the best solution in a space of candidate solutions. Since its inspiration by Dervis Karaboga in 2005 [76], the artificial bee colony (ABC) algorithm has been considered a cornerstone in SI. The ABC algorithm mimics the behavior of foraging honeybees, consisting of three categories: employed bees, onlookers, and scouts.

The algorithm governs the exploration and exploitation processes, defining the search engine as globally optimal in a search landscape [77]. The Artificial Bee Colony (ABC) algorithm adopts a population-based approach for optimization as shown in Figure (2.7), the process commences with the random initialization of a population of bees solutions (employed bees). For each bee, a function called the fitness function is applied to each bee solution to get a value, which serves as the fitness value. The fitness function is a function to evaluate the solutions. The population is subsequently partitioned into two primary groups based on fitness: high-performing bees and low-performing bees. The high-performing bees are further divided using the average of fitness values of its bees for distinguishing between elite bees and moderate bees.

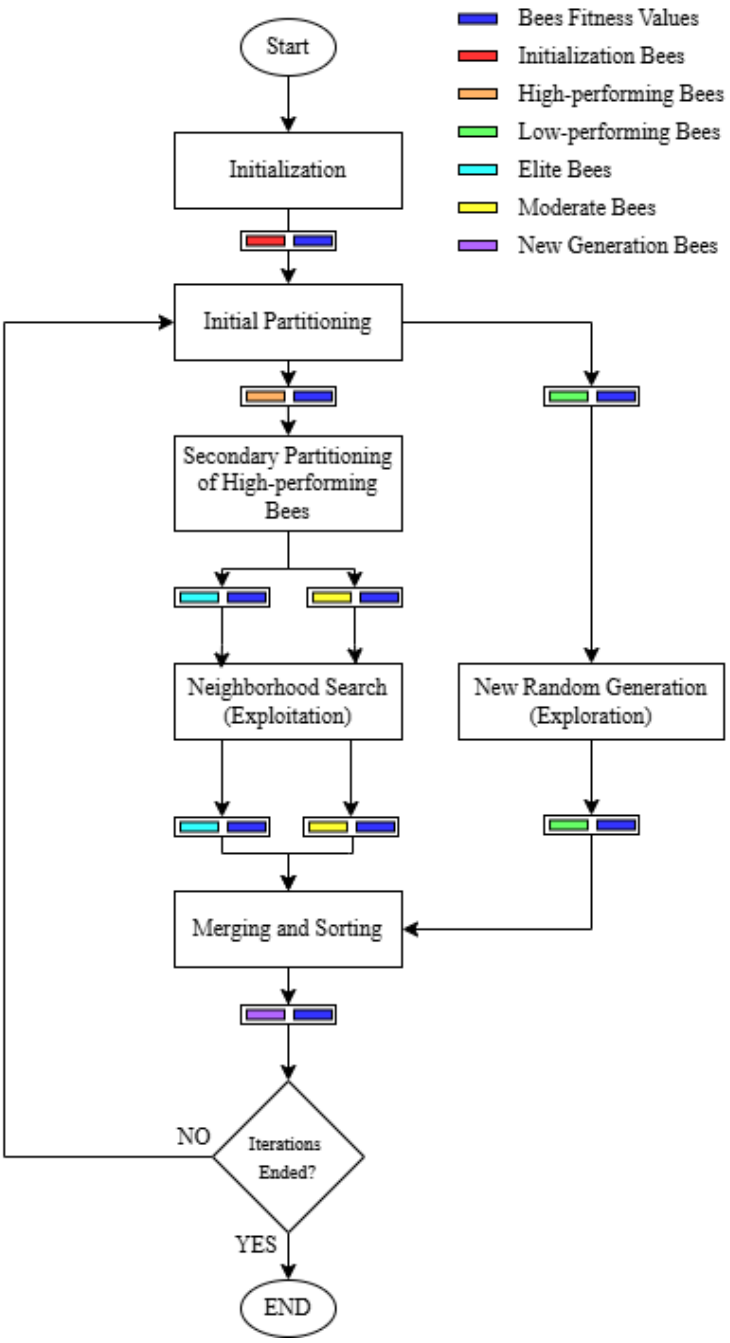


Figure 2.7: Artificial bee colony steps.

Increasing the number of colonies in the ABC algorithm enhances exploration of the search space and improves the likelihood of finding the global optimum, but it also increases computational cost and processing time. Both elite and moderate bees are subjected to a neighborhood

exploration procedure (onlooker bees), wherein new bee solutions are generated for each bee within a defined neighborhood range, thus facilitating local exploitation of promising regions in the search space.

The number of neighborhood explorations allocated to elite bees is higher than the number of neighborhood explorations allocated to moderate bees, favoring the elite bees with a more intense local search. In contrast, the low-performing bees undergo complete replacement via random generation of new bee solutions (scout bees) to encourage global exploration and maintain population diversity.

The newly generated bees, moderate bees, and elite bees are merged and then sorted according to their fitness values. This iterative process of partitioning, neighborhood exploitation, random regeneration, merging, and sorting is repeated across a predefined number of iterations.

Throughout the iterations, the algorithm continuously monitors and records the best bee solution discovered thus far, ensuring that when the iterations are ended, the best bee solution represents the optimal solution.

2.11 Evaluation Performance

To assess a predictive model's performance, it is essential to employ rigorous metrics that comprehensively analyze its effectiveness. This section outlines the primary metrics utilized to evaluate the model's classification performance: accuracy, precision, recall, and F1-score [78, 79]. The employed evaluation metrics comprise:

1. Accuracy measures the overall correctness of predictions as demonstrated by equation 2.11.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (2.11)$$

2. Precision indicates how often the model correctly identifies OME when predicted as demonstrated by equation 2.12.

$$Precision = \frac{TP}{TP+FP} \qquad (2.12)$$

3. Recall evaluates the model's ability to detect actual OME cases, reducing false negatives as demonstrated by equation 2.13.

$$Recall = \frac{TP}{TP+FN} \qquad (2.13)$$

4. F1-score balances precision and recall, providing a comprehensive measure of classification performance as demonstrated by equation 2.14.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \qquad (2.14)$$

Where:

TP = True Positives

TN = True Negatives

FP = False Positives

FN = False Negatives

In a binary classification problem, the confusion matrix is a 2×2 table that summarizes the model's performance by comparing predicted and actual outcomes, containing four key elements: true positives, true negatives, false positives, and false negatives [80]. It provides a clear understanding of how well the model distinguishes between the two classes, allowing for the computation of key metrics such as accuracy, precision,

recall, and F1-score. When extended to multi-class classification, the confusion matrix becomes a $n \times n$ table, where (n) represents the number of classes. Each row corresponds to the actual class, and each column represents the predicted class, enabling a detailed analysis of misclassifications across multiple categories [81]. This multi-class confusion matrix helps identify which specific classes are frequently confused, offering deeper insights into model weaknesses and guiding improvements in training and feature extraction.

2.12 Software Tools

The programming language utilized to apply and test the various deep learning models is Python, which is considered an elegant and high-level programming language. Its popularity within scientific computing and machine learning stems from its readability and expansive ecosystem. Python's features can be supplemented with a number of libraries which simplify the handling of data, model creation, and the evaluation of models.

For numerical operations and data handling, Python libraries NumPy and Pandas are used and data visualization is performed using Matplotlib and Seaborn. Model construction and training is performed in TensorFlow and Keras, which are sophisticated tools for training deep neural networks.

The code was run in Google Colaboratory (Colab), a cloud-based interactive platform that provides a mounted drive for files, GPU-enabled Python environment and real-time collaboration with other users. Colab integrates seamlessly with Google Drive which streamlines access to datasets as well as collaborative code writing.

2.13 Literature Review

AI encompasses algorithmic frameworks that mimic human cognition to automate complex tasks. In OM, AI-driven analysis leverages otoscopic images, tympanometry, and acoustic signals to distinguish normal and pathological states.

Choi et al. (2022) constructed a multi-class CNN using EfficientNet-B4 that classifies primary middle-ear disorders (OME, COM, normal) and secondary findings (attic cholesteatoma, meningitis, tube insertion, otomycosis) with an accuracy of 95.32% for primary classes [82]. Sundgaard et al. (2022) developed a CNN for detecting otitis media based on wide-band tympanometry using 1,014 measurements, obtaining 92.6% accuracy, explaining model decisions with saliency maps [83].

Sundgaard et al. (2021) conducted a comparative study on five loss functions the used for OM classification, determining that deep metric (triplet) loss yielded the highest precision for AOM detection while maintaining an adequate recall [84]. In order to differentiate CSOM, cholesteatoma, and normal anatomy, Wang et al. (2022) [85] developed two deep learning networks (CNN with VGG16), a classification model, and an Region Of Interest (ROI) localization model, on 973 CT-scanned ears. In examining CNN-based OM screening, Sandström et al. (2022) utilized digitized otoscopic images that had been expertly labeled, dividing the 347 images into three screening categories [86].

Mehedi et al. (2025) implemented Fuzzy Restricted Boltzmann Machine (FRBM) for the classification of infections involving the eardrum and ear canal [87]. Crowson et al. (2023) performed an evaluation of a

pediatric middle-ear deep learning algorithm against clinician practitioners for myringotomy/tube-indication diagnostic comparisons [88]. Cha et al. (2019) developed an ensemble model consisting of Inception-V3 and ResNet101 to classify six categories of ear diseases [89]. Başaran, Cömert, and Çelik (2020) developed a Faster R-CNN based detector for the tympanic membranes and evaluated it against a set of noise conditions using different pre-trained models [90].

Wu et al. (2021) aimed for home-based screening of AOM, OME, and normal ears employing Xception and MobileNet-V2 with transfer learning on 12,203 pediatric otoscopic images, supplemented by 102 images obtained via smartphone [91].

While some works fuse ML and DL methods, others combine multiple modalities of a single diagnostic type, such as imaging. The combination of the random forest analysis of tympanometry data and otoscopy image Inception-ResNet-v2 predictions gave accuracy of 84.9% when they applied the majority voting technique (Binol et al., 2020) [92]. Çalışkan (2022) [93] combined VGG16 with Support Vector Machine (SVM) to classify tympanometry images into normal and abnormal classes using a dataset of 956 images, having an accuracy of 82.17%.

Akyol et al. (2025) [94] constructed an ensemble with soft voting from several pretrained CNNs on otoscopy images, noting an accuracy of 98.8%, sensitivity of 97.5%, and specificity of 99.1%. Lee et al. (2025) [95] noted that with EfficientNet-B7 backbones, the fulfilling of classification and regression tasks for TM diseases and hearing loss in children surpassed 93.59% accuracy.

Myburgh et al. (2016) [96] trained a decision tree algorithm on high quality preprocessed images of eardrums taken with digital video of otoscopes, which then uses predetermined indicators to categorize images that are not diagnosed into five OM groups. Images taken on-site with a cheap, custom-made video-otoscope had an accuracy of 78.7%. The dataset includes otoscopy images different cases of tympanic membrane where the authors removed 73 images due to insufficient image quality.

Ting et al. (2023) [97] developed and verified OME detection with the aid of in-ear microphones and a machine learning model. Two commercial microphones were inserted into each ear canal to record the sound produced by participants as they continuously uttered five three-vowel vowels. Table (2.1) presents a comparison of literature using AI algorithms for the diagnosis of otitis media.

Table 2.1: Comparison of literature in diagnosing otitis media.

Ref. / Year	**Diagnose Method**	**Dataset**	**Size of Dataset**	**Types of Categories**	**Models**	**Main Results**
[82] 2022	Otoscopy Images	Images from otologic clinic in Asan Medical Center (Private)	1,630	OME, COM, None	EfficientNet -B4	Accuracy 95.32%
[83] 2022	Tympano metry	Images from Kamide ENT clinic, Japan (Private)	1014	OME, AOM	CNN	Accuracy 92.6% Recall 92.2% F1-Score 92.6%

[84] 2021	Otoscopy Images	Images from Kamide ENT clinic, Japan (Private)	1336	AOM, OME, and No Effusion	Deep neural network learning	Accuracy 85%
[85] 2022	CT images	Images from Xiangya Hospital (Private)	973	Middle Ear Cholesteatoma (MEC), CSOM, Normal	CNN with VGG-16	F1-score: 87.2% Precision: 90.1% Recall: 85.4%
[86] 2022	Otoscopy Images	Images from New York, USA (Public)	347	Normal, Pathological, Wax	CNN	Accuracy: 90%
[87] 2025	Otoscopy Images	dataset of ear images from Saudi Arabia (Private)	200	Infected vs. not infected	FRBM	Accuracy: 98.65%
[88] 2023	Otoscopy Images	Images from Massachusetts General Brigham (Private)	639	Normal, OME, AOM	Neural network	Accuracy 80.8%
[89] 2019	Otoscopy Images	Images from Severance Hospital (Private)	10,544	Tympanic perforation, Attic retraction, Otitis externa ±	Inception-V3 and ResNet101	accuracy 93.67%

				myringitis, Tumor, Normal		
[90] 2020	Otoscopy Images	Images from Hospital in Turkey between 10/2018 and 1/2019. (Public)	282	Normal and Abnormal	R-CNN, AlexNet, VGGNets, GoogLeNet, and ResNets	high Accuracy: 90.48% for VGG-16
[91] 2021	Otoscopy images	Images from India (Private)	12,203	AOM, OME, Normal	Xception MobileNet-V2	Accuracy: 95.72%
[92] 2020	Tympano metry and Otoscopy videos	Dataset from Ohio State University (OSU) and Nationwide Children's Hospital (Private)	73	Normal vs. Abnormal	Random forest , Inception-ResNet-v2	Accuracy: 84.9%
[93] 2022	Tympano metry	Images from Hospital in Turkey (Private)	956	Normal vs. Abnormal	VGG16 + SVM	Accuracy 82.17%. sensitivity 71.43%, specificity 90.62% f-score 77.92%

[94] 2025	Otoscopy Images	Images from the Clinical Hospital of Universidad de Chile (Public)	880	Normal, Earwax plug, Myringosclerosis, COM	Deep learning-based ensemble method	Accuracy: 98.8%, Sensitivity 97.5%, Specificity 99.1%
[95] 2025	Otoscopy Images	Images from Soonchunhyang University Hospital (Public)	757	normal, COM, AOM, and otitis externa	EfficientNet B7 model using MLP and drop connect	Accuracy: 93.59% sensitivity 87.19%, specificity 95.73%.
[96] 2016	Digital video-otoscopes	Various video-otoscopes from Pretoria, South Africa (Private)	562	CSOM, O/W, TM, OME, AOM	Decision Tree	Accuracy: 78.7%
[97] 2023	microphones with vowel sound recordings	Japan at Taipei Veterans General Hospital (Nov 2020 – Aug 2021) (Private)	62	OME, Normal	SVM, Naive Bayes, AdaBoost, Random Forest, and CNN	Accuracy for CNN: 80.65%

2.14 Research Gaps

A review of the existing literature reveals that most studies in Otitis Media diagnostic have focused on visual or acoustic data, such as studies (Esteva et al., 2021; Wu et al., 2021). While models that combine image and

tympanometry-derived features further improved diagnostic performance (Binol et al., 2020; Akyol et al., 2025), but these methods still rely on processed or transformed data rather than raw device outputs. Although there have been studies that have analyzed tympanometry data using derived features or images, raw output from devices has not been used directly in AI models. The raw outputs from tympanometry devices, such as compliance, pressure, gradient, etc., can be immediately entered into the AI model. By doing this, the AI can discern the pattern, which aids in mitigating the bias wherein the raw outputs are less complex than images or audio. To address limitations in previous studies, a new collection of tympanogram types instead of image- and audio-based approaches, the proposed method overcomes the lack of real-world datasets.

Chapter Three
Research Methodology

3.1 Introduction

In this chapter, the proposed methodology is explained which include all necessary algorithms and illustrations. The methodology starts with data collecting and concludes with assessing the results.

3.2 Research Methodology

The research methodology is shown in Figure (3.1). Initially, the data collection and preprocessing phase. The second phase data splitting divided into training, validation, and testing, class imbalance is addressed for the training set. The third phase, the model building and training phase, the Artificial Bee Colony (ABC) algorithm is utilized to determine the optimal learning rate and the optimal activation function for the LSTM model. Multiple independent colonies operate in parallel, each initialized with a unique set of candidate learning rates and activation functions.

The best-performing solution from each colony is used to train the final LSTM model. Finally, the resulted model, that trained using the optimal learning rate and the optimal activation function is subjected to a thorough evaluation using a set of pre-defined performance metrics, including accuracy, precision, recall, and F1-score. These metrics provide a multidimensional assessment of model performance, ensuring a comprehensive understanding of the classifier's predictive capability. In realm of SI, a variety of superior algorithms have been developed, including the firefly algorithm (FA), genetic algorithm (GA), differential evolution

(DE), and particle swarm optimization (PSO). These strategies demonstrate strong efficacy in addressing optimization difficulties. In comparison to other algorithms, ABC possesses a straightforward structure, fewer control parameters, and enhanced search capabilities. Consequently, it has been extensively examined by numerous experts and employed to address various complicated issues.

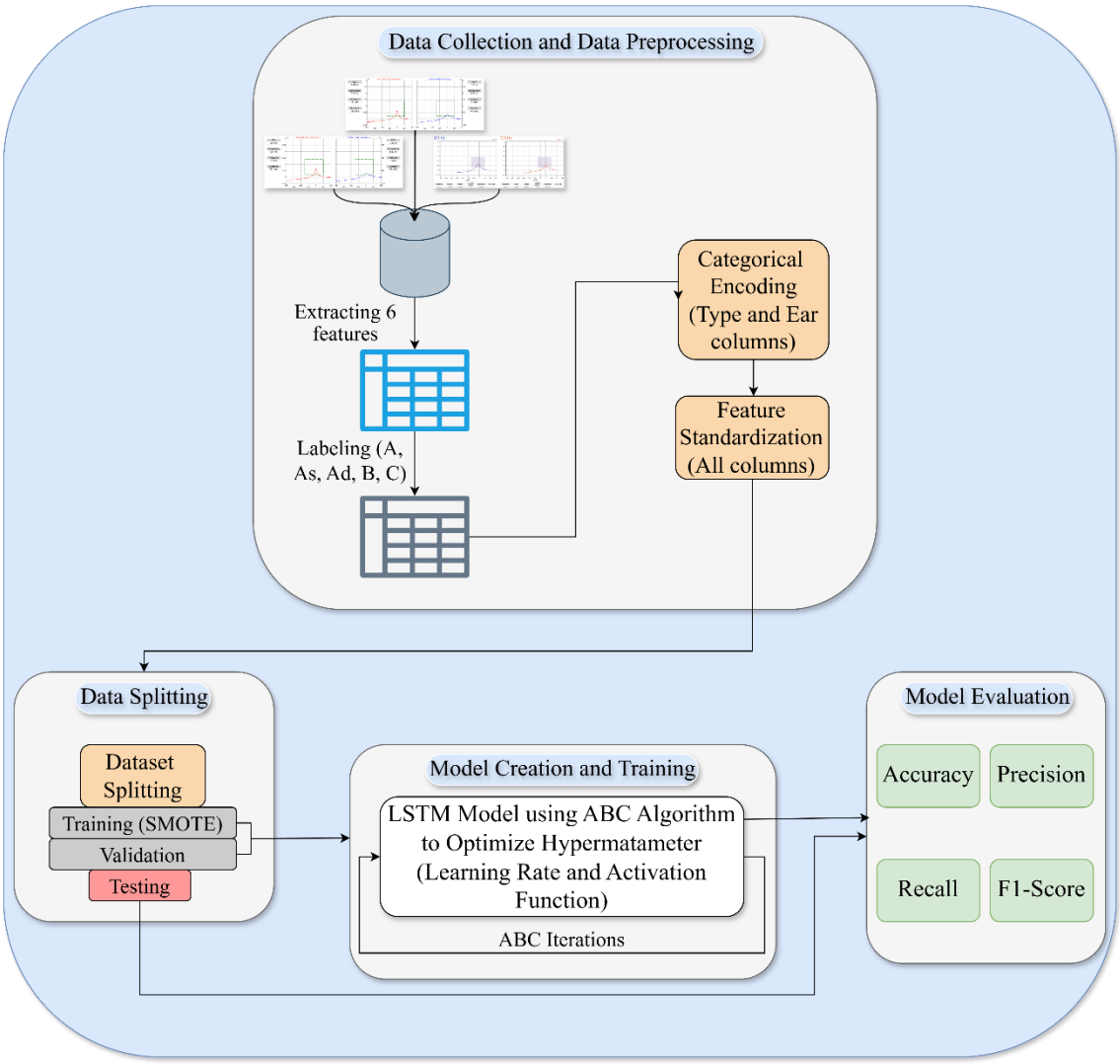


Figure 3.1: The Study Methodology.

LSTM is typically designed for sequential or temporal data, but they are also effective in modeling relationships between dependent features when those features exhibit ordered or correlated patterns — as in tympanometry data. The tympanometry data contain pressure, gradient, and compliance values that vary across a continuous range and have functional dependencies similar to time-series behavior. The following sections will detail the steps of the proposed methodology for the study.

3.3 Data Collection and Preprocessing

This section offers a detailed examination of the dataset collection, organization, visualization, and preprocessing methods employed for OM diseases classification. The initial section outlines the methodology for collecting tympanometry data from medical clinics in Iraq and provides visual interpretations of the dataset. The second portion provides preprocessing steps that are applied on the dataset.

3.3.1 Data Collection

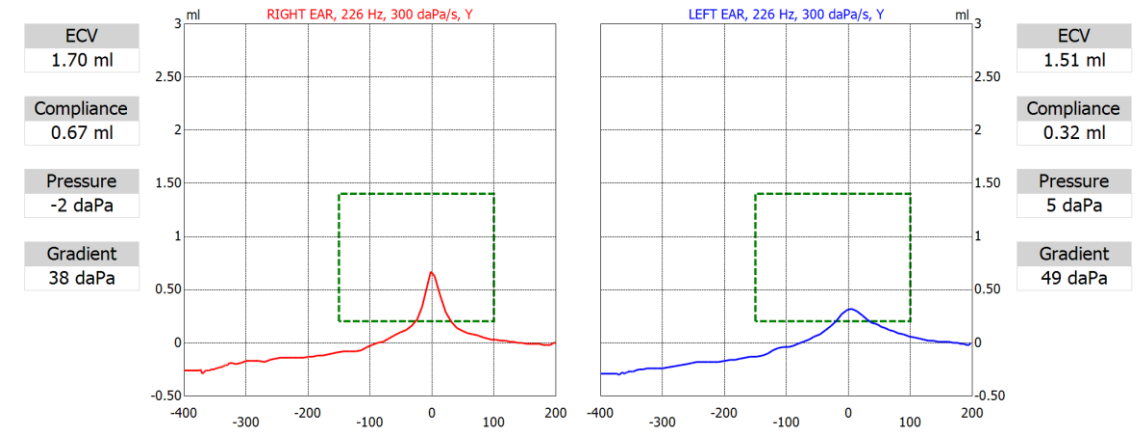
This thesis focuses on addressing the issue of categorizing OM diseases based on tympanometry. The criterion is established to determine the dataset that will be utilized in the thesis. A comprehensive database from 892 patients at various medical centers in Iraq, including hospitals and hearing service centers as shown in Table (3.1), which contains four diseases and one health condition. The dataset includes two formats tympanometry tests: graphical images and electronic APX format. There are 278 images for graphical images, where some images have more than one tympanometry test, and there are 614 tympanometry tests as electronic APX.

Table 3.1: Sources from which tympanometry tests were collected.

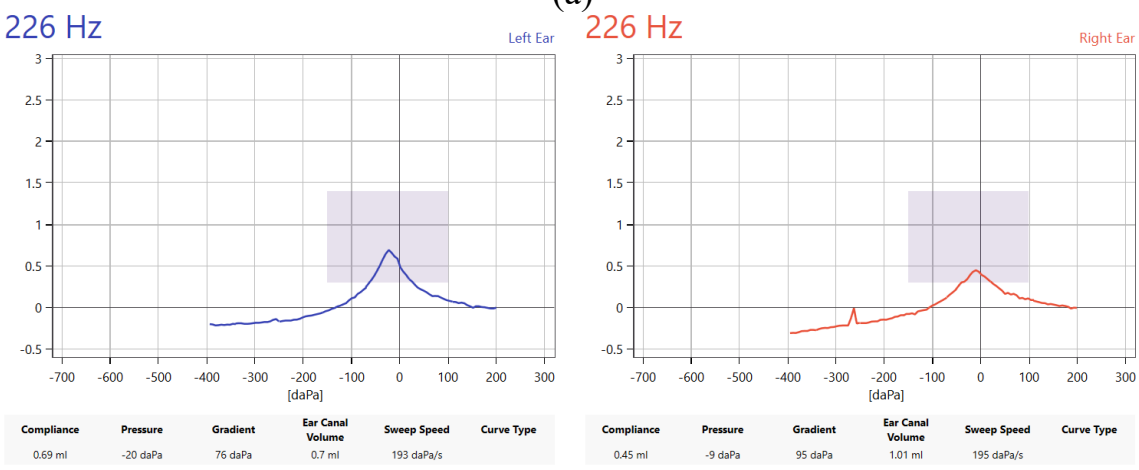
Name	Location	Image/APX
Tikrit Teaching Hospital	Salah al-Din, Iraq	46
Al-Jamhuri Hospital	Nineveh Governorate	15
Puretone Hearing Services Center	Baghdad Governorate - Al-Harithiya	614
The First Center for Hearing Services	Mosul - Al-Masarif neighborhood	150
Al Rahaf Center for Hearing Services	Salah al-Din Governorate - Tikrit	67

The graphical data were obtained directly from the medical facility in image format, while the APX format were retrieved and visualized using the Amplisuite V2.1 tool, a Windows-based desktop software intended for audiological data handling. Two types of format made up the combined dataset, called the tymp-OM dataset conains 1,808, where a sample of two formats of tympanometry tests employed in this dataset is shown in Figure (3.2). An extraction of six clinical and demographic attributes was applied from each tympanometry test: Age, Ear (left or right), Pressure (daPa), Gradient (daPa), ECV (ml), and Compliance (ml).

Each ear (left or right) in the tympanometry tests, where in the images or APX formats, was dealt with as an independent sample, because there was no statistical relationship between each ear for the same patient. Each image or APX format contains tympanometry data for the left and right ears. In simple calculation there are (278 + 614) x 2 = 1784 samples. For the image format, as mentioned lastly, there are 12 images that have two tympanometry tests for the same patient, so the total samples are (12 x 2) + 1784 = 1808 samples.



(a)



(b)

Figure 3.2: The two formats of tympanometry test for different patients, where (a) is the graphical image and (b) is the electronic APX.

3.3.2 Data Preprocessing

Efficient data preprocessing is a critical step in building an effective and generalizable deep learning model. In this stage, some critical preprocessing methods in the classification of the tympanometry dataset were carried out. These operations involved, as shown in Figure (3.3), labeling, transforming categorical data into numerical data, and feature standardizing.

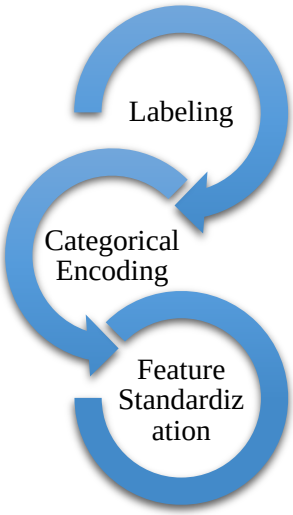


Figure 3.3: Preprocessing steps.

3.3.2.1 Labeling

The dataset was initially examined for missing values, and none were found. Each sample was reviewed and classified by an otolaryngologist (Dr. Mona Jassim Mohammed) using the Jerger system, resulting in a final dataset of 1,808 labeled records categorized into five tympanogram types, one healthy (A) and four diseases (As, Ad, B, and C), as illustrated in Figure (3.4). The Age, Ear, Pressure, Gradient, ECV, and Compliance columns are the features and the Type column is the target.

Figure (3.5) presents a horizontal bar chart depicting the distribution of the Type column in the final dataset. The primary classification of the samples was Type A (n = 1196), indicating normal middle ear function. This was followed by Type As (n = 254), signifying reduced compliance often associated with rigidity in the middle ear system. Type Ad (n = 170) indicates markedly increased compliance, often linked to ossicular chain discontinuity or tympanic membrane hypermobility.

	Age	Ear	Pressure	Gradient	ECV	Compliance	Type
0	64	R	17	97.0	1.00	0.79	A
1	64	L	-77	92.0	0.98	0.88	A
2	13	R	-4	58.0	0.00	1.00	Ad
3	13	L	13	39.0	1.00	0.00	As
4	27	R	3	78.0	0.00	1.00	A
...	...	...	...	...	...	...	...
1803	25	R	-64	120.0	0.99	0.57	A
1804	25	L	72	146.0	1.51	0.89	A
1805	7	R	-14	150.0	0.56	0.14	As
1806	7	L	-1	171.0	0.51	0.20	As
1807	31	L	-260	53.0	4.50	0.22	B

1808 rows × 7 columns

Figure 3.4: An example of the dataset.

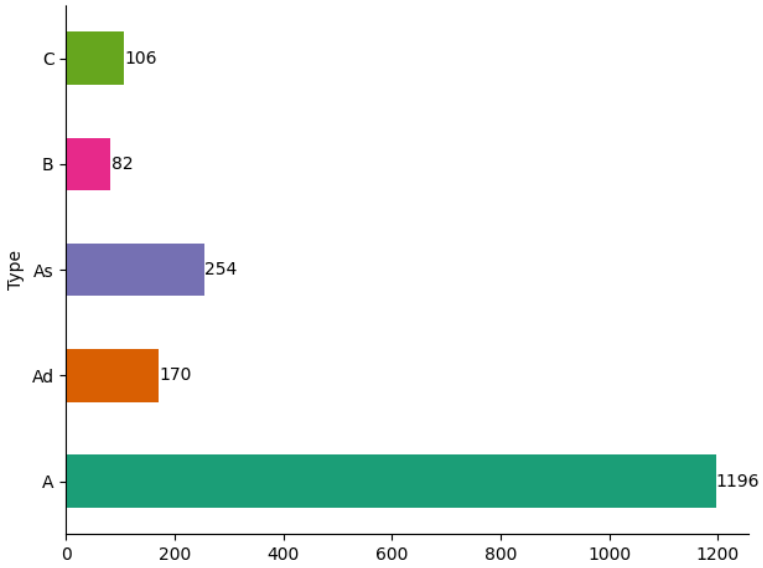


Figure 3.5: Distribution of column Type.

Type C (n = 106) signifies negative middle ear pressure, indicative of Eustachian tube dysfunction. Figure (3.6) depicts the distribution of middle ear pressure among the five tympanogram categories using a scatter plot. Every type of tympanogram has distinct pressure characteristics that correspond with clinical anticipations.

Type A tympanograms, indicative of normal middle ear function, display pressure values centered around 0 daPa, defined by a narrow distribution. Type Ad and Type As tympanograms exhibit pressures near atmospheric levels, however with slightly more variability. Conversely, Type C exhibits a notable shift towards negative pressures, frequently descending below -100 daPa, signifying Eustachian tube dysfunction.

Type B tympanograms exhibit a broad and uneven range of pressure values. This image highlights the importance of pressure measurements in tympanogram categorization and demonstrates the diagnostic heterogeneity within the dataset.

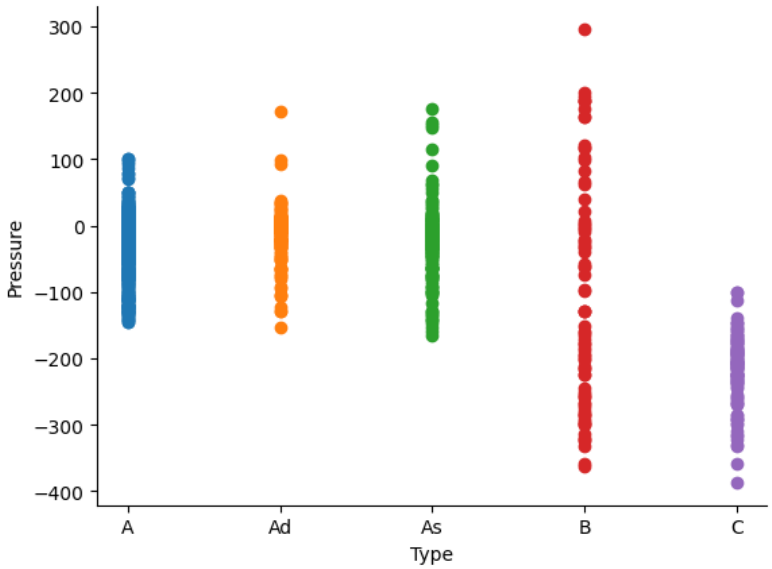


Figure 3.6: Scatter plot for pressure and Type columns.

3.3.2.2 Categorical Encoding

Initially, categorical variables were converted into numerical format to facilitate deep learning model training. The Type column was encoded as follows: A=0, Ad=1, As=2, B=3, and C=4. Similarly, the Ear column was

transformed by assigning 0 to L (left) and 1 to R (right). This encoding enabled the models to process these features effectively.

3.3.2.3 Feature Standardization

To normalize the numerical features and enhance model convergence, StandardScaler was applied. This method adjusts the features to have a mean of zero and a standard deviation of one. Standardization is particularly important for AI models, as it ensures that all input features contribute equally during model optimization. Without standardization, features with large numeric values (like pressure or gradient) would dominate model learning, while smaller-scale features (like compliance) would have little influence.

3.4 Dataset Splitting

As shown in Figure (3.7), the dataset was divided into training (64%), validation (16%), and testing (20%) subsets to support robust model evaluation. The train and validation subsets are used in the training process, where train data is responsible for training the LSTM model after the ABC algorithm generated a random learning rate and a random activation function. The validation data is used to assess the LSTM model after training by calculating the validation accuracy. The model that had the best validation accuracy found is considered the final trained LSTM model. The test data used to assess the final model by calculating the accuracy, precision, recall, and F1 score. The training dataset exhibited an issue of class imbalance. The representation of the training data was balanced via the Synthetic Minority Over-sampling Technique (SMOTE) to address this issue.

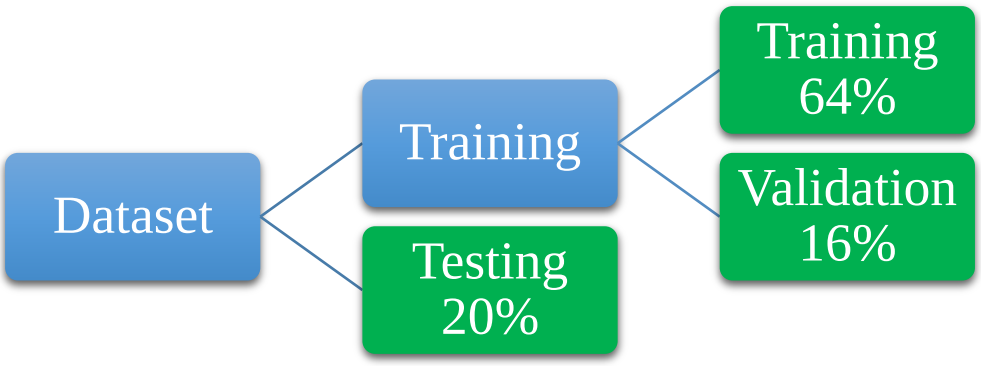


Figure 3.7: Dataset splitting phase

SMOTE works by selecting a minority sample, finding its nearest neighbors, and creating new samples along the line segments connecting them. The rationale for implementing SMOTE subsequent to standardization is to guarantee that SMOTE for all characteristics contributes uniformly in the generation of synthetic samples. As SMOTE generates new data points through interpolation of existing samples, disparate feature sizes might skew distance computations and yield implausible synthetic data. Standardizing the data initially ensures that each feature is translated to a uniform scale, enabling SMOTE to produce more balanced and significant synthetic samples that accurately represent the underlying structure of the data. The dataset splitting led to 3324 samples for training, 832 samples for validation, and 1040 samples for testing.

3.5 Model Creation and Training

This section describes the process of designing and training the classification model for OM diseases based on tympanometry data. The approach integrates a LSTM with the ABC optimization algorithm to enhance model performance.

The classification task was organized as a multiclass problem five classes, each representing one of the five types of tympanograms. After many experimental tests, the LSTM was built with these layers: the input from the tymp-OM dataset are first passed through the LSTM layer, which consists of 64 units designed to capture patterns inherent in the data. The output from the LSTM layer is then forwarded to a Dense layer with 32 units, where higher-level feature abstraction begins. This is followed by a subsequent Dense layer with 16 units that further refines the representation. Finally, the output is passed to the last Dense layer containing 5 neurons with a softmax activation function.

The model computes the sparse categorical cross-entropy loss between the predicted probabilities and the true class labels, and the backpropagation algorithm adjusts the weights throughout the network using gradients calculated via the Adam optimizer. This process is repeated iteratively for 50 epochs to minimize the loss and improve classification accuracy.

Table (3.2) and Table (3.3) show the hyperparameters of ABC and the hyperparameter values that ABC will find the best values for LSTM, respectively. In this work, the ABC algorithm is employed to optimize the hyperparameters of LSTM model (learning rate and activation function). The learning rate is an important hyperparameter in neural network training because it determines the step size in updating weights. Activation functions control the nonlinear nature of information flow within the network. These two hyperparameters were chosen to reduce computational complexity. However, if all hyperparameters were included, the search would be very large. Therefore, focusing on these two hyperparameters quickly yields

tangible results. The approach simulates the collective foraging behavior of honey bee colonies to effectively balance exploration and exploitation during the search process.

Four independent colonies are executed in parallel, each initialized with a random population of 10 candidate learning rates, 10 candidate activation functions, and with a number of iterations of 5. The fitness function used is a LSTM model where the fitness value is the validation accuracy after the LSTM model trains on the training dataset using a candidate of learning rate and activation function.

Table 3.2: The hyperparameters of ABC algorithm.

Name	**Values**
Number of Colonies Working in Parallel	4
Population of learning rates and activation functions	10 for each
ABC iterations	5
Number of Generated Bees for Elite (N2)	3
Number of Generated Bees for Moderate (N1)	2
Neighbourhood Value (N)	0.01
Middle location of bee population (M)	10
Fitness Function	Validation Accuracy

Table 3.3: The hyperparameter values that ABC finds for an LSTM.

Name	**Values**
Learning Rate	[0.0001, 0.1]
Activation Function	(relu, linear, leaky_relu, tanh)

Within each colony, the search process iteratively partitions the population based on fitness. The top-performing half of the population undergoes a secondary partitioning based on the average fitness to distinguish elite and moderate candidates. Elite candidates are subjected to a more intensive local search, generating multiple new learning rates within a defined neighborhood range and multiple new activation functions, while moderate candidates undergo a less intensive local search. Meanwhile, the lower-performing half of the population is entirely replaced by newly generated random candidates to maintain diversity and avoid premature convergence.

After each round of exploitation and exploration, all candidates are merged and sorted according to their updated fitness, and the best solution is continuously tracked across iterations.

This iterative process is repeated for a predefined number of cycles, with the parallel execution of multiple colonies further enhancing the global search capability and robustness of the optimization process. The steps of the approach, as illustrated in Figure (3.8), represent the stages of hybrid methods, which integrated LSTM with ABC. The output of these steps is a trained LSTM having the best learning rate and activation function.

1. **Initialization Phase:**
    - Randomly generate an initial population (bees) of learning rates in range of [0.0001, 0.1] and activation functions in four types (relu, linear, leaky_relu, tanh).
    - Train the LSTM model using each learning rate and activation function.

- Evaluate each model on a validation dataset and record the accuracy (fitness values).

2. **Population Partitioning:**
    - Sort all learning rate and activation function solutions based on validation accuracy.
    - Divide the population into two groups:
        - High-performing bees (top M bees, where M is the centre location of the list resulted from the initialization phase).
        - Low-performing bees (remaining bees).
3. **Secondary Partitioning of High-performing Bees:**
    - Compute the average validation accuracy of the high-performing bees.
    - Split high-performing bees into:
        - Elite bees (above-average accuracy).
        - Moderate bees (below-average accuracy)
4. **Neighborhood Search (Exploitation):**
    - Parameters used are:
    - Neighbourhood value (N): It defines the range of each learning rate solution within which new learning rate solutions are generated during the local search process. The range is in shape [LR - N, LR + N] where N = 0.01 is the Neighbourhood value and LR is the learning rate that responsible to generate new solutions.

- N2: It specifies the number of new learning rates and new activation functions to generate performed for each in elite bees during exploitation.
- N1: It specifies the number of new learning rates and new activation functions to generate performed for each in moderate bees during exploitation, where the N2 > N1
- For each bee in the elite bees, generate N2 = 3 learning rates within the neighbourhood range and N2 = 3 activation functions.
- For each bee in the moderate bees, generate N1 = 2 learning rates within the neighbourhood range and N1 = 2 activation functions.

5. **New Random Generation (Exploration):**
    - Completely replace the low-performing bees with new randomly generated bees (learning rates in range of [0.0001, 0.1] and activation functions in the types of relu, linear, leaky_relu, or tanh) to encourage exploration.
6. **Merging and Sorting:**
    - Merge elite bees, moderate bees, and newly generated bees.
    - Sort the entire population based on validation accuracy.
7. **Iteration Control:**
    - Repeat steps 2–6 for a predefined number of iterations.
8. **Best Solution Tracking:**
    - Continuously update and record the best-performing bee that get the best validation accuracy across all iterations.

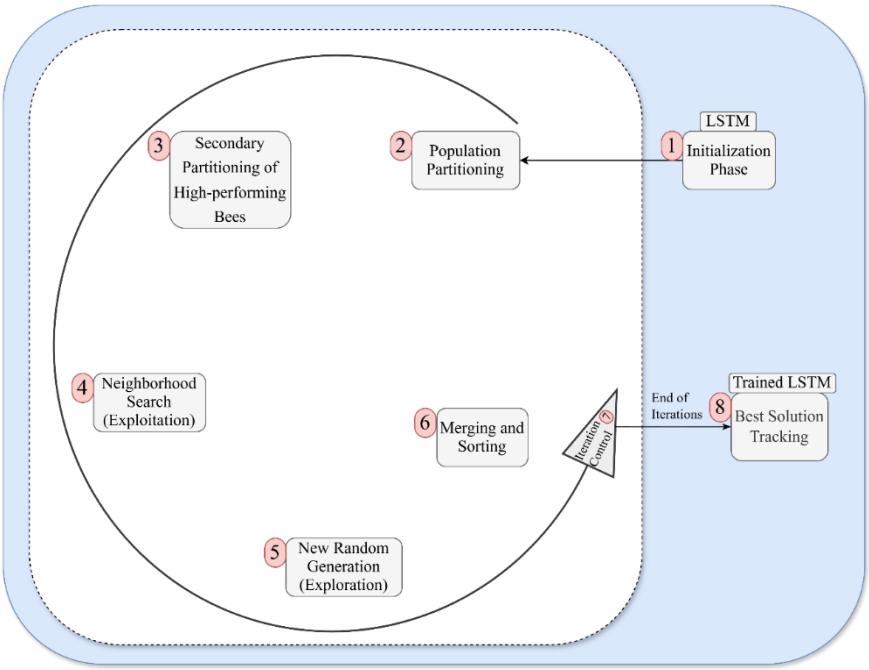


Figure 3.8: steps of combine ABC and LSTM

The advantage of using ABC with LSTM is that it is more intelligent for exploration and exploitation, yielding better results than using random values or manually tuning hyperparameters. The reason for choosing this range of values for the learning rate and activation function in the initialization phase is to explore the widest possible range of learning rate values and the most commonly used activation functions in neural networks.

3.6 Model Evaluation

After concluding up the work technique, the suggested model's performance evaluation is presented. This thesis evaluates the proposed system based on a variety of standard metrics such as accuracy, precision, recall, and F1 were covered in detail in Chapter Two.

Chapter Four

Results and Discussion

4.1 Introduction

This chapter presents the experimental and evaluation results of different deep learning models for the classification of OM diseases. At first, the thesis displays the result of the proposed model (ABC-LSTM) performance. The experimental findings for several methods, such MLP, CNN, LSTM, ABC-MLP, and ABC-CNN on (Tymp-OM) dataset are shown in section (4.3). Section (4.4) displays the comparison the proposed model (ABC-LSTM) with MLP, CNN, LSTM, ABC-MLP, and ABC-CNN. In section (4.5) the suggested method ABC-LSTM is also compare with previous works for classification OM diseases.

4.2 Results Analysis of Proposed Model

After the ABC ended its iterations, the Colony ID = 4 had the trained LSTM with the optimal learning rate and optimal activation function.

This configuration yielded a learning rate of 0.0991, with the best activation function being ReLU. The outcome of this configuration yielded a validation accuracy of 96.15%. The training process's performance is illustrated in Figure (4.1), displaying the trends of training loss and training accuracy over 50 epochs, respectively.

The initial training loss is elevated, but it experiences a significant decline in the early epochs. The decline persists gradually, ultimately reaching a stable low value. The ongoing decrease in loss demonstrates the

model's capacity to efficiently reduce error and enhance prediction accuracy as time progresses.

Simultaneously, the training accuracy demonstrates a steady rise across the epochs. The model demonstrates significant enhancement in the early stages of training and continues to show consistent advancement. The observed increase in accuracy, along with the declining loss curve, suggests robust model convergence and consistent learning stability.

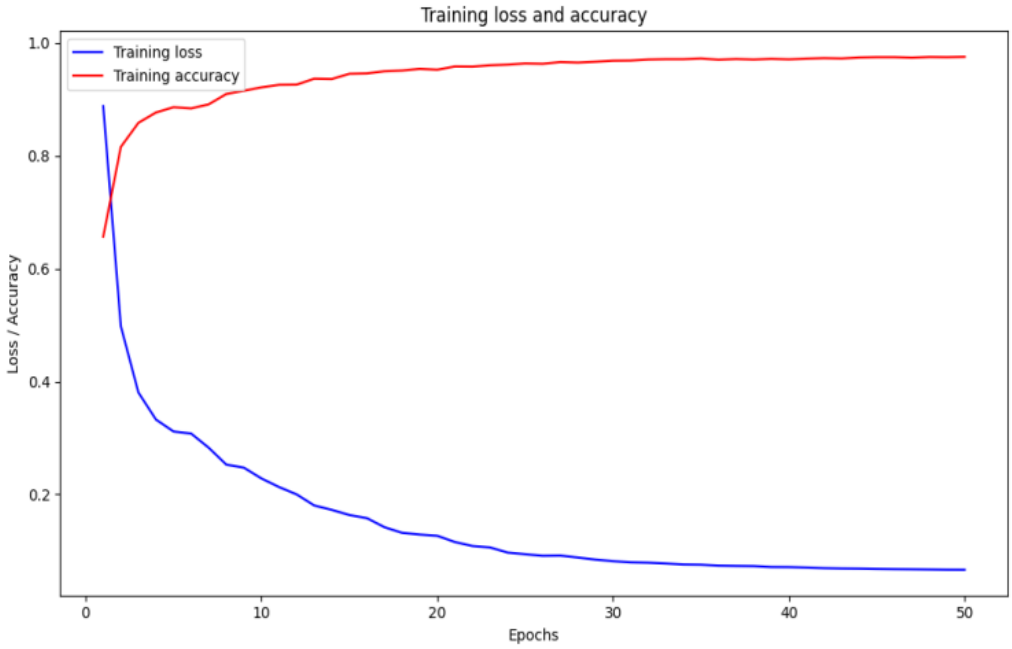


Figure 4.1: Training plot for accuracy and loss for the ABC-LSTM.

The model underwent evaluation on the independent test dataset to determine its generalization capability. The model demonstrated an impressive overall accuracy of 95.96% on the test set, reflecting robust predictive capabilities. The precision stood at 96.11%. The recall achieved was 96.32%. The F1-score reached 96.16%, indicating reliable performance across imbalanced classes. The training and testing accuracies of the ABC-LSTM model are very close. This small difference indicates that the model

generalizes well to unseen data and does not exhibit overfitting, suggesting that the ABC optimization successfully tuned the learning rate and activation function to balance model complexity and generalization. Figure (4.2) shows the confusion matrix of ABC-LSTM. The confusion matrix shows that most errors occur between class A (normal) and class As (otosclerosis).

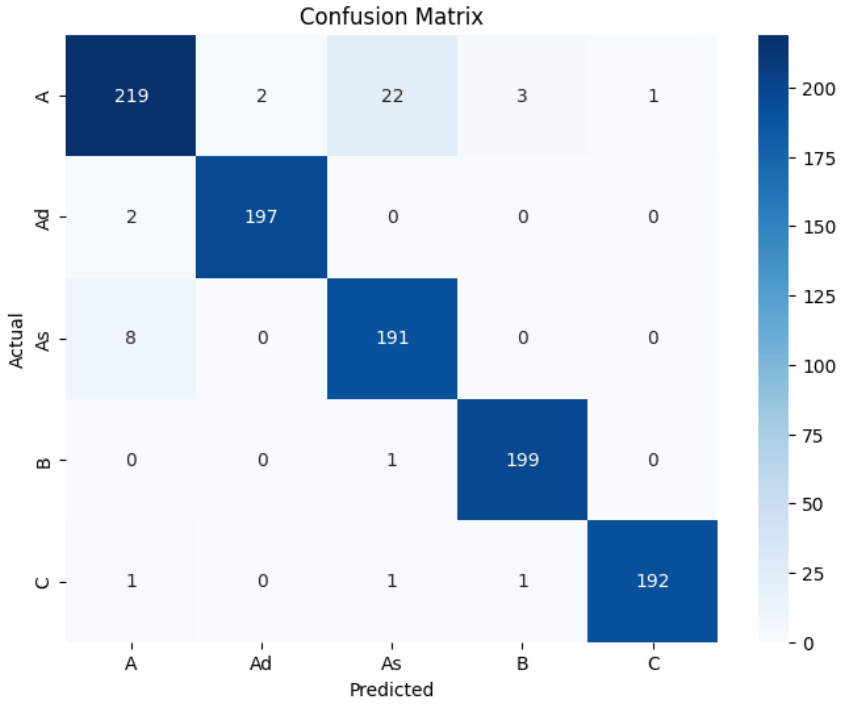


Figure 4.2: The confusion matrix of ABC-LSTM.

This is expected because these two tympanogram types have similar pressure distributions and peak shapes. In clinical terms, As is a slightly flattened version of A, so the extracted numerical features, like pressure, overlap in range. As a result, the model occasionally confuses borderline cases where the tympanogram curve exhibits mild stiffness but still retains near-normal pressure behavior. This reflects the real diagnostic challenge

even for human specialists. The proposed model (ABC-LSTM) performs well, with most predictions correctly aligned along the diagonal.

4.3 Experimental Results of Other Techniques

The section focuses on the performance of various deep learning architectures (MLP, CNN, and LSTM), and ABC optimization with MLP and CNN.

4.3.1 MLP Results

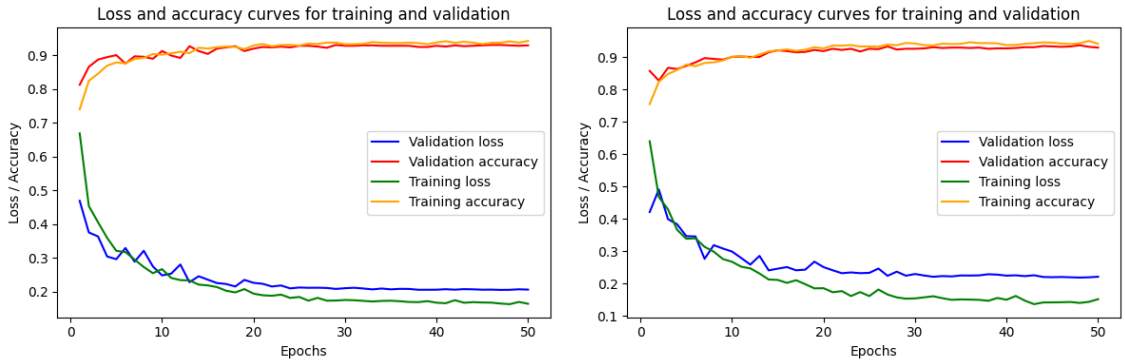
In Table (4.1) illustrates the hyperparameters used of 4-layer MLP architecture. SGD and Adam were used because they are the most widely used optimizers in neural networks. For SGD with the using of the learning rate of 0.09, a batch size of 32, and training over 50 epochs, the model achieved test accuracy of 95.77%, precision of 95.86%, recall of 95.99%, and F1-score of 95.92%. while, MLP with Adam achieved test accuracy of 95.48%, precision of 95.57%, recall of 95.72%, and F1-score of 95.63% by using a learning rate of 0.05, a batch size of 32, and training over 50 epochs.

The performance disparity between SGD and Adam in the MLP outcomes is negligible. This minor enhancement of SGD can be ascribed to its stable convergence characteristics, which may have aligned more well with the data distribution compared to Adam's adaptive updates. Nonetheless, considering the minimal disparity and the overall convergence resemblance, the difference signifies ordinary random fluctuation rather than a significant performance superiority. The confusion matrix of MLP with SGD and MLP with Adam are shown in appendix. The loss and accuracy curves for training and validation of the MLP with SGD and MLP

with Adam are illustrated in Figure (4.3). The loss decrease sharply during the initial epochs and gradually converge.

Table 4.1: Configurations of MLP.

Name	Values / Details
Loss Function	Sparse Categorical Cross-Entropy
Optimizers	SGD (momentum = 0.9), Adam
Learning Rates	0.09, 0.05
Epochs	50
Batch Sizes	32
Learning Rate Decay	Multiply by 0.9 each epoch
4-Layer MLP	1. Dense (64 units, ReLU), 2. Dropout (rate = 0.2), 3. Dense (32 units, ReLU), 4. Dense (5 units, Softmax)



(a) SGD (b) Adam

Figure 4.3: The training and validation plot for accuracy and loss for the MLP.

4.3.2 CNN Results

The Table (4.2) illustrates the hyperparameters used of 6-layer CNN architecture with training over 150 epochs. The CNN that used SGD achieved test accuracy of 76.06% by using a learning rate of 0.09, a batch size of 32, the model achieved precision of 76.74%, recall of 77.01%, and

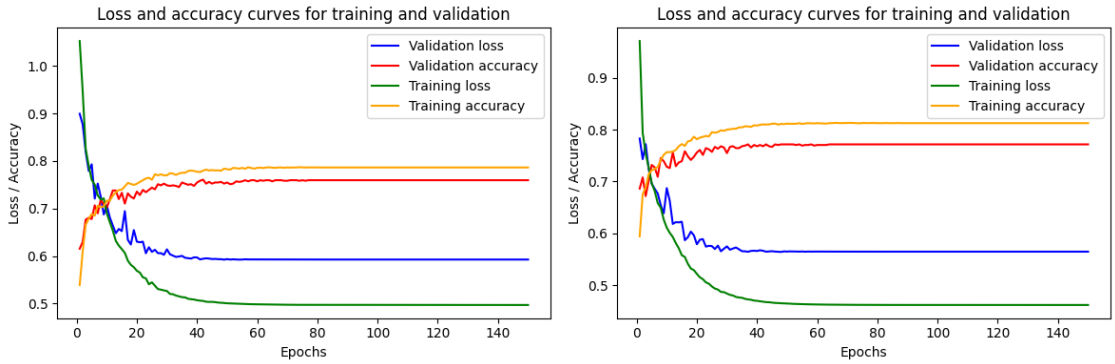
F1-score of 76.69%. The CNN that used Adam achieved test accuracy of 76.92% by using a learning rate of 0.01, a batch size of 16. The model achieved precision of 77.22%, recall of 77.89%, and F1-score of 77.36.

Table 4.2: Configurations of CNN.

Name	Values / Details
Loss Function	Sparse Categorical Cross-Entropy
Optimizers	SGD (momentum = 0.9), Adam
Learning Rates	0.09, 0.01
Epochs	150
Batch Sizes	32, 16
Learning Rate Decay	Multiply by 0.9 each epoch
6-Layer CNN	1. Conv1D (32 filters, ReLU) 2. MaxPooling1D (2 pool_size) 3. Conv1D (16 filters, ReLU) 4. Flatten() 5. Dense (64 units, ReLU) 6. Dense (5 units, Softmax)

The low CNN accuracy in SGD and Adam may be due to a learning rate, which, using this learning rate, likely caused unstable weight updates, preventing proper convergence. The explanation for Adam's better results than SGD is that Adam adapts to the characteristics of the data and gradients during training.

The confusion matrix of CNN with SGD and CNN with Adam are shown in the appendix. The loss and accuracy curves for training and validation of the CNN with SGD and CNN with Adam models are illustrated in Figure (4.4). The loss decrease sharply during the first epochs. The loss and accuracy curves stabilize around epochs 50, indicating that the models have reached their maximum learning ability.



(a) SGD (b) Adam

Figure 4.4: The training and validation plots for accuracy and loss for the CNN.

4.3.3 LSTM Results

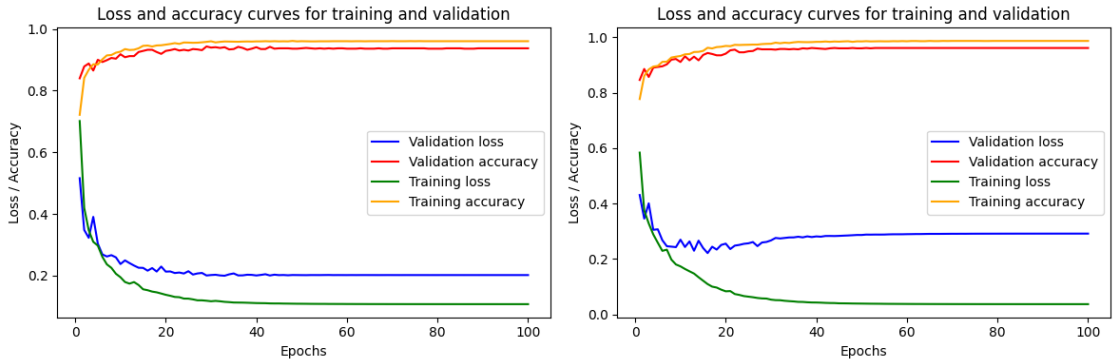
The Table (4.3) introduces the hyperparameters used of 3-layer LSTM architecture with training over 100 epochs. The LSTM that used SGD achieved test accuracy of 95.77% by using a learning rate of 0.09, a batch size of 16.The model achieved precision of 95.92%, recall of 96.16%, and F1-score of 95.97%. The LSTM that used Adam achieved test accuracy of 95.77% by a learning rate of 0.05, a batch size of 32. The model achieved precision of 95.91%, recall of 96.09%, and F1-score of 95.98%. The confusion matrix of LSTM with SGD Adam are shown in appendix. The loss and accuracy curves for training and validation of the LSTM with SGD and LSTM with Adam models are illustrated in Figure (4.5). The loss and accuracy curves over the course of 100 training epochs.

The loss decreases sharply during the initial epochs and gradually converges in LSTM with SGD, while the validation loss in LSTM with Adam has increases in the last epochs, which means there is an overfitting. The rise in validation loss post-epoch 20 using the Adam optimizer signifies the commencement of overfitting, as the model enhances performance on

training data but deteriorates in generalization to unseen data. This transpires when Adam aggressively adjusts learning rates, facilitating swift convergence while simultaneously increasing the risk of overfitting. Early stopping can be utilized to terminate training when validation loss ceases to improve. The low training loss and a high validation loss are typical relationships for overlearning where the model memorizes the training data.

Table 4.3: Configurations of LSTM.

Name	Values / Details
Loss Function	Sparse Categorical Cross-Entropy
Optimizers	SGD (momentum = 0.9), Adam
Learning Rates	0.09, 0.05
Epochs	100
Batch Sizes	16, 32
Learning Rate Decay	Multiply by 0.9 each epoch
3-Layer LSTM	1. LSTM (64 units, ReLU) 2. Dense (32 units, ReLU) 3. Dense (5 units, Softmax)



(a) SGD (b) Adam

Figure 4.5: The training and validation plots for accuracy and loss for the LSTM.

4.3.4 ABC-MLP Results

The trained MLP model, utilizing the optimal learning rate and activation function identified through the ABC optimization algorithm, demonstrated its peak performance in Colony ID = 1, achieving a learning rate of 0.03925 and employing the ReLU activation function. The configuration achieved a validation accuracy of 94.71%, showcasing the efficacy of the ABC algorithm in optimizing hyperparameters for deep learning applications.

The training process's performance is illustrated in Figure (4.6), showcasing the trends of training loss and training accuracy over 50 epochs, respectively.

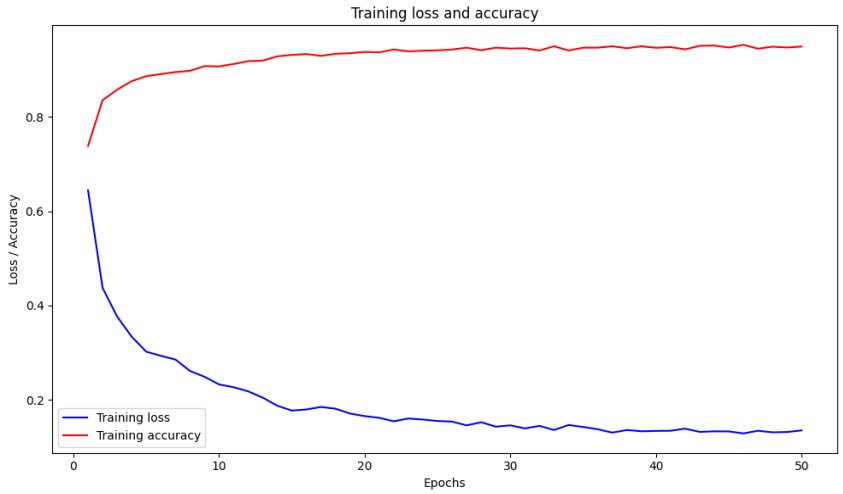


Figure 4.6: Training plots for accuracy and loss for the ABC-MLP.

The initial training loss is elevated, followed by a significant decline in the early epochs. The decline persists gradually, ultimately reaching a stable low value. The ongoing decrease in loss demonstrates the model's proficiency in minimizing error and enhancing prediction accuracy as time progresses.

Simultaneously, the training accuracy demonstrates a steady rise across the epochs. The model demonstrates significant enhancement in the early stages of training and continues to show consistent advancement. The increase in accuracy alongside the declining loss curve suggests robust model convergence and consistent learning stability.

Following the training phase, the ABC-MLP was evaluated on the independent test dataset to assess its generalization ability. The model achieved an overall accuracy of 95.48% on the test set, indicating strong predictive performance. The model achieved precision of 95.57%, recall of 95.82%, and F1-score of 95.66%. The confusion matrix of ABC-MLP is shown in appendix.

Although ABC has not shown significant superiority in terms of performance indicators, its real value lies in automating the process of selecting hyperparameters, which reduces the need for lengthy manual experimentation and saves effort and time in building models. Therefore, ABC is a complement to MLP in improving tuning, rather than a means of achieving a quantum leap in performance accuracy.

4.3.5 ABC-CNN Results

The trained CNN model using the optimal learning rate and optimal activation function found by the ABC optimization algorithm achieved its highest performance in Colony ID = 2, which produced the best learning rate of 0.01349 and best activation function is ReLU. This configuration resulted in a validation accuracy of 93.99%, demonstrating the effectiveness of the ABC algorithm in fine-tuning hyperparameters for deep learning tasks. The performance of the training process is depicted in Figure (4.7),

which illustrate the training loss and training accuracy trends over 50 epochs, respectively. The training loss starts at a high value and sharply decreases during the early epochs. It then continues to decline gradually, eventually stabilizing at a low value.

This consistent reduction in loss reflects the model's ability to effectively minimize error and improve prediction accuracy over time. In parallel, training accuracy shows a continuous increase in throughout the epochs.

The model exhibits rapid improvement during the initial training phase and maintains steady progress. This upward trend in accuracy, coupled with the decreasing loss curve, indicates strong model convergence and learning stability.

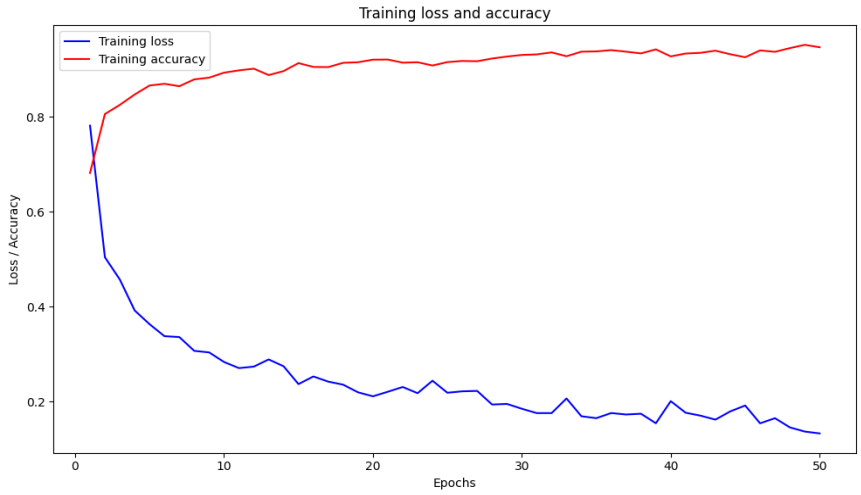


Figure 4.7: Training plots for accuracy and loss for the ABC-CNN.

Following the training phase, the ABC-CNN was evaluated on the independent test dataset to assess its generalization ability. The model achieved an overall accuracy of 94.23% on the test set, indicating strong predictive performance. The model achieved precision of 94.54%, recall of

94.79%, and F1-score of 94.46%. The confusion matrix of ABC-CNN is shown in Appendix section.

After reviewing the ABC-CNN results, the importance of ABC lies in increasing the accuracy of CNN. Without ABC, the accuracy and results of CNN showed poor results. However, ABC worked on finding the best hyperparameters, especially the learning rate, to greatly improve the results.

4.4 Comparison ABC-LSTM vs Other Techniques

Table (4.4) presents the comparison classification of the proposed model (ABC-LSTM) with MLP, CNN, LSTM—trained using two different optimization algorithms: SGD and Adam and the against the performance of ABC-MLP, ABC-CNN using different measures. The MLP model achieved its highest accuracy of 95.77% when trained with SGD, slightly outperforming the Adam optimizer, which yielded 95.48%. In contrast, the CNN model showed lower performance overall, reaching 76.06% accuracy with SGD and a marginally higher 76.92% with Adam.

Table 4.4: Comparison results between ABC-LSTM with other techniques

Model	Accuracy	Precision	Recall	F1-Score	Training Time
MLP with SGD	95.77 %	95.86%	95.99%	95.92%	25 Second
MLP with Adam	95.48 %	95.57%	95.72%	95.63%	30 Second
CNN with SGD	76.06 %	76.74%	77.01%	76.69%	92 Second
CNN with Adam	76.92 %	77.22%	77.89%	77.36%	94 Second
LSTM with	95.77 %	95.92%	96.16%	95.97%	90

SGD					Second
LSTM with Adam	95.77 %	95.91%	96.09%	95.98%	84 Second
ABC-MLP	95.48 %	95.57%	95.82%	95.66%	4418 Second
ABC-CNN	94.23 %	94.54%	94.79%	94.46%	6640 Second
Proposed model ABC-LSTM	**95.96 %**	**96.11%**	**96.32%**	**96.16%**	**5724 Second**

The LSTM model demonz2102strated strong and consistent performance, achieving 95.77% accuracy with both optimizers. The ABC-MLP and ABC-CNN models achieved an accuracy of 95.48 % and 94.23%, respectively. Meanwhile, the ABC-LSTM model reached the highest accuracy of 95.96%, outperforming all other default models, ABC-MLP, and the ABC-CNN. This is indicating the effectiveness of the ABC algorithm in tuning deep learning models. In addition, this comparison demonstrates that integrating swarm intelligence, such as ABC, into the training process can enhance model performance, particularly for recurrent architectures like LSTM.

ABC did not markedly enhance MLP performance, as MLP models exhibit reduced sensitivity to minor hyperparameter fluctuations due to their simplistic structure and absence of intricate interconnections. Consequently, ABC optimization has little capacity for additional performance improvement. Conversely, models such as CNN and LSTM encompass a greater number of hyperparameters and non-linear interactions, rendering them more amenable to ABC's optimization method, which can more effectively fine-tune learning rates and activation functions

for these intricate architectures. It can be concluded that using ABC is more time-consuming and computationally complex due to the large search operations required to obtain the best hyperparameters for each model.

4.5 Comparison with Previous Studies

Table (4.5) offers a comparative analysis for the proposed ABC-LSTM model using the Tymp-OM dataset of 5196 samples and 5 classes with previously published studies on tympanometry classification. The proposed model achieved a markedly higher accuracy of 95.96%, outperforming all prior approaches in the literature. For instance, studies such as Sundgaard et al. (2022) [83] and Binol et al. (2020) [92], which employed CNN and ensemble methods, used a dataset of 1014 images (2 classes) and 73 videos (2 classes), respectively, and reported accuracies of 92.6% and 84.9%, respectively. The VGG16 model by Çalışkan (2022) [93], which used 956 images (2 classes), reached 82.17%, still significantly below the proposed model's performance. From the results, the proposed model (ABC-LSTM) in our study achieved better results when comparing with previous studies.

Table 4.5: Comparison the proposed ABC-LSTM with previous studies.

Paper	Model	No. of classes	Dataset Size	Metrics
[83]	CNN	2	1014 images	Accuracy: 92.6% F1-Score: 92.6%
[92]	Majority voting	2	73 videos	Accuracy: 84.9%
[93]	VGG16	2	956 images	Accuracy: 82.17% F1-Score: 77.92%
This study	**ABC-LSTM**	**5**	**5196 records**	**Accuracy: 95.96%** **F1-Score: 96.16%**

Chapter Five

Conclusion and Future Works

5.1 Introduction

This chapter encapsulates the research findings of the thesis and delineates the prospective avenues for further inquiry in this domain. Section (5.2) examines the benefits of the thesis; the research findings are delineated in Section (5.3), and future work are shown in Section (5.4).

5.2 Research Benefit

This thesis offers significant contributions to both the medical and computational fields by presenting a reliable classification for OM diseases based on tympanometry data to help in the diagnosis of OM. The thesis bridges the gap between accessible healthcare and advanced diagnostic tools by enabling accurate classification of OM diseases using deep learning model.

By achieving high classification accuracy, particularly with lightweight architecture by using low number of layers for the LSTM, the methodology is well-suited for deployment in low-resource or non-specialist environments, such as rural clinics or primary care settings. Furthermore, the novel dataset and the methodology establish a foundation for future research and development in medical AI for OM.

5.3 Conclusion

This thesis set out to address the critical need for an accessible, reliable, and automated method to classify OM diseases to help in

diagnosing OM, particularly in non-specialist and resource-limited clinical settings. By leveraging deep learning models, the ABC optimization algorithm, and a novel dataset of 1,808 tympanometric readings with 5 classes from 892 patients, this thesis demonstrated the efficacy of a ABC-LSTM based classification model using tabular tympanometric data.

Preprocessing techniques were applied on the final dataset (name of data and size), comprising categorical encoding, standardization, and class balancing. The proposed methodology introduced a hybrid method of LSTM with ABC algorithm to identify the optimal learning rate and optimal activation function, addressing a key challenge in deep learning model optimization.

The proposed ABC-LSTM achieved strong convergence behavior and high accuracy (95.96%) on test data, with performance metrics such as precision, recall, and F1-score further confirming the model’s robustness and ability to generalize across all five tympanogram classes.

The results show the extent and strength of the proposed model (ABC-LSTM) with the ability to classify OM diseases. The proposed methodology exhibited robust performance; yet, some restrictions must be recognized.

The dataset, while clinically significant, is moderate in size (1,808 records), perhaps constraining the model's exposure to uncommon tympanogram patterns and marginally hindering generalization. Although preprocessing and SMOTE enhanced data balance, augmenting the dataset with a broader array of clinical cases would bolster robustness.

5.4 Future work

Future research can extend this work in several promising directions:

- Increasing the numbers of test samples will enable the evaluation of a higher number of tympanometry data with good results and good performance.
- Deploying the proposed model as a mobile application or a web browser to increase its use by ordinary users.
- Developing the proposed method to make it a powerful tool for classifying the largest possible number of categories.
- Implement the suggested model to any extra local datasets.

References

[1] C. J. Plack, The Sense of Hearing, 4th ed. London, 2023. doi: 10.4324/9781003303329.

[2] P. Homøe et al., "Panel 5: Impact of otitis media on quality of life and development," Int. J. Pediatr. Otorhinolaryngol., vol. 130, p. 109837, 2020.

[3] P. P. Nguyen, "Developing a Deep Learning Algorithm to Improve Diagnosis of Otitis Media," Ph.D. dissertation, Flinders University, College of Science and Engineering, 2021.

[4] G. Zhang et al., "Artificial intelligence-enabled innovations in cochlear implant technology: Advancing auditory prosthetics for hearing restoration," Bioeng. Transl. Med., vol. 10, no. 3, p. e10752, 2025.

[5] R. Zablotni et al., "Sound-Induced Round Window Vibration—Experiment and Numerical Simulations of Energy Transfer Through the Cochlea of the Human Ear," Appl. Sci., vol. 15, no. 1, p. 301, 2024.

[6] R. Bal and P. Deshmukh, "Management of Eustachian tube dysfunction: A review," Cureus, vol. 14, no. 11, p. e31432, 2022.

[7] K. M. Harmes et al., "Otitis media: diagnosis and treatment," Am. Fam. Physician, vol. 88, no. 7, pp. 435–440, 2013.

[8] A. G. Schilder et al., "Otitis media," Nat. Rev. Dis. Primers, vol. 2, no. 1, pp. 1–18, 2016.

[9] H. Wang et al., "Global, regional, and national epidemiology of otitis media in children from 1990 to 2021," Front. Pediatr., vol. 13, p. 1513629, 2025.

[10] S. A. A. Mehdi et al., "Diagnostic Accuracy of Tympanometry for Diagnosis of Fluid the Middle Ears of Children with Otitis Media with Effusion staking Myringotomy as Gold Standard," Pak. Armed Forces Med. J., vol. 73, no. 1, p. 21, 2023.

[11] C. Suh, D. Z. Zhao, L. Ganti, and D. Zhao, "History and Evolution of the Otoscope," Cureus, vol. 17, no. 1, 2025.

[12] O. Kose, W. R. J. Funnell, and S. J. Daniel, "Vibration measurements of the gerbil eardrum under quasi-static pressure steps," J. Assoc. Res. Otolaryngol., vol. 21, pp. 287–302, 2020.

[13] S. A. Alowais et al., "Revolutionizing healthcare: the role of artificial intelligence in clinical practice," BMC Med. Educ., vol. 23, no. 1, p. 689, 2023.

[14] X. Chen et al., "Recent advances and clinical applications of deep learning in medical image analysis," Med. Image Anal., vol. 79, p. 102444, 2022.

[15] A. Bell, "The curious type C tympanogram: contraction of the tensor tympani masquerades as negative middle ear pressure," J. Hear. Sci., vol. 11, no. 2, 2021.

[16] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: transforming the practice of medicine," Future Healthc. J., vol. 8, no. 2, pp. e188–e194, 2021.

[17] L. M. Haile et al., "Hearing loss prevalence and years lived with disability, 1990–2019: findings from the Global Burden of Disease Study 2019," The Lancet, vol. 397, no. 10278, pp. 996–1009, 2021.

[18] S. A. A. Mehdi et al., "Diagnostic Accuracy of Tympanometry for Diagnosis of Fluid the Middle Ears of Children with Otitis Media with Effusion staking Myringotomy as Gold Standard," Pak. Armed Forces Med. J., vol. 73, no. 1, p. 21, 2023.

[19] A. Zahid, J. C. Wilson, I. D. Grice, and I. R. Peak, "Otitis media: recent advances in otitis media vaccine development and model systems," Front. Microbiol., vol. 15, p. 1345027, 2025.

[20] C. Dubois et al., "Development and validation of a smartphone-based deep-learning-enabled system to detect middle-ear conditions in otoscopic images," NPJ Digit. Med., vol. 7, no. 1, p. 162, 2025.

[21] S. Y. Jung et al., "Toll-like receptors: Expression and roles in otitis media," Int. J. Mol. Sci., vol. 22, no. 15, p. 7868, 2021.

[22] T. Marom, J. Nokso-Koivisto, and T. Chonmaitree, "Viral–bacterial interactions in acute otitis media," Curr. Allergy Asthma Rep., vol. 12, no. 6, pp. 551–558, 2012.

[23] A. B. Bayoumy et al., "The natural course of tympanic membrane retractions in the posterosuperior quadrant of pars tensa: a watchful waiting policy," Otol. Neurotol., vol. 42, no. 1, pp. e50–e59, 2021.

[24] V. S. Liang, "AI in the determination of middle ear diseases and clinicians' perception," ResearchSpace@ Auckland, 2024.

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

[26] O. Feussner, R. Haase, and J. Baier, "Case report: Otitis media with subsequent mastoiditis and cerebral herniation in a patient with Arnold chiari malformation," Front. Pediatr., vol. 10, p. 1013300, 2023.

[27] A. Arason and J. A. Sigurdsson, "The problems of antibiotic overuse," Scand. J. Prim. Health Care, vol. 28, no. 2, pp. 65–66, 2010.

[28] Shahnaz, H. AlMakadma, and C. A. Sanford, "The rise and fall of aural acoustic immittance assessment tools," in Seminars in Hearing, vol. 44, no. 01, pp. 005–016, 2023.

[29] T. Hein, S. Hatzopoulos, P. Skarzynski, and M. Colella-Santos, "Wideband tympanometry," in Advances in Clinical Audiology, 2017, pp. 29–42.

[30] K. Ramakrishnan, R. A. Sparks, and W. E. Berryhill, "Diagnosis and treatment of otitis media," Am. Fam. Physician, vol. 76, no. 11, pp. 1650–1658, 2007.

[31] K. Bright, C. O. Greeley, J. Eichwald, C. O. Loveland, and G. Tanner, American Academy of Audiology childhood hearing screening guidelines. Reston, VA: American Academy of Audiology Task Force, 2011.

[32] J. Lous, C. T. Ryborg, J. J. Damsgaard, and A. P. Munck, "Tympanometry in general practice: use, problems and solutions," Fam. Pract., vol. 29, no. 6, pp. 726–732, 2012.

[33] X. Meng, K. Zhu, J. Yue, and C. Han, "The role of wideband tympanometry in the diagnosis of Meniere's disease," Front. Neurol., vol. 13, p. 808921, 2022.

[34] S. Al-Salim, R. M. Tempero, H. Johnson, and G. R. Merchant, "Audiologic profiles of children with otitis media with effusion," Ear Hear., vol. 42, no. 5, pp. 1195–1207, 2021.

[35] J. Chan et al., "Performing tympanometry using smartphones," Commun. Med., vol. 2, no. 1, p. 57, 2022.

[36] D. Othman, M. Alashkar, and M. A. Bitar, "Accuracy of Video Otoscopy in Predicting the Presence of Middle Ear Effusion in Children Compared to Tympanometry: A Diagnostic Study," Cureus, vol. 16, no. 11, 2024.

[37] J. M. Górriz et al., "Artificial intelligence within the interplay between natural and artificial computation: Advances in data science, trends and applications," Neurocomputing, vol. 410, pp. 237–270, 2020.

[38] T. Bates, C. Cobo, O. Mariño, and S. Wheeler, "Can artificial intelligence transform higher education?," Int. J. Educ. Technol. High. Educ., vol. 17, no. 1, p. 42, 2020.

[39] J. Chubb, P. Cowling, and D. Reed, "Speeding up to keep up: exploring the use of AI in the research process," AI & Soc., vol. 37, no. 4, pp. 1439–1457, 2022.

[40] M. G. Abdolrasol et al., "Artificial neural networks based optimization techniques: A review," Electronics, vol. 10, no. 21, p. 2689, 2021.

[41] K. R. Chowdhary, "Introducing artificial intelligence," in Fundamentals of Artificial Intelligence. New Delhi: Springer India, 2020, pp. 1–23.

[42] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," J. Ind. Inf. Integr., vol. 23, p. 100224, 2021.

[43] L. Markauskaite et al., "Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI?," Comput. Educ. Artif. Intell., vol. 3, p. 100056, 2022.

[44] S. Sah, "Machine Learning: A Review of Learning Types," Preprints, 2020, doi: 10.20944/preprints202007.0230.v1.

[45] S. Schmidgall et al., "Brain-inspired learning in artificial neural networks: a review," APL Mach. Learn., vol. 2, no. 2, 2024, doi: 10.1063/5.0186054.

[46] S. Setiowati, E. L. Franita, and I. Ardiyanto, "A review of optimization method in face recognition: Comparison deep learning and non-deep learning methods," in 2017 9th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE), 2017, pp. 1–6.

[47] Bianchini and F. Scarselli, "On the complexity of neural network classifiers: A comparison between shallow and deep architectures," IEEE Trans. Neural Netw. Learn. Syst., vol. 25, no. 8, pp. 1553–1565, 2014.

[48] P. Van et al., "Deep learning convolutional neural network in rainfall-runoff modelling," J. Hydroinformatics, vol. 22, no. 3, pp. 541–561, 2020.

[49] H. A. Afan et al., "Modeling the fluctuations of groundwater level by employing ensemble deep learning techniques," Eng. Appl. Comput. Fluid Mech., vol. 15, no. 1, pp. 1420–1439, 2021.

[50] J. Gu et al., "Recent advances in convolutional neural networks," Pattern Recognit., vol. 77, pp. 354–377, 2018.

[51] W. Fang, P. E. Love, H. Luo, and L. Ding, "Computer vision for behaviour-based safety in construction: A review and future directions," Adv. Eng. Inform., vol. 43, p. 100980, 2020.

[52] D. Palaz, M. Magimai-Doss, and R. Collobert, "End-to-end acoustic modeling using convolutional neural networks for HMM-based automatic speech recognition," Speech Commun., vol. 108, pp. 15–32, 2019.

[53] H. C. Li, Z. Y. Deng, and H. H. Chiang, "Lightweight and resource-constrained learning network for face recognition with performance optimization," Sensors, vol. 20, no. 21, p. 6114, 2020.

[54] S. F. Ahmed et al., "Deep learning modelling techniques: current progress, applications, advantages, and challenges," Artif. Intell. Rev., vol. 56, no. 11, pp. 13521–13617, 2023.

[55] M. A. H. Rony et al., "Artificial intelligence-driven advancements in otitis media diagnosis: a systematic review," IEEE Access, vol. 12, pp. 99282–99307, 2024.

[56] L. Alzubaidi et al., "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," J. Big Data, vol. 8, no. 1, pp. 1–74, 2021.

[57] D. X. Zhou, "Theory of deep convolutional neural networks: Downsampling," Neural Netw., vol. 124, pp. 319–327, 2020.

[58] S. Y. Jhong et al., "An automated biometric identification system using CNN-based palm vein recognition," in 2020 Int. Conf. Adv. Robot. Intell. Syst. (ARIS), 2020, pp. 1–6.

[59] A. Al-Azzawi et al., "DeepCryoPicker: fully automated deep neural network for single protein particle picking in cryo-EM," BMC Bioinformatics, vol. 21, no. 1, p. 509, 2020.

[60] T. Wang, C. Lu, M. Yang, F. Hong, and C. Liu, "A hybrid method for heartbeat classification via convolutional neural networks, multilayer perceptrons and focal loss," PeerJ Comput. Sci., vol. 6, p. e324, 2020.

[61] G. Li, M. Zhang, J. Li, F. Lv, and G. Tong, "Efficient densely connected convolutional neural networks," Pattern Recognit., vol. 109, p. 107610, 2021.

[62] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Int. Conf. Mach. Learn., 2015, pp. 448–456.

[63] S. Ruder, "An overview of gradient descent optimization algorithms," arXiv:1609.04747, 2016.

[64] J. Wang and G. Joshi, "Cooperative SGD: A unified framework for the design and analysis of local-update SGD algorithms," J. Mach. Learn. Res., vol. 22, no. 213, pp. 1–50, 2021.

[65] H. Qi, F. Wang, and H. Wang, "Statistical analysis of fixed mini-batch gradient descent estimator," J. Comput. Graph. Stat., vol. 32, no. 4, pp. 1348–1360, 2023.

[66] Z. Zhang, "Improved adam optimizer for deep neural networks," in 2018 IEEE/ACM 26th Int. Symp. Qual. Serv. (IWQoS), 2018, pp. 1–2.

[67] J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long short term memory recurrent neural network classifier for intrusion detection," in 2016 Int. Conf. Platform Technol. Service (PlatCon), 2016, pp. 1–5.

[68] P. S. Muhuri, P. Chatterjee, X. Yuan, K. Roy, and A. Esterline, "Using a long short-term memory recurrent neural network (LSTM-RNN) to classify network attacks," Information, vol. 11, no. 5, p. 243, 2020.

[69] F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall–runoff modelling using long short-term memory (LSTM) networks," Hydrol. Earth Syst. Sci., vol. 22, no. 11, pp. 6005–6022, 2018.

[70] S. M. Al-Selwi et al., "RNN-LSTM: From applications to modeling techniques and beyond—Systematic review," J. King Saud Univ.-Comput. Inf. Sci., vol. 36, no. 5, p. 102068, 2024.

[71] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," Artif. Intell. Rev., vol. 53, no. 8, pp. 5929–5955, 2020.

[72] P. Nachev, D. Herron, N. McNally, G. Rees, and B. Williams, "Redefining the research hospital," NPJ Digit. Med., vol. 2, no. 1, p. 119, 2019.

[73] T. P. Quinn, M. Senadeera, S. Jacobs, S. Coghlan, and V. Le, "Trust and medical AI: the challenges we face and the expertise needed to overcome them," J. Am. Med. Inform. Assoc., vol. 28, no. 4, pp. 890–894, 2021.

[74] W. Sun, M. Tang, L. Zhang, Z. Huo, and L. Shu, "A survey of using swarm intelligence algorithms in IoT," Sensors, vol. 20, no. 5, p. 1420, 2020.

[75] C. Kolias, G. Kambourakis, and M. Maragoudakis, "Swarm intelligence in intrusion detection: A survey," Comput. Secur., vol. 30, no. 8, pp. 625–642, 2011.

[76] S. Alsamia, E. Koch, H. Albedran, and R. Ray, "Adaptive Exploration Artificial Bee Colony for Mathematical Optimization," AI, vol. 5, no. 4, pp. 2218–2236, 2024.

[77] J. Wang, Y. Liu, S. Rao, X. Zhou, and J. Hu, "A novel self-adaptive multi-strategy artificial bee colony algorithm for coverage

optimization in wireless sensor networks," Ad Hoc Netw., vol. 150, p. 103284, 2023.

[78] S. K. Abdulateef, "Evolutionary Optimization of Geometrical Image Contour Detection," Int. J. Intell. Eng. Syst., vol. 15, no. 2, pp. 287–297, 2022.

[79] A. K. Mustfa, S. K. Abdulateef, Q. A. Hameed, M. A. Ahmed, and Z. Al-Qaysi, "Classification Flower Images Based on Deep Learning and Machine Learning," in 2025 8th Int. Symp. Multidiscip. Stud. Innov. Technol. (ISMSIT), 2025, pp. 1–5.

[80] S. Sathyanarayanan and B. R. Tantri, "Confusion matrix-based performance evaluation metrics," Afr. J. Biomed. Res., vol. 27, no. 4S, pp. 4023–4031, 2024.

[81] I. Markoulidakis, G. Kopsiaftis, I. Rallis, and I. Georgoulas, "Multi-class confusion matrix reduction method and its application on net promoter score classification problem," in Proc. 14th PErvasive Technol. Relat. Assist. Environ. Conf., 2021, pp. 412–419.

[82] Y. Choi et al., "Automated multi-class classification for prediction of tympanic membrane changes with deep learning models," PLoS One, vol. 17, no. 10, p. e0275846, 2022.

[83] J. V. Sundgaard et al., "A deep learning approach for detecting otitis media from wideband tympanometry measurements," IEEE J. Biomed. Health Inform., vol. 26, no. 7, pp. 2974–2982, 2022.

[84] J. V. Sundgaard et al., "Deep metric learning for otitis media classification," Med. Image Anal., vol. 71, p. 102034, 2021.

[85] Z. Wang et al., "Structure-aware deep learning for chronic middle ear disease," Expert Syst. Appl., vol. 194, p. 116519, 2022.

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

[87] I. M. Mehedi et al., "Artificial Intelligence with Deep Learning Based Automated Ear Infection Detection," IEEE Access, vol. 12, pp. 48335–48348, 2024.

[88] 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, 2023.

[89] D. Cha, C. Pae, S. B. Seong, J. Y. Choi, and H. J. Park, "Automated diagnosis of ear disease using ensemble deep learning with a big

otoendoscopy image database," EBioMedicine, vol. 45, pp. 606–614, 2019.

[90] E. Başaran, Z. Cömert, and Y. Çelik, "Convolutional neural network approach for automatic tympanic membrane detection and classification," Biomed. Signal Process. Control, vol. 56, p. 101734, 2020.

[91] Z. Wu et al., "Deep learning for classification of pediatric otitis media," The Laryngoscope, vol. 131, no. 7, pp. E2344–E2351, 2021.

[92] H. Binol et al., "Decision fusion on image analysis and tympanometry to detect eardrum abnormalities," in Med. Imaging 2020: Comput.-Aided Diagn., vol. 11314, pp. 375–382, 2020.

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

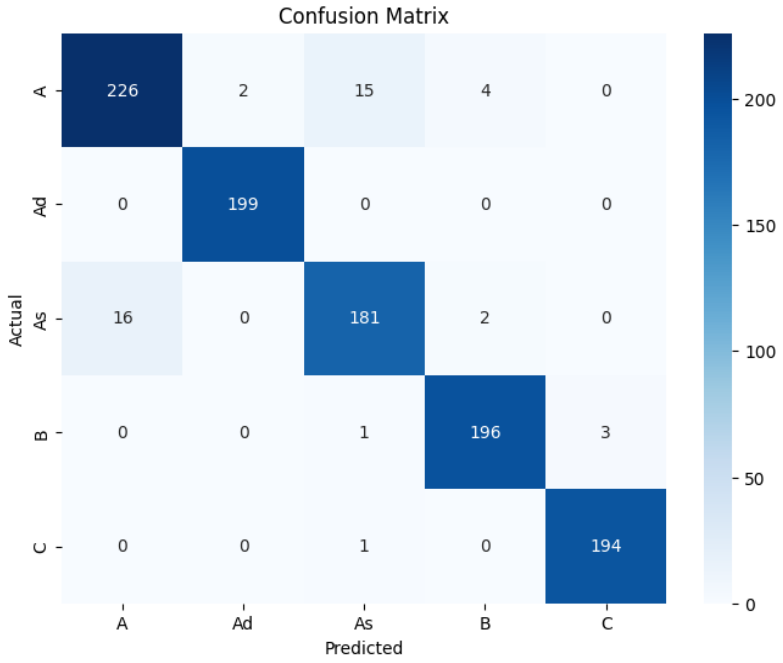
[94] K. Akyol, E. Uçar, Ü. Atila, and M. Uçar, "An ensemble approach for classification of tympanic membrane conditions using soft voting classifier," Multimed. Tools Appl., vol. 83, no. 32, pp. 77809–77830, 2024.

[95] H. Lee, H. Jang, W. Jeon, and S. Choi, "Diagnosis of Tympanic Membrane Disease and Pediatric Hearing Using Convolutional Neural Network Models with Multi-Layer Perceptrons," Appl. Sci., vol. 14, no. 13, p. 5457, 2024.

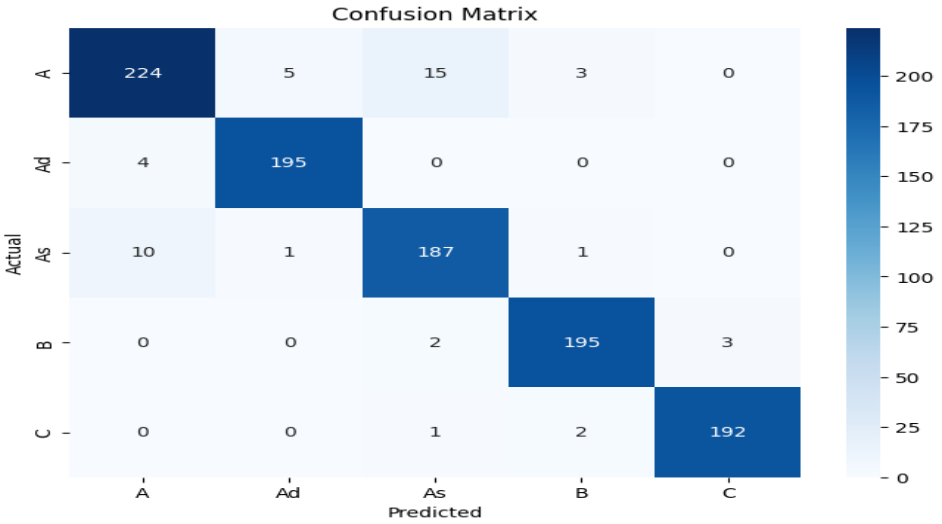
[96] H. C. Myburgh, W. H. Van Zijl, D. Swanepoel, S. Hellström, and C. Laurent, "Otitis media diagnosis for developing countries using tympanic membrane image-analysis," EBioMedicine, vol. 5, pp. 156–160, 2016.

[97] K. C. Ting et al., "Detection of Otitis Media With Effusion Using In-Ear Microphones and Machine Learning," IEEE Sens. J., vol. 23, no. 22, pp. 28411–28420, 2023.

Appendix

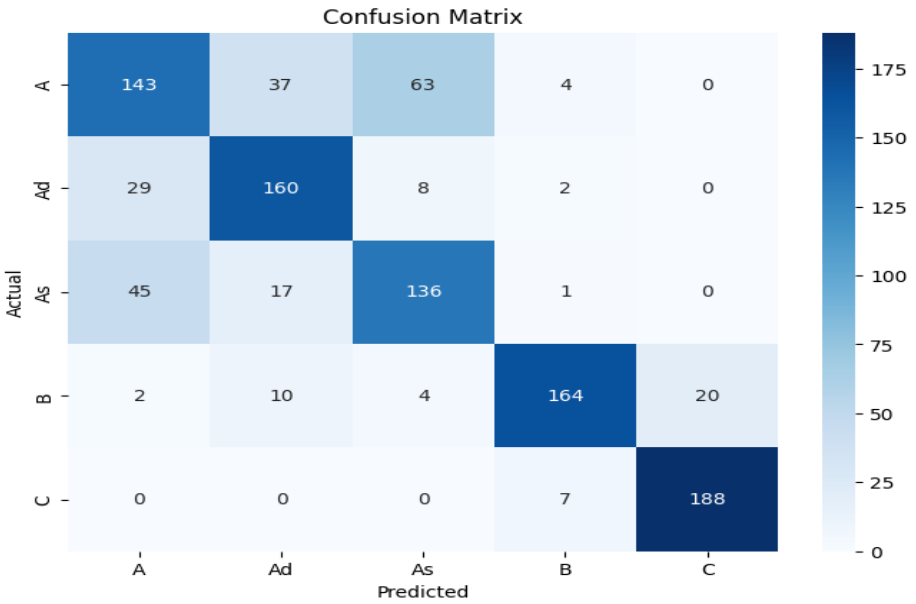


(a) SGD

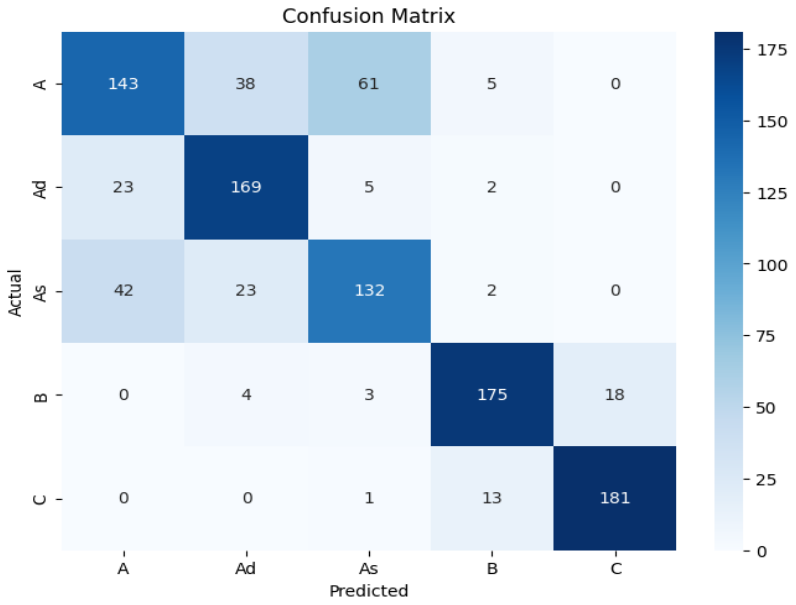


(b) Adam

The confusion matrix of MLP.

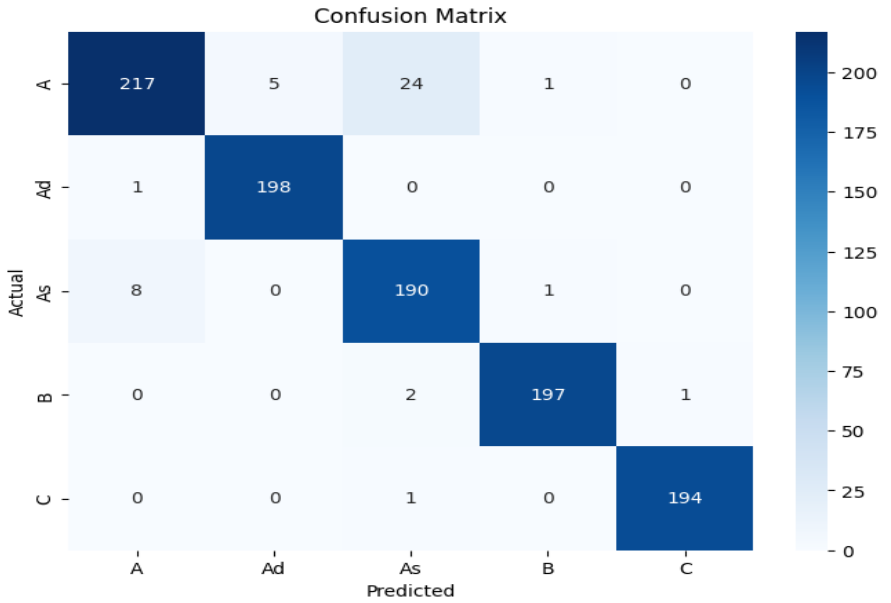


(a) SGD

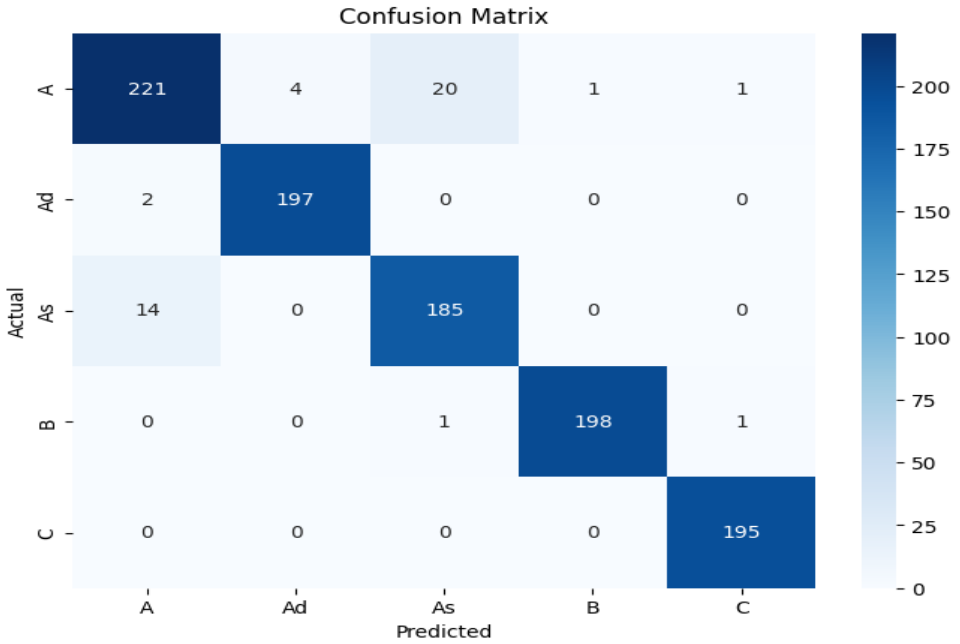


(b) Adam

The confusion matrix of CNN.

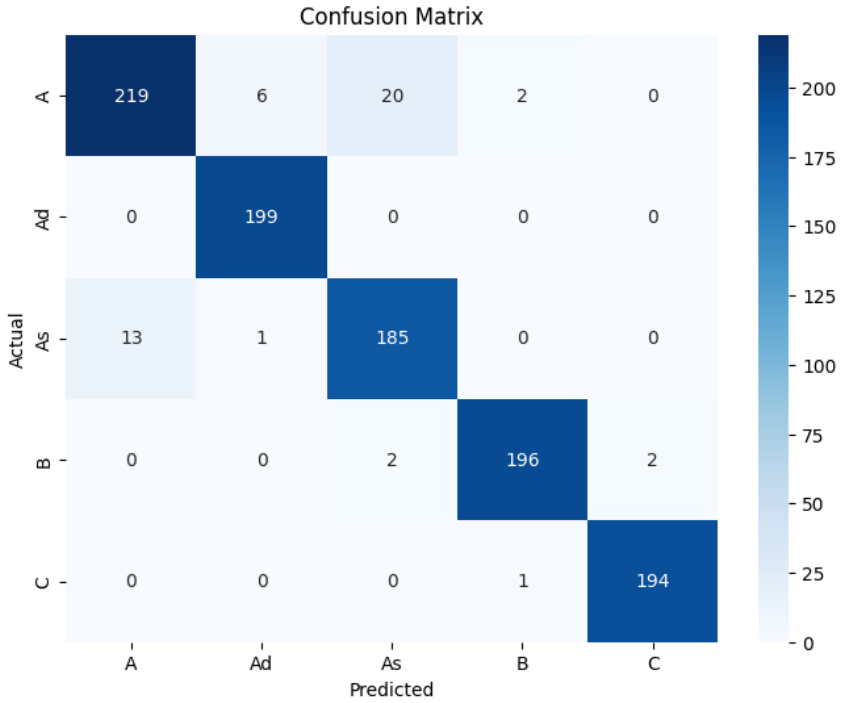


(a) SGD

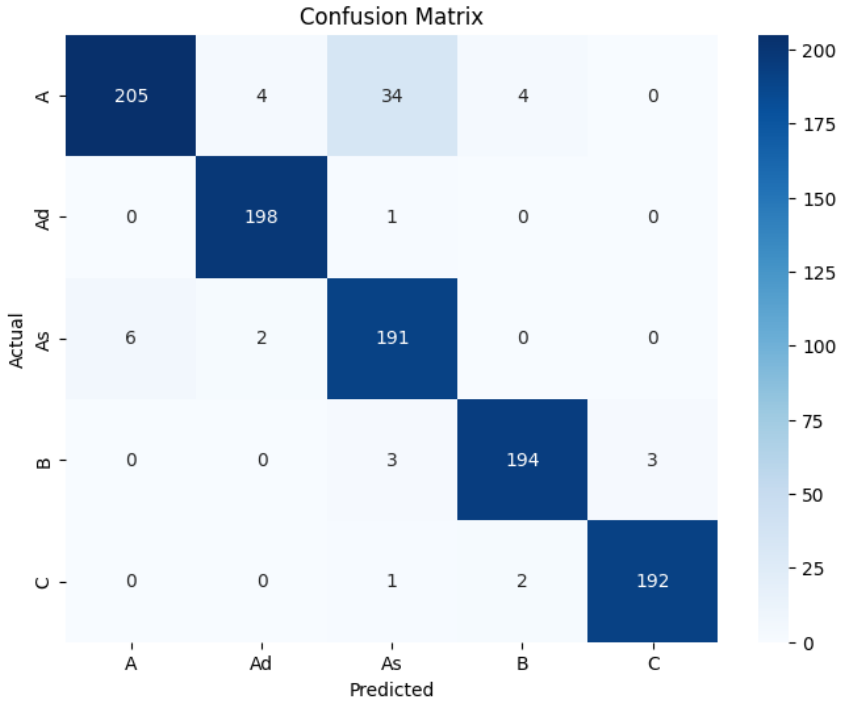


(b) Adam

The confusion matrix of LSTM.



The confusion matrix of ABC-MLP



The confusion matrix of ABC-CNN

الخلاصة

يتميز التهاب الأذن الوسطى بالتهاب وتراكم السوائل في المنطقة خلف طبلة الأذن، والتي تحتوي على العظام الصغيرة المسؤولة عن السمع. يُعد التهاب الأذن الوسطى أحد الأسباب الرئيسية لضعف السمع عالميًا. يُعد التشخيص الدقيق لالتهاب الأذن الوسطى أمرًا بالغ الأهمية في المجال الطبي، إلا أن تشخيصه في الرعاية الصحية الأولية يُعيقه نقص المعدات والخبرة المتخصصة، واختلاف تفسيرات الأطباء، والتشخيصات الخاطئة. في الآونة الأخيرة، تُقدم تقنيات التعلم العميق نهجًا عمليًا لأتمتة تصنيف الأمراض في التهاب الأذن الوسطى. تُركز هذه الرسالة على طريقة تصنيف أمراض التهاب الأذن الوسطى بناءً على بيانات قياس طبلة الأذن. تتكون بيانات قياس طبلة الأذن الوسطى من منحنيات الضغط والامتثال التي تُمثل وظيفة الأذن الوسطى.

تتكون الطريقة المقترحة من أربع مراحل: جمع مجموعة بيانات جديدة من خمسة مراكز سريرية لـ 892 مريضًا، ومعالجة البيانات مسبقًا، وتقسيم مجموعة البيانات، وإنشاء النموذج وتدريبه، وبعد ذلك مرحلة التقييم. تُقدّم هذه الرسالة نموذجًا جديدًا يجمع بين خوارزمية مستعمرة النحل الاصطناعي والذاكرة طويلة المدى قصيرة المدى (ABC-LSTM) لتحسين المعاملات الفائقة لتصنيف خمس فئات (A, B, C, Ad, As) من أمراض الأذن الوسطى. يستطيع النموذج المُقترح التعامل مع البيانات الزمنية بفضل خوارزمية LSTM، وتحسين معاملاتها الفائقة بفضل خوارزمية ABC، مما يجعله مناسبًا جدًا لتصنيف أمراض الأذن الوسطى باستخدام بيانات قياس طبلة الأذن.

أظهر النموذج المُقترح (ABC-LSTM) تقاربًا قويًا أثناء التدريب، وحقق دقةً بلغت 95.96%، ودقةً بلغت 96.11%، وتذكرًا بلغ 96.32%، ودرجة F بلغت 96.16% في مجموعة بيانات الاختبار. بالإضافة إلى ذلك، تفوق النموذج بشكل ملحوظ على جميع النماذج الأخرى في التجارب التي أُجريت على مجموعة متنوعة من المقاييس لمجموعة البيانات الجديدة، بالإضافة إلى الأعمال المنشورة سابقًا. تشير هذه النتائج إلى تصنيف قابل للتطبيق لأمراض التهاب الأذن الوسطى بناءً على بيانات قياس طبلة الأذن للمساعدة في تصنيف التهاب الأذن الوسطى، والمساهمة في التخصصات الطبية والحسابية، وتظهر اتفاقًا مماثلاً مع الأطباء المتخصصين في تشخيص التهاب الأذن الوسطى في المراحل المبكرة.

جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة تكريت /كلية علوم الحاسوب والرياضيات

قسم علوم الحاسوب

تحليل بيانات قياس طبلة الأذن استناداً إلى نموذج ABC-LSTM الهجين لتصنيف التهاب الأذن الوسطى

رسالة ماجستير مقدمة الى

مجلس كلية علوم الحاسوب والرياضيات في جامعة تكريت

وهي جزء من متطلبات نيل شهادة الماجستير

في علوم الحاسوب

من قبل الطالب

محمد محمود حسين

بإشراف

أ.م. سلوى خالد عبد اللطيف

1447 هـ

2025 م