



Article

Performance Comparison of Convolutional Neural Network-Based Hearing Loss Classification Model Using Auditory Brainstem Response Data

Jun Ma ¹ , Seong Jun Choi ², Sungyeup Kim ³ and Min Hong ^{4,*} 

¹ Department of Software Convergence, Soonchunhyang University, Asan 31538, Republic of Korea; ringring369@gmail.com

² Department of Otorhinolaryngology—Head and Neck Surgery, College of Medicine, Soonchunhyang University Cheonan Hospital, Cheonan 31151, Republic of Korea; akas9238@hanmail.net

³ Insitute for Artificial Intelligence and Software, Soonchunhyang University, Asan 31538, Republic of Korea; sungyeup.kim@gmail.com

⁴ Department of Computer Software Engineering, Soonchunhyang University, Asan 31538, Republic of Korea

* Correspondence: mhong@sch.ac.kr

Abstract: This study evaluates the efficacy of several Convolutional Neural Network (CNN) models for the classification of hearing loss in patients using preprocessed auditory brainstem response (ABR) image data. Specifically, we employed six CNN architectures—VGG16, VGG19, DenseNet121, DenseNet-201, AlexNet, and InceptionV3—to differentiate between patients with hearing loss and those with normal hearing. A dataset comprising 7990 preprocessed ABR images was utilized to assess the performance and accuracy of these models. Each model was systematically tested to determine its capability to accurately classify hearing loss. A comparative analysis of the models focused on metrics of accuracy and computational efficiency. The results indicated that the AlexNet model exhibited superior performance, achieving an accuracy of 95.93%. The findings from this research suggest that deep learning models, particularly AlexNet in this instance, hold significant potential for automating the diagnosis of hearing loss using ABR graph data. Future work will aim to refine these models to enhance their diagnostic accuracy and efficiency, fostering their practical application in clinical settings.



Citation: Ma, J.; Choi, S.J.; Kim, S.; Hong, M. Performance Comparison of Convolutional Neural Network-Based Hearing Loss Classification Model Using Auditory Brainstem Response Data. *Diagnostics* **2024**, *14*, 1232.

<https://doi.org/10.3390/diagnostics14121232>

Academic Editor: Zhuhuang Zhou

Received: 8 May 2024

Revised: 6 June 2024

Accepted: 9 June 2024

Published: 12 June 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: auditory brainstem response; ABR; deep learning; VGG16; VGG19; DenseNet121; Densenet201; Alexnet; image processing; hearing loss

1. Introduction

Convolutional Neural Networks (CNNs) are a specialized category of deep learning algorithms predominantly utilized in the fields of image and video recognition. Characteristically, CNNs automate the process of learning and classifying image features through a structured network comprising convolutional layers, pooling layers, and fully connected layers. The convolutional layer primarily serves to extract pertinent features from images, while the pooling layer reduces computational load by diminishing the spatial dimensions of the data. The fully connected layer then performs the final task of classification. These networks are extensively applied across various tasks in computer vision, including image classification, object detection, and face recognition, due to their robustness in handling complex visual inputs [1,2]. In the realm of medical imaging, CNNs assume a critical role given the intricate nature of most medical datasets. They provide effective mechanisms for processing and interpreting such data swiftly, which is indispensable in clinical settings. Consequently, CNNs are employed in diverse medical imaging applications encompassing disease classification, tissue categorization, and image segmentation, among others [3]. This versatility underlines the significance of CNNs in advancing medical image analysis and improving diagnostic methodologies.

The Auditory Brainstem Response (ABR) is an electrophysiological measurement reflecting the brainstem's activity in response to auditory stimuli. This response involves the transmission of a neuroelectric signal from the cochlea through the auditory pathways to the auditory cortex of the brain. The ABR test, a diagnostic procedure used to assess hearing functionality, measures the waveform of this electrical response. This test is particularly valuable in clinical settings for evaluating hearing impairment. Its non-invasive nature and independence from patient consciousness—being unaffected by sleep or anesthesia—make it particularly suitable for use in populations unable to provide reliable auditory feedback. These include newborns, infants, young children, the elderly, and individuals with congenital disabilities. Consequently, the ABR test provides a robust and objective method for assessing auditory function across a diverse patient demographic [4].

ABR measurement is a neurophysiological method used to record changes in brain waves triggered by auditory stimulation. This technique involves the application of click sounds at intervals of approximately 0.8 ms combined with energy modulation during auditory transmission to stimulate brain wave activity. When auditory stimuli ranging from 10 dB to 100 dB are administered, typically, five to seven waves are detectable. In normal adults, a waveform responsive to the stimulus typically emerges within approximately 10 milliseconds following the onset of the click sound [5]. Among the identifiable waves, wave number 5 (V wave) is particularly significant for clinical assessments. The threshold of hearing is determined by analyzing the latency periods of the V wave across various dB levels. Hearing loss is subsequently diagnosed based on these latency values within the specified dB range [6].

During the ABR testing procedure, as illustrated in Figure 1, small electrodes marked by red circles are affixed to the subject's forehead and behind the ears. These electrodes detect electrical activity within the auditory nerve and brainstem in response to auditory stimuli. The test involves the administration of a series of click sounds delivered through an eartip inserted into the ear, with the brain's responses—essentially, brain waves—being detected and automatically recorded by a computer system. This method offers an objective assessment of hearing functionality, contrasting with other methods that rely on subjective patient responses. In the process of measurement, the audiologist identifies and records the V wave, which is critical for hearing and occurs between 6 ms and 8 ms, from among wave information labeled from 1 to 5, corresponding to each decibel (dB) stimulus level. The measurement concludes once the waveforms for all dB stimulus levels have been successfully recorded [7].



Figure 1. Auditory brainstem response test scene.

In our previous study [8], we conducted preprocessing to standardize the auditory brainstem response (ABR) graph outputs across various manufacturers. ABR graph data from five different manufacturers—Audera, Navigator, Eclipse, Viking Select, and Interacoustics—were collected. Each ABR graph was normalized as depicted in Figure 2, resulting in a dataset comprising 10,000 data entries. Furthermore, during the preprocessing phase, a total of 2010 images were filtered out due to significant reductions in

graph resolution or improper graph outputs. Consequently, the analysis was conducted on 7990 images using normalized ABR data from the remaining valid datasets.

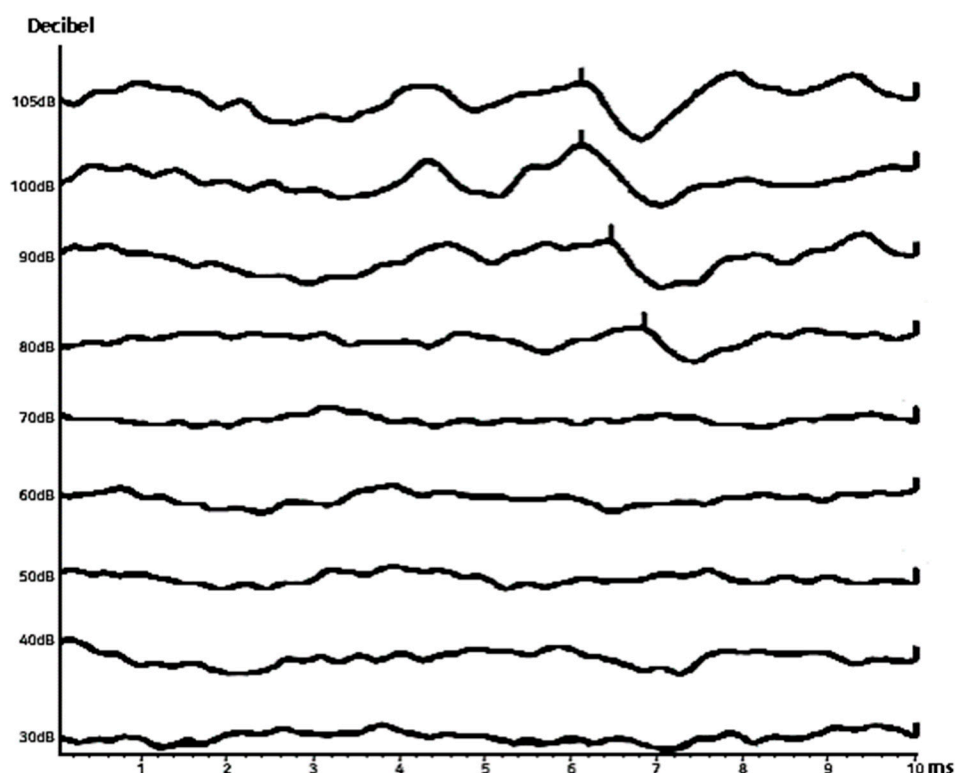


Figure 2. Auditory brainstem response data after pre-processing.

2. Materials and Methods

2.1. Auditory Brainstem Response Data

The ABR graph serves as a crucial diagnostic tool for evaluating the auditory system by visualizing changes in electrical activity over time. An analysis of the ABR graph can reveal significant insights into the auditory system's condition through various characteristic features described below.

Multi-wave form: the ABR graph typically displays a series of waves, each sequentially representing electrical activity in different parts of the brain at specific times.

The size and spacing of waves: the initial wave usually appears as the largest and most distinct wave, with subsequent waves diminishing progressively in size and becoming closer together.

Latency and amplitude: these parameters are critical; latency refers to the timing of the wave's occurrence post stimulus, and amplitude denotes the wave's magnitude.

Baseline: the baseline of the graph indicates normal brain activity levels; any deviation from this baseline may suggest abnormalities in the auditory pathway.

Axes: the ABR graph is typically oriented with time (ms) on the horizontal axis and amplitude (μV) on the vertical axis, where sound pressure levels (decibels, dB) may also be considered.

By systematically assessing these features—particularly changes in latency and amplitude—abnormalities such as hearing impairment or disorders within the central auditory pathway can be detected. Thus, the ABR graph not only aids in assessing the patient's hearing status but also contributes to the formulation of an appropriate treatment plan [9–12]. The methodology for analyzing an ABR graph, as depicted in Figure 3, is essential for comprehensively evaluating both hearing function and the broader state of the auditory system.

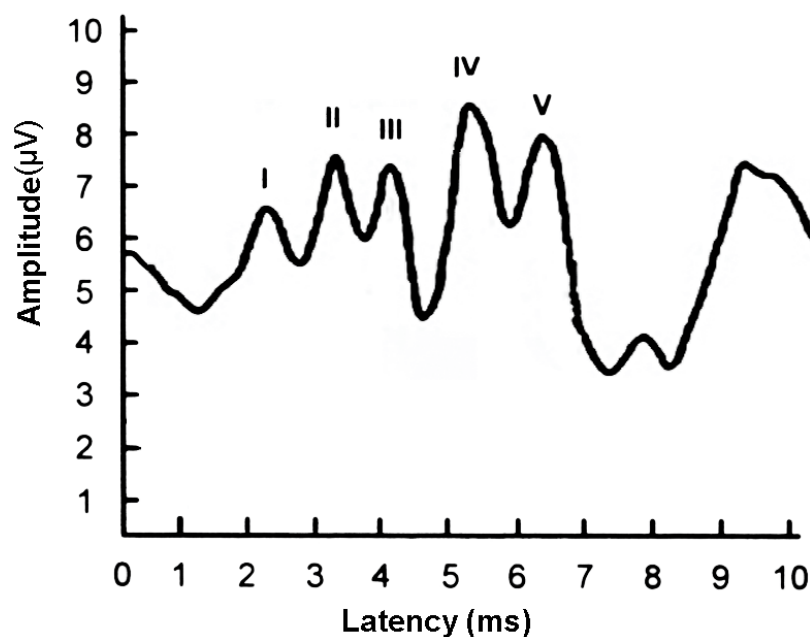


Figure 3. Auditory brainstem response example graph.

Wave I: This initial wave is indicative of the acoustic signal's arrival at the auditory nerve. The latency for Wave I, representing the time taken for the stimulus to reach the auditory nerve, typically ranges from 1 to 4 ms.

Wave II: occurring in the auditory brainstem, specifically in the region associated with the auditory pathway's initial processing stages, the latency from Wave I to Wave II measures the transmission time to the auditory brainstem and is generally observed between 2 and 4 ms.

Wave III: this wave is generated in the cochlear nuclei—the Dorsal Cochlear Nucleus (DCN) and Ventral Cochlear Nucleus (VCN)—located in the lower auditory brainstem. The latency from Wave II to Wave III, which measures the passage of stimulus through the auditory brainstem, typically ranges between 3 and 5 ms.

Wave IV: Representing signals generated en route to the Medial Superior Olive (MSO) at the upper part of the auditory brainstem; the latency between Wave III and Wave IV usually spans 4 to 5.5 ms.

Wave V: The largest of the acoustic signals, Wave V emanates from the output region of the auditory brainstem, reaching the Inferior Colliculus (IC). The latency from Wave IV to Wave V is noted between 5.5 and 7 ms.

Wave latency: the latency of each wave quantifies the time required for its generation. Within a normal auditory system, these latencies fall within specific ranges; however, abnormalities may manifest as delayed latencies.

Wave amplitude: The amplitude of each wave reflects its magnitude. Typically, a healthy auditory system produces waves of a large and consistent amplitude. Reduced amplitude may indicate auditory abnormalities.

Interpeak latency: This metric illustrates the latency differences between consecutive waves, reflecting the conduction time along the central auditory pathway. Normal auditory systems exhibit consistent interpeak latencies, whereas increased values may suggest central auditory pathway dysfunction.

These metrics provide a comprehensive framework for assessing the integrity and functionality of the auditory system through ABR testing, facilitating the identification and characterization of potential auditory impairments.

Hearing loss: Hearing loss is characterized as a reduction in the ability to perceive or interpret sounds; the loss is attributable to anomalies within the auditory system, which may involve the external or internal ear structures or the auditory nerve. This condition

can be either temporary or permanent and may affect one or both ears. Within the context of this study, the severity of hearing loss is assessed based on the detection of the V wave in the ABR graph data, as illustrated in Figure 4. Typically, V waves are elicited by sound stimuli ranging from 10 to 100 dB, presented in 10 dB increments. An absence of V waves in waveforms at or below 40 dB typically leads otolaryngologists to diagnose hearing loss.

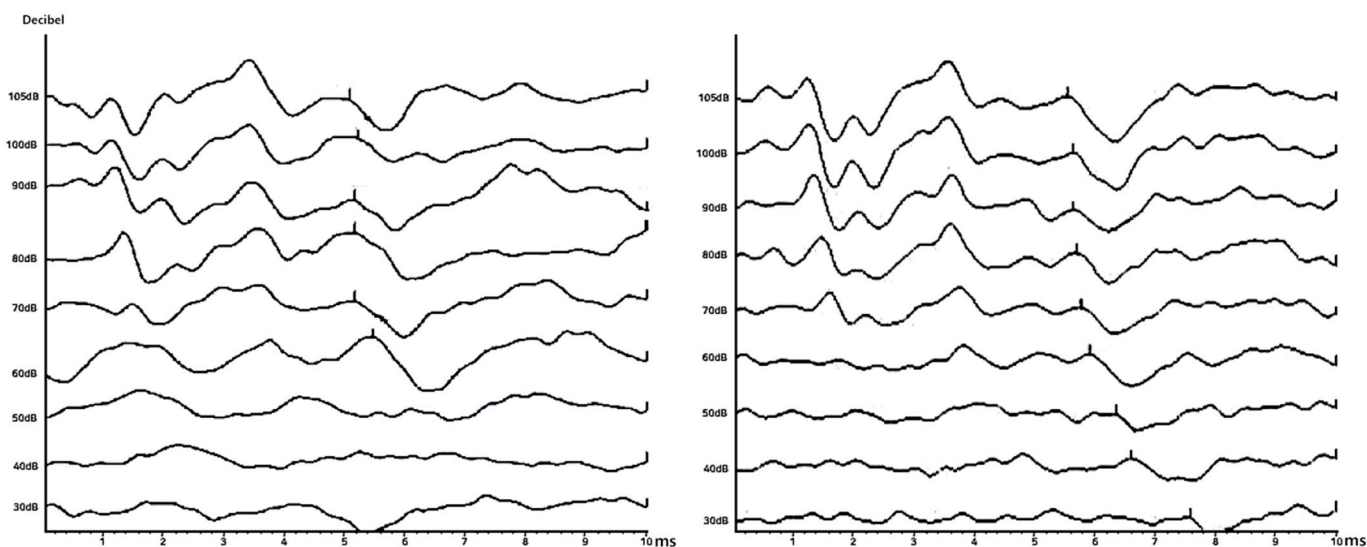


Figure 4. An ABR graph of a patient with hearing loss (left) and an ABR graph of a normal person (right).

The analysis of an ABR graph involves a detailed evaluation of various parameters to ascertain the auditory system's status. Key indicators include the latency and amplitude of waves; a delay in wave latency or a reduction in amplitude may signal auditory impairments or issues within the central auditory pathway. By examining these attributes, clinicians can effectively gauge a patient's hearing condition and devise appropriate treatment strategies. Such diagnostic practices are crucial for the early detection and management of hearing loss, thereby enhancing the quality of life and communication abilities of affected individuals [13–16].

2.2. CNN Classification Model

In our previous study, we evaluated the classification of hearing loss using solely the VGG16 model [8]. Expanding upon this initial approach, the current study incorporates a broader array of convolutional neural network models, specifically VGG16, VGG19, DenseNet121, DenseNet201, AlexNet, and InceptionV3, to perform more comprehensive learning and classification tests. Each model, recognized for its unique strengths and limitations, has previously been utilized across a variety of medical image classification tasks. For the purposes of this study, we tailored the hyperparameters, specifically the batch size and layer configurations, to optimize the learning process for ABR image classification. This part details the architectural nuances and characteristics of each model and describes the specific modifications made to the hyperparameters to enhance model performance for this application. These adjustments are pivotal in refining our approach to accurately classifying hearing loss through deep learning techniques.

2.2.1. VGG16 and VGG19

The VGG model, developed by the Visual Geometry Group at Oxford, includes two primary configurations: VGG16 and VGG19. VGG19 extends the architecture of VGG16 by adding three additional convolutional layers positioned before the 3rd, 4th, and 5th max pooling layers, enhancing its depth and complexity. In our experiments with the VGG16 model, the original images, sized 573×505 pixels, were initially resized to 224×224 pixels.

Subsequently, the images were further scaled down to dimensions of 286×252 , 143×126 , 71×63 , 35×31 , and 17×15 to facilitate object recognition. The learning process involved adjustments in the dense layer configurations, with neuron counts set to 1024, 512, and 2, optimizing the network's ability to discern features at various scales. The VGG19 model, with its additional convolutional layers, retains the number of layers for sizes 286×252 and 143×126 but adds a layer each at smaller scales (71×63 and smaller), comprising a total of 19 layers to enhance detail recognition [17–21]. In a related study conducted by Dey et al., a pneumonia detection model utilizing VGG19 applied to chest X-ray images demonstrated a high classification accuracy of up to 97.94% [22]. Similarly, Mateen et al. reported that the VGG19 model was effectively utilized in medical image analysis, achieving an impressive classification accuracy of 98.13% in a retinopathy classification system using fundus images [23]. For the purpose of this study, which was tailored to ABR image classification, the VGG19 model and the VGG16 model were enhanced by incorporating two additional dense layers and a dropout layer in each configuration to prevent overfitting. The learning framework was structured with dense layers of 1024, 512, 256, 128, and 2 neurons, with a batch size of 8, serving as a robust hyperparameter setup. This architecture was designed to maximize the model's ability to accurately classify ABR images, leveraging deeper layers for more nuanced feature extraction.

2.2.2. DenseNet121 and DenseNet201

DenseNet is a CNN model engineered to enhance training efficiency by integrating the concept of shorter connections. This design enables direct links between the input and output layers, fostering a deeper and structurally more efficient network capable of delivering precise performance outcomes. Unlike traditional CNNs, which feature connections primarily to the immediately subsequent layer, DenseNet boasts a comprehensive connection structure with $L(L + 1)/2$ direct connections, greatly enriching the flow of information across the network. To efficiently manage down-sampling, the architecture is segmented into three distinct dense blocks. Each block is separated by a transition layer which performs both convolution and pooling operations, thus maintaining the network's depth while progressively reducing its dimensionality [24–28]. In a related study by Chauhan et al., a DenseNet model was employed to differentiate COVID-19 patients from healthy individuals using chest X-ray images, achieving an impressive accuracy rate of 98.45% [29]. Within the context of this paper, the DenseNet121 and DenseNet201 models, comprising 121 and 201 layers, respectively, were utilized. Tailored specifically for ABR image classification, the learning process was conducted using dense layers configured with 256 and 2 neurons and a hyperparameter setting of a batch size of 8. This configuration was designed to optimize the network's capability for high-accuracy classification in ABR imaging.

2.2.3. AlexNet

AlexNet, a CNN model, significantly impacted the field of deep learning after securing victory in the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). Named after Alex Krizhevsky, the lead author of the seminal paper "ImageNet Classification with Deep Convolutional Neural Networks", AlexNet's architecture has been instrumental in advancing CNN development. Its structure features a sequential layout comprising an input layer and five convolutional layers—each accompanied by a max pooling layer and a normalization layer—culminating in a dense layer dedicated to classification tasks [30–33]. In research conducted by Chen et al., the efficacy of various models including 3DAlexNet, ResNet50, and InceptionV4 was evaluated for the classification of magnetic resonance images to diagnose prostate cancer, yielding classification accuracies of 92.1%, 87.6%, and 85.7%, respectively [34]. Another study by Titoriya and Sachdeva utilized the AlexNet model to classify breast cancer tissue images, demonstrating a high classification accuracy of 95.7%, thereby underscoring its potential for medical imaging applications [35]. In this study, modifications were made to the original AlexNet architecture to enhance its suitability for Auditory Brainstem Response (ABR) image classification. Adjustments

included the integration of max pooling layers at the first and fifth convolutional layers and the insertion of dropout layers within each dense layer to mitigate overfitting. The learning process was optimized by setting the hyperparameters of the dense layers to 4096, 4096, and 2, with a batch size of 8, facilitating an improved learning rates and robust classification performance in ABR image analysis.

2.2.4. InceptionV3

Generally, there is a correlation between increased model size and both accuracy and computational effort. For instance, the DenseNet architecture enhances performance by deepening the model with skip connections, yet this also escalates computational demands, resulting in longer training durations due to the increased depth. Similarly, enlarging model size augments computational requirements, which presents a limitation when operating within memory constraints. The Inception model, devised by Google, addresses this challenge by employing convolutional decomposition to expand the model size while minimizing computational costs. The InceptionV3 model, which was utilized in this research, stands out among the Inception series with its 42-layer deep network, which is optimized to maintain a balance between a low parameter count and computational efficiency, akin to that of the VGG models [36,37]. In research conducted by Wang, Cheng, et al., the InceptionV3 model was applied to develop a classification system for lung nodules using chest X-ray images, achieving a classification accuracy of up to 86.4% [38]. In the context of this study, the InceptionV3 model was adapted for ABR image classification. Modifications were made to the model's configuration, setting the hyperparameters of the dense layers to 256 and 2, and the batch size to 8, to tailor the learning process specifically for ABR image analysis. This strategic adjustment aims to leverage the model's efficiency and deep learning capabilities for precise ABR image classification.

3. Results

Model Training and ABR Data Classification Results

Using 7990 ABR data excluding impure data, learning and classification tests were conducted with 4794, 1598, and 1598 train, validation, and test data at a ratio of 6:2:2, respectively. The accuracy, loss results, and test classification confusion matrix results of each model's learning are as follows.

The results in Figure 5 highlight the performance metrics for the VGG16 model during training, showing an accuracy of 91.58% and a loss of 6.52%. The confusion matrix for the test dataset for this model indicated 769 true negatives (tn), 52 false negatives (fn), 707 true positives (tp), and 70 false positives (fp). The VGG19 model demonstrated improved training performance, achieving an accuracy of 94.84% and a loss of 4.64%. The test data for this model produced a confusion matrix with 770 tn, 12 fn, 735 tp, and 81 fp. Additionally, the DenseNet121 model recorded a training accuracy of 92.52% and a loss of 5.77%, with its confusion matrix displaying 727 tn, 49 fn, 753 tp, and 69 fp.

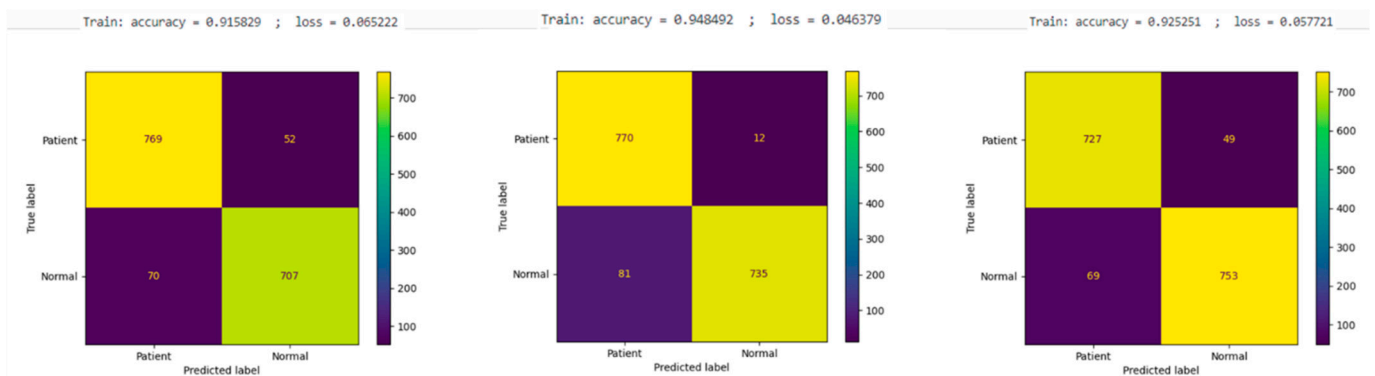


Figure 5. Learning results of VGG16 model (left), VGG19 model (middle), and DenseNet121 model (right).

And the results in Figure 6, the DenseNet201 model's training performance exhibited an accuracy of 93.09% and a loss of 5.19%, with the confusion matrix for the test dataset indicating 752 tn, 34 fn, 739 tp, and 73 fp. The AlexNet model, on the other hand, achieved a training accuracy of 96.54% and a remarkably lower loss of 2.99%. The corresponding confusion matrix demonstrated its high precision with 748 tn, 51 fn, 785 tp, and only 14 fp, underscoring its efficacy in accurately classifying the conditions with minimal misclassifications. Lastly, the InceptionV3 model registered a training accuracy of 91.64% and a loss of 6.59%, with its test data confusion matrix revealing 760 tn, 56 fn, 685 tp, and 97 fp.

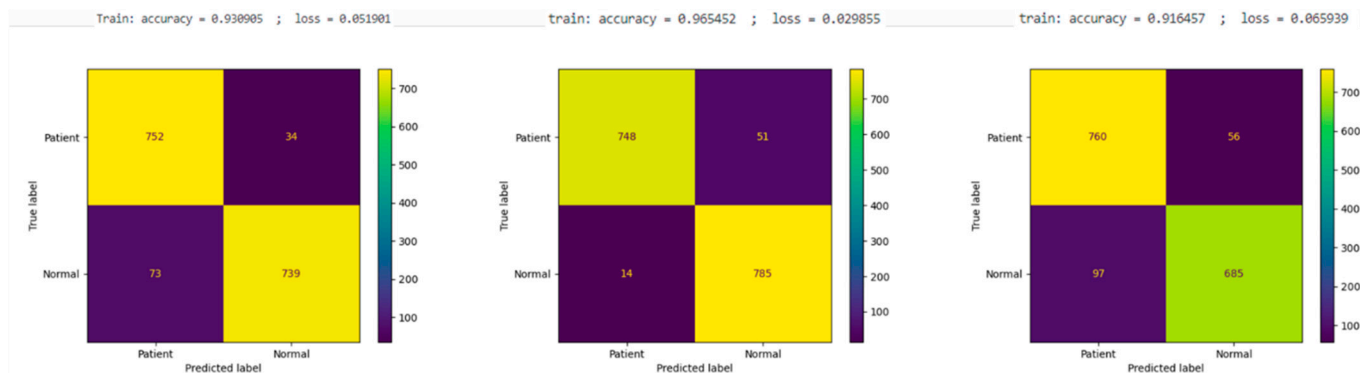


Figure 6. Learning results of DenseNet201 model (left), AlexNet model (middle), and InceptionV3 model (right).

Based on the confusion matrix results from the test data for each model, various performance metrics were calculated and systematically tabulated. These metrics include accuracy, the true negative rate (TNR), the true positive rate (TPR), the false positive rate (FPR), the false negative rate (FNR), precision, and the F1 score. The formulas for each of these metrics are outlined below, with their respective results being derived from Equations (1)–(7):

$$\text{Accuracy} = \frac{\text{tp} + \text{tn}}{\text{tp} + \text{tn} + \text{fp} + \text{fn}} \quad (1)$$

Accuracy: this is calculated as the ratio of correctly predicted observations (both true positives and true negatives) to the total observations in the dataset.

$$\text{TNR} = \frac{\text{tn}}{\text{tn} + \text{fp}} \quad (2)$$

True negative rate (TNR), also known as specificity: this measures the proportion of actual negatives that are correctly identified.

$$\text{TPR} = \frac{\text{tp}}{\text{tp} + \text{fn}} \quad (3)$$

True positive rate (TPR), also known as sensitivity or recall: this metric indicates the proportion of actual positives that are correctly identified.

$$\text{FPR} = \frac{\text{fp}}{\text{fp} + \text{tn}} \quad (4)$$

False positive rate (FPR): this is calculated as the ratio of the number of false positives to the sum of the false positives and true negatives.

$$\text{FNR} = \frac{\text{fn}}{\text{fn} + \text{tp}} \quad (5)$$

False negative rate (FNR): this measures the proportion of positives which yield negative test outcomes with the model.

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \quad (6)$$

Precision is also known as the positive predictive value: this is the ratio of true positives to the combined total of true positives and false positives.

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{TPR}}{\text{precision} + \text{TPR}} \quad (7)$$

F1 score: this is the harmonic mean of Precision and Recall, providing a balance between the two when their rates may vary.

The calculated values for these metrics are compiled in a Table 1 within this paper, providing a comprehensive assessment of each model's performance on the test dataset. This structured approach allows for a detailed comparison and evaluation of the effectiveness of each classification model in the context of hearing loss detection.

Table 1. Total data validity calculation results.

| | Accuracy | TNR | TPR | FPR | FNR | Precision | F1 Score |
|-------------|----------|--------|--------|-------|--------|-----------|----------|
| VGG16 | 92.37% | 93.67% | 90.99% | 6.33% | 9.01% | 93.15% | 0.9206 |
| VGG19 | 94.18% | 98.47% | 90.07% | 1.53% | 9.93% | 98.39% | 0.9405 |
| DenseNet121 | 92.62% | 93.69% | 91.61% | 6.31% | 8.39% | 93.89% | 0.9273 |
| DenseNet201 | 93.30% | 95.67% | 91.01% | 4.33% | 8.99% | 95.60% | 0.9325 |
| AlexNet | 95.93% | 93.62% | 98.25% | 6.38% | 1.75% | 93.90% | 0.9602 |
| InceptionV3 | 90.43% | 93.14% | 87.60% | 6.86% | 12.40% | 92.44% | 0.8995 |

The result of the classification models utilized in this research were evaluated based on outputs of classification scores. Figure 7 presents the classification score results for the AlexNet model, which demonstrated the highest accuracy among the models tested. This figure displays image data that was randomly selected from the test dataset. It includes the filename of the data—where “napa” denotes an image associated with hearing loss and “tupa” indicates a normal hearing image. Additionally, the outcomes of the classification process are indicated, with the number 0 representing hearing loss and the number 1 representing normal hearing. The scores leading up to these classifications are also documented to provide a comprehensive view of the model's performance in distinguishing between the two categories. This detailed display of results facilitates an understanding of the model's efficacy in accurately classifying auditory conditions based on ABR image data.

These findings illustrate the varying levels of performance and accuracy across the models tested, offering insights into their respective strengths and areas for improvement in the classification of conditions based on the training and validation datasets.

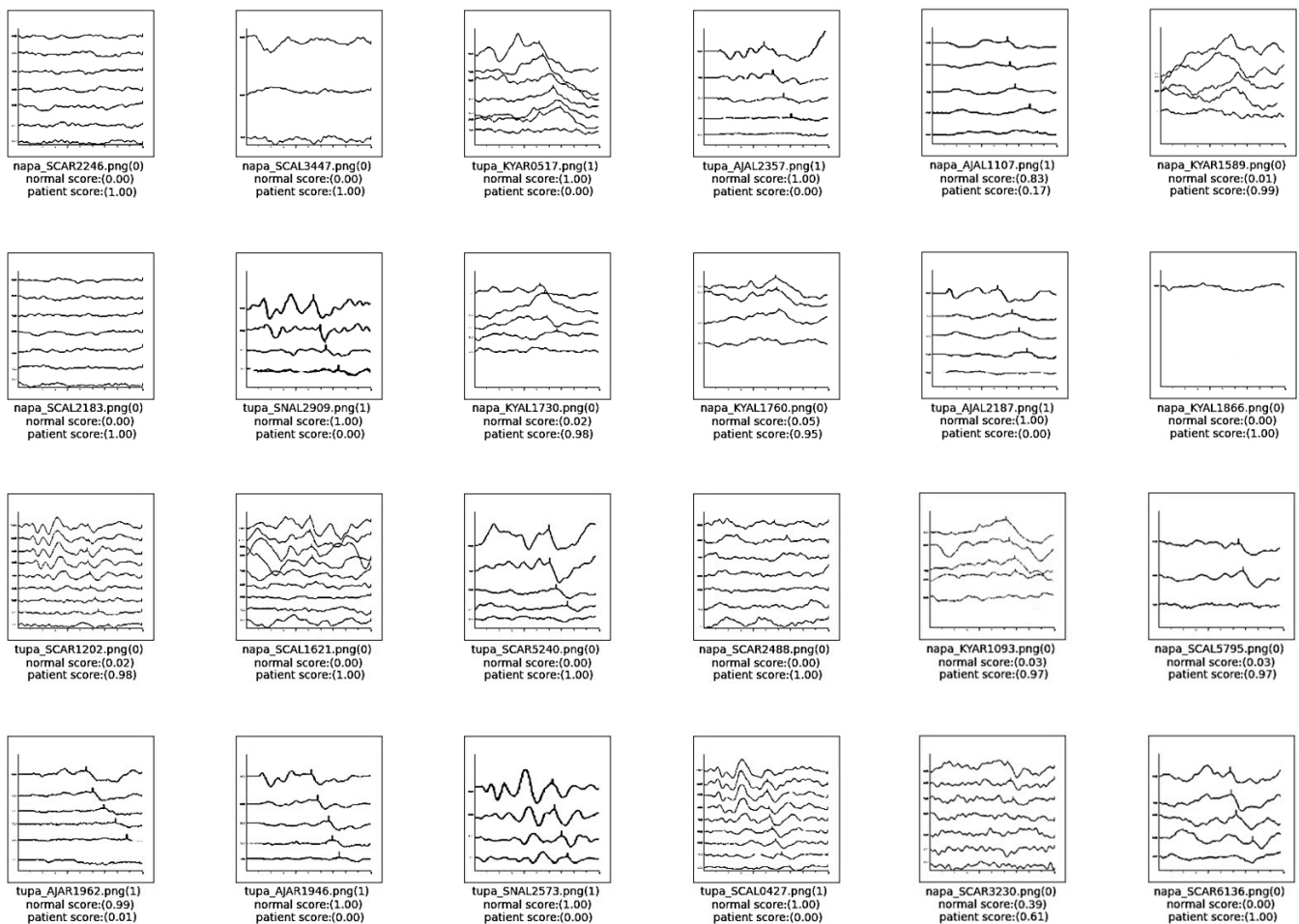


Figure 7. Classification result scores of AlexNet model.

4. Discussion

4.1. Classification Data Analysis

In the current study, the classification learning revealed that AlexNet achieved the highest overall accuracy, recording 95.93%. However, when focusing specifically on the accuracy of hearing loss classification, as measured by the TNR, VGG19 excelled with a TNR of 98.47%, making it the most effective model for this particular objective. Given that the primary aim of this research is to accurately classify hearing loss, the VGG19 model emerges as the superior performer in this context. Nevertheless, it is important to consider the structural differences between the models. AlexNet, with its comparatively shallower layer depth, consumes fewer temporal resources during the learning classification process. Thus, for applications involving the classification of ABR data on a scale larger than the current dataset of 7990 cases, AlexNet presents a viable option due to its efficiency in handling larger datasets without a significant increase in computational demand. This balance between accuracy and efficiency is crucial for scaling the application of these models to larger datasets in future studies.

4.2. Analysis of ABR Data That Are Not Classified Correctly

In this part, we will discuss the results that were not classified correctly among the classification results from various models.

4.2.1. False Negative: In Case the Data Are actually Normal but Are Classified as a Patient with Hearing Loss

Figure 8 presents a selection of misclassified cases identified through the application of the AlexNet, VGG19, and VGG16 models within the classification analyses of this study. These instances involve subjects who, despite having normal hearing, were erroneously classified as suffering from hearing loss. The graph on the left represents ABR data from a 1-year-old infant, with V waves detected at 60 dB, 40 dB, and 30 dB. The analysis suggests that the model's misclassification may stem from the limited number of data points in this sample, contrasting with the more comprehensive ABR data typically gathered from the general population, which is measured in 10 dB increments from 30 dB to 90 dB. The middle graph displays ABR results for a 52-year-old individual; no V wave was detected at 30 dB. This subject, diagnosed with normal hearing by a medical professional, was analyzed in comparison to typical public ABR data, which usually exhibits V waves across all tested decibels. The absence of a V wave at 30 dB in this case led to a model misrecognition, highlighting a deviation from expected patterns observed in broader datasets. The graph on the right documents ABR measurements for an 18-year-old individual. The analysis determined that the model misclassification occurred because the final 30 dB graph plotted very close to the x-axis. This proximity likely influenced the model's perception, causing it to misidentify the presence of a V wave, which deviates from the normative data where V waves are consistently present across all measurements. These illustrations underscore challenges with model accuracy when faced with atypical data representations, emphasizing the need for the continuous refinement of classification algorithms to enhance diagnostic precision in clinical settings.

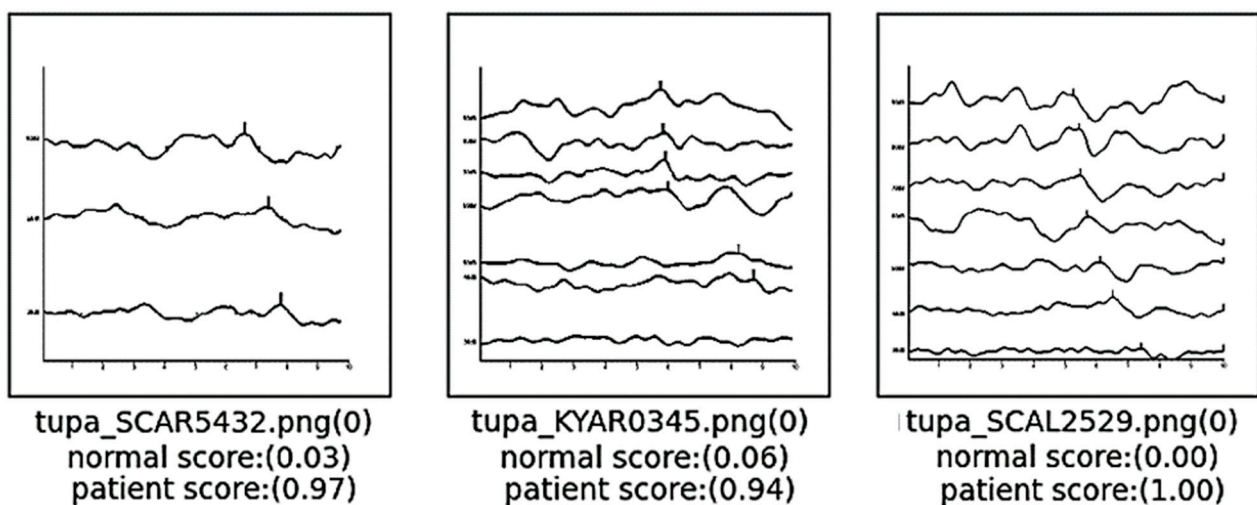


Figure 8. Cases of false negatives.

4.2.2. False Positive: In Case the Data Represent Patients with Actual Hearing Loss but Classified as Normal

Figure 9 presents images indicative of hearing loss that were inaccurately classified as normal by the model. For the images on the left and right, prior to the existing preprocessing steps, the resolution was significantly compromised, necessitating further preprocessing to enhance resolution quality. Despite these efforts, the resolution remained comparatively lower than that of typical ABR graphs, which likely impeded the model's ability to accurately classify these cases. Concerning the image in the middle, the analysis suggests that the misclassification occurred due to the proximity of the wave in the bottom graph to the x-axis and its closeness to the graph directly above it. This spatial arrangement may have confused the model, leading to an incorrect interpretation of the data. This observation underscores the sensitivity of classification models to variations in graphical representation and highlights the need for robust preprocessing techniques to ensure consistent image

quality across all data inputs. Such improvements are critical for enhancing the accuracy of diagnostic models in clinical applications.

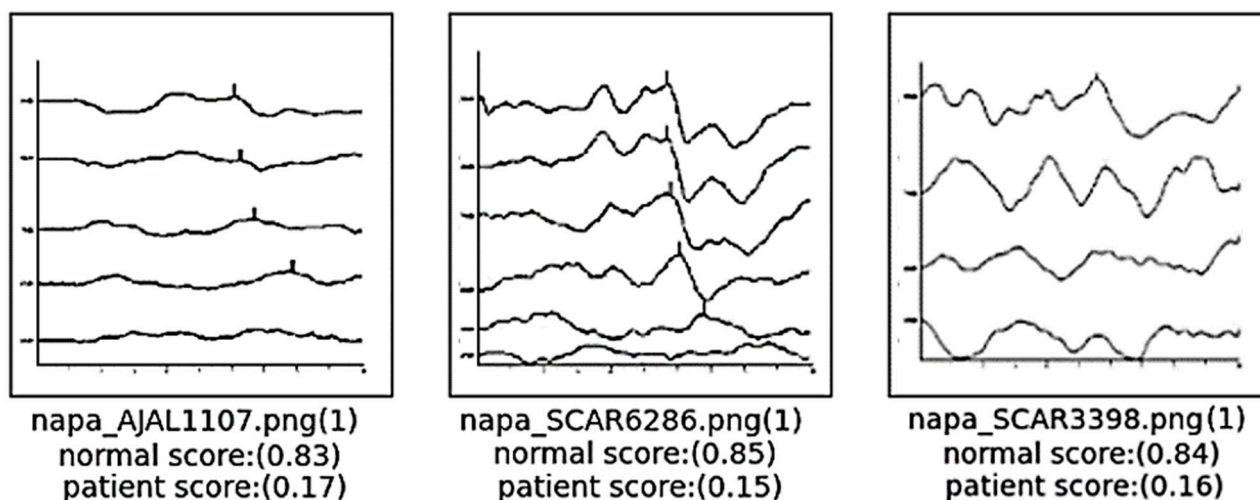


Figure 9. Cases of false positives.

5. Conclusions

In this study, we developed multiple models to classify hearing loss using preprocessed ABR graph data and evaluated their comparative performances. The AlexNet model exhibited the highest accuracy with a value of 95.93%, while the VGG19 model demonstrated the best TNR at 98.47%. Among the six evaluated models, AlexNet showed the quickest learning speed, followed by InceptionV3, VGG16, VGG19, DenseNet121, and DenseNet201 in terms of processing time. For instances in which images are incorrectly classified, future work will involve exploring further supplementation and preprocessing strategies. These will aim to enhance image quality without compromising the integrity of the original data including measures such as increasing resolution, augmenting the X and Y axes, and adjusting the wave positioning for each decibel level.

Previously, the process of diagnosing hearing loss using ABR involved audiologists and otolaryngologists manually reviewing each ABR graph, which was time-consuming. This study was conducted to address this issue and improve the efficiency of the diagnostic process. The findings of this research pave the way for the development of a robust model that can support the preliminary automatic classification of ABR data, assisting in the pre-diagnostic stages before clinical evaluation by a physician. Additionally, the study plans to extend into the creation of an automatic V-latency detection algorithm which will be designed for universal application across various devices rather than being confined to specific equipment. This advancement is anticipated to simplify the diagnosis of hearing loss and related auditory conditions, thereby enhancing patient care and diagnostic efficiency.

Author Contributions: Conceptualization, M.H. and S.J.C.; methodology, J.M. and M.H.; software, J.M.; validation, J.M., S.K., and S.J.C.; formal analysis, J.M. and S.K.; investigation, J.M.; resources J.M. and S.J.C.; data curation, J.M., S.K., and S.J.C.; writing original draft preparation, J.M.; writing review and editing, M.H., S.K., and S.J.C.; visualization, J.M.; supervision, M.H.; project administration, S.J.C.; funding acquisition, M.H. and S.J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the BK21 FOUR (Fostering Outstanding Universities for Research) Grant, No. 5199990914048, and was supported by the Soonchunhyang University Research Fund.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board of Soonchunhyang University Cheonan Hospital (Cheonan, Korea) (IRB number: 2021-06-040; approval date: 4 June 2021).

Informed Consent Statement: Because of the retrospective design of the study, patient consent was waived.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Li, Q.; Cai, W.; Wang, X.; Zhou, Y.; Feng, D.D.; Chen, M. Medical image classification with convolutional neural network. In Proceedings of the 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), Singapore, 10–12 December 2014; pp. 844–848.
- Mu, R.; Zeng, X. A review of deep learning research. *KSII Trans. Internet Inf. Syst. (TIIS)* **2019**, *13*, 1738–1764.
- Yadav, S.S.; Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **2019**, *6*, 1–18. [[CrossRef](#)]
- Eggermont, J.J. Auditory Brainstem Response. In *Handbook of Clinical Neurology*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2019; Volume 160, pp. 451–464.
- Sun, J.; Liu, H.; Wang, X.; Li, Y.; Ni, X. Application of auditory brainstem response to different types of hearing loss in infants. *J. Clin. Otorhinolaryngol. Head Neck Surg.* **2022**, *36*, 120–125.
- Aldè, M.; Binda, S.; Primache, V.; Pellegrinelli, L.; Pariani, E.; Pregliasco, F.; Berardino, F.D.; Cantarella, G.; Ambrosetti, U. Congenital cytomegalovirus and hearing loss: The state of the art. *J. Clin. Med.* **2023**, *12*, 4465. [[CrossRef](#)]
- Elberling, C.; Parbo, J. Reference data for ABRs in retrocochlear diagnosis. *Scand. Audiol.* **1987**, *16*, 49–55. [[CrossRef](#)]
- Ma, J.; Seo, J.H.; Moon, I.J.; Park, M.K.; Lee, J.B.; Kim, H.; Ahn, J.H.; Jang, J.H.; Lee, J.D.; Choi, S.J.; et al. Auditory Brainstem Response Data Preprocessing Method for the Automatic Classification of Hearing Loss Patients. *Diagnostics* **2023**, *13*, 3538. [[CrossRef](#)] [[PubMed](#)]
- Hood, L.J. Principles and applications in auditory evoked potentials. *Ear Hear.* **1996**, *17*, 178. [[CrossRef](#)]
- Sininger, Y.S. Auditory brain stem response for objective measures of hearing. *Ear Hear.* **1993**, *14*, 23–30. [[CrossRef](#)]
- Sininger, Y.S.; Abdala, C.; Cone-Wesson, B. Auditory threshold sensitivity of the human neonate as measured by the auditory brainstem response. *Hear. Res.* **1997**, *104*, 27–38. [[CrossRef](#)] [[PubMed](#)]
- Aiyer, R.G.; Parikh, B. Evaluation of auditory brainstem responses for hearing screening of high-risk infants. *Indian J. Otolaryngol. Head Neck Surg.* **2009**, *61*, 47–53. [[CrossRef](#)]
- Verhulst, S.; Jagadeesh, A.; Mauermann, M.; Ernst, F. Individual differences in auditory brainstem response wave characteristics: Relations to different aspects of peripheral hearing loss. *Trends Hear.* **2016**, *20*, 2331216516672186. [[CrossRef](#)] [[PubMed](#)]
- Galambos, R.; Despland, P.A. The auditory brainstem response (ABR) evaluates risk factors for hearing loss in the newborn. *Pediatr. Res.* **1980**, *14*, 159–163. [[CrossRef](#)] [[PubMed](#)]
- McCreery, R.W.; Kaminski, J.; Beauchaine, K.; Lenzen, N.; Simms, K.; Gorga, M.P. The impact of degree of hearing loss on auditory brainstem response predictions of behavioral thresholds. *Ear Hear.* **2015**, *36*, 309. [[CrossRef](#)] [[PubMed](#)]
- Stapells, D.R.; Gravel, J.S.; Martin, B.A. Thresholds for auditory brain stem responses to tones in notched noise from infants and young children with normal hearing or sensorineural hearing loss. *Ear Hear.* **1995**, *16*, 361–371. [[CrossRef](#)] [[PubMed](#)]
- Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**. [[CrossRef](#)]
- Qassim, H.; Verma, A.; Feinzimer, D. Compressed residual-VGG16 CNN model for big data places image recognition. In Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018; pp. 169–175.
- Mascarenhas, S.; Agarwal, M. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In Proceedings of the 2021 International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON), Bengaluru, India, 19–21 November 2021; Volume 1, pp. 96–99.
- Carvalho, T.; De Rezende, E.R.; Alves, M.T.; Balieiro, F.K.; Sovat, R.B. Exposing computer generated images by eye's region classification via transfer learning of VGG19 CNN. In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017; pp. 866–870.
- Yin, J.; Qu, J.; Huang, W.; Chen, Q. Road Damage Detection and Classification based on Multi-level Feature Pyramids. *KSII Trans. Internet Inf. Syst.* **2021**, *15*, 786.
- Dey, N.; Zhang, Y.D.; Rajinikanth, V.; Pugalenth, R.; Raja, N.S.M. Customized VGG19 architecture for pneumonia detection in chest X-rays. *Pattern Recognit. Lett.* **2021**, *143*, 67–74. [[CrossRef](#)]
- Mateen, M.; Wen, J.; Nasrullah; Song, S.; Huang, Z. Fundus image classification using VGG-19 architecture with PCA and SVD. *Symmetry* **2018**, *11*, 1. [[CrossRef](#)]
- Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
- Singh, D.; Kumar, V.; Kaur, M. Densely connected convolutional networks-based COVID-19 screening model. *Appl. Intell.* **2021**, *51*, 3044–3051. [[CrossRef](#)]
- Jaiswal, A.; Gianchandani, N.; Singh, D.; Kumar, V.; Kaur, M. Classification of the COVID-19 infected patients using DenseNet201 based deep transfer learning. *J. Biomol. Struct. Dyn.* **2021**, *39*, 5682–5689. [[CrossRef](#)]

27. Chhabra, M.; Kumar, R. A Smart Healthcare System Based on Classifier DenseNet 121 Model to Detect Multiple Diseases. In *Mobile Radio Communications and 5G Networks, Proceedings of the Second MRCN 2021*; Springer Nature: Singapore, 2022; pp. 297–312.
28. Frimpong, E.A.; Qin, Z.; Turkson, R.E.; Cobbinah, B.M.; Baagyere, E.Y.; Tenagyei, E.K. Enhancing Alzheimer’s Disease Classification using 3D Convolutional Neural Network and Multilayer Perceptron Model with Attention Network. *KSII Trans. Internet Inf. Syst.* **2023**, *17*, 2924–2944.
29. Chauhan, T.; Palivela, H.; Tiwari, S. Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100020. [[CrossRef](#)]
30. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [[CrossRef](#)]
31. Yuan, Z.W.; Zhang, J. Feature extraction and image retrieval based on AlexNet. In *Proceedings of the 8th International Conference on Digital Image Processing (ICDIP 2016)*, Chengdu, China, 20–22 May 2016; SPIE: Bellingham, WA, USA, 2016; Volume 10033, pp. 65–69.
32. Alippi, C.; Disabato, S.; Roveri, M. Moving convolutional neural networks to embedded systems: The alexnet and VGG-16 case. In *Proceedings of the 2018 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, Porto, Portugal, 11–13 April 2018; pp. 212–223.
33. Abd Almisreb, A.; Jamil, N.; Din, N.M. Utilizing AlexNet deep transfer learning for ear recognition. In *Proceedings of the 2018 Fourth International Conference on Information Retrieval and Knowledge Management (CAMP)*, Kota Kinabalu, Malaysia, 26–28 March 2018; pp. 1–5.
34. Chen, J.; Wan, Z.; Zhang, J.; Li, W.; Chen, Y.; Li, Y.; Duan, Y. Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet. *Comput. Methods Programs Biomed.* **2021**, *200*, 105878. [[CrossRef](#)]
35. Tituriya, A.; Sachdeva, S. Breast cancer histopathology image classification using AlexNet. In *Proceedings of the 2019 4th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 21–22 November 2019; pp. 708–712.
36. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
37. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826.
38. Wang, C.; Chen, D.; Hao, L.; Liu, X.; Zeng, Y.; Chen, J.; Zhang, G. Pulmonary image classification based on inception-v3 transfer learning model. *IEEE Access* **2019**, *7*, 146533–146541. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.