Neuroscience Informatics 2 (2022) 100037

Contents lists available at ScienceDirect

Neuroscience Informatics

www.elsevier.com/locate/neuri

Original article

A hybrid scheme for AEP based hearing deficiency diagnosis: CWT and convoluted k-nearest neighbour (CKNN) pipeline

Md Nahidul Islam *, Norizam Sulaiman, Bifta Sama Bari, Mamunur Rashid, Mahfuzah Mustafa

Faculty of Electrical & Electronics Engineering Technology, Universiti Malaysia Pahang, 26600 Pekan, Pahang, Malaysia

ARTICLE INFO

Article history:
Received 8 November 2021
Received in revised form 10 December 2021
Accepted 11 December 2021

Keywords:
Brain-computer interface (BCI)
Auditory evoked potentials (AEP)
Machine learning (ML)
Deep learning (DL)
Hybridization

ABSTRACT

The auditory evoked potential (AEP) has been considered a standard clinical instrument for hearing and neurological evaluation. Although several approaches for learning EEG signal characteristics have been established earlier, the hybridization concept has rarely been explored to produce novel representations of AEP features and achieve further performance enhancement for AEP signals. Moreover, the classification of auditory attention within a concise time interval is still facing some challenges. To address this concern, this study has proposed a hybridization scheme, represented as a hybrid convoluted k-nearest neighbour (CKNN) algorithm, consisting of concatenating the convolutional layer of CNN with k-nearest neighbour (k-NN) classifier. The proposed architecture helps in improving the accuracy of KNN from 83.23% to 92.26% with a 3-second decision window. The effect of several concise decision windows is also investigated in this analysis. The proposed architecture is validated by a publicly benchmark AEP dataset, and the outcomes indicate that the CKNN significantly outperforms other state-of-the-art techniques with a concise decision window. The proposed framework shows superior performance in a concise decision window that can be effectively used for early hearing deficiency diagnosis. This paper also presents several discoveries that could be helpful to the neurological community.

1. Introduction

Hearing deficiency is the most common sensation of impairment nowadays, and over 5% of the world's population (432 million adults and 34 million children) require rehabilitation to address this issue. Hearing loss becomes more prevalent as people age; about a quarter of persons over the age of 60 suffer from hearing loss. The world health organization (WHO) estimates that the hearing impairment issue will exceed 630 million by 2030 and more than 900 million by 2050 [1]. So, addressing this deficiency is one of the major concerns; thus, early and appropriate hearing screening tests could be an effective solution for the vast population concern. Moreover, conventional hearing test techniques are time-consuming and require sufficient clinical time and expertise to interpret and maintain since it desires the person to respond directly.

Furthermore, in the application of hearing aid, the effect of hearing loss [2], the stimulus's background noise and the stimulus's position [3][4]) is still an open question. Researchers have developed several approaches [5][6][7][8] to detect early hearing ability in recent years. Among the wide range of control signals of EEG (non-invasive and user-friendly), EEG-based AEP signals have become more popular and are widely used for hearing impairment diagnosis [9]. Indeed, the AEP signal is extensively used to determine hearing capabilities, measure hearing ability, and detect neurological hearing impairments. The AEP signals elicited from the brain are acoustic stimuli and recorded from electrodes attached to the scalp. However, neural activities produced by the brain through auditory stimuli are challenging in neuroscience.

Typically, the diagnosis of hearing deficiency involves four main stages: acquisition of data, data pre-processing, feature extraction, and classification. Traditionally, feature extraction has been accomplished through the analysis of time-domain [10], frequency-domain [7], and time-frequency domain [5] approaches, whereas the classification has been conducted with different machine learning [11][12] and deep learning techniques [13]. Generally, support vector machine (SVM) [6], k-NN [12], artificial neural network (ANN) [14], linear discriminant analysis (LDA) [15] and Naïve Bayesian (NB) [16] is widely used in neurological response classification. Convolutional neural networks (CNNs) have recently become the most favourable technique for a wide variety of identification and detection applications, particularly in image classi-

* Corresponding author.
E-mail address: MEG20002@stdmail.ump.edu.my (M.N. Islam).

https://doi.org/10.1016/j.neuri.2021.100037

fication [17]. Recent research on CNNs has also yielded encouraging results in the categorization of EEG signals: in seizure detection [18], depression detection [19], and sleep stage classification [20]. Moreover, the CNN has been extensively utilized for the following purpose: Alzheimer's disease diagnosis using MR imaging [21], multiclass skin cancer classification [22], identification of coronavirus disease using X-ray images [23], detecting the character of the license plate to minimize the traffic problems [24], face mask detection [25], for recognizing the handwritten documents [26], brain tumors detection using MRI images [27], and secure Blockchain Security Module (BSM) for BCI with Multimedia Life Cycle Framework (MLCF) (BSM-BCIMLCF) that safely connects wearables while investigating the present-day BCI life cycle (BCILC) protection [28]. Besides, some other methods are also used for particular task classification, such as liver tumor detection using artificial intelligence-based *K*-means clustering (KMC) algorithm [29] and recognition of Alzheimer's patients using support vector machine (SVM) [30].

Deckers et al. [13] proposed an auditory attention detection (AAD) system where they used a neural network algorithm to decode the locus of interest for the competing speaker scenario. They achieved 76.1% accuracy on 1 s of the decision period in the experiment. However, this CNN-based approach has a high degree of inter-subject variability and involves vast amounts of training data for each subject to form a subject-independent decoder. In [31], short window lengths during dynamic change are demonstrated to be beneficial, even if they have a low precision in the context of robust AAD-dependent gain controls. With a shorter decision-making window, particularly below 10 s, the AAD accurateness drops significantly [32], [33]. In terms of EEG-based AAD, Ciccarelli et al. [34] recently showed that a (subject-dependent) CNN based classifier could outperform the linear methods for the decision windows of 10 s. Zhang et al. [5] proposed an auditory brainstem response (ABRs) identification system. They used the proposed combined system using wavelet analysis and Bayesian networks for classification purposes. Here, five-level discrete wavelet transforms (DWT) extract the features by thresholding and matching the wavelet coefficients. The classification accuracy (78.80% for the 64 averaged ABRs and 84.17% for the 128 averaged ABRs) is insufficient for the system implementation in a real-life application. Gronfors [35] developed an automatic peak detection system using the pedestal peak method. This study shows the affection of performance with different dB stimulation intensity. McCullagh et al. [14] reported a 73.7% testing accuracy using the artificial neural network to classify 166 auditory brainstem responses (ABRs) with 2000 repetitions which is insufficient to use the approach in real-time implementation. McKearney et al. [36] proposed a deep neural network approach to classify auditory brainstem responses. Although they achieved an excellent performance (92.9% testing accuracy), the testing set is too small (42 paired waveforms), and more dataset is required to test the robustness and applicability of their approach.

The proposed study aims to develop an intelligent system for early hearing disorder diagnosis in a short time interval. Although the concise decision window is considered one of the prerequisites to developing real-time and real-life applications, relatively a few research studies have explored this concept [13]. The concise decision window reduces the impact of additional features and computational complexity for the additional feature. Moreover, the shorter decision window makes the system faster, which will help to detect the hearing disorder earlier. Despite the several advantages of selecting the concise decision window, the hearing screening strategy is still very challenging [31]. The shorter decision windows contain less information and sometimes provide inadequate accuracy [13] [31] [33].

For the hearing condition diagnosis, some recent studies [12] [37] [38] [39] have widely used the K-NN algorithm. Despite some key benefits of the k-NN algorithm in the neurological field, including few parameters and a faster training process, several limitations exist. Firstly, the k-NN algorithm does not work well with a high dimensional dataset since, in such a case, it becomes difficult for the algorithm to calculate the distance in each dimension. The cost of calculating the distance between the new and existing points is vast, degrading the algorithm's performance. Moreover, feature scaling (standardization or normalization) is needed before applying the k-NN algorithm to any dataset; otherwise, k-NN may generate wrong predictions. To address the limitations of some single classifier, integrating two or more classifiers could be an effective alternative to perform error-free classification tasks rather than a single classifier. For this purpose, the researchers have developed several hybrid techniques where they combine more than one algorithm. In [40] [41] [42] [43], the researchers have developed the hybrid architectures and effectively concatenated two algorithms strength which provides a high performance than a single classifier, and in some cases, it helps to overcome the single classifier limitations. This paper proposes a novel hybrid architecture with time-frequency images which can detect the hidden local information in neural activity to reflect this concept. To overcome the mentioned issues of k-NN, the hybrid architecture is built with CNN's convolutional blocks and the k-NN algorithm. This hybridization's objective is that several hidden layers of the blocks can select discriminatory representations from the higher dimensional data. This process helps the k-NN algorithm to provide more precise predictions by reducing the mentioned issues of k-NN. The key contributions of the paper are summarized as below:

- In the proposed approach, we have designed a hybrid architecture (CKNN) based on concatenating the convolutional layers of CNN with the k-NN algorithm, which help in improving the overall performance of the k-NN algorithm for the AEP dataset.
- To mitigate the impact of the additional features, this study has investigated concise decision windows (1 s, 2 s, and 3 s), which help in reducing the time consumption and show the robustness and applicability of the proposed system in real-time implementation.
- Furthermore, the dataset has been tested with six popular transfer learning algorithms, including InceptionResNetV2 [44], MobileNet [45], ResNet50 [46], VGG16 [47], VGG19 [47] and Xception [48]), in which the proposed hybrid architecture shows superior performance to detect the early hearing ability than other existing approaches.

The rest of the manuscript is arranged as follows: Section 2 presents the detailed data description, data pre-processing, transformation process of raw signals to time-frequency images, and the proposed hybrid architecture building procedure. The experimental result to test the validation of the model is represented in Section 3. Section 4 exhibits a comprehensive discussion and compares the proposed model with related studies. Section 5 describes the outcomes with the future directions of the proposed research.

2. Materials and methods

2.1. Data description

To verify the effectiveness of the proposed hybrid architecture for hearing condition classification, systematic experiments have been conducted on a publicly available online dataset (provided by ExpORL, Dept. Neurosciences, KULeuven, and Dept. Electrical Engineering (ESAT), KULeuven) [49]. This signals acquisition technique used a 64-channel BioSemi Active Two system with a sampling rate of 8196 Hz. The dataset included 16 normal-hearing subjects, each with 20 repetitions in which every story lasted 12 minutes

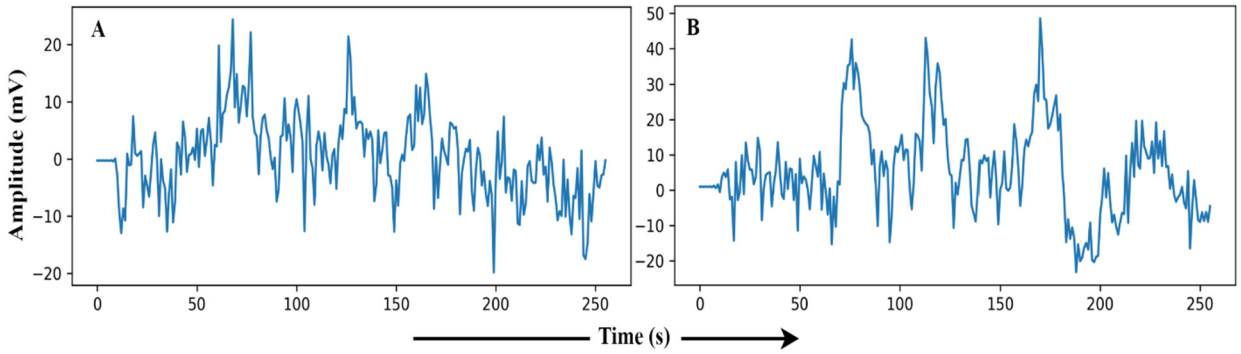

Fig. 1. AEP raw data plotting in 2 s time window: (A) hear auditory stimulus with the left ear (B) hear auditory stimulus with the right ear.

and was divided into two segments of 6 minutes each. For providing the sound stimuli, an Etymotic ER3 insert earphone was provided where the sound intensity was 60 db. The entire data capturing procedure was conducted in a soundproof, electromagnetically shielded space. Simulation software (APEX3) was used in this procedure [50]. Three male Flemish speakers had participated where they narrated four Dutch stories as auditory stimulation [51]. Each story lasted 12 minutes and was divided into two segments (six minutes each). The stimuli had identical root-mean-square intensities and were assessed as equally loud. In each trial, two parts of two storylines had been provided for the subjects where the left ear received one part of the storylines, and the right ear received the other part. The attended ear was alternated over successive trials to ensure that each ear received an equal volume of stimuli, preventing the lateralization bias [4]. So, every subject received stimuli in the same order, either dichotically or after head-related transfer function (HRTF) filtering (simulating sound coming from ±90°). The HRTF/dichotic condition was randomized and equally distributed within and over participants as with the attended ear.

2.2. Data pre-processing

The initial phase following data collecting is pre-processing the AEP data. Each trial was filtered with a high pass frequency (0.5 Hz cut off) and downsampled from the sampling rate of 8192 Hz to 128 Hz in the pre-processing phase. In the experiment, we have studied sixteen participants, and each trial was divided into the same duration. To investigate the robustness of the proposed architecture, the dataset has been segmented into three different short decision windows (1 s, 2 s, and 3 s). The reason for selecting the shorter decision windows is to reduce the computational complexity, minimize the use of additional features for a single decision, and make the system faster. In the experimental analysis, we have randomly picked 200 observations from each decision window, and a total of 3200 observations have been achieved from each window. Fig. 1 illustrates the filtered AEP data of subject-1, channel-1, in the 2 s decision window, whereas the subject hears auditory stimulus through headphones defined as left and right labels.

2.3. CWT for time-frequency analysis

CWT is a technique for extracting time-frequency features that enable multi-scale signal refining via scaling and translation procedures. After the pre-processing phase, the segmented dataset transforms from the time domain to the time-frequency domain using the CWT [52]. In this study, CWT has been utilized as a feature extraction method that converts the AEP raw signal into time-frequency images. The wavelet transforms execute an internal signal action and generate a series of wavelets. The mother wavelet is scaled and translated to create the wavelet set, which is a family of wavelets $\psi(t)$, shown as

$$\psi_{S,\tau}(t) = \frac{1}{\sqrt{S}}\psi\left(\frac{t-\tau}{S}\right) \tag{1}$$

Here, S represents the scale parameter inversely related to frequency, and τ represents the translation parameter. A CWT of a signal $x(t)$ can be achieved by a complex conjugate convolution operation, mathematically defined as follows [53]:

$$W(s,\tau) = \langle x(t), \psi_{S,\tau} \rangle = \frac{1}{\sqrt{s}} \int x(t)\, \psi^*\left(\frac{t-\tau}{S}\right) dt \tag{2}$$

Where $\Psi^*(\cdot)$ denotes the complex conjugate of the above function $\Psi(\cdot)$ and this operation decomposes the signal $x(t)$ in a series of wavelet coefficients, in which the base function is the wavelet family. In the equation, the s and τ are two types of parameters in the family wavelets. The signal $x(t)$ is transformed and projected to the time and scale dimensions of the family wavelets.

In this study, we used wavelet basis functions (Mother Wavelets). The time-frequency images are then used as the input of the proposed hybrid architecture. The transformation process of CWT is shown in Fig. 2.

Since the dataset was collected by 64-channel BioSemi Active Two system, we first segmented the dataset in different (1 s, 2 s, and 2 s) and then converted the raw AEP signals into time-frequency images. Finally, we have concatenated the 64 channels data in (M*M) square in a single image, where the value of M has been set to 8. So, each image provides the time-frequency information of 64 channels. Fig. 3 shows the time-frequency image of 64 channels.

2.4. The proposed hybrid CKNN scheme

The staggering growth of hearing impairment poses a significant threat and creates a persuasive need for automatic classification techniques. In this paper, we comprise two algorithms (convolutional layers and the k-NN) in an effective way to enhance the early hearing impairment detection ability. The proposed hybrid system overcomes the existing limitations of k-NN, including lower efficacy in high dimensions, need feature scaling, high sensitivities to noisy data, missing values, and outliers. This study used two convolutional blocks with different kernel sizes to significantly deal with these issues before the features input at k-NN.

The proposed hybrid architecture is composed of two convolution layers and a k-NN algorithm. The convolutional layers consist of several types of kernels, pooling layers and one dropout layer. The convolution layer is the fundamental building element of the CNN, and it is responsible for preserving the connection between pixels by learning image features from tiny squares of input data. Convolution is a linear process in which weights are multiplied by

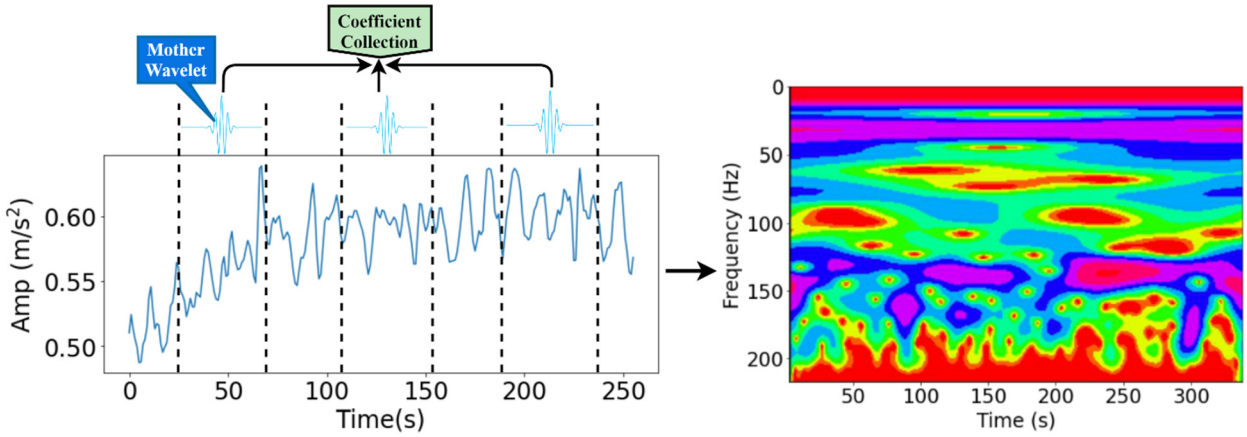

Fig. 2. The transformation process from time-domain signal to time-frequency domain image.

the input. This layer conducts a dot product on two matrices, one of which contains the set of learnable parameters referred to as the kernel, and the other includes the restricted portion of the receptive field. Kernels are geographically smaller than images but have a greater in-depth. This implies that if the picture has three (RGB) channels, the kernel's height and width will be smaller in comparison to the depth, which will span all three channels. Two inputs are required for a mathematical action, such as an image matrix and a filter or kernel. The image matrix dimension is $(h \times w \times d)$, and after applying the filter, the dimension is $(f_h \times f_w \times \mathrm{d})$. The output dimension of the matrix is $(h - f_h + 1) \times (w - f_w + 1) \times 1$.

Several fixed-size filters in the convolution layer allow the complex functions to be used in the input image [54]. This is done by sliding the locally trained filters over the image. Each filter is applied to the entire image with the same weight and bias values. The process is called the weight-sharing mechanism, and it enables the representation of the whole image with the same characteristic. Considering an l-layers CNN architecture, the l-layer convoluted feature maps can be expressed as:

$$y_{l,j}^{conv} = \sum_{i}^{k} W_{i,j}^{l} * y_{l-1,i}^{pool} + b_j^l \tag{3}$$

$$y_{l,j}^{ReLU} = f\left(y_{l,j}^{conv}\right) = max\left[0, y_{l-1,j}^{conv}\right] \tag{4}$$

Where $y_{l,j}^{conv}$ represents the output of the l-th layer, and W^l is the kernel of the l-th layer, k represents the number of kernels, b^l is the bias and $f(\cdot)$ is the activation function which transforms the input to the output map to increase the nonlinear property. An activation function ReLU (Rectified Linear Unit) expressed as $\max(0, x)$ is usually used instead of a sigmoid. This study utilized the ReLU function, and it proves far better outcomes in most of the classification tasks, accelerating convergence and alleviating the vanishing gradient problem [55].

The next stage is the application of the pooling layer to decrease the image stack. The straightforward reason for using this layer is to reduce the spatial size of the input image. By reducing the size and calculation of the input image feature, pooling regulates the overfitting issue. This study used the one dropout layer after the first pooling layer. The reason behind adding the dropout layer is to minimize the overfitting problem and reduce the model complexity. Mele et al. [56] reported that in the case of the CIFAT-10 dataset, the error rate was 16.6% when CNN had been employed. They improved the model's performance with an error rate of 15.6% when the dropout layer was utilized in the last hidden layer. The neuron is temporarily dropped with the probability p at each iteration. Then, at every training step, the dropped-out neuron is resampled with the probability p, and a dropped-out neuron will be active at the next step. Here, the hyperparameter p is the dropout rate. Prior to KNN functionality, all the features are flattened, which is a suitable size for the k-NN algorithm.

The k-NN algorithm is a machine learning approach that classifies a data point based on the majority of its neighbours [57]. The k-NN method executes in two steps; first, it determines the value of the nearest neighbour, and then, using the value nearest neighbour, it classifies the data point into a certain class. K is the number of nearest neighbours in k-NN. The k-NN algorithm is fundamentally determined by the number of neighbours (k). The k-NN method has a disadvantage in terms of determining the value of k. Although several studies have been conducted on this issue, selecting the k value for the k-NN method remains difficult and challenging [58]. This study utilized the 'Gridsearchcv' function in Python to address this issue, where we set the value of k as [range (1-15)]. This strategy helps in finding the optimal value of k for our task. This method calculates the distance between them using distance measures such as Euclidean, Hamming, Manhattan, and Minkowski to determine the nearest comparable points. For selecting the optimal distance metric, this approach used metric is [Euclidean, Manhattan] with the help of the 'Gridsearchcv' function. This study achieved the maximum performance with the K value of 10, Euclidean distance, and uniform weight.

The Euclidean distance is defined as the length of the line segment connecting the two instances. It is determined by squaring the difference between the points' x and y coordinates. Suppose $X = (X_1, X_2, \ldots, X_N)$ and $Y = (Y_1, Y_2, \ldots\ldots, Y_N)$ are two points in Euclidean N-space, then the distance from X to Y is calculated by the following equation,

$$Euclidean\ distance: (X, Y) = \sqrt{\sum_{i=1}^{N} (X_i - Y_i)^2} \tag{5}$$

Fig. 4 illustrates the detailed diagram of the proposed hearing loss diagnosis system. First of all, the raw AEP data is transformed into the time-frequency image with the help of CWT.

Then, the images are resized into 224 × 224 × 3. In the first convolutional layer, the kernel size is 5 × 5, where we have used 32 kernels. In the next step, this study has used the Max-pooling layer to reduce the size of the input. Max pooling is used to pass the window through a filtered image that is the output of a previous convolution layer, storing the window's maximum value. The pooling layer reduces the image size from 224 × 224 × 32 to 74

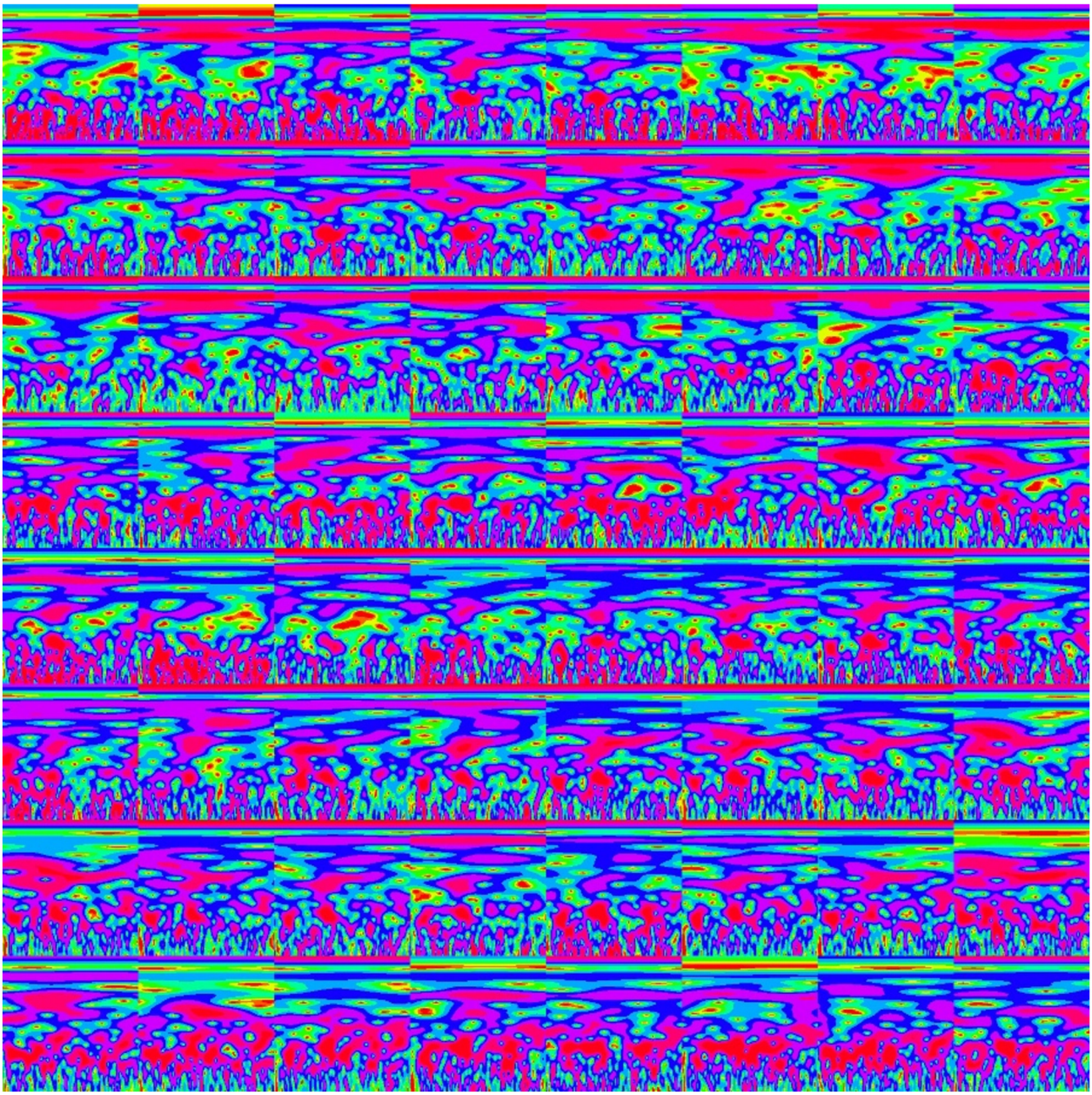

Fig. 3. The time-frequency image of 64 channels.

Table 1
The proposed hybrid architecture building procedure.

Step-1	Design some convolutional blocks with suitable parameters (used to filter the given time-frequency images). This study utilized the 'Keras-Tuner' library, which helps to select the most optimal set of parameters to build the blocks.
Step-2	Add the Flatten layer after the last convolutional block, which helps in preparing the suitable input size for the k-NN model (as shown in Fig. 4).
Step-3	Build a k-NN model with suitable hyperparameters for the task classification (this study used the auto hyperparameter tuning process with the help of the 'GridSearchCV' library to figure out the suitable hyperparameter).
Step-4	Integrate the convolutional blocks with the k-NN model to build the proposed hybrid architecture (CKNN).
Step-5	Train the proposed hybrid architecture with the dataset.

$\times$ 74 $\times$ 32. Then one dropout layer is added after the pooling layer, where the value is set to 0.6. After that, another convolutional layer with the kernel size 3 $\times$ 3$\times$ 64 is combined. Next, a 3 $\times$ 3 Max pooling layers are added, where the output size is 24 $\times$ 24 $\times$ 64. After flattening the features, those features are used as the input of k-NN. The proposed hybrid architecture building procedure is illustrated step-by-step in Table 1.

3. Results

This experimental analysis has been carried out in Python, with the help of Google Colab, Windows 10, Intel(R) Xeon(R) CPU @ 2.30GHz, Tesla K80, and CUDA Version 10.1. The experiment is conducted on the publicly available reputed AEP dataset collected from 16 subjects (S) to verify the proposed architecture's effective-

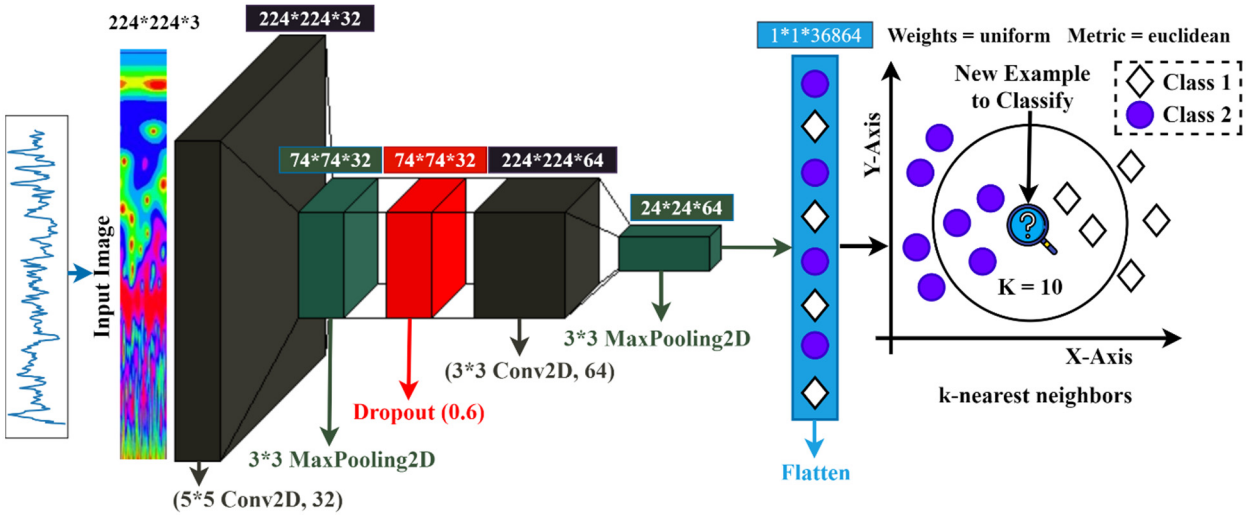

Fig. 4. Schematic diagram of proposed hybrid Classification System.

ness. This study examines three distinct decision windows (1 s, 2 s, and 3 s), which is considered as the shortest time window among the state-of-art methods. This term refers to the amount of information necessary to make a single left/right choice. The practical reason for selecting the concise decision windows (DW) is to decrease computing complexity and speed up the system that aids to detect early hearing disorders. We randomly selected 200 observations from each subject with each decision window and finally achieved 3200 observations to conduct the experiment. Then, the entire dataset is randomly split, and 70% dataset has been used to train the architecture. In contrast, the rest 30% dataset has been used to test the robustness of the proposed architecture for hearing impairment diagnosis. A total of 2240 observations has been used for training the model and 960 observations for testing the performance. Then, the observations have been converted into time-frequency images using CWT. Finally, the extracted features have been classified by the k-NN and the CKNN architectures.

3.1. Performance of this experiment over the decision windows

To evaluate the performance of AEP for hearing condition classification, this study has been conducted with three different concise decision windows. The performance of the proposed system for hearing condition diagnosis is shown in Table 2, which illustrates the effect of decision windows on the overall performance. To demonstrate the proficiency and robustness of our system, we evaluate several types of performance measurement techniques (accuracy, precision, recall, f1 score, and Cohen's kappa score) [59][60] for each subject (S). To check the performance of our architecture, the performance reported in Table 2 is based on ten different runs of the architecture.

Among the different decision windows, the 1 s decision window's performance is lower. In the decision windows (2 s and 3 s), the proposed architecture shows higher average accuracy and larger inter-subject variability. In all decision windows, two subjects (subjects 9 and 14) performed poorer than the others (therefore not considered in the subsequent analysis). As reported in Table 2, our proposed architecture outperforms the k-NN algorithm in all decision windows (10% improvement in 1 s decision window, 6% improvement in 2 s decision window, and 6.87% improvement in 3 s decision window). Therefore, further analysis has not discussed the performance of the k-NN algorithms.

For 1 s decision window, the average (AVG) performance of the system, including accuracy, precision, recall, f1 score, and Cohen's kappa score are 88.21%, 92.98%, 84.10%, 87.05%, 76.56%, respectively (avoid two subjects that performed poorer). The performance achieved with this decision window is not enough to use this system in real-time applications. Before we move on to the next decision window, the possible reason for the unsatisfactory performance with this concise decision window has been investigated. Firstly, the shorter decision window contains less information, and sometimes this few information is not enough to make the decision. Secondly, if the algorithms do not get enough information, it provides unsatisfactory outcomes. For this reason and to enhance the performance, we move on to the 2 s decision windows.

For the 2 s decision window, we have achieved the average value of accuracy, precision, recall, f1 score, Cohen's kappa score is 91.95%, 93.30%, 91.27%, 92.12%, 84.56%, respectively where 3.74%, 0.32%, 7.17%, 5.07%, and 8% improvement has been achieved, respectively, compared to the 1 s decision window length. Despite the impressive performance of the proposed system with a 2 s decision window, some subject's analysis shows unsatisfactory performance, such as comparatively lower classification accuracy has been noticed by Subjects-3 (86.67%), Subject-10 (88.33%), Subject-12 (85%), Subject-16 (88.33%). However, the classification accuracy of other subjects is more than or equal to 90%. Some impressive performance has been noticed by Subject-6 (100%), Subject-7 (95%), Subject-8 (95%) and Subject-15 (95%). To study the window effect more, the 3 s decision window has also been investigated with the proposed system; however, the emphasis of our study is on a concise decision window.

For the 3 s decision window, in the average of all the subjects, the proposed CKNN has achieved an accuracy, precision, recall score, f1 score and Cohen's kappa score of 92.26%, 92.57%, 92.86%, 92.51%, and 84.50% respectively. Firstly, in this case, we have improved 4.05% accuracy, 0.41% precision, 8.76% recall, 5.46% f1 score and 7.94% Cohen's kappa score than 1 s decision window.

Secondly, the improvement has been achieved compared to the 2 s decision window with 0.31% accuracy, 0.73% precision, 1.59% recall, 0.39% f1 score, and 0.06% Cohen's kappa score. From this comparison, it is clear that we have achieved a significant improvement with the 3 s decision window compared to the 1 s decision window. On the other hand, compared to the 2 s decision window, the improvement is not so high. For this reason, further analysis is not conducted with more decision window; however, the main goal of this experiment is early hearing condition diagnosis. The concise decision window with high-performance help to meet this achievement.

Table 2
The effect of decision windows on the overall performance.

DW	Method	Performance (%)	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16	AVG
1 s	KNN	Accuracy	76.67	88.33	65	91.67	80	65	88.33	75	51.67	73.33	70	91.67	75	65	86.67	55	74.90
		Precision	79.31	96.15	65.62	90.62	88	65.62	85.29	86.36	54.17	94.12	74.07	93.33	80.77	77.78	79.49	83.33	80.88
		Recall	74.19	80.65	67.74	93.55	70.97	67.74	93.55	61.29	41.94	51.61	64.52	90.32	67.74	45.16	100	16.13	67.94
		F1 Score	76.67	87.72	66.67	92.06	78.57	66.67	89.23	71.70	47.27	66.67	68.97	91.80	73.68	57.14	88.57	27.03	71.90
		Cohen Kappa	53.39	76.77	29.84	83.30	60.22	29.84	76.56	50.44	3.970	47.43	40.20	83.33	50.22	30.92	73.06	12.34	50.11
1 s	CKNN	Accuracy	95	96.67	86.67	96.67	93.33	70	98.33	93.33	63.33	88.33	80	98.33	81.67	73.33	91.67	65	85.73
		Precision	96.67	100	89.66	96.77	96.55	70.97	96.88	100	80	100	82.76	96.88	88.46	80	86.11	100	91.36
		Recall	93.55	93.55	83.87	96.77	90.32	70.97	100	87.10	38.71	77.42	77.42	100	74.19	64.52	100	32.26	80.04
		F1 Score	95.08	96.67	86.67	96.77	93.33	70.97	98.41	93.10	52.17	87.27	80	98.41	80.70	71.43	92.54	48.78	83.90
		Cohen Kappa	90	93.34	73.36	93.33	86.68	39.93	96.66	86.71	27.87	76.82	60.04	96.66	63.50	46.96	83.22	31.52	71.66
2 s	KNN	Accuracy	88.33	83.34	78.33	91.67	78.33	96.67	93.33	93.33	78.33	78.33	88.33	71.67	83.34	80	95	76.66	85
		Precision	85.29	100	80	88.24	80	93.94	100	100	78.13	84.62	90	76.92	76.93	88	91.18	81.48	87
		Recall	93.55	67.75	77.42	96.77	77.42	100	87.10	87.10	80.65	70.97	87.10	64.52	96.77	70.97	100	70.97	83
		F1 Score	89.23	80.77	78.69	92.31	78.69	96.88	93.10	93.10	79.37	77.19	88.52	70.18	85.71	78.57	95.38	75.86	85
		Cohen Kappa	76.56	67	56.67	83.26	56.67	93.31	86.71	86.71	56.57	56.86	76.67	43.58	66.33	60.22	89.96	53.49	69
2 s	CKNN	Accuracy	93.33	91.67	86.67	96.67	90	100	95	95	86.67	88.33	91.67	85	90.67	85	95	88.33	91
		Precision	100	100	84.85	93.94	90.32	100	100	100	82.86	90	93.33	86.67	90.67	100	91.18	85.29	93
		Recall	87.10	83.87	90.32	100	90.32	100	90.32	90.32	93.55	87.10	90.32	83.87	90.67	70.97	100	93.55	90
		F1 Score	93.10	91.23	87.50	96.88	90.32	100	94.92	94.92	87.88	88.52	91.80	85.25	90.67	83.02	95.38	89.23	91
		Cohen Kappa	86.71	83.41	73.24	93.31	79.98	100	90.02	90.02	73.18	76.67	83.34	70	90.67	70.26	89.96	76.56	83
3 s	KNN	Accuracy	91.67	91.67	86.67	95	85	86.67	91.67	83.33	66.67	81.67	83.33	85	78.33	71.67	86.67	66.67	83.23
		Precision	90.62	100	84.85	96.67	84.38	81.08	100	81.82	76.19	88.46	88.89	95.83	87.50	85	89.66	72	87.68
		Recall	93.55	83.87	90.32	93.55	87.10	96.77	83.87	87.10	51.61	74.19	77.42	74.19	67.74	54.84	83.87	58.06	78.63
		F1 Score	92.06	91.23	87.50	95.08	85.71	88.24	91.23	84.38	61.54	80.70	82.76	83.64	76.36	66.67	86.67	64.29	82.38
		Cohen Kappa	83.30	83.41	73.24	90	69.93	73.12	83.41	66.56	33.99	63.50	66.78	70.20	56.95	43.96	73.36	33.70	66.59
3 s	CKNN	Accuracy	100	96.67	88.33	98.33	95	90	95	86.67	68.33	95	85	93.33	91.67	81.67	90	86.67	90.10
		Precision	100	100	85.29	100	91.18	83.78	100	84.85	77.27	93.75	89.29	90.91	90.62	91.67	90.32	96	91.56
		Recall	100	93.55	93.55	96.77	100	100	90.32	90.32	54.84	96.77	80.65	96.77	93.55	70.97	90.32	77.42	89.11
		F1 Score	100	96.67	89.23	98.36	95.38	91.18	94.92	87.50	64.15	95.24	84.75	93.75	92.06	80	90.32	85.71	89.95
		Cohen Kappa	100	93.34	76.56	96.67	89.96	79.84	90.02	73.24	37.22	89.98	70.07	86.62	83.30	63.58	79.98	73.48	80.24

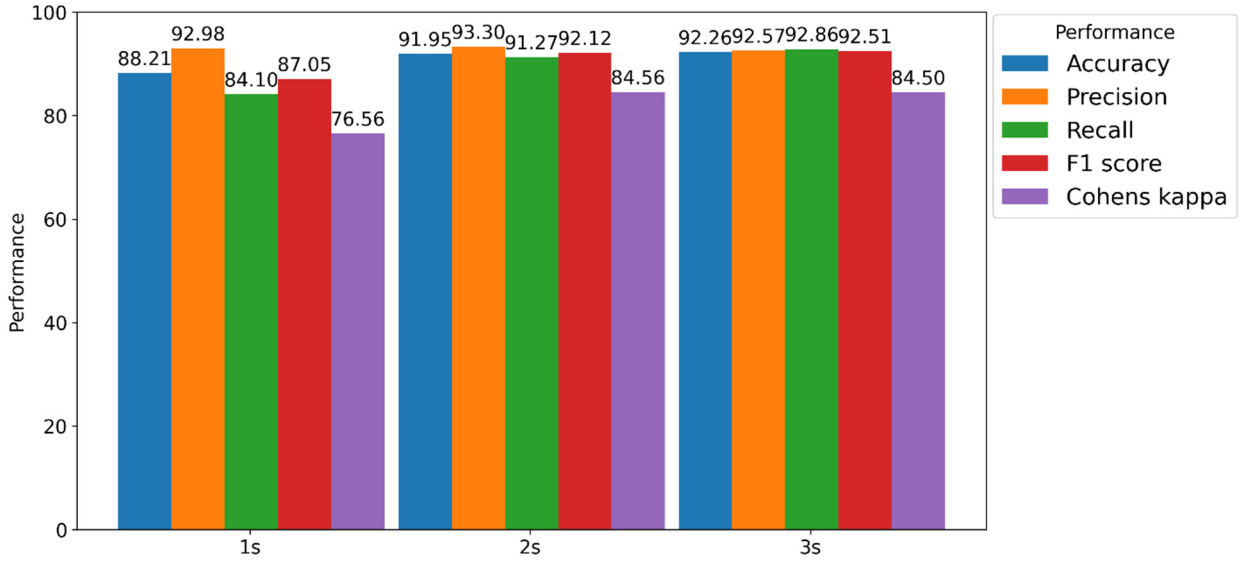

Fig. 5. The effect of decision windows over the proposed architecture.

Despite the significant performance of our system, some subjects' analysis provides unsatisfactory performance. There are several possible reasons for the poorer performance. For example, EEG-based AEP signals come from the brain area, and the signals are non-linear, non-stationary and have artifact prone characteristics. So, it is challenging to achieve high performance with the dataset. Sometimes, the brain produces unwanted noisy signals due to the lack of the subject's attention or muscle movement. Since the signal acquisition procedure (publicly available dataset) was conducted in a soundproof place, there is a low chance of external artifacts. Another factor that was selecting a concise decision window contains less information may decrease the overall performance [13] [31] [33]. However, a concise decision window is one of the prerequisites factors for developing the real-time application, but a few studies have been conducted with a shorter decision window [13]. It reduces the computational complexity for the additional features and makes the system faster, but it is still very challenging to achieve adequate performance [31]. These phenomena could be responsible for the lower performance in some cases.

However, the proposed system achieved an impressive performance with a shorter decision window. The experimental outcomes show that it is possible to decode the hearing condition with concise decision windows. This study has presented a comparison to analyze the effect of different decision windows on overall performance. Fig. 5 visualizes the average performance of our system with different decision windows.

3.2. Confusion matrix

When the classifier deals with multiple classes of similar features, it can be confused in some cases. A confusion matrix can be used to estimate the classification accuracy of a model visually. Based on the auditory stimuli in the ear, two types of AEP data have been used to conduct this experiment. When the subject hears auditory stimuli in the left ear is denoted Class1, and when the subject hears auditory stimuli in the right ear is indicated Class2, as shown in Fig. 6. Since the AEP signals are produced from the brain area and are very unstable, sometimes it produces noisy signals due to the artifacts or muscle movement, making the classifier confused in some cases. Fig. 6 illustrates the confusion matrix of our proposed architecture with the AEP dataset in the 3 s decision window.

All correct predictions are on the diagonal, and all incorrect predictions are off the diagonal; hence, the classes that have confused the system. Based on these results, the classifier's performance can be visually evaluated. According to the analysis, the KNN algorithm correctly predicts 813 observations out of 960 observations, whilst the proposed CKNN algorithm correctly predicts 872 observations out of the same number of observations. So, the confusion matrix also indicates that the proposed CKNN architecture is more accurate than the k-NN and helps to overcome the limitations of k-NN algorithms.

3.3. Comparison of pre-network recognition accuracy

Furthermore, to study the robustness and applicability of the proposed architecture, this experiment is conducted with six popular widely used transfer learning (TL) architectures, namely, InceptionResNetV2 [44], MobileNet [45], ResNet50 [46], VGG16 [47], VGG19 [47] and Xception [48]). In the comparison, we only consider the 3 s decision windows (provide the best performance among other decision windows). For the fair comparison, the input sizes are the same (height-224 * width-224 * depth-3) as the proposed architecture input so that these architectures get the exact dimension features illustrated in Table 3. During the training procedure, a popular 'Adam' optimizer is used to determine the set of weights and biases of the neural network that reduce the loss function. In this study, we set the batch size of 64, and the learning rate is 0.0001, which is small in size but help to achieve more precise results. We replace the input and output layer in all the transfer learning architectures based on the targeted class. The results reported here are based on 100 epochs, and accuracy is reported on the best outcomes of five different runs of all the TL architecture. The TL architecture used pre-trained 'ImageNet' weights for the hearing condition diagnosis. As seen in Fig. 7, the accuracy curve is presented along the X-axis with the number of epochs and the associated testing accuracy on the Y-axis.

Furthermore, the testing accuracies of various pre-trained networks with our proposed architecture are represented in Fig. 8, whereas Table 3 demonstrated the recognition accuracy along with the suitable input size of all architectures. During the transformation procedure (1D signals to 2D time-frequency images) using CWT, we resized all the images into 224 * 224 * 3 which is the suitable input size of all the TL architecture. As illustrates in Fig. 8, the height testing accuracy of 60.25% is achieved with the Mo-

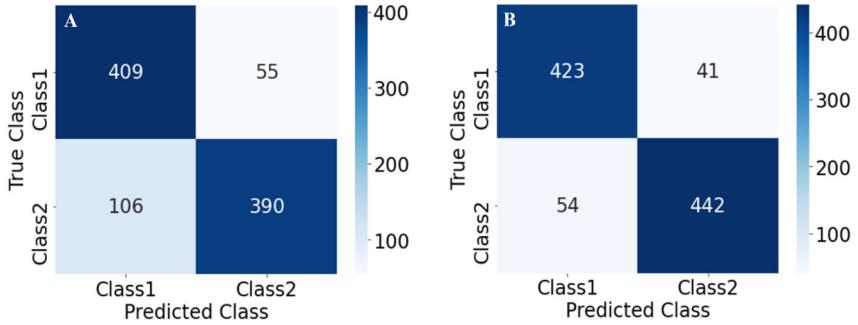

Fig. 6. Confusion matrix A) KNN approach B) CKNN approach.

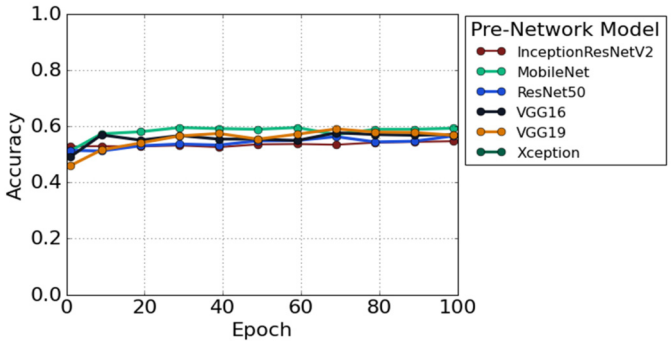

Fig. 7. The accuracy curve of six pre-trained architectures.

Table 3
The performance comparison with six pre-trained models.

Algorithm	Input size	Recognition accuracy
InceptionResNetV2	224 × 224 × 3	54%
MobileNet	224 × 224 × 3	60.25%
ResNet50	224 × 224 × 3	54.88%
VGG16	224 × 224 × 3	57.38%
VGG19	224 × 224 × 3	56.63%
Xception	224 × 224 × 3	57.63%
Proposed approach (CKNN)	**224 × 224 × 3**	**92.26%**

bileNet architecture. The rest of the architecture provides less than 60% accuracy, whereas our hybrid network provides 92.26% recognition accuracy; improved more than 30% accuracy compared to the other TL architectures. The most straightforward reason behind the poorer accuracy of these TL architectures is that the TL architecture trains with the natural images, and our time-frequency images are not the same, whereas the proposed architecture trains from scratch (train with time-frequency images). On the other hand, this study has efficiently developed an integrated technique to integrate the convolutions block with a k-NN algorithm. Some researchers reported that the integration of two algorithms can empower the recognition performance for the specific task classification [40][41][42][43] and can overcome the limitations of the single classifier. For this reason, our architecture has achieved state-of-the-art performance for hearing condition detection.

4. Discussion

In this paper, a hearing condition diagnosis system based on time-frequency images and hybrid architecture (CKNN) is proposed. The experimental outcomes show that it is possible to decode the hearing condition within 2 s to 3 s, with an average accuracy of 92.26%. This section represents the impact of different parameters of the proposed hybrid architecture, making the parameter's selecting procedure clearer. Finally, related work of hearing disorder diagnosis has been included in this section which helps to verify the applicability and effectiveness of the proposed architecture.

4.1. Parameter's selection

This study has used two convolutional layers, consisting of 32 kernels with the size of (5×5) and another one is 64 kernels with the size of (3×3). The convolution layer ensures the spatial relationship between pixels by learning image features using small squares of input data. After every convolutional layer, we adopt two (3×3) max-pooling layers, which is a down-sampling strategy and help to reduce the size of the extracted feature maps. The straightforward reason behind adding the max-pooling is to reduce the number of parameters and computational cost in the network. Max-pooling provides fundamental translation invariance to the internal representation and reduces the massive amount of time for training. To reduce the model's complexity dropout regularization technique has been added where randomly selected neurons are ignored during training. In [56], the error rate was improved to 15.6% from 16.6%, while the dropout layer was added in the last hidden layer. Basically, the dropout resolved the overfitting issue in large networks. Dropout compels a neural network to acquire more robust characteristics, which are beneficial when combined with a large number of distinct random subsets of the other neurons. The major benefit of this approach is that it prevents all neurons in a layer from improving their weights simultaneously. This adaptation, which is carried out in random groupings, prevents all neurons from converging on the same objective, therefore decorating the weights [56]. This study has been used the dropout rate at $p = 0.6$. This strategy helps in reducing the model's complexity, computational time and prove a more robust architecture to identify the hearing deficiency diagnosis. Another key contribution of this study is to tune the hyperparameters. Hilde J.P. et al. [61] experimented with 59 datasets, and in this analysis, they showed the importance of tuning hyperparameters of machine learning algorithms. To classify the finally extracted features from convolutional architecture, we have used the KNN algorithm. Most machine learning algorithms' performance highly depends on the hyperparameters setting. Manually selecting hyperparameters is very time consuming for the analysis, and any wrong selection may hamper the architecture's performance. So, to obtain the optimal hyperparameters for the hearing deficiency task classification, this study has used the hyperparameters tuning strategy with the help of the 'Gridsearchcv' function. In this study, we have passed the predefined values for the hyperparameters to the 'Gridsearchcv' function, and this function tries all the combinations of the values. Then, the function evaluates the model for each combination and provides the most optimal hyperparameters which help in improving the model's performance for the task classification. In the k-NN algorithm, selecting the most optimal k value is still challenging [58].

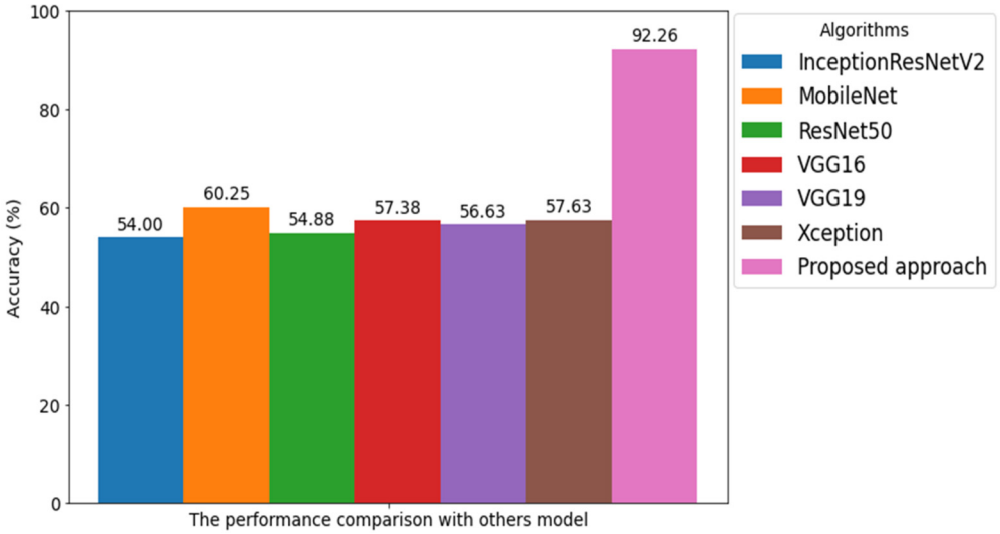

Fig. 8. The accuracy comparison with other popular models.

For this reason, this study optimizes the k value (1 to 15). Another hyperparameter of k-NN is choosing the distance metric. This study has been conducted with two widely used distance matric from the several types of distance matric (Euclidean, Manhattan). The maximum accuracy for all the subjects has been achieved with the K value of 10, Euclidean distance, and uniform weight. Finally, the convolutional blocks have been integrated with the KNN algorithm for building the proposed architecture (CKNN). So, the following strategies mentioned above help to achieve the optimal parameters for this task classification and enhance the overall performance of the proposed architecture.

4.2. Related work

In recent years, through the development of computer technology, machine learning has been widely utilized to train and diagnosis early hearing deficiency and has become a satisfactory alternative for the detection of hearing deficiency. Previous efforts have been made to detect early hearing disorders using several machine learning or deep learning approaches.

Gronfors [35] developed an automatic peak detection system and advanced pattern recognition using the pedestal peak method. They used 212 ABRs recorded from 32 anonymous subjects and achieved a maximum of 86% testing accuracy. Hoppe et al. [62] performed an automatic sequential recognition method for cortical auditory evoked potentials signals. They used several mother wavelets functions, whilst they did not find any significance in the recognition accuracy. Effern et al. [63] conducted a single-trial analysis of event-related potentials. Their study did not specify which mother wavelet was used in their experiment. McCullagh et al. [14] reported 73.7% accuracy using the ANN algorithm to classify 166 auditory brainstem responses (ABRs) with 2000 repetitions. Habraken et al. [64] also used an ANN-based system with a feature extraction algorithm and achieved an accuracy of about 80%. Alpsan and Ozdamar [65] reported accuracy of around 76% for the ABR classification by using an ANN-based classifier for threshold determination. However, in [66] and [67], authors employed a CNN-based classification approach. They achieved a very encouraging performance with a maximum of 94.1% testing accuracy with the raw AEP dataset and 90.74% testing accuracy with time and frequency domain features. Although these studies used different AEP datasets, very few observations were utilized to validate their system. To test the performance of their approach, only 671 observations were used [66], whereas 280 observations were used [67]. More observations should be used to validate any model due to the non-stationery and artifact-prone characteristics of the AEP dataset. Considering this matter, our study has been validated with 3200 observations. Table 3 lists the summary of some related studies to compare the performance of the proposed study. In [68][6] [69] [7], the SVM classifier was used to classify the AEP dataset. These studies achieved 78.80%, 85.71, 87%, and 78.7% accuracy. The obtained overall performance is not enough to apply the models in real-life applications. Tang et al. [70] proposed a TS-PSO hybrid model to classify the two-class AEP dataset. They used Wavelet entropy as a feature extraction method and achieved 86.17% testing accuracy. Zhang et al. [5] proposed an auditory brainstem response (ABR) signals classification technique where they used wavelet analysis and Bayesian networks and achieved an overall accuracy of 78.80%. The ANN-based system for ABR detection was utilized in [14][64][65], and the maximum recognition accuracies were 73.7%, 80%, and 76%, respectively. These accuracies still need significant improvement to apply in real-life hearing disorder identification systems.

Further looking at Table 4, the proposed architecture significantly improves hearing disorder diagnosis compared to other studies. Moreover, the proposed system (CWT+CKNN) acquires a significant improvement (10% in 1 s, 6% in 2 s, and 6.87% in 3 s DW) with the concise decision window over the KNN algorithm. Furthermore, compared to the six popular transfer learning algorithms described in Table 3, our proposed architecture has improved more than 30% for hearing disorder identification with the AEP dataset. The experimental outcomes clearly explain the effectiveness and applicability of our proposed hybrid architecture. Thus, our architecture has achieved an impressive performance than the other related studies for hearing deficiency diagnosis reported in the literature.

Despite the impressive performance with a short decision window, some difficulties have been faced during the experimental analysis. For example, a wide range of similar datasets is required to check the validity and prove the feasibility of our proposed hybrid architecture, but during the experiment, we have not found any similar dataset for further validation. Next, several types of EEG headsets were used to perform the analysis, which contains the different number of electrodes between (1-256). So, the number of electrodes and electrodes that can acquire more quality signals for hearting disorder detection is also an open question and should be determined [71][72].

Table 4
The performance comparison with related previous studies.

Reference	Data		Feature extraction	Classification method	Classification accuracy
	Subject	Class			
[70]	180	2	WE	TS-PSO	86.17%
[6]	32	2	Global and nodal graph	SVM	85.71%
[14]	166 ABRs with 2000 repetitions	2	BAEP waveforms.	ANN	73.7%
[64]	366 ABRs with 1000 repetitions		Multi decision neural network	ANN	80%.
[68]	200	3	WPT	SVM	74.7%
[5]	8	2	DWT	Bayesian network classification	78.80%
[69]	39	2	SIFT	SVM	87%
[7]	Observation: 671	2	FFT	SVM	78.7%
[66]	Observation:671	2	Raw AEP	CNN	94.1%
[67]	Observation: 280 Subjects: 151	2	latency, FFT and DWT	A feed-forward multilayer perceptron	90.74%
Proposed Work	**Observations: 3200**	**2**	**CWT**	**CKNN**	**92.26%**

For hearing conditions diagnosis, most of the studies are carried out with ordinary machine learning algorithms [6] [7] and deep learning algorithms [55] [73] [74]. Despite some successful applications of hybrid architecture for task classification [40] [41] [42] [43], hybrid architectures are rarely investigated [70] for hearing condition detection. However, the recognition performance of most of the studies is insufficient to utilize the system in real-life applications. To reduce the computational complexity and make the system faster, experiment with a concise decision can provide the alternative solution; however limited studies [13] have been explored this issue. A more accurate and faster approach could be an effective tool for future hearing devices and provide a great application in real-life uses. The proposed hybrid architecture shows superior performance with time-frequency distribution for early hearing condition diagnosis, which would provide potential clinical utility in aiding clinicians to interpret the auditory response.

5. Conclusion

This paper presents a novel hearing ability detection system based on CWT and a hybrid CKNN approach. The CWT is used to transform the time-series data into time-frequency images, whereas the proposed CKNN architecture is used to classify the extracted features. The proposed system has been evaluated on a reputed publicly available AEP dataset recorded from 16 subjects. Our experimental outcomes demonstrated that the proposed architecture achieved an accuracy of 92.26%, precision of 92.57%, recall score of 92.86%, f1 score of 92.51%, and Cohen's kappa score of 84.50% with a 3 s decision window. This improvement with a concise decision window indicates the effectiveness and acceptability of the proposed hybrid architecture and helps in the faster analysis and reducing the computational complexity of the additional feature. To sum up, these satisfactory results within a short time interval may provide an outstanding reference for auditory loss diagnosis in clinical applications. Further study will be conducted with more data variance to test the proposed architecture's robustness to make the system more effective in real-time medical applications.

CRediT authorship contribution statement

Md Nahidul Islam: Conceptualization, Methodology, Software, Data analysis, Visualization, Investigation, Writing - Original Draft. **Norizam Sulaiman:** Methodology, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition. **Bifta Sama Bari:** Data analysis, Writing - Review & Editing, Visualization. **Mamunur Rashid:** Methodology, Data analysis, Visualization, Writing - Review & Editing. **Mahfuzah Mustafa:** Writing - Review & Editing, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Ministry of Higher Education under Fundamental research grant No. FRGS/1/2018/TK04/UMP/02/3 (University reference RDU190109) and Universiti Malaysia Pahang for laboratory facilities as well as additional financial support under Internal Research grant RDU190109.

References

[1] Deafness and hearing loss, (n.d.), https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss. (Accessed 18 June 2020).
[2] E. Holmes, P.T. Kitterick, A.Q. Summerfield, Peripheral hearing loss reduces the ability of children to direct selective attention during multi-talker listening, Hear. Res. 350 (2017) 160–172, https://doi.org/10.1016/j.heares.2017.05.005.
[3] N. Das, A. Bertrand, T. Francart, EEG-based auditory attention detection: boundary conditions for background noise and speaker positions, J. Neural Eng. 15 (2018), https://doi.org/10.1088/1741-2552/aae0a6.
[4] N. Das, W. Biesmans, A. Bertrand, T. Francart, The effect of head-related filtering and ear-specific decoding bias on auditory attention detection, J. Neural Eng. 13 (2016), https://doi.org/10.1088/1741-2560/13/5/056014.
[5] R. Zhang, G. McAllister, B. Scotney, S. McClean, G. Houston, Combining wavelet analysis and Bayesian networks for the classification of auditory brainstem response, IEEE Trans. Inf. Technol. Biomed. 10 (2006) 458–467, https://doi.org/10.1109/TITB.2005.863865.
[6] M.S. Mahmud, M. Yeasin, D. Shen, S.R. Arnott, C. Alain, G.M. Bidelman, What brain connectivity patterns from EEG tell us about hearing loss: a graph theoretic approach, in: ICECE 2018 - 10th Int. Conf. Electr. Comput. Eng., Institute of Electrical and Electronics Engineers Inc., 2019, pp. 205–208.
[7] P.Z. Li, L. Huang, C.D. Wang, C. Li, J.H. Lai, Brain network analysis for auditory disease: a twofold study, Neurocomputing 347 (2019) 230–239, https://doi.org/10.1016/j.neucom.2019.04.013.
[8] L. Wen, L. Gao, X. Li, A new deep transfer learning based on sparse auto-encoder for fault diagnosis, IEEE Trans. Syst. Man Cybern. Syst. 49 (2019) 136–144, https://doi.org/10.1109/TSMC.2017.2754287.
[9] N. Sriraam, EEG based automated detection of auditory loss: a pilot study, Expert Syst. Appl. 39 (2012) 723–731, https://doi.org/10.1016/j.eswa.2011.07.064.
[10] P. Xue, J. Bai, Q. Wang, X. Zhang, P. Feng, Analysis and classification of the nasal finals in hearing-impaired patients using tongue movement features, Speech Commun. 104 (2018) 57–65, https://doi.org/10.1016/j.specom.2018.09.008.

[11] M. Rashid, B.S. Bari, M.J. Hasan, M.A.M. Razman, R.M. Musa, A.F.A. Nasir, A.P.P.A. Majeed, The classification of motor imagery response: an accuracy enhancement through the ensemble of random subspace k-NN, PeerJ Comput. Sci. 7 (2021) 1–31, https://doi.org/10.7717/peerj-cs.374.
[12] B. Thorpe, T. Dussard, Classification of speech using MATLAB and K-nearest neighbour model: aid to the hearing impaired, in: Conf. Proc. - IEEE SOUTHEASTCON, Institute of Electrical and Electronics Engineers Inc., 2018.
[13] L. Deckers, N. Das, A.H. Ansari, A. Bertrand, T. Francart, EEG-based detection of the attended speaker and the locus of auditory attention with convolutional neural networks, BioRxiv (2018) 475673, https://doi.org/10.1101/475673.
[14] P.J. Mccullagh, G. King, H.G. Mcallister, H.G. Houston, Classification of brainstem auditory evoked potentials using artificial neural networks, Stud. Health Technol. Inform. 34 (1996) 547–550, https://doi.org/10.3233/978-1-60750-878-6-547.
[15] T. Grent-'t-Jong, R. Gajwani, J. Gross, A.I. Gumley, R. Krishnadas, S.M. Lawrie, M. Schwannauer, F. Schultze-Lutter, P.J. Uhlhaas, 40-Hz auditory steady-state responses characterize circuit dysfunctions and predict clinical outcomes in clinical-high-risk participants: a MEG study, Biol. Psychiatry (2021), https://doi.org/10.1016/j.biopsych.2021.03.018.
[16] Z. Shirzhiyan, E. Shamsi, A.S. Jafarpisheh, A.H. Jafari, Objective classification of auditory brainstem responses to consonant-vowel syllables using local discriminant bases, Speech Commun. 114 (2019) 36–48, https://doi.org/10.1016/j.specom.2019.09.003.
[17] Y. Lecun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444, https://doi.org/10.1038/nature14539.
[18] A.H. Ansari, P.J. Cherian, A. Caicedo, G. Naulaers, M. De Vos, S. Van Huffel, Neonatal seizure detection using deep convolutional neural networks, Int. J. Neural Syst. 29 (2019), https://doi.org/10.1142/S0129065718500119.
[19] L. He, C. Cao, Automated depression analysis using convolutional neural networks from speech, J. Biomed. Inform. 83 (2018) 103–111, https://doi.org/10.1016/j.jbi.2018.05.007.
[20] A.H. Ansari, O. De Wel, M. Lavanga, A. Caicedo, A. Dereymaeker, K. Jansen, J. Vervisch, M. De Vos, G. Naulaers, S. Van Huffel, Quiet sleep detection in preterm infants using deep convolutional neural networks, J. Neural Eng. 15 (2018), https://doi.org/10.1088/1741-2552/aadc1f.
[21] X. Chen, L. Li, A. Sharma, G. Dhiman, S. Vimal, The application of convolutional neural network model in diagnosis and nursing of MR imaging in Alzheimer's disease, Interdiscip. Sci. (2021), https://doi.org/10.1007/S12539-021-00450-7.
[22] K. Ali, Z.A. Shaikh, A.A. Khan, A.A. Laghari, Multiclass skin cancer classification using EfficientNets – a first step towards preventing skin cancer, Neurosci. Inform. (2021) 100034, https://doi.org/10.1016/J.NEURI.2021.100034.
[23] G. Dhiman, V. Vinoth Kumar, A. Kaur, A. Sharma, DON: deep learning and optimization-based framework for detection of novel coronavirus disease using X-ray images, Interdiscip. Sci. Comput. Life Sci. 13 (2021) 260–272, https://doi.org/10.1007/S12539-021-00418-7/FIGURES/13.
[24] C. Cheng, A. Sharma, Improved CNN license plate image recognition based on shark odor optimization algorithm, Int. J. Syst. Assur. Eng. Manag. 2021 (2021) 1–8, https://doi.org/10.1007/S13198-021-01309-2.
[25] G. Kaur, R. Sinha, P.K. Tiwari, S.K. Yadav, P. Pandey, R. Raj, A. Vashisth, M. Rakhra, Face mask recognition system using CNN model, Neurosci. Inform. (2021) 100035, https://doi.org/10.1016/J.NEURI.2021.100035.
[26] S.D. Pande, P.P. Jadhav, R. Joshi, A.D. Sawant, V. Muddebihalkar, S. Rathod, M.N. Gurav, S. Das, Digitization of handwritten Devanagari text using CNN transfer learning – a better customer service support, Neurosci. Inform. 2 (2022) 100016, https://doi.org/10.1016/J.NEURI.2021.100016.
[27] A. Deshpande, V.V. Estrela, P. Patavardhan, The DCT-CNN-ResNet50 architecture to classify brain tumors with super-resolution, convolutional neural network, and the ResNet50, Neurosci. Inform. 1 (2021) 100013, https://doi.org/10.1016/J.NEURI.2021.100013.
[28] A.A. Khan, A.A. Laghari, A.A. Shaikh, M.A. Dootio, V.V. Estrela, R.T. Lopes, A blockchain security module for brain-computer interface (BCI) with multimedia life cycle framework (MLCF), Neurosci. Inform. (2021) 100030, https://doi.org/10.1016/J.NEURI.2021.100030.
[29] L. Liu, L. Wang, D. Xu, H. Zhang, A. Sharma, S. Tiwari, M. Kaur, M. Khurana, M.A. Shah, CT image segmentation method of liver tumor based on artificial intelligence enabled medical imaging, Math. Probl. Eng. 2021 (2021), https://doi.org/10.1155/2021/9919507.
[30] A. Sharma, S. Kaur, N. Memon, A. Jainul Fathima, S. Ray, M.W. Bhatt, Alzheimer's patients detection using support vector machine (SVM) with quantitative analysis, Neurosci. Inform. 1 (2021) 100012, https://doi.org/10.1016/J.NEURI.2021.100012.
[31] S. Geirnaert, T. Francart, A. Bertrand, An interpretable performance metric for auditory attention decoding algorithms in a context of neuro-steered gain control, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2020) 307–317, https://doi.org/10.1109/TNSRE.2019.2952724.
[32] J.A. O'Sullivan, A.J. Power, N. Mesgarani, S. Rajaram, J.J. Foxe, B.G. Shinn-Cunningham, M. Slaney, S.A. Shamma, E.C. Lalor, Attentional selection in a cocktail party environment can be decoded from single-trial EEG, Cereb. Cortex 25 (2015) 1697–1706, https://doi.org/10.1093/cercor/bht355.
[33] W. Biesmans, N. Das, T. Francart, A. Bertrand, Auditory-inspired speech envelope extraction methods for improved EEG-based auditory attention detection in a cocktail party scenario, IEEE Trans. Neural Syst. Rehabil. Eng. 25 (2017) 402–412, https://doi.org/10.1109/TNSRE.2016.2571900.
[34] G. Ciccarelli, M. Nolan, J. Perricone, P.T. Calamia, S. Haro, J. O'Sullivan, N. Mesgarani, T.F. Quatieri, C.J. Smalt, Comparison of two-talker attention decoding from EEG with nonlinear neural networks and linear methods, Sci. Rep. 9 (2019) 1–10, https://doi.org/10.1038/s41598-019-47795-0.
[35] T.K. Grönfors, Computer analysis of auditory brainstem responses by using advanced pattern recognition, J. Med. Syst. 18 (1994) 191–199, https://doi.org/10.1007/BF00996703.
[36] R.M. McKearney, R.C. MacKinnon, Objective auditory brainstem response classification using machine learning, Int. J. Audiol. 58 (2019) 224–230, https://doi.org/10.1080/14992027.2018.1551633.
[37] C.H. Chen, W.T. Huang, T.H. Tan, C.C. Chang, Y.J. Chang, Using K-nearest neighbor classification to diagnose abnormal lung sounds, Sensors (Switzerland) 15 (2015) 13132–13158, https://doi.org/10.3390/s150613132.
[38] K.V. Park, K.H. Oh, Y.J. Jeong, J. Rhee, M.S. Han, S.W. Han, J. Choi, Machine learning models for predicting hearing prognosis in unilateral idiopathic sudden sensorineural hearing loss, Clin. Exp. Otorhinolaryngol. 13 (2020) 148–156, https://doi.org/10.21053/ceo.2019.01858.
[39] H.S. Sanjay, B.V. Hiremath, B.S. Prithvi, P.A. Dinesh, Machine learning based assessment of auditory threshold perception in human beings, SN Appl. Sci. 2 (2020) 1–10, https://doi.org/10.1007/s42452-019-1929-7.
[40] H. Wu, P. Gurram, H. Kwon, S. Prasad, A hybrid CSVM-HMM model for acoustic signal classification using a tetrahedral sensor array, in: Proc. IEEE Sensors, 2014-Decem., 2014, pp. 1352–1355.
[41] A. Kundu, G.C. Chen, An integrated hybrid neural network and hidden Markov model classifier for sonar signals, in: 1995 International Conference on Acoustics, Speech, and Signal Processing, vol. 45, 1997, pp. 2566–2570.
[42] S. Rasheed, D.W. Stashuk, M.S. Kamel, A hybrid classifier fusion approach for motor unit potential classification during EMG signal decomposition, IEEE Trans. Biomed. Eng. 54 (2007) 1715–1721, https://doi.org/10.1109/TBME.2007.892922.
[43] X. Li, L. Cui, S. Tao, J. Chen, X. Zhang, G.Q. Zhang, HyCLASSS: a hybrid classifier for automatic sleep stage scoring, IEEE J. Biomed. Health Inform. 22 (2018) 375–385, https://doi.org/10.1109/JBHI.2017.2668993.
[44] M. Längkvist, L. Karlsson, A. Loutfi, Inception-v4, inception-ResNet and the impact of residual connections on learning, Pattern Recognit. Lett. 42 (2014) 11–24, http://arxiv.org/abs/1512.00567.
[45] H. Pan, Z. Pang, Y. Wang, Y. Wang, L. Chen, A new image recognition and classification method combining transfer learning algorithm and MobileNet model for welding defects, IEEE Access 8 (2020) 119951–119960, https://doi.org/10.1109/ACCESS.2020.3005450.
[46] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., IEEE Computer Society, 2016, pp. 770–778.
[47] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 2015, pp. 1–14.
[48] F. Chollet, Xception: deep learning with depthwise separable convolutions, in: Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017, Institute of Electrical and Electronics Engineers Inc., 2017, pp. 1800–1807.
[49] N. Das, T. Francart, A. Bertrand, Auditory attention detection dataset KULeuven, https://doi.org/10.5281/ZENODO.3997352, 2020.
[50] T. Francart, A. van Wieringen, J. Wouters, APEX 3: a multi-purpose test platform for auditory psychophysical experiments, J. Neurosci. Methods 172 (2008) 283–293, https://doi.org/10.1016/j.jneumeth.2008.04.020.
[51] Radio books for children, (n.d.), http://www.radioboeken.eu/kinderradioboeken.php?lang=NL. (Accessed 27 March 2021).
[52] R. Yan, R.X. Gao, X. Chen, Wavelets for fault diagnosis of rotary machines: a review with applications, Signal Process. 96 (2014) 1–15, https://doi.org/10.1016/j.sigpro.2013.04.015.
[53] L. Huang, J. Wang, Forecasting energy fluctuation model by wavelet decomposition and stochastic recurrent wavelet neural network, Neurocomputing 309 (2018) 70–82, https://doi.org/10.1016/j.neucom.2018.04.071.
[54] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, G.Z. Yang, Deep learning for health informatics, IEEE J. Biomed. Health Inform. 21 (2017) 4–21, https://doi.org/10.1109/JBHI.2016.2636665.
[55] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84–90, https://doi.org/10.1145/3065386.
[56] B. Mele, G. Altarelli, Dropout: a simple way to prevent neural networks from overfittin, Phys. Lett. B 299 (1993) 345–350, https://doi.org/10.1016/0370-2693(93)90272-J.
[57] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27, https://doi.org/10.1109/TIT.1967.1053964.
[58] P. Kang, S. Cho, Locally linear reconstruction for instance-based learning, Pattern Recognit. 41 (2008) 3507–3518, https://doi.org/10.1016/j.patcog.2008.04.009.
[59] M. Rashid, M. Islam, N. Sulaiman, B.S. Bari, R.K. Saha, M.J. Hasan, Electrocorticography based motor imagery movements classification using long short-

term memory (LSTM) based on deep learning approach, SN Appl. Sci. 2 (2020), https://doi.org/10.1007/s42452-020-2023-x.
[60] Y. Te Wu, T.H. Huang, C. Yi Lin, S.J. Tsai, P.S. Wang, Classification of EEG motor imagery using support vector machine and convolutional neural network, in: 2018 Int. Autom. Control Conf. CACS 2018, 2019, https://doi.org/10.1109/CACS.2018.8606765.
[61] H.J.P. Weerts, A.C. Mueller, J. Vanschoren, Importance of tuning hyperparameters of machine learning algorithms, https://arxiv.org/abs/2007.07588v1, 2020. (Accessed 2 May 2021).
[62] U. Hoppe, S. Weiss, R.W. Stewart, U. Eysholdt, An automatic sequential recognition method for cortical auditory evoked potentials, IEEE Trans. Biomed. Eng. 48 (2001) 154–163, https://doi.org/10.1109/10.909636.
[63] A. Effern, K. Lehnertz, G. Fernández, T. Grunwald, P. David, C.E. Elger, Single trial analysis of event related potentials: non-linear de-noising with wavelets, Clin. Neurophysiol. 111 (2000) 2255–2263, https://doi.org/10.1016/S1388-2457(00)00463-6.
[64] J.B.A. Habraken, M.J. van Gils, P.J.M. Cluitmans, Identification of peak V in brainstem auditory evoked potentials with neural networks, Comput. Biol. Med. 23 (1993) 369–380, https://doi.org/10.1016/0010-4825(93)90134-M.
[65] D. Alpsan, O. Ozdamar, Auditory brainstem evoked potential classification for threshold detection by neural networks. II. Effects of input coding, training set size and composition and network size on performance, Automedica 15 (1992) 83–93, https://miami.pure.elsevier.com/en/publications/auditory-brainstem-evoked-potential-classification-for-threshold--2. (Accessed 23 April 2021).
[66] R.R. Hallac, J. Lee, M. Pressler, J.R. Seaward, A.A. Kane, Identifying ear abnormality from 2D photographs using convolutional neural networks, Sci. Rep. 9 (2019) 1–6, https://doi.org/10.1038/s41598-019-54779-7.
[67] S. Dass, M.S. Holi, K. Soundararajan, Classification of brainstem auditory evoked potentials using artificial neural network based on time and frequency domain features, J. Clin. Eng. 41 (2016) 72–82, https://doi.org/10.1097/JCE.0000000000000148.
[68] H. Dietl, S. Weiss, Detection of cochlear hearing loss applying wavelet packets and support vector machines, in: Conf. Rec. - Asilomar Conf. Signals, Syst. Comput., 2004, pp. 1575–1579.
[69] L. Tan, Y. Chen, T.C. Maloney, M.M. Caré, S.K. Holland, L.J. Lu, Combined analysis of sMRI and fMRI imaging data provides accurate disease markers for hearing impairment, NeuroImage Clin. 3 (2013) 416–428, https://doi.org/10.1016/j.nicl.2013.09.008.
[70] C. Tang, E. Lee, Hearing loss identification via wavelet entropy and combination of Tabu search and particle swarm optimization, in: Int. Conf. Digit. Signal Process DSP, Institute of Electrical and Electronics Engineers Inc., 2019.
[71] J. Montoya-Martínez, A. Bertrand, T. Francart, Optimal number and placement of EEG electrodes for measurement of neural tracking of speech, Dynamical systems, signal processing and data analytics oto-rhino-laryngology optimal number and placement of EEG electrodes for measurement of neural tracking of speech 2 BioRxiv (2019) 800979, https://doi.org/10.1101/800979.
[72] A.M. Narayanan, A.A. Bertrand, The effect of miniaturization and galvanic separation of EEG sensor devices in an auditory attention detection task, in: Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 77–80.
[73] S.A. Nossier, M.R.M. Rizk, N.D. Moussa, S. el Shehaby, Enhanced smart hearing aid using deep neural networks, Alex. Eng. J. 58 (2019) 539–550, https://doi.org/10.1016/j.aej.2019.05.006.
[74] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Ind. Inform. 15 (2019) 2446–2455, https://doi.org/10.1109/TII.2018.2864759.