VOL. 23, NO. 2, 2024, 123-129
elektrika.utm.my
ISSN 0128-4428

ELEKTRIKA
Journal of Electrical Engineering

# An Integrated Hybrid Convolutional-Support Vector Machine (CSVM) Model with the CWT Method for Hearing Disorder Detection using AEP Signals

**Md Nahidul Islam**[1*], **Norizam Sulaiman**[1] and **Mahfuzah Mustafa**[2]

[1]Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang, Pekan, Pahang, 26600, Malaysia.

[*]Corresponding author: nahidul76.edu@gmail.com

**Abstract:** Hearing disorder is the most widespread sensory disability worldwide, impairing human communication and learning. An early and accurate hearing disorder detection system using an electroencephalogram (EEG) is the appropriate technique for dealing with this concern. The most significant modality for diagnosing hearing deficiency among EEG control signals is the auditory evoked potential (AEP), which is generated in the cortical region of the brain through auditory stimulus. This study aims to develop an efficient approach for detecting hearing disorders. For this purpose, this study has designed a hybrid model based on the convolutional operation of CNN and the SVM classifier. Initially, the CWT method was utilized to transform the raw AEP signals into time-frequency images. Then, the extracted features were classified using the proposed CSVM model. To test the robustness of the proposed model, this study also implemented a convolutional neural network (CNN) and support vector machine (SVM) with the same parameters. The experimental results with the hybrid CSVM model showed superior performance on the publicly available AEP dataset by achieving 94.48% testing accuracy, 96.40% precision, 92.96% recall, 94.65% F1 score, 88.95% Cohen Kappa score, which indicates that the proposed hybrid model could be used for early hearing disorder detection. Future enhancements will concentrate on identifying different hearing-signal-based data and the cloud-based, automated classification of AEP signals.

**Keywords:** Hearing disorder, EEG, AEP, Deep Learning, hybrid CSVM model



*Article History: received 17 February 2024; accepted 9 July 2024; published 29 August 2024.*

## 1. INTRODUCTION

Hearing disorder is the most common physical impairment in humans, characterized by a partial or full inability to hear a sound, resulting in communication and learning difficulties. The World Health Organization (WHO) reported that 466 million individuals had hearing loss globally in 2018. By 2030, it is estimated to reach 630 million and more than 900 million by 2050 [1]. Therefore, an efficient diagnostic system is crucial to address this issue and prevent hearing loss through timely intervention. However, traditional testing methods are time-consuming, require clinical expertise, and rely on direct responses from individuals.

Researchers nowadays have provided a range of testing methods for hearing disorders. In these, auditory evoked potential (AEP) signals are commonly utilized for the diagnosis of hearing disorders [2][3]. Currently, the AEP signal's outcomes are often employed in a variety of brain-computer interface (BCI) technologies [4][5] and brain hearing issues [6]. Developing an intelligent system with high performance in BCI depends on some standard procedures. This approach is often divided into four stages: data collecting, pre-processing, feature extraction and selection, and classification [7]. Traditionally, feature extraction has been achieved by extracting information from raw signals using time, frequency, or time-frequency domain techniques. After that, the extracted characteristics are utilized to train the deep learning (DL) or machine learning (ML) model. Extraction of significant and discriminative features from EEG data is also critical for characterizing and categorizing patterns of brain activity [8]. Moreover, conventional hearing disorder diagnosis methods have some drawbacks. For instance, previous hearing deficiency detection methods are frequently focused on manual feature selection [9]. As a result, if the manually selected characteristics are inadequate for the task classification, the performance of hearing deficiency will decrease significantly. Moreover, a concise decision (DW) provides less data for categorization, challenging high performance. However, it aids in early hearing loss detection and accelerates architecture by reducing computing complexity. However, a few studies have explored the necessity of concise DW for real-world applications [10].

DL strategies, on the other hand, may be able to overcome the limitations because of their excellent capabilities to learn features [11][12]. Several hidden layers in the DL architecture enable them to learn hierarchical representations explicitly. However, some issues remain with DL models, although they have been employed effectively in hearing loss diagnostic tasks previously. Only a few studies have been performed for identifying hearing disorders using DL algorithms with more than ten hidden layers [13] [14]. For the hearing condition diagnosis, some recent studies [3][15] have widely used the SVM algorithm. Despite several fundamental advantages of the SVM algorithm in the neurological sector, most research produced poorer results in identifying hearing disorders. In addition, feature scaling (normalization or standardization) is required prior to applying the SVM algorithm to any dataset; otherwise, the SVM algorithm may generate incorrect predictions. Integrating two or more classifiers might be an effective alternative to a single classifier for performing error-free classification tasks, given the limitations of some single classifiers. The researchers have created several hybrid strategies that integrate many algorithms for this goal. In [16][17], the research employed hybrid models and effectively integrated the strengths of two models, which deliver higher performance than a single classifier in some situations.

In this study, a hybrid model with time-frequency images is presented for hearing disorder detection. In the hybrid model, the convolutional operation is used to detect the hidden local information in neural activity, whereas the SVM is used to classify and extract the features. The following is a summary of the paper's key contributions:

- In the proposed approach, a hybrid model (CSVM) based on integrating the convolutional operation of CNN with an SVM classifier is designed to improve hearing disorder detection performance.
- This experiment is performed within a shorter decision window, which minimizes the effect of additional features and decreases time consumption, demonstrating the robustness and adaptability of the system in real-time scenarios.
- In addition, the AEP signals have been evaluated with CNN and SVM classifiers, and the proposed model demonstrates higher performance in detecting hearing disorders.

The remaining sections of the paper are organized as follows: Section 2 represents the Materials and methods, including data description, pre-processing, feature extraction technique, and proposed hybrid classification model. Section 3 explores the findings of the experimental analysis and discussion. Finally, Chapter 4 summarizes the study's findings, contribution, and recommendations for future work.

## 2. METHODOLOGY

This study aims to design a high-performance hearing disorder detection system. For this purpose, a publicly available AEP dataset is used. First, the raw AEP signals are segmented into a 2s decision window to prepare the observations. Then, time-frequency images are generated using the CWT method. Finally, the CWT images are classified using the proposed CSVM model. The overall procedure of the proposed approach is illustrated in Figure 1.

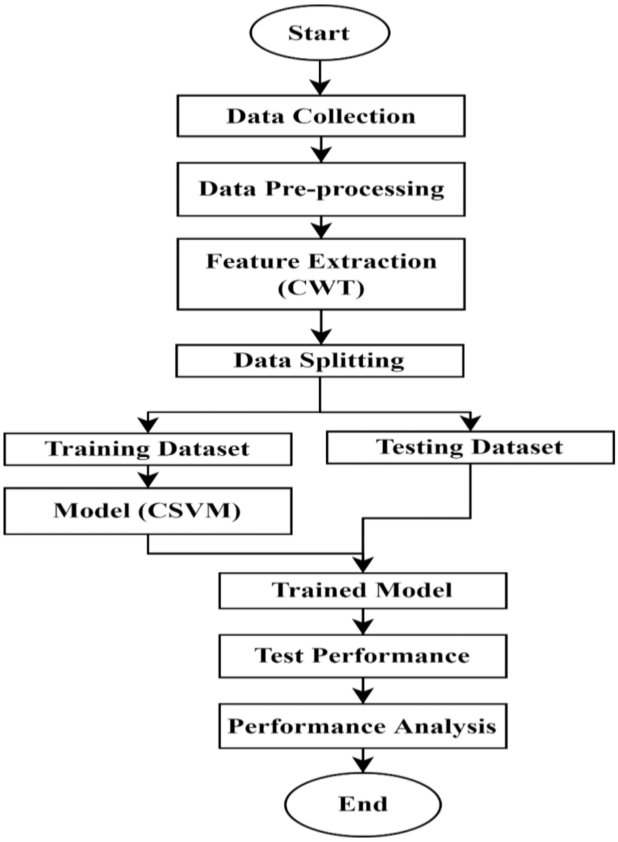


Figure 1. Flow chart of the proposed approach for hearing disorder detection.

### 2.1 Data Description

Experimental AEP datasets are provided by ExpORL, Dept. Neurosciences, KULeuven, and Dept. Electrical Engineering (ESAT), KULeuven [18]. The AEP data were gathered at an 8196 Hz sampling rate utilizing a 64-channel BioSemi Active Two device. The whole dataset was collected from 16 subjects, and each individual underwent the trial 20 times. Using Etymotic ER3, earphones with a 4 kHz cut-off frequency were inserted, and the auditory stimuli were delivered at 60 dBA. APEX 3 was used as simulation software [19]. In each trial, participants were given two different stories to hear. To overcome the lateralization bias described by [20], the attended ear was switched between trials to ensure that each ear received equal data. The dataset contains (72-minutes × 16-subjects) minutes of EEG recording. The trials in this investigation were down-sampled from 8192 Hz to 128 Hz.

### 2.2 Data Preprocessing

After collecting the publicly available online dataset, the dataset has been segmented to prepare the observations. To prepare the observations, we segmented each subject's data into a two-second decision window (DW). Analyzing this shorter DW and achieving high performance is challenging but highly effective for real-time applications. The raw AEP segmented data for subject-1 channel-1 is shown in Figure 2.

After preparing the observations, a time-frequency feature extraction method, continuous wavelet transform (CWT), was used to extract time-frequency characteristics that facilitate multi-scale signal amplification using scaling and translating techniques [21]. The segmented dataset is transferred from raw signals to the time-frequency images.

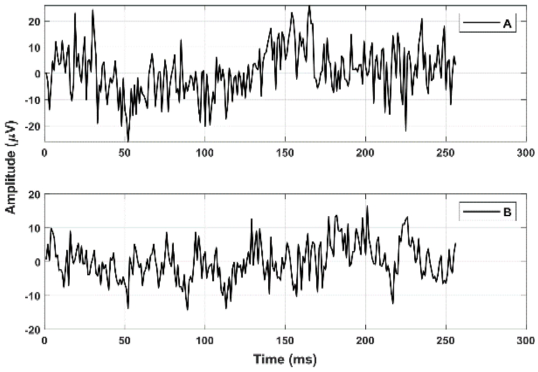


Figure 2. Raw AEP signals: (A) hear auditory stimulus with the left ear and (B) hear auditory stimulus with the right ear.

The wavelet set is created by scaling and translating the mother wavelet, which is a family of wavelets ψ(t), shown in equations 1 and 2 [21].

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

In this study, the symbol S is used to denote the scale parameter, which is inversely proportional to frequency, while the symbol $\tau$ represents the translation parameter. The variable "t" represents the time domain. It denotes the continuous variable over which the wavelet function is defined and applied. The signal x(t) can be achieved by a complex conjugate convolution operation, mathematically defined by equation 2 [22].

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

Where, $\psi^*$ denotes the complex conjugate of the above function, and this operation decomposes the signal in a series of wavelet coefficients.

This study utilized the wavelet basis functions (Mother Wavelets). After that, the proposed model is fed with time-frequency images. Figure 3 illustrates the time-frequency CWT images. This study is conducted using the publicly available dataset. The device used for collecting the experimental dataset has 64 channels, producing the data from 64 channels simultaneously. For that reason, each channel's raw data is first converted into time-frequency images using the CWT. Then, all images were concatenated and considered as observations. After concatenation, each observation provides the time-frequency information of 64 channels.

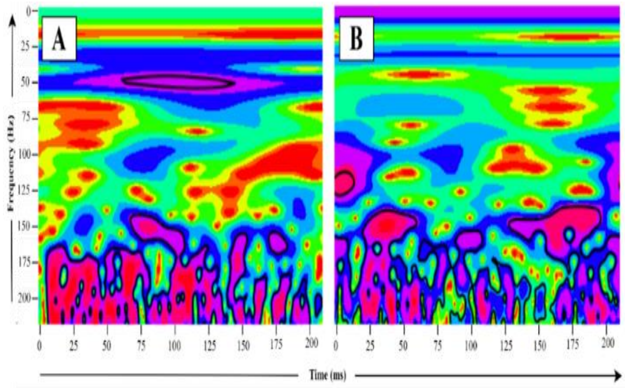


Figure 3. The CWT images: (A) hear auditory stimulus with the left ear, (B) hear auditory stimulus with the right ear.

## 2.3 The Proposed Hybrid CSVM Model

This study effectively integrated two methods to construct the presented CSVM model, improving the performance for the identification of hearing disorders. The developed CSVM model includes two convolution layers and an SVM classifier. The convolutional layers include a variety of kernel sizes, pooling layers, and one dropout layer. The discriminating features are extracted using the convolutional process, and the extracted features are then classified using the SVM.

The architecture begins by applying the CWT to raw AEP data, transforming these signals into time-frequency images. This image serves as the input for the subsequent convolutional processes. Figure 4 in the paper visually represents this sophisticated model, illustrating the seamless flow from data input to diagnostic output, showcasing the synergy between advanced convolutional processing and machine learning classification.

The initial convolutional layer employs 32 kernels of size (5 x 5), which are essential for capturing fine-grained patterns and features from the time-frequency images. The convolutional process extracts vital features by applying these kernels, and the subsequent max-pooling layer significantly reduces the dimensionality of the data from (224 x 224 x 32) to (74 x 74 x 32), ensuring that only the most salient features are retained. This reduction preserves essential information and decreases computational load.

Following the dimensionality reduction, a dropout layer with a dropout rate of 0.6 is strategically placed to prevent overfitting. This layer randomly deactivates a portion of the neurons during training, forcing the network to learn more robust features that are not reliant on any specific set of neurons. After this regularization step, another convolutional layer with 64 kernels of size (3 x 3 x 64) further refines the feature extraction process.

A second max-pooling layer follows, using a (3 x 3) kernel to reduce the data dimensions to (24 x 24 x 64). This step intensifies the focus on the most influential features, preparing the data for the final classification stage. The processed features are then flattened into a single vector, effectively serving as the SVM classifier's input. The SVM's role is pivotal as it classifies these features into categories indicative of specific hearing disorders. The integration of CNN and SVM in the CSVM model represents a potent approach to biomedical signal processing, leveraging the strengths to achieve superior accuracy and reliability in diagnosing hearing impairments.

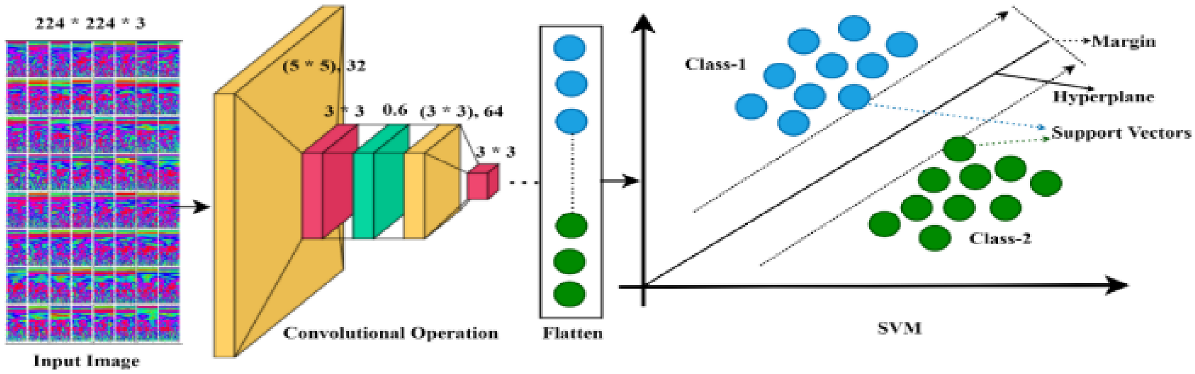


Figure 4. Schematic diagram of the proposed CSVM approach.

### 2.4 Performance Evaluation

In this experimental analysis, we employ five performance evaluation metrics [23] (equations 3 to 7) to measure the effectiveness of our proposed model. Accuracy is determined by computing the proportion of correctly classified observations to the total instances, offering an overall measure of the model's performance. Precision captures the quality of positive predictions, emphasizing the ratio of true positives to all predicted positives. This metric is essential when minimizing false positives is crucial.

Recall assesses the model's ability to identify actual positives by measuring the proportion of true positives out of all actual positive instances, providing valuable insight when missing positive cases is costly. The F1 score combines precision and recall in a single value, reflecting a balanced measure that emphasizes both concerns equally. Cohen's Kappa score evaluates the agreement between predicted and actual labels, considering both observed and expected agreement to deliver a more nuanced understanding of performance beyond what random chance would predict. These metrics offer a comprehensive framework for evaluating the performance of classification models in this study. Each model implemented in this study has been evaluated using these five performance evaluation metrics.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

$$Kappa\ score = \frac{P_o - P_e}{1 - P_e} \quad (7)$$

Here, TP is the True Positive Rate, TN is the True Negative Rate, FP is the False Positive Rate, FN is the False Negative Rate, $P_o$ is Observed Agreement, and $P_e$ is Expected Agreement.

## 3. RESULTS AND DISCUSSION

In this study, we employ the publicly available AEP dataset to evaluate the effectiveness of our proposed model. The raw signals from the dataset are first converted into time-frequency images using the CWT, then concatenated across different channels as outlined in Section 2.2. These processed images form the basis for both training and testing our model.

The dataset, which includes 6400 images, is allocated into training and testing categories at ratios of 70% and 30%, respectively. This allocation provides 4480 images for the training set and 1920 images for the testing set. To ensure an unbiased approach in both the training and evaluation phases, the dataset maintains a balanced class distribution. Specifically, Classes 1 and 2 are equally represented, each contributing 2240 images to the training set and 960 images to the testing set. This balanced allocation is imperative to mitigate any potential model bias and to facilitate a comprehensive assessment of the model's ability to generalize across varied types of auditory evoked responses.

This experimental investigation was conducted in Python with the assistance of Google Colab, Windows 10, Intel(R) Xeon(R) CPU @ 2.30GHz, Tesla K80, CUDA Version K80, and CUDA Version 1.

### 3.1 The Experimental Results of the Model

This study has implemented three distinct models to assess the effectiveness of AEP for categorizing hearing conditions. In this study, the SVM, CNN, and hybrid CSVM models have been implemented. Table 2 compares the performance of the three models for diagnosing hearing conditions and contrasts the presented model's performance with that of the other two models. For each subject (s), five performance measuring techniques [23][24] (accuracy, precision, recall, F1 score, and Cohen's kappa score) have been evaluated to show the effectiveness and robustness of the strategy.

Table 1 illustrates the performance of the three different models for hearing disorder detection. In this experimental analysis, the performance of individual subject (s) has been calculated. For the experimental analysis with the SVM model, the average (AVG) performance, including accuracy, precision, recall, F1 score, and Cohen's kappa score, are 89.22%, 87.28%, 90.79%, 88.89%, and 78.43%, respectively. For the analysis with the CNN model, an AVG of 91.98% accuracy, 93.16% precision, 89.80% recall score, 91.30% F1 Score, and 83.88% Cohen kappa score were achieved.

Table 1. The performance of three different models.

| Method | Accuracy (%) | Precision (%) | Recall (%) | F1 (%) | Cohen Kappa (%) |
|---|---|---|---|---|---|
| SVM | 89.22 | 87.28 | 90.79 | 88.89 | 78.43 |
| CNN | 91.98 | 93.16 | 89.80 | 91.30 | 83.88 |
| CSVM | 94.48 | 96.40 | 92.96 | 94.65 | 88.95 |

For the proposed CSVM model, the AVG value of accuracy, precision, recall, F1 score, and Cohen's kappa score is 94.48%, 96.40%, 92.96%, 94.65%, 88.95%, respectively has been achieved. From the analysis result, it is demonstrated that the proposed CSVM model has achieved 5.26%, 9.12%, 2.17%, 5.76%, and 10.52% improvement compared to the SVM model, and 2.50%, 3.24%, 3.16%, 3.35%, and 5.07% improvement compared to the CNN model.

From Table 1, it is clearly demonstrated that the advantages of integrating convolutional layers with SVM techniques within our CSVM framework highlight its superior performance in detecting auditory disorders. This methodological choice utilizes the strengths of SVMs in creating robust classification models, particularly advantageous in high-dimensional spaces typical of medical diagnostic imaging.

SVMs are preferred for their ability to effectively maximize the margin between different classes, a critical factor in achieving high classification accuracy. This feature is crucial in our application, where precise distinction between auditory conditions is necessary. Furthermore, SVMs inherently prevent overfitting through their regularization approach, ensuring that our model generalizes well to new, unseen data—a key requirement for reliable clinical diagnostics.

Our study's performance metrics clearly demonstrate the CSVM model's efficacy with higher performance. This demonstrates not only the capability of SVMs in handling complex datasets but also their suitability for enhancing model reliability and diagnostic precision in clinical settings.

We also calculate the confusion matrix on the testing set (shown in Figure 5), which provides a detailed representation of the classification model's performance in identifying auditory stimuli in different ears. This matrix organizes predictions into four quadrants: true positives, false negatives, true negatives, and false positives, illustrating the model's accuracy in distinguishing between auditory stimuli in the left and right ears.

Our CSVM model is evaluated against a dataset consisting of 1,008 instances where subjects are expected to hear stimuli in the left ear (Class 1) and 912 instances in the right ear (Class 2). The model successfully identifies 937 instances as true positives, confirming its high efficiency in detecting auditory stimuli in the left ear.

Furthermore, it correctly recognizes 877 instances as true negatives for stimuli in the right ear, demonstrating its robustness in accurately identifying stimuli in the right ear. This capability is crucial for applications requiring precise lateralization of auditory detection.

Despite these successes, the matrix also records 71 false negatives, where the model fails to detect stimuli in the left ear when it is present, and 35 false positives, where it incorrectly identifies the presence of stimuli in the left ear when the actual stimuli are in the right ear. These errors highlight potential refinement in the model’s sensitivity and specificity.

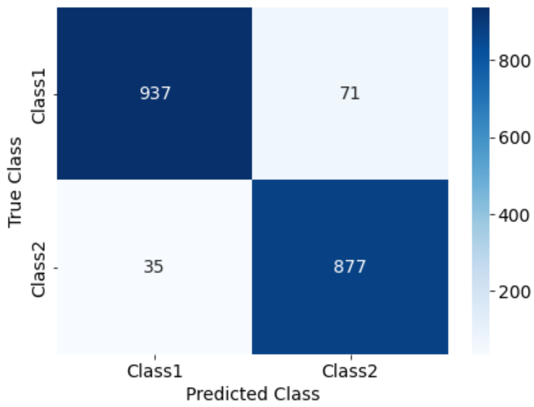


Figure 5. The confusion matrix using the proposed model.

### 3.2 Performance Comparison with the Related Studies

Machine learning (ML) algorithms have seen extensive utilization in recent years for identifying hearing impairments, spurred by advancements in computer technology. As computational capabilities have expanded, both machine learning and deep learning (DL) techniques have been increasingly applied to the early detection of hearing abnormalities. This application of technology has enabled researchers and clinicians to leverage sophisticated analytical methods to improve diagnostic accuracy, providing significant benefits in medical interventions and patient care.

In this section, we benchmark the efficacy of our proposed Convolutional Support Vector Machine (CSVM) model against a series of studies showcasing its superior performance in detecting auditory stimuli. The study [2] featured a Bayesian Network classifier that processed features extracted using Discrete Wavelet Transform, achieving an accuracy of 78.80% with only 8 subjects across two classes. Another relevant study [3] achieved an 85.71% accuracy by employing a Support Vector Machine to analyze global and nodal graph features from 32 subjects. Similarly, a study [15] utilized Wavelet Packet Transform with an SVM, reporting a lower accuracy of 74.7% from a larger cohort of 200 subjects classified into three categories.

A more comparable study [25] used raw Auditory Evoked Potential data with a Convolutional Neural Network, reaching 94.1% accuracy across 671 observations. Study [26] applied Scale-Invariant Feature Transform with an SVM, achieving an 87% accuracy from 39 subjects.

Our CSVM model stands out by integrating Continuous Wavelet Transform and a CSVM classifier, reaching a top accuracy of 94.48% on a substantial dataset of 6400 observations in two classes. This extensive testing and the high performance show our model's robustness and highlight its potential for clinical application, establishing a new standard in auditory detection technology. This comparative analysis illustrates our model’s leading position in the field and its ability to improve existing methods significantly.

Table 2. The performance comparison of the proposed model with related studies.

| References | Data | | Feature Extraction | Methods | Accuracy |
|---|---|---|---|---|---|
| | Subject | Class | | | |
| [2] | 8 | 2 | DWT | Bayesian Network | 78.80% |
| [3] | 32 | 2 | Global and nodal graph | SVM | 85.71% |
| [15] | 200 | 3 | WPT | SVM | 74.7% |
| [25] | Observation: 671 | 2 | Raw-AEP | CNN | 94.1% |
| [26] | 39 | 2 | SIFT | SVM | 87% |
| Proposed Model (CSVM) | Observation: 6400 | 2 | CWT | CSVM | 94.48% |

## 4. CONCLUSION

A hearing disorder detection approach based on the CWT and proposed CSVM model has been proposed in this paper. Firstly, the raw AEP signals have been segmented and transferred to time-frequency images using the CWT method. Then, the images were classified using the proposed CSVM model. The proposed model has been evaluated on the publicly available dataset and achieved 94.48% testing accuracy. This enhancement with a shorter decision window (two seconds) demonstrates the efficacy and acceptability of the suggested hybrid architecture, expedites the analysis, and reduces the computational cost of additional features. The focus of our future study will be categorizing many AEP-signal-based data. This research intends to improve the model's generalization capability, the instructional effectiveness of neural networks, and therapeutic benefits.

## ACKNOWLEDGMENT

The authors would like to thank Universiti Malaysia Pahang Al-Sultan Abdullah for providing financial support under the internal research grant PDU233212 and laboratory facilities to conduct the research.

## REFERENCES

[1] World Health Organization (WHO), “Deafness and hearing loss.” Accessed: Feb. 11, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss

[2] R. Zhang, G. McAllister, B. Scotney, S. McClean, and G. Houston, “Combining wavelet analysis and Bayesian networks for the classification of auditory brainstem response,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 3, pp. 458–467, 2006, doi: 10.1109/TITB.2005.863865.

[3] M. S. Mahmud, M. Yeasin, D. Shen, S. R. Arnott, C. Alain, and G. M. Bidelman, "What brain connectivity patterns from EEG tell us about hearing loss: A graph theoretic approach," in *ICECE 2018 - 10th International Conference on Electrical and Computer Engineering*, Institute of Electrical and Electronics Engineers Inc., Feb. 2019, pp. 205–208. doi: 10.1109/ICECE.2018.8636698.

[4] S. Gao, Y. Wang, and X. Gao, "Visual and Auditory Brain – Computer Interfaces," vol. 61, no. 5, pp. 1436–1447, 2014.

[5] S. Alam, E. Technology, B. S. Bari, E. Technology, M. S. Jadin, and E. Technology, "Empirical Mode Decomposition Coupled with Fast Fourier Transform based Feature Extraction Method for Motor Imagery Tasks Classification," no. November, pp. 10–15, 2020.

[6] N. Sriraam, "EEG based automated detection of auditory loss: A pilot study," *Expert Syst Appl*, vol. 39, no. 1, pp. 723–731, 2012, doi: 10.1016/j.eswa.2011.07.064.

[7] M. N. Islam, N. Sulaiman, M. Rashid, B. S. Bari, and M. Mustafa, "Hearing Disorder Detection using Auditory Evoked Potential (AEP) Signals," in *ETCCE 2020 - International Conference on Emerging Technology in Computing, Communication and Electronics*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020. doi: 10.1109/ETCCE51779.2020.9350918.

[8] Q. Wei, Y. Wang, X. Gao, and S. Gao, "Amplitude and phase coupling measures for feature extraction in an EEG-based brain-computer interface," *J Neural Eng*, vol. 4, no. 2, pp. 120–129, 2007, doi: 10.1088/1741-2560/4/2/012.

[9] M. Vavrečka and L. Lhotská, "EEG Feature Selection Based on Time Series Classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7988 LNAI, pp. 520–527, 2013, doi: 10.1007/978-3-642-39712-7_40.

[10] L. Deckers, N. Das, A. H. Ansari, A. Bertrand, and T. Francart, "EEG-based detection of the attended speaker and the locus of auditory attention with convolutional neural networks," *bioRxiv*, p. 475673, 2018, doi: 10.1101/475673.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017, doi: 10.1145/3065386.

[12] S. A. Nossier, M. R. M. Rizk, N. D. Moussa, and S. el Shehaby, "Enhanced smart hearing aid using deep neural networks," 2019. doi: 10.1016/j.aej.2019.05.006.

[13] G. Ciccarelli *et al.*, "Comparison of Two-Talker Attention Decoding from EEG with Nonlinear Neural Networks and Linear Methods," *Sci Rep*, vol. 9, no. 1, pp. 1–10, 2019, doi: 10.1038/s41598-019-47795-0.

[14] R. M. McKearney and R. C. MacKinnon, "Objective auditory brainstem response classification using machine learning," *Int J Audiol*, vol. 58, no. 4, pp. 224–230, 2019, doi: 10.1080/14992027.2018.1551633.

[15] H. Dietl and S. Weiss, "Detection of cochlear hearing loss applying wavelet packets and support vector machines," in *Conference Record - Asilomar Conference on Signals, Systems and Computers*, 2004, pp. 1575–1579. doi: 10.1109/acssc.2004.1399421.

[16] H. Wu, P. Gurram, H. Kwon, and S. Prasad, "A hybrid CSVM-HMM model for acoustic signal classification using a tetrahedral sensor array," *Proceedings of IEEE Sensors*, vol. 2014-Decem, no. December, pp. 1352–1355, 2014, doi: 10.1109/ICSENS.2014.6985262.

[17] S. Rasheed, D. W. Stashuk, and M. S. Kamel, "A hybrid classifier fusion approach for motor unit potential classification during EMG signal decomposition," *IEEE Trans Biomed Eng*, vol. 54, no. 9, pp. 1715–1721, 2007, doi: 10.1109/TBME.2007.892922.

[18] N. Das, T. Francart, and A. Bertrand, "Auditory Attention Detection Dataset KULeuven," Aug. 2020, doi: 10.5281/ZENODO.3997352.

[19] T. Francart, A. van Wieringen, and J. Wouters, "APEX 3: a multi-purpose test platform for auditory psychophysical experiments," *J Neurosci Methods*, vol. 172, no. 2, pp. 283–293, 2008, doi: 10.1016/j.jneumeth.2008.04.020.

[20] N. Das, W. Biesmans, A. Bertrand, and T. Francart, "The effect of head-related filtering and ear-specific decoding bias on auditory attention detection," *J Neural Eng*, vol. 13, no. 5, Sep. 2016, doi: 10.1088/1741-2560/13/5/056014.

[21] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: A review with applications," *Signal Processing*, vol. 96, no. PART A, pp. 1–15, Mar. 2014, doi: 10.1016/j.sigpro.2013.04.015.

[22] L. Huang and J. Wang, "Forecasting energy fluctuation model by wavelet decomposition and stochastic recurrent wavelet neural network," *Neurocomputing*, vol. 309, pp. 70–82, Oct. 2018, doi: 10.1016/j.neucom.2018.04.071.

[23] M. Rashid, M. Islam, N. Sulaiman, B. S. Bari, R. K. Saha, and M. J. Hasan, "Electrocorticography based motor imagery movements classification using long short-term memory (LSTM) based on deep learning approach," *SN Appl Sci*, vol. 2, no. 2, 2020, doi: 10.1007/s42452-020-2023-x.

[24] Y. Te Wu, T. H. Huang, C. Yi Lin, S. J. Tsai, and P. S. Wang, "Classification of EEG Motor Imagery Using Support Vector Machine and Convolutional Neural Network," *2018 International Automatic Control Conference, CACS 2018*, 2019, doi: 10.1109/CACS.2018.8606765.

[25] R. R. Hallac, J. Lee, M. Pressler, J. R. Seaward, and A. A. Kane, "Identifying Ear Abnormality from 2D Photographs Using Convolutional Neural Networks," *Sci Rep*, vol. 9, no. 1, pp. 1–6, 2019, doi: 10.1038/s41598-019-54779-7.

[26] L. Tan, Y. Chen, T. C. Maloney, M. M. Caré, S. K. Holland, and L. J. Lu, "Combined analysis of sMRI and fMRI imaging data provides accurate disease markers for hearing impairment," *Neuroimage Clin*, vol. 3, pp. 416–428, 2013, doi: 10.1016/j.nicl.2013.09.008.