Driver Drowsiness Detection Using Gray Wolf Optimizer Based on Voice Recognition

Sarah S. Jasim[1], Alia K. Abdul Hassan[2], and Scott Turner[3]

[1]Department of IT, Technical College of Management-Baghdad, Middle Technical University, Baghdad, Iraq

[2]Department of Computer Science, University of Technology, Baghdad, Iraq

[3]School of Engineering, Design, and Technology, Church Christ Church University, Kent, UK

Abstract—**Globally, drowsiness detection prevents accidents. Blood biochemicals, brain impulses, etc., can measure tiredness. However, due to user discomfort, these approaches are challenging to implement. This article describes a voice-based drowsiness detection system and shows how to detect driver fatigue before it hampers driving. A neural network and Gray Wolf Optimizer are used to classify sleepiness automatically. The recommended approach is evaluated in alert and sleep-deprived states on the driver tiredness detection voice real dataset. The approach used in speech recognition is mel-frequency cepstral coefficients (MFCCs) and linear prediction coefficients (LPCs). The SVM algorithm has the lowest accuracy (71.8%) compared to the typical neural network. GWOANN employs 13-9-7-5 and 30-20-13-7 neurons in hidden layers, where the GWOANN technique had 86.96% and 90.05% accuracy, respectively, whereas the ANN model achieved 82.50% and 85.27% accuracy, respectively.**

Index Terms—**Drowsiness, Artificial neural network, Feature extraction, Gray Wolf Optimizer, Normalization, Mel-frequency cepstral coefficients, Linear prediction coefficients**

I. Introduction

Transportation safety officials have been concerned about distractions while driving as a significant contributor to traffic accidents (Dasgupta, et al., 2015; Zhang, et al., 2017). Such incidents necessitate research into on-board monitoring of the attentiveness level of the driver. Human alertness can be measured using various methods, including an electroencephalogram (EEG), ocular characteristics, blood samples, speech (Ooi, et al., 2016; Jasim and Hassan, 2022), and skin conductance. The EEG, electrooculography (EOG) signals, and blood biochemical techniques have been deemed the most accurate for determining the state of alertness. However, because it relies on direct physical contact, they are challenging to put into widespread practice (Alzu'bi, Al-Nuaimy and Al-Zubi, 2013; Huo, Zheng and Lu, 2016). Driving can be viewed as a road, vehicle, and driver system. Roads and cars are already more stable thanks to research, but the driver is still the weakest link. Therefore, it has been argued that determining drivers' emotions should be a present study emphasis (Zhang, 2019). Emotions are a complex phenomenon that is difficult to define. It is a combination of mental and bodily sensations induced by thoughts. Some emotions cause avoidance or approach.

In contrast to avoidance, approach conduct is accompanied by good emotions (happiness) (Ooi, et al., 2016). Drivers' emotions strongly influence their driving behavior, which is closely linked to road safety. Therefore, understanding and monitoring driver emotions are believed to promote better driving ethics and the need to develop emotion-based accident prevention systems. Exhaustion, rage, and stress driving are all linked to traffic crashes (Dasgupta, et al., 2015).

Since the early 1960s, Automatic Speech Recognition (ASR) has been the most researched topic in speech processing, translating human spoken words into computer recognized words. It mainly entails extracting speech information and storing it in a data model. ASR comprises two phases (training and testing). In the training phase, known speech is recorded and stored in a speech database as a parametric representation (Badr and Abdul-Hassan, 2020; Abdul-Hassan and Hadi, 2020). In the testing phase, the ASR system compares the retrieved features to trained reference templates to identify the utterance (Gamit and Dhameliya, 2015).

The sympathetic nerve system includes the hippocampus, amygdala, and cingulate gyrus; these organs respond to attentiveness, showing a link with emotion. We can also employ sympathetic nervous system information from sweat glands to recognize emotions (Yoshida, et al., 2014). The present study aims to identify the emotion of drivers during speaking in two scenarios (drowsy and non-drowsy). These authors recognize that a lone driver would not necessarily be

ARO-The Scientific Journal of Koya University
Vol. X, No. 2 (2022), Article ID: ARO.11000. 5 pages
DOI: 10.14500/aro.11000
Received: 6 June 2022; Accepted: 27 November 2022
Regular research paper: Published: 05 December 2022
Corresponding author's email: sara-sm@mtu.edu.iq

speaking whilst driving, but this could be partially remedied by the system that requires the driver to speak every so often. However, in my study took this challenge, which is to rely on the driver's voice and analyze it based on the energy property of the human voice because the level of energy of his voice in a state of drowsiness decreases, therefore, this factor is essential in analyzing the characteristic of drowsiness.

The paper is organized as follows. Section II discusses the related work. Section III presents the proposed system and the voice-based algorithm. Section IV presents the results and discussion of the experiment. Finally, the paper is concluded in Section V.

II. Related Work

Recent advances in artificial intelligence (AI) have revolutionized image, signal, and voice detection (Yu, Wang and Zhang, 2021). Here are some recent works on emotion recognition:

An overarching paradigm for identifying drowsiness states using prosody, articulation, and other speech quality-related aspects were described by Krajewski, Batliner and Golz, 2009. Speech, speaker, and affect acoustic recognition were all used (frame-level-based speech features). In addition, features from perceptual and signal processing (such as fundamental frequency [FF], intensity [I], pause patterns [PP], formants [F], and cepstral coefficients [CC]) were utilized in the feature computation, used four different classifiers (SVM, KNN, MLP, and DT). The SVM classifier was 86.1% accurate.

Dasgupta, et al. (2015) introduced a voice and vision-based sleepiness detection system suitable for car drivers. It was done using an image-based algorithm that computes PERCLOS and a voice-based method that computes voiced speech ratio (VSR). In addition, face detection was performed with 94% accuracy using a support vector machine classifier (SVM).

Ooi, et al. (2016) used electrodermal activity (EDA) to study stress and anger as critical emotions. A simulated driving assignment containing neutral, stress, and anger scenarios was designed for emotional stimulation. First, the acquired EDA signals were filtered bandpass at 0.5–2 Hz and Fourier transformed. After that, the power spectral density mean, median, and standard deviation were calculated, and an f-test of two samples was used to examine the statistical significance of the data. A support vector machine classifier (SVM) was employed for classification and achieved 85% each for neutral stress, neutral anger, and 70% for stress anger, respectively.

Wankhade and Kharat (2017) presented a novel approach for emotion recognition using EEG signals and using a two-tier classifier based on both K-nearest neighbor (K-NN) and neural network (NN) classifiers. The proposed KNN-NN classifier achieved an accuracy of 97%.

Using an EEG-based emotion classification and a decision tree algorithm, Pane, et al., 2017, developed emotion categorization models for four states of mind: Happy, sad, angry, and relaxed. To divide the EEG signal into gamma, beta, alpha, and theta bands, researchers utilized an IIR bandpass filter with a Chebyshev type II window. Three algorithms were employed to classify emotions: Repeated Incremental Pruning to Produce Error Reduction (RIPPER), J4.8, and SVM. On average, RIPPER achieved 92.01% accuracy in the binary assessment of sad versus relaxed emotional state.

Greco, et al. (2019) investigated the possibility of combining electrodermal activity (EDA) and voice data to recognize human arousal levels during single effective word pronunciation. The support vector machine with a recursive feature elimination (SVM-RFE) was trained and tested on three datasets, using the two channels (speech and EDA) independently and combined. The results suggest that combining EDA and speech data improve the classifier marginally (+11.64%).

The three methods presented by Martin, et al., 2021, are based on the acoustic quality of the speech, reading faults, and a completely new method that relies on Automatic Speech Recognition systems' errors. Unweighted average recall (UAR) values of 74.2% were achieved with a classification system built on a principal components analysis (PCA) and logistic regression (LR) (Table I).

III. The Proposed System

At present, researchers in pattern recognition and machine learning are increasingly recognizing the importance of swarm optimization for reducing data dimensionality and improving classification accuracy. Features can be selected using various swarm optimization techniques, such as particle swarm optimization (PSO) or ant colony optimization (ACO). New swarm-inspired algorithms, such as Gray Wolf Optimization (GWO), have recently appeared (Hassan and Mohammed, 2020). GWO is a novel metaheuristic algorithm based on how Gray Wolves behave in the wild. The group used to have two species: Wolf males and wolf females to keep order (Xu, et al., 2019). Any pack has a social hierarchy that looks like this:

1. Alpha wolves (decide [α]) are dominating wolves. Wolves are at the forefront and represent the herd's leadership and the issuance of decisions.
2. Beta wolves (β) are second-level wolves. Beta's support the dominant decisions.

Table 1
Summary of Artworks Utilizing Hybrid Techniques

References	Hybrid features	Machine learning methods	Accuracy
Krajewski, Batliner and Golz, 2009	FF, I, PP, F, and CC	SVM	86.1%
Dasgupta, et al., 2015	VSR+PERCLOS	SVM	94%
Ooi, et al., 2016	EDA signals	SVM	85% and 70%, respectively
Wankhade and Kharat	EEG signals	KNN-NN	97%
Pane, et al., 2017	EEG signal	RIPPER	92.01%
Greco, et al., 2019	Speech+EDA	SVM-RFE	+11.64%
Martin, et al., 2021	AQS, RF, ASRse	PCA+LR	74.2%

3. Delta wolves (δ) are the third-level wolves who obey the alphas and betas.
4. The omegas wolves (ω) signify the pack's smallest alpha, beta, and delta wolves' scheme. The GWO algorithm ranks alpha wolves first, followed by beta and delta wolves. Therefore, omegas (ω) comprise this population cluster (Heidari and Pahlavani, 2017).

The general stages of applying GWO are as follows:

Start the Gray Wolf population Xi, in which I = 1, 2, 3, 4... n; Start a, A and C, where a is a vector which on linearly decreasing from 2 to 0, A and C are a coefficient matrix. N is the number of total iterations for optimization. The fitness of each candidate solution is computed through equations:

$$X(t+1) = X(t) - A.D\ldots\ldots\ldots \tag{1}$$

$$D = \left|C.X_p(t) - X(t)\right|\ldots\ldots\ldots \tag{2}$$

$$A = 2a.r_1 - a\ldots\ldots\ldots \tag{3}$$

Where, X and $X\alpha$ are $the\ first$ fitness $search\ agent,$ $X\beta\ is\ the\ second$ fitness $search\ agent$, and $X\delta\ is$ the third fitness $search\ agent$. The optimal Gray Wolf position ($X\alpha$) has the best fitness value f($X\alpha$).

The operation starts by randomly generating the Gray Wolf population Xi and repeating operations until the stop condition is not satisfied. The primary step is to modify all search agents (each current search agent's position through equation) using Equation (2.45) and find the fitness values of all modified search agents.

$$X(t+1) = \frac{(X\alpha + X\beta + X\delta\ )}{3}\ldots\ldots\ldots \tag{4}$$

The stop condition is represented by the maximum iteration until there is no update of the current search agent (best result).

A. Proposed Gray Wolf Optimizer (GWO) with artificial neural network (GWOANN) based drowsiness detection system

The proposed method Gray Wolf Optimizer with Artificial Neural Network (GWO-ANN) was used for training the speech signal features for these reasons:

- Finding the best starting weights that are valid proved to be a challenging task with other approaches.
- There was a concern about potential overfitting in neural networks
- The trend toward minimizing the mean square error in the training stage.
- The initial range of the initial weights may put some constraints on the search space.

To overcome the above problems, the proposed classifier method is the GWO-ANN based on the relationship between input and output using GWO. The method used a specific weight for neurons obtained by GWO. These weights are efficient in the training process. First, the ANN is trained with GWO algorithms to find the optimum weights and biases. The neural network then employs one-step-forward ANN. Finally, the fitness function (Equation 1) is used to evaluate the results. The total MSE is used as a fitness function.

First stage: Using GWO for ANN training:

Selection: Size of individuals (Packs),

Improvement: Largest iterations are identified,

Formation: Using back-propagation algorithm in NN, implementation: GWO in the following equation uses to get the better value of weights and bias,

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(T_i - Y_i\right)^2} \tag{5}$$

Where, n shows the number of test samples, T is the target value, and Y is the predicted output. It means a better model.

Restoring: Optimum of weights and bias,

Second stage: GWO used to optimize back-propagation algorithms for training ANN:

Selection: GWO outputs as the beginning of weights and bias,

Restoring: Optimum training of ANN model.

B. Speech-Based Method

The GWO-ANN described above represents the training model of this method. A total of 18 features of the recorded speech are obtained from the previous steps and then used for training. The size of input features differs here depending on the specified energy threshold, which neglects the silence segments. Finally, each frame is input to the proposed method for training with its class label (drowsy or not). The general steps of the proposed method for detecting the driver's drowsiness are explained in Fig. 1.

Dataset

In this work, a speech dataset was recorded through (human interaction) as a large vowel phonation, single words, connected speech, reading the speech, and impromptu speaking as all forms of oral communication. The samples were selected because they were common phrases and had a mixture of required frequency components. A Samsung mobile device was used for 16 subjects (eight men and eight women), and each person had two samples for 12 phrases (in the sleepy and normal states). Examples of the speech data are shown in Fig. 2 for the normal and abnormal datasets. Signal energy has been used with edge detection algorithms to distinguish between voiced and silent clips because the level of energy of a person's voice in drowsiness decreases. Therefore, this factor is essential in analyzing the characteristic of drowsiness. Each frame has several samples (ω), where, ω < n where n is all sample size. Frame-by-frame, the intensity of a speaker's words is measured. Each sample has been squared, and the total of all squared samples has been computed. The following is the estimated energy equation:

$$Energy = \sum_{i=1}^{\omega} x_i^2 \tag{6}$$

Feature extraction

It is essential to execute a standard pre-processing step to reduce noise (small frames) and remove noise from voice signals to improve speech recognition results; the intensity of the speaker's words is measured frame by frame. Each

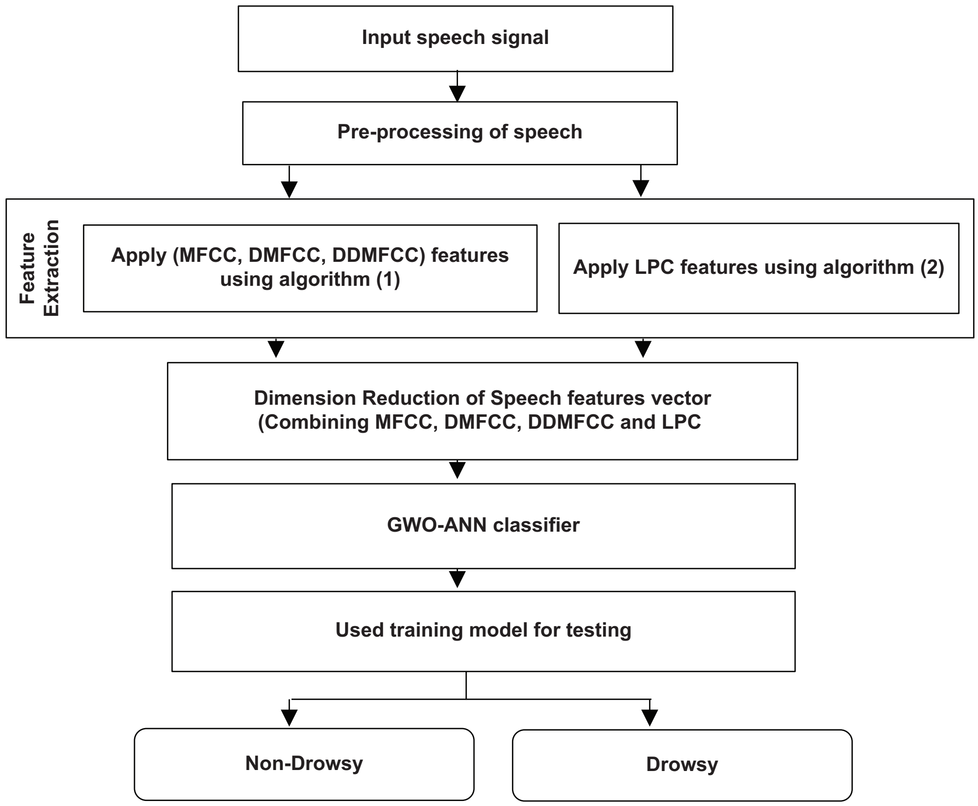

Fig. 1. The proposed speech drowsiness detection sub-system.

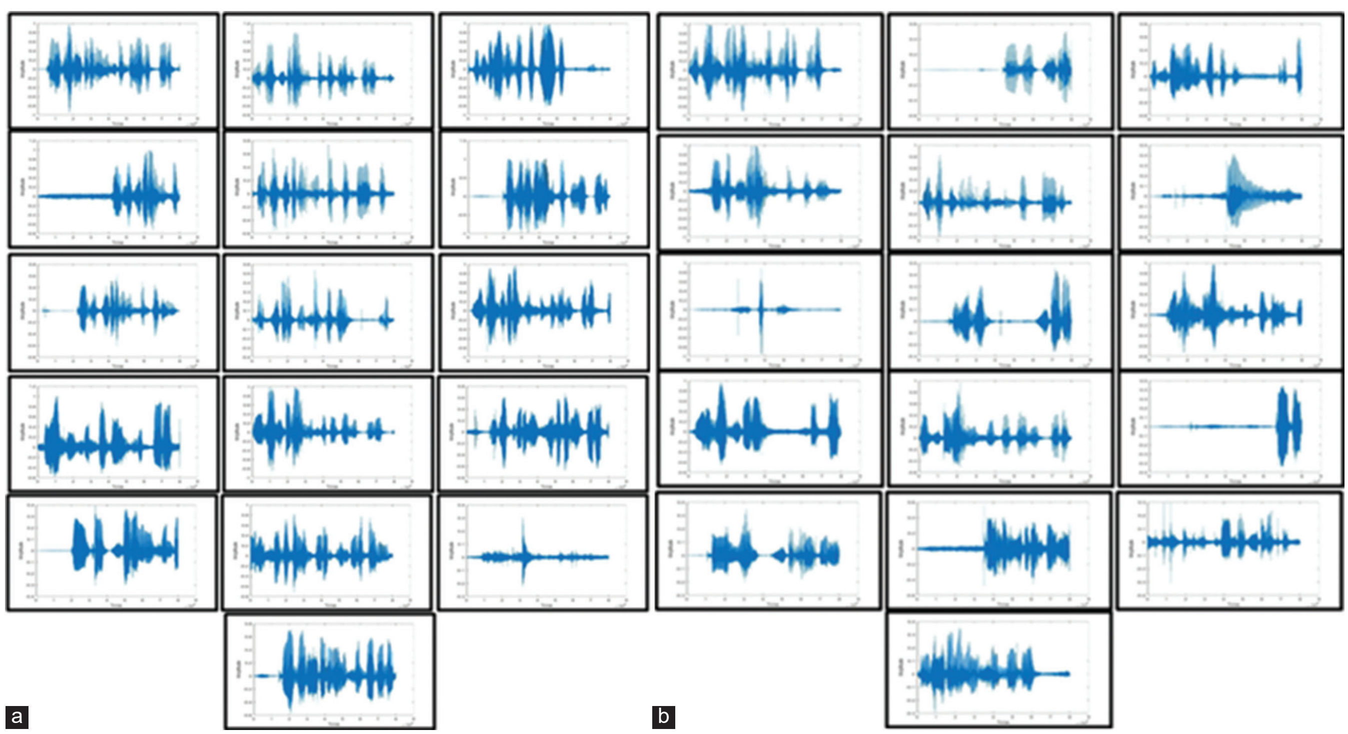

Fig. 2. (a) The samples of normal voice signal. (b) The samples of abnormal voice signal.

sample was squared, and the total of all squared samples was calculated using a specific threshold limit chosen by the experiment to reduce noise from a signal.

The speech stream is transformed into measurable values with differentiating characteristics. In this approach, two features are extracted from each utterance frame: Mel-frequency cepstrum coefficients (MFCCs) and linear predictive coding (LPC). The MFCCs characteristics exhibit promising performance in speech-related challenges, due to their capacity to accurately represent the speech spectrum's amplitude. On the other hand, LPCs exhibit qualities comparable to formant frequencies and are resistant to noise. These two groups of characteristics were combined because LPCs features are intended to be complementary to MFCCs features. Each frame will produce two distinct groupings of features: 42 MFCCs with their first and second derivatives and 12 LPCs. The number of MFCC features is chosen to be 14 dimensions for the first derivative and 14 dimensions for the second derivative. These dimensions provide the best results found through trial and error. The algorithms (1 and 2) show how this feature is extracted. The samples of these features explain in Tables II and III, respectively.

Dimension reduction of features

The total features of speech signals are represented by MFCC, delta of MFCC (1st derivative), delta-delta of MFCC (2nd derivative), and LPC. This feature is a total of 54 features for each frame. The dimension reduction is applied by getting each 54 features per frame together and applying statistical values for each block, such as mean, standard deviation, skewness, kurtosis, root mean squared, energy, power, peak amplitude, and crest factor, for all these features obtained the mean and standard deviation.

Algorithm (1): The MFCC feature extraction

Input: Speech (phonocardiogram) signal
Output: Total MFCC features, total delta MFCC features, total delta-delta MFCC features, and total class

Begin
Step 1: Assign no of speech files (normal and abnormal)
Step2: For each speech file
- Split into frames, phonocardiogram signals
- Apply humming windowing to frames
- Get spectrum by applying FFT to all frames
- Determine matrix for a Mel spaced filter bank
- Transform spectrum to Mel spectrum
- Obtain the MFCC vector for each frame by applying DCT (a feature I)
- Obtain delta MFCC (Features II)
- Obtain delta-delta MFCC (features III)

Step 3: Assign a class label for each frame
Step 4: Append all features into a new matrix called total MFCC features
Step 5: Append all features into a new matrix called total delta MFCC features
Step 6: Append all features into a new matrix called total delta-delta MFCC features
Step 7: Append class label into a new vector called total class
End Algorithm

Algorithm (2): The LPC feature extraction
Input: Speech (phonocardiogram) signal
Output: Total LPC features and total class
Begin
Step 1: Assign no of speech files (normal and abnormal)
Step 2: For each speech file
- Split into frames, phonocardiogram signals
- Apply humming windowing to frames
- Predicate frame and compute the error
- Obtain LPC coefficients

Step 3: Assign a class label for each frame
Step 4: Append all features into a new matrix called total LPC features
Step 5: Append class label into a new vector called total class
End Algorithm

The samples of these features are explained in Table IV; the remaining features are reduced to 18, replaced by the total features.

Classification of drowsiness

Only two states, namely, the state of sleepiness (1) and the state of alertness (0), are used to classify binary drowsiness. They employ machine learning technologies, for example, to identify driver drowsiness detection using machine learning methods (SVM, KNN, ANN, DT, etc.). The proposed approach (GWOANN) was developed for this study because it successfully solved problems of non-linearity, high-dimensional classifications, and even classification in speech signals.

Table II
(42) Features of MFCCs, DMFCC, and DDMFCC

#	fet1	fet2	fet3	fet4	….	fet13	fet14
1	3.577749	−9.94938	3.016517	0.643449	….	−0.26558	1.048812
2	4.207627	−9.54046	2.638826	0.571662	….	−0.30679	1.009354
3	4.412491	−8.99577	2.50445	0.42222	….	−0.21795	0.994256
4	4.342696	−8.77655	2.650918	0.226735	….	−0.2938	1.037622
5	4.164026	−8.66088	2.47254	0.201869	….	−0.36561	0.933478
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
868	2.445446	−10.4559	3.48321	0.270007	….	0.203008	−0.411
#	fet1	fet2	fet3	fet4	….	fet13	fet14
1	0.477033	−1.32658	0.402202	0.085793	….	−0.03541	0.139842
2	0.918792	−2.267	0.653495	0.140566	….	−0.06746	0.239462
3	1.247611	−2.81677	0.79891	0.156359	….	−0.07744	0.303424
4	1.420042	−3.03746	0.880373	0.132012	….	−0.09027	0.340026
5	1.423893	−2.95017	0.849688	0.096793	….	−0.10288	0.328155
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
868	−0.1662	0.245301	−0.14333	−0.11122	….	−0.00524	−0.14969
#	fet1	fet2	fet3	fet4	….	fet13	fet14
1	0.063604	−0.17688	0.053627	0.011439	….	−0.00472	0.018646
2	0.170209	−0.43492	0.127353	0.027322	….	−0.01254	0.045912
3	0.29003	−0.69071	0.198684	0.040624	….	−0.01943	0.073725
4	0.391254	−0.88202	0.254247	0.045468	….	−0.02546	0.096305
5	0.445657	−0.96045	0.276373	0.041216	….	−0.03016	0.105967
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
868	−0.11303	0.109999	−0.08671	−0.03681	….	−0.00865	−0.06352

Table III
(12) Features of LPC

#	fet1	fet2	fet3	fet4	….	fet11	fet12
1	−1.56953	0.375995	0.196997	0.124915	….	0.097282	−0.01144
2	−1.65702	0.435827	0.259976	0.134094	….	0.060285	−0.01609
3	−1.54567	0.317511	0.239578	0.13383	….	0.066661	0.009763
4	−1.81261	0.584168	0.331195	0.122206	….	−0.01225	−0.02677
5	−1.77142	0.546968	0.331306	0.098397	….	0.014741	−0.03164
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
868	−1.5446	0.514186	0.085213	−0.07562	….	0.093095	−0.02346

TABLE IV
DIMENSION REDUCTION OF FEATURES

#	fet1	fet2	fet3	fet4	….	fet17	fet18
1.	−0.04673	1.484917	0.392288	0.391199	….	0.298653	0.070647
2.	−0.04347	1.498126	0.27122	0.225178	….	0.330448	0.135811
3.	−0.05821	1.513109	0.305602	0.274714	….	0.322942	0.14225
4.	−0.02647	1.520109	0.24549	0.214877	….	0.340287	0.167434
5.	−0.03354	1.499883	0.251718	0.197142	….	0.346273	0.163717
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
.	.	.	.	.	….	.	.
1840	−0.04655	1.638488	0.167963	0.167685	….	0.27592	0.102743

Support vector machine classifier

SVM is a prevalent machine learning technique that produces a high degree of accuracy that was used successfully in many areas (Okfalisa, 2021; Hassan and Alawi, 2017). An efficient classifier changes the “kernel” function used to carry out the classifications in a linear and non-linear manner (Zeng, et al., 2018) (Tao, Sun and Sun, 2018). The fundamental principle of SVM is to create separate hyperplanes for classification in high-dimensional spaces. The hyperplane that is the most distant from each class with a kernel function to the closest training data point achieves optimal separation. SVM has, therefore, been widely used in the field of EOG in recent years (Chen, Wang, and Hua, 2018). In SVM, four standard kernels are used, linear, polynomial, RBF, and sigmoid. The proposed method uses the first three, and specific results are obtained after the classification. Then, recognition (drowsy or non-drowsy) is performed (Salam and Hassan, 2019).

Artificial neural networks (ANNs)

A neural network is a collection of processing units designed to produce human-like outcomes (ADWAN, et al., 2022; Hassan and Hadi, 2016). On the other hand, one subdivision performs its computations and sends them to a second (Yusiong, 2012; Abdulwahed, 2018; Rashid and Abdullah, 2018). The network has three layers: Input, hidden, and output (Huang, Cheng, and Zhang, 2022; Nwobi-Okoye and Ochieze, 2018). ANN models have been developed for a long time. Even when people or other computer algorithms cannot find patterns and trends in complex or confusing data, neural networks can disclose them; ANN has good classification and performance approximation performance (Abed, 2019). As a result, a qualified neural network can be seen as an “Expert” in the knowledge category for which it has received only assessment. This expert can forecast new situations of concern and respond to “what if” queries (Bati and Adam, 2006; Hassan and Jasim, 2010).

The proposed Gray Wolf Optimizer with artificial neural network (GWO-ANN)

In the starting, GWO algorithms are used to train ANN to identify the best weights and biases, as shown in Algorithms 3. An effective back-propagation network is then used to equip the neural network. Finally, check to see if the network has reached the correct error rate or if the number of generations required to finish the algorithm has been surpassed. Fig. 1 depicts the proposed method. For the depiction of the ANN, consider using a two-layered network:

Algorithm (3): Find the best ANN weight using GWO

Input: Extracted speech feature vector (18 features) for normal/abnormal and class label
Output: Best ANN weight

Begin
Step 1: Initialized GWO parameters (pop size and initial weights)
Step 2: Specifying the number of hidden layers of ANN
Step 3: Specifying the same number of each node in each layer in the algorithm ANN
classifier
Step 4: For I = 1 to pop size
- Randomly generating initial weights for all neurons in ANN
- Convert weights into vector
- Find the fitness function for all populations that MSE represents
- Sort MSE of pops in ascending order
- Find best weights to population (minimum)
Step 4: Find the alpha (α) wolf, beta (β) wolf, and delta (δ) wolf (value and position)
Step 5: Find the average X av of X_α, X_β, and X_δ,
Step 6: Update (X_α, X_β, and X_δ)
Step 7: Depending on the new values (X_α, X_β, and X_δ), update all weights in pops
Step 8: Reorder the weights of neurons
Step 9: Find the MSE of inputs with new weights as a fitness function
Step 10: Sort the fitness function in ascending order
Step 11: Find new best for (α), (β), and (δ) wolves (X_1, X_2, and X_3, respectively)
Step 12: If the new best is less MSE than the old best, replace; otherwise, continue
Step 13: Applying steps from 8 to 13 until Max-iteration No.
Step 14: Return the best ANN weights (optimum weights)
End

$$\sum_{k=1}^{N} w_k f\left(\sum_{i=1}^{m} w_i x_i + b\right) \tag{7}$$

Where, N represents the number of neurons in the hidden layer, w represents the weight of a net, b represents the bias value, and f represents the activation function of each neuron in this example. The following is the mean squared error (MSE):

$$MSE = \frac{\sum_t^z \left(d^t - y^t\right)^2}{z} \tag{8}$$

If d is the desired output and y is the actual output, z is the number of testing outcomes, T is the goal value, and Y is the projected output, then Equation (3) indicates a superior model.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(T_i - Y_i\right)^2} \tag{9}$$

IV. EXPERIMENTS RESULTS AND DISCUSSION

The recorded dataset has been used in current experiments to detect driver drowsiness. The following stages are included in each experiment:

- Pre-processing.
- Features extraction.

- Normalization features.
- Classification by GWOANN.

Discovering the parameters that work well for each classifier is critical. There were a set of predetermined values used. Three trials were run for each classifier, with a different training-testing proportion. These were 90%-10%; 80-20%; and 70-30%. This research tested two classification categories (drowsy or not).

A. Results for SVM classifier

The SVM classifier score is expected to have higher when the following parameters are used, outlier = (0.001), K-Fold = (2), TrPer = (90%, 80%, and 70%), counter = 1000, and kernel = Linear. Table V shows the results of applying the SVM classifier. It can be noticed that the SVM classifier achieved the best accuracy when 90-10% training-testing percentages.

Table VI where shows that the accuracy decreased when using the same parameters outlier = (0.001), K-Fold = (2), TrPer = (90%, 80%, and 70%), Counter = 1000, but the kernel is used polynomial.

B. Using ANN Classifier

It is used to classify the driver's sleepiness using MLP. Unsupervised learning is used to build an ANN that can classify driver drowsiness. The parameters used in Table VII determine the standalone ANN classifier.

Table V
Accuracy of the SVM Model Training Data (90%, 80% and 70%) with Kernel of Linear

Run number	Training 90%, testing 10%	Training 80%, testing 20%	Training 70%, testing 30%
	Acc.	Acc.	Acc.
1	72.83%	73.37%	71.56%
2	68.48%	70.65%	74.28%
3	69.57%	66.58%	70.83%
4	71.20%	69.84%	71.38%
5	71.74%	71.20%	69.38%
6	71.20%	69.02%	72.28%
7	72.83%	77.17%	69.57%
8	73.91%	72.28%	74.64%
9	70.11%	71.74%	71.92%
10	76.46%	71.20%	71.20%

Table VI
Accuracy of the SVM Model Training Data (90%, 80% and 70%) with Kernel of polynomial

Run number	Training 90%, testing 10%	Training 80%, testing 20%	Training 70%, testing 30%
	Acc.	Acc.	Acc.
1	67.39%	65.49%	65.40%
2	60.33%	68.21%	64.49%
3	60.33%	59.51%	67.39%
4	68.48%	62.23%	66.67%
5	60.33%	66.30%	64.86%
6	65.76%	63.86%	64.49%
7	65.22%	68.75%	59.96%
8	59.24%	68.21%	69.20%
9	66.30%	64.40%	67.93%
10	63.59%	66.30%	58.88%

Tables VIII and IX illustrate the results of utilizing a standalone ANN classifier with 90-10%, 80-20%, and 70-30% training-testing percentages. The ANN classifier had the best accuracy when the training-testing percentages were 70-30%. Table III shows that the accuracy percentages have reduced to 90-10% and 80-20%.

C. Using GWO-ANN

GWO swarm is used to find an accurate weights of ANN. The suggested method involves first training the network with starting weights and biases and then updating the findings. Hence, the global optima back-propagation is sped up. The suggested technique includes weights and biases. It is based on RMSE. This classification is more accurate than the traditional ANN classifier, as shown in Table X.

Tables XI and XII demonstrate the outcomes of implementing the proposed hybrid approach (GWO-ANN) with 90-10%, 80-20%, and 70-30% training-testing

Table VII
Parameters Used for a Standalone ANN Model

Parameters	Value
Max iteration	15,000
Number of (neurons in i/p layer)	18
Number of (hidden layer)	4
Number of (neurons in each hidden layer)	(13, 9, 7, 5) and (30, 20, 13, 7)
Number of (neurons in o/p layer)	2

Table VIII
The Results Provided by Standalone ANN in Case the No. of Neurons in the Hidden Layer Are (13, 9, 7, 5)

No. of hidden of neurons	13, 9, 7, 5		
Run number	Acc. (90-10%)	Acc. (80-20%)	Acc. (70-30%)
1	80.43%	80.98%	77.72%
2	86.41%	80.98%	78.80%
3	83.70%	82.07%	80.62%
4	78.26%	80.16%	78.99%
5	86.41%	84.24%	78.80%
6	84.78%	77.45%	77.36%
7	84.24%	81.79%	79.17%
8	81.52%	82.88%	78.44%
9	79.35%	81.52%	82.61%
10	79.89%	81.25%	81.16%

Table IX
The Results Provided by Standalone ANN in Case the No. of Neurons in the Hidden Layer Are (30,20,13,7)

No. of hidden of neurons	30,20,13,7		
Run number	Acc. (90-10%)	Acc. (80-20%)	Acc. (70-30%)
1	86.41%	82.61%	83.51%
2	86.96%	85.05%	79.35%
3	86.96%	85.60%	80.62%
4	83.70%	82.07%	82.25%
5	85.33%	79.08%	79.89%
6	82.07%	83.42%	83.33%
7	88.04%	86.14%	80.80%
8	83.15%	83.70%	84.24%
9	85.33%	85.05%	83.70%
10	84.78%	82.88%	79.71%

percentages. The GWO-ANN approach, once again, achieved balanced accuracy in all percentages of training and testing. To avoid overfitting, the algorithm's lower number of parameters was used; it achieved higher simplicity and reasonably lowered the risk of overfitting. They observed that the experiment determined the number of neurons in the hidden layers.

Tables XIII show the best, worst, mean, and standard deviation of three classifiers employing 90-10%, 80-20%, and 70-30% training-testing percentages with the number of neurons in hidden layers (13, 9, 7, 5) and (30,20,13,7). In 90-10% training-testing percentages, the proposed method's STD value (0.01650) was the lowest compared to the STD value of the ANN classifier (0.01793). On the other hand, compared to the ANN classifier and SVM, the value of best, worst, and mean (93.48%, 87.50%, and 90.05%) achieved the highest value. Thus, the proposed approach (GWO-ANN) is stable and specific.

In addition, all classifications considered in this work have reasonable precision, but the GWO-ANN classification reaches 90.05% with the highest rating accuracy. Therefore, the proposed solution may be essential for future studies or future "systems" vehicles as reference work. Table XIV

Table X
Parameters Based on GWO and ANN

Parameters based on GWO	
Parameters	Value
Iteration no.	100
Population size	100
Parameters based on ANN	
Parameters	Value
Number of (neurons in i/p layer)	18
Number of (hidden layer)	4
Number of (neurons in each hidden layer)	(13, 9, 7, 5) and (30, 20, 13, 7)
Number of (neurons in o/p layer)	2
Max iteration	15000

Table XI
The Results Provided by GWO-ANN in Case the No. of Neurons in the Hidden Layer Are (13, 9, 7, 5)

No. of hidden of neurons	13, 9, 7, 5		
Run number	Acc. (90-10%)	Acc. (80-20%)	Acc. (70-30%)
1	85.33%	87.50%	84.06%
2	87.50%	85.87%	85.51%
3	86.41%	86.14%	85.51%
4	88.04%	86.68%	86.41%
5	88.04%	87.23%	84.42%
6	85.87%	85.33%	82.61%
7	86.96%	85.33%	84.42%
8	86.96%	84.78%	86.96%
9	88.04%	81.25%	86.78%
10	86.41%	87.50%	84.06%

Table XII
The Results Provided by GWO-ANN in Case the No. of Neurons in the Hidden Layer Are (30,20,13,7)

No. of hidden of neurons	30,20,13,7		
Run number	Acc. (90-10%)	Acc. (80-20%)	Acc. (70-30%)
1	89.67%	90.22%	87.50%
2	90.76%	87.50%	87.50%
3	89.13%	88.59%	87.14%
4	88.04%	85.60%	87.50%
5	91.30%	89.13%	86.59%
6	89.67%	87.77%	86.59%
7	93.48%	88.32%	88.41%
8	91.30%	89.67%	89.31%
9	89.67%	87.23%	85.69%
10	87.50%	88.04%	87.32%

Table XIII
The Results of Classifiers Models

No. of hidden of neurons	Name classifier	(90% training, 10% testing)			
		Best Acc.	Worst Acc.	Mean	STD.
13, 9, 7, 5	ANN	86.41%	78.26%	82.50%	0.02832
	GWO-ANN	88.04%	85.33%	86.96%	0.00908
30,20,13,7	ANN	88.04%	82.07%	85.27%	0.01793
	GWO-ANN	93.48%	87.50%	90.05%	0.01650
------------	SVM	76.50%	68.50%	71.80%	0.02188
No. of hidden of neurons	**Name classifier**	**(80% training, 20% testing)**			
		Best Acc.	Best Acc.	Best Acc.	Best Acc.
13, 9, 7, 5	ANN	84.24%	77.45%	81.33%	0.01684
	GWO-ANN	87.50%	81.25%	85.76%	0.01758
30,20,13,7	ANN	86.14%	79.08%	83.56%	0.01974
	GWO-ANN	90.22%	85.60%	88.21%	0.01252
------------	SVM	77.20%	66.60%	71.30%	0.02645
No. of hidden of neurons	**Name classifier**	**(70% training, 30% testing)**			
		Best Acc.	Worst Acc.	Mean	STD.
13, 9, 7, 5	ANN	82.61%	77.36%	79.37%	0.01541
	GWO-ANN	86.96%	82.61%	85.07%	0.01327
30,20,13,7	ANN	84.24%	79.35%	81.74%	0.01772
	GWO-ANN	89.31%	85.69%	87.36%	0.00948
------------	SVM	74.60%	69.40%	71.70%	0.01635

Table XIV
Improvement Rate of GWO-ANN Over Standalone ANN

Classifier	90-10%		
	IR of best	IR of worst	IR of mean
Hidden neurons in ANN (13, 9, 7, 5)	2%	9%	5%
Hidden neurons in ANN (30,20,13,7)	6%	7%	6%
SVM	22%	28%	25%
Classifier	**80-20%**		
	IR of best	IR of worst	IR of mean
Hidden neurons in ANN (13, 9, 7, 5)	4%	5%	6%
Hidden neurons in ANN (30,20,13,7)	5%	8%	6%
SVM	17%	29%	24%
Classifier	**70-30%**		
	IR of best	IR of worst	IR of mean
Hidden neurons in ANN (13, 9, 7, 5)	5%	7%	7%
Hidden neurons in ANN (30,20,13,7)	6%	8%	7%
SVM	20%	23%	22%

Table XV
Comparative Analysis of the Recorded Dataset to Other Recent State-of-the-Arts

Method	Accuracy %
Krajewski, Batliner and Golz, 2009 SVM	86.1
Ooi, et al., 2016 SVM	85
Greco, et al., 2019 SVM-RFE	+11.64
Martin, et al., 2021 PCA+LR	74.2
Proposed ANN	85.27
Proposed GWO-ANN	90.05

compares the proposed method's improvement rate with the ANN classifier in 90-10%, 80-20%, and 70-30% training-testing percentages.

$$Improvement\,rate = \frac{A1 - A2}{A2} \tag{10}$$

Where, A1 represents the accuracy of the proposed algorithm and A2 represents the accuracy of another classifier.

D. Comparative analysis of results

The results of the ANN and GWOANN algorithms are compared to the findings of previous works utilizing the recorded dataset for each work. Table XV displays the outcomes of this evaluation. The proposed paper's results were very accurate compared to those of other papers.

V. Conclusion

This paper presents a system for drowsiness detection of vehicle drivers based on voice analysis by monitoring the alertness level periodically. The drowsiness detection is less accurate in a noisy environment, and it is a challenge to get high accuracy in detecting the abnormal status. The performance of the ANN classification method is enhanced by the swarm intelligent GWO. The method is based on speech signals for drowsiness detection through voice. The results of the experiments showed that efficiency is achieved when using a neural network with (4) hidden layer networks of (13, 9, 7, 5) and (30,20,13,7) neurons. The GWOANN technique had an 86.96% and 90.05% accuracy. The method is flexible in combining with other detection methods to build an integrated system for actual use. A discrete wavelet transform could be a future direction to consider.

References

Abdul-Hassan, A.K. and Hadi, I.H., 2020. A proposed authentication approach based on voice and fuzzy logic. In: *Recent Trends in Intelligent Computing, Communication and Devices.* Springer, Berlin, Heidelberg.

Abdulwahed, M.N., 2018. Analysis of image noise reduction using neural network. *Engineering and Technology Journal,* 36, pp.76-87.

Abed, I.S., 2019. Lung Cancer Detection from X-ray images by combined Backpropagation Neural Network and PCA. *Engineering and Technology Journal,* 37, pp.166-171.

Adwan, I., Milad, A., Abdullah, N.H., Widyatmoko, I., Mubaraki, M., Yazid, M.R. and Yusoff, N.I., 2022. Predicting asphalt pavement temperature by using neural network and multiple linear regression approach in the Eastern Mediterranean region. *Journal of Engineering Science and Technology,* 17, pp.0015-0032.

Alzu'bi, H.S., Al-Nuaimy, W. and Al-Zubi, N.S., 2013. EEG-based driver fatigue detection. In: *2013 Sixth International Conference on Developments in eSystems Engineering*. IEEE, New Jersey, United States. pp.111-114.

Badr, A.A. and Abdul-Hassan, A.K., 2020. A review on voice-based interface for human-robot interaction. *Iraqi Journal for Electrical and Electronic Engineering*, 16, pp.91-102.

Bati, A.F. and Adam, N.E., 2006. Hybrid neuro-genetic based controller of power system. *Iraqi Journal of Computers, Communication, Control and Systems Engineering,* 6, pp.1-115.

Chen, J., Wang, H. and Hua, C., 2018. Assessment of driver drowsiness using electroencephalogram signals based on multiple functional brain networks. *International Journal of Psychophysiology,* 133, pp.120-130.

Dasgupta, A., Kabi, B., George, A., Happy, S. and Routray, A., 2015. A drowsiness detection scheme based on fusion of voice and vision cues. *arXiv preprint arXiv:1509*.

Gamit, M.R. and Dhameliya, K., 2015. Isolated words recognition using MFCC, LPC and neural network. *International Journal of Research in Engineering and Technology,* 4, pp.146-149.

Greco, A., Marzi, C., Lanata, A., Scilingo, E.P. and Vanello, N., 2019. Combining electrodermal activity and speech analysis towards a more accurate emotion recognition system. In: *41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* IEEE, New Jersey, United States, pp.229-232.

Hassan, A. and Hadi, M., 2016. Sense-based information retrieval using artificial bee colony approach. *International Journal of Applied Engineering Research,* 11, pp.8708-8713.

Hassan, A.K. and Alawi, M., 2017. Proposed handwriting Arabic words classification based on discrete wavelet transform and support vector machine. *Iraqi Journal of Science,* 58, pp.1159-1168.

Hassan, A.K. and Jasim, S.S., 2010. Integrating neural network with genetic algorithms for the classification plant disease. *Engineering and Technology Journal,* 28, pp.686-702.

Hassan, A.K. and Mohammed, S.N., 2020. A novel facial emotion recognition scheme based on graph mining. *Defence Technology,* 16, pp.1062-1072.

Heidari, A.A. and Pahlavani, P., 2017. An efficient modified grey wolf optimizer with Lévy flight for optimization tasks. *Applied Soft Computing,* 60, pp.115-134.

Huang, X., Cheng, C. and Zhang, X.B., 2022. Machine learning and numerical investigation on drag reduction of underwater serial multi-projectiles. *Defence Technology,* 18, pp.229-237.

Huo, X.Q., Zheng, W.L. and Lu, B.L., 2016. Driving fatigue detection with a fusion of EEG and forehead EOG. In: *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, New Jersey, United States. pp.897-904.

Jasim, S.S. and Hassan, A.K., 2022. Modern drowsiness detection techniques: A review. *International Journal of Electrical and Computer Engineering*, 12, pp.2986-2995.

Krajewski, J., Batliner, A. and Golz, M., 2009. Acoustic sleepiness detection: Framework and validation of a speech-adapted pattern recognition approach. *Behavior Research Methods,* 41, pp.795-804.

Martin, V.P., Rouas, J.L., Boyer, F. and Philip, P., 2021. Automatic Speech Recognition systems errors for accident-prone sleepiness detection through voice. In: *2021 29th European Signal Processing Conference (EUSIPCO)*, IEEE, New Jersey, United States. pp.541-545.

Nwobi-Okoye, C.C. and Ochieze, B.Q., 2018. Age hardening process modelling and optimization of aluminium alloy A356/Cow horn particulate composite for brake drum application using RSM, ANN and simulated annealing. *Defence Technology,* 14, pp.336-345.

Okfalisa, Handayani, L., Juwita, P.D., Affandes, M., Fauzi, S.S. and Saktioto., 2021. Coronary heart disease using support vector machine. *Journal of Engineering Science and Technology,* 16, p.16.

Ooi, J.S., Ahmad, S.A., Chong, Y.Z., Ali, S.H., Ai, G. and Wagatsuma, H., 2016. Driver emotion recognition framework based on electrodermal activity measurements during simulated driving conditions. In: *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, IEEE, New Jersey, United States. pp.365-369.

Pane, E.S., Hendrawan, M.A., Wibawa, A.D. and Purnomo, M.H., 2017. Identifying rules for electroencephalograph (EEG) emotion recognition and classification. In: *2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME),* IEEE, New Jersey, United States. pp.167-172.

Rashid, T.A. and Abdullah, S.M., 2018. A hybrid of an artificial bee colony, genetic algorithm, and neural network for diabetic Mellitus diagnosing. *ARO-The Scientific Journal of Koya University,* 6, pp.55-64.

Salam, M. and Hassan, A.A., 2019. Offline isolated Arabic handwriting character recognition system based on SVM. *International Arab Journal of Information Technology,* 16, pp.467-472.

Tao, P., Sun, Z. and Sun, Z., 2018. An improved intrusion detection algorithm based on GA and SVM. *IEEE Access,* 6, pp.13624-13631.

Wankhade, S.B. and Kharat, P.A., 2017. A novel two-tier classifier based on K-nearest neighbour and neural network classifier for emotion recognition using EEG signals. *International Journal of Latest Technology in Engineering, Management and Applied Science (IJLTEMAS),* 6, p.7.

Xu, L., Wang, H., Lin, W., Gulliver, T.A. and Le, K.N., 2019. GWO-BP neural network-based OP performance prediction for mobile multiuser communication networks. *IEEE Access,* 7, pp.152690-152700.

Yoshida, R., Nakayama, T., Ogitsu, T., Takemura, H., Mizoguchi, H., Yamaguchi, E., Inagaki, S., Takeda, Y., Namatame, M., Sugimoto, M. and Kusunoki, F., 2014. Feasibility study on estimating visual attention using electrodermal activity. *Proceedings of the International Conference on Sensing Technology*, 2014, pp.589-595.

Yu, X., Wang, S.H. and Zhang, Y.D., 2021. CGNet: A graph-knowledge embedded convolutional neural network for detection of pneumonia. *Information Processing and Management,* 58, p.102411.

Yusiong, J.P., 2012. Optimizing artificial neural networks using cat swarm optimization algorithm. *International Journal of Intelligent Systems and Applications,* 5, p.69.

Zeng, N., Qiu, H., Wang, Z., Liu, W., Zhang, H. and Li, Y., 2018. A new switching-delayed-PSO-based optimized SVM algorithm for diagnosis of Alzheimer's disease. *Neurocomputing,* 320, pp.195-202.

Zhang, F., Su, J., Geng, L. and Xiao, Z., 2017, Driver fatigue detection based on eye state recognition. In: *2017 International Conference on Machine Vision and Information Technology (CMVIT)*, IEEE, New Jersey, United States. pp.105-110.

Zhang, L., 2019. *Analysis of Machine Learning Algorithms for the Recognition of Basic Emotions: Data Mining of Psychophysiological Sensor Information*. Ulm Universität, Germany.