Iraqi Journal for Computer Science and Mathematics

Volume 6 | Issue 1 Article 13

2025

WAR Strategy Algorithm- based Hybrid Optimization for Accurate and Rapid Speech Recognition

Shahad Thamear Abd Al-Latief
College of Graduate Studies (COGS), Universiti Tenaga Nasional (National Energy University), Selangor, Malaysia, PT21299@student.uniten.edu.my

Salman Yussof
Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional (National Energy University), Selangor, Malaysia, salman@uniten.edu.my

Azhana Ahmad
College of Computing and Informatics, Universiti Tenaga Nasional (National Energy University), Selangor, Malaysia, azhana@uniten.edu.my

Saif Mohanad Khadim
College of Graduate Studies (COGS), Universiti Tenaga Nasional (National Energy University), Selangor, Malaysia, PE21093@student.uniten.edu.my

Ahmed Alkhayyat
Islamic University, Najaf, Iraq, ahmedalkhayyat85@iunajaf.edu.iq

Recommended Citation

Al-Latief, Shahad Thamear Abd; Yussof, Salman; Ahmad, Azhana; Khadim, Saif Mohanad; and Alkhayyat, Ahmed (2025) "WAR Strategy Algorithm- based Hybrid Optimization for Accurate and Rapid Speech Recognition," *Iraqi Journal for Computer Science and Mathematics*: Vol. 6: Iss. 1, Article 13.
DOI: https://doi.org/10.52866/2788-7421.1243
Available at: https://ijcsm.researchcommons.org/ijcsm/vol6/iss1/13

IRAQI JOURNAL FOR COMPUTER SCIENCE AND MATHEMATICS 2025;6:169–184

ORIGINAL STUDY

WAR Strategy Algorithm-based Hybrid Optimization for Accurate and Rapid Speech Recognition

Shahad Thamear Abd Al-Latief [a,*], **Salman Yussof** [b], **Azhana Ahmad** [c], **Saif Mohanad Khadim** [d], **Ahmed Alkhayyat** [e]

[a] College of Graduate Studies (COGS), Universiti Tenaga Nasional (National Energy University), Selangor, Malaysia
[b] Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional (National Energy University), Selangor, Malaysia
[c] College of Computing and Informatics, Universiti Tenaga Nasional (National Energy University), Selangor, Malaysia
[d] College of Graduate Studies (COGS), Universiti Tenaga Nasional (National Energy University), Selangor, Malaysia
[e] Islamic University, Najaf, Iraq

ABSTRACT

Speech recognition-based applications increased and developed as a result of artificial intelligence's rapid growth, particularly Machine Learning, which play a crucial role in many aspects of daily life, such as applications related to human-computer interaction, and natural language processing. The complexity and diversity of speech signals provides challenges in maximizing the rate of accuracy and efficiency of speech recognition systems. Hyperparameter tuning is a crucial step in machine learning that has a significant role in optimizing the performance and generalization by determining the optimal values for the model's hyperparameters. This paper employed the recently developed WAR Strategy optimization algorithm for optimizing the features related to the speech signal and tuning the hyperparameters of machine learning typical models for accurate and rapid speech recognition. Two types of features are extracted from the speech signal including the spectral feature using the Mel-Frequency Cepstral Coefficients (MFCCs) technique and the statistical features. Afterward these features are optimized using the WAR Strategy optimization algorithm to obtain the optimum features set that describe the speech signal important information. Finally, the hyperparameters of six classical machine learning models are tuned to serve as newly designed classifiers in the final classification phase of the proposed system. Three different language speech datasets are used to evaluate the proposed system (i.e. English, Arabic, Malaysian) to prove the high generalization property of the proposed system. The obtained recognition accuracy that was ranging from 98.38% to 100% in a training time between 0.001 to 19.8 second demonstrate the high effectiveness of the proposed speech recognition system in dealing with the many obstacles facing the recognition of speech signal within high accuracy, low resources requirements, and minimum training time.

Keywords: Speech recognition, Feature extraction, Feature optimization, Hyperparameters tuning, WAR strategy algorithm, Machine learning

1. Introduction

Communication between people is necessary for the sharing of information, expressing their feelings and needs, and transferring knowledge. The most fundamental, widespread, efficient, and natural form of human communication is speech, that relies on the vocal tract with the help of tongue, and muscles to produce it [1]. Speech is a dynamic and complex signal that undergoes continual changes in

Received 24 September 2024; revised 22 November 2024; accepted 29 December 2024.
Available online 27 March 2025

* Corresponding author.
E-mail addresses: PT21299@student.uniten.edu.my (S. T. A. Al-Latief), salman@uniten.edu.my (S. Yussof), azhana@uniten.edu.my (A. Ahmad), PE21093@student.uniten.edu.my (S. M. Khadim), ahmedalkhayyat85@iunajaf.edu.iq (A. Alkhayyat).

https://doi.org/10.52866/2788-7421.1243

both frequency spectrum and strength over time [2]. Currently, there is a strong emphasis on creating user-friendly platforms that allow humans to interact with computers utilizing their natural communication abilities. In which, people are so comfortable with speech and have the desire to interact with computers using it, rather than having to resort to primitive interfaces such as keyboards and pointing devices [3]. This has led to the appearance of the automatic speech recognition systems that are considered as one of the main fields of acoustical speech signal analysis [4]. Speech recognition is a fundamental method in human-machine interaction that aims to convert spoken words into text using algorithmic rules executed by machine software. However, the wide range of human speech patterns and dialects poses a major obstacle in creating a standalone speech recognition system [5]. Whereas the speech signal is subject to various fluctuations based on factors such the gender, societal pattern/dialect, speaking manner, and velocity. Moreover, the acoustic environment variations, like using different acquisition devices, noisy signals, mispronunciations, and speech overlapping can influence on the system effectiveness of a speech recognition [6]. Establishing an automatic speech recognition system capable of effectively addressing all the complexities associated with speech signals, while achieving optimal performance and minimizing time and resource requirements, poses a significant challenge. Currently, such speech recognition systems are available for a limited number of languages, from the total 6500 languages spoken worldwide [7]. The speech recognition is one of the problems in the field of pattern recognition, and many Machine Learning (ML) based models including the traditional and the advanced one like deep learning has been adopted widely in solving it and achieved an acceptable recognition rate [8]. However, these models have some limitations that have not been solved yet including the required time, the computational cost, the generalization, and the vast amount of data required for training [9]. There has been a recent uptick in the use of metaheuristic optimization methods to fine-tune ML for tasks like speech recognition, biometric identification, and natural language analysis. Due to, these algorithms can handle high-dimensional spaces problems and improve the search ability in finding the best parameters and features for representing the data, so reducing the time and improving the systems efficiency [10]. Many metaheuristic algorithms have been employed for either feature selection or hyperparameters tuning of ML models [11]. Feature selection is crucial in developing an efficient speech recognition system since spoken language encompasses numerous aspects that enable us to convey information beyond mere words [12]. Additionally, Finding the optimal values for the model's hyperparameters allows to further optimize the model's performance and get better outcomes from ML [13]. While there have been previous efforts in the subject of speech recognition, to our knowledge, there is a lack of substantial study on optimizing the extracted features and hyperparameters of the speech classification model. Researchers often encounter challenges while attempting to discern the pertinent characteristics from a feature vector with a high dimension and exclude the irrelevant or less significant features that have less impact on the performance to enhance the accuracy of the learning model. Moreover, determining the intended parameters of the ML model that need to optimize to make the model perform in more efficient manner is a job that required many experiences.

To this end, a new metaheuristic algorithm so called WAR Strategy is adopted in this paper in order to optimize the features extracted set and perform a hyperparameters tuning for six ML algorithms for effective speech recognition. The feature optimization within the WAR Strategy algorithm reduces the dimensionality and helps to identify the best set of features relevant to represent the speech signal. Moreover, tuning the hyperparameters of the ML makes the training time reduced along with the recognition time and overcomes the generalization problem. In which, the introduced system has achieved a high rate of accuracy in recognizing the speech signal in three different languages.

2. Related works

Recently, the metaheuristic optimization algorithms utilization has been extended in speech recognition problems. However, there is little research in using these algorithms in optimizing the recognition process of the speech in different languages. The main utilization of these optimization algorithms is either for selecting the best feature set representing the speech signal or tuning the classifier parameters. In [14], utilized a hybrid bioinspired algorithm that depends on two algorithms including the Artificial Bee Colony (ABC) and Particle Swarm Optimization (PSO) to select the best set of features. This system is mainly composed of three main phases including preprocessing, extracting and selecting the best features, and recognition using the Support Vector Machine (SVM) model. It has been evaluated on three datasets of natives of India categorized on names of fruits, animals, and a combined signal of recognition accuracies of 92.32%, 94.47%, and 91.55% and error rate of 0.076, 0.055, 0.084 sequentially. In [15], a proposed a hybrid optimization algorithm so

Table 1. The gap analysis of the related works.

Study	Methodology	Limitation
[14]	Hybrid algorithms (ABC) and (PSO) for features selection and SVM for classification	High computational complexity, time consuming, and poor generalization due it has not been tested on different language or big dataset.
[15]	Combined PPO and CSO for optimizing ANN weights	High computational cost, time consuming, and limited scalability
[16]	Adjust the inner layers and neurons of ANN using the Opposition ABC algorithm	The evaluation was done on a specific dataset with controlled conditions, so it has poor generalization
[17]	Uses the PSO for feature optimizing along with SVM for classification	Cannot be generalizable to larger vocabularies and did not address the variations in the speakers' voice or accent.
[18]	Combines BBA and LAHC for feature selection and random forest for classification	Limited generalizability beyond Indian languages, and lacks testing in noisy conditions.
[19]	Combines HS and NMR algorithms for features selection and applied the random forest for classification	Did not address the time required, has a computational complexity due to the use of two optimization algorithms and lacks testing in noisy data.
[20]	The parameters of HMM model is adjusted using the BFOA.	High computational cost, limited scalability to larger vocabulary tasks and lacks multilingual applications, so it shows poor generalization
[21]	The inner layers of DNN are adjusted using WOA	Limited testing with real-time, noisy, or low-resource language conditions
[22]	The RNN has been optimized GOA.	Shows a poor generalization, in which it lacks cross-linguistic adaptability and have not been tested on noisy data

called Predator-Inuenced Civilized Swarm optimization, which integrate the Civilized Swarm Optimization (CSO) in addition to Predator Prey Optimization (PPO) and use it to adjust the weights related to biases of the Artificial Neural Network (ANN). It has been tested using two datasets: TI-46 isolated spoken word, and a recorded Hindi numeral, with 0.321 Mean Square Error. In [16], the focus was on designing a speech recognition system that can recognize the vocabularies having many similar sounding words. First the structure of ANN is adjusted using the algorithm of Levenberg–Marquardt, second the inner layers and neurons are farther more adjusted using the Opposition ABC algorithm. 30 speech signals from 36 persons (16 females and 20 males) are considered for evaluation and give a 99.36% accuracy. In [17], uses the PSO to optimize the extracted features using the MFCC and apply the SVM for recognition aiming to minimize the complexity of computing and processing time. the utilized dataset is composed of voice signal commands in Brazilian Portuguese language. The system has achieved 92% for actions commands and 99% for digits commands. While the execution time of the system with one speaker and a 2×2 matrix was equal to 3.6 seconds. In [18], a new feature selection procedure has been developed by hybridizing Binary Bat Algorithm (BBA) with Late Acceptance Hill-Climbing (LAHC). The main aim is to develop a model having a reduced complexity and time for identifying different Indian languages. The Random Forest (RF) classier has been adopted and achieved a 92.35% accuracy on Indic TTS database and 100% accuracy on the Indic Speech database. In [19], present a system to identify the spoken language type to so can be adopted in the applications related to Human-Computer Interaction (HCI). With the use of the two algorithms including Harmony Search (HS) and Naked Mole-Rat (NMR) the. It has been tested on three datasets: CSS10, VoxForge, Madras, using many ML classifiers. The random forest has achieved the highest results which equal to 99.89%, 98.22%, and 99.75% on the three datasets. In [20] the parameters of the Hidden Markov Model (HMM) including observation and transition probabilities are adjusted with the use of algorithm of Bacterial Foraging Optimization (BFOA). The results exhibit a high improvement of accuracy rate at signal/noise ratios of 15 dB for Arabic speech with a recognition rate equal to 92%. The inner layers along with the neurons of Deep Neural Network (DNN) in [21], are adjusted using the Whale Optimization algorithm (WOA). The main goal is to reduce the computational time when recognizing a dataset having large vocabularies. The validation process utilized real-time data captured under uncontrolled conditions for both isolated and continuous signals. The system shows accuracy rates equal to 99.6%, and 98.1% for isolated and continuous signals in sequence. In [22], the Recurrent Neural Network (RNN) has been optimized for recognizing the Marathi language using Grasshopper Optimization Algorithm (GOA). It has been tested on a dataset recorded using many individuals and shows a 96% 12 accuracy of recognition. Table 1, summarize the gap analysis of the related works.

Taking into account the aforementioned related works of using the metaheuristic optimization algorithms for speech recognition, the main problems are the time, generalization, and noisy data handling. In this paper a new speech recognition system is introduced that depends on the use of the newly developed WAR Strategy algorithm within ML algorithms with the following contributions:

- Apply the WAR Strategy algorithm to optimize the features extracted from speech signals to acquire the most realistic and powerful feature set to overcome the variations and noise in speech signal.
- Utilize the WAR Strategy optimization algorithm to optimize six classical ML models by tuning the hyperparameters for accurate speech recognition with a minimized training time.
- Assess the suggested speech recognition system's excellent generalization capacity on three publicly available speech datasets in three languages: English, Arabic, and Malaysian, that exhibit significant differences.

The following parts of this paper are organized in the following order: In Section 3, an overview of the WAR Strategy optimization method is given. The proposed speech recognition system phases include extraction of features, feature optimization, hyperparameters optimization of six classifiers, and classification, are outlined in Section 4. In Section 5 the outcomes obtained by implementing the suggested approach on three speech datasets of variant languages from three case studies, along with a discussion of these outcomes. Section 6 presents a succinct of the conclusions and potential future research.

3. WAR strategy optimization algorithm

Metaheuristics are a collection of optimization algorithms that have been developed to address the challenging and time-consuming problems in numerous fields and applications of today's world. The functionality of these algorithms is basically inspired by the natural or social phenomena and this what makes it offer an effective and adaptable strategies for addressing challenges that traditional optimization methods may find challenging, whereas the main intent of the metaheuristic algorithms is to maximize the systems performance by identifying the optimal or near-optimal solutions in a specific problem domain. many problems have been solved and optimized using these algorithms such as problems related to scheduling, routing, optimizing of function, features selection, and tuning of ML hyperparameters [23]. The WAR Strategy Optimization Algorithm is a recently developed metaheuristic algorithm that has been designed to address the various optimization challenges. Its main source of inspiration is derived from the strategic maneuvers employed by armed units during times of conflict in which every soldier independently moves towards the most favorable value. This algorithm incorporates two generally acknowledged military tactics, specifically attack and defensive strategies in which the position of each soldier in the battlefield is adjusted based on the currently implemented strategy. To improve the algorithm's convergence and durability, a new method for updating weights and a strategy for moving underperforming individuals are developed. The introduced WAR Strategy algorithm is very effective in balancing the exploration and exploitation phases, with a rapid convergence speed in multiple search domains. Algorithm 1 exhibits the steps of the WAR Strategy optimizer in detail [24].

4. Methodology framework

The speech recognition system described in this work is made up of several steps, and each one is very important for reaching the goals. The speech dataset is split into two different groups at the beginning, as shown in Fig. 1, before the system processes begin. Seventy percent (70%) of the information is used to train the system, and the other thirty percent (30%) is used for testing and evaluation. Two types of features are extracted as the first phase of the proposed system, including the statistical and spectral features. Afterward, the extracted features are optimized within the use of the metaheuristic algorithm (i.e. WAR Strategy) to obtain the optimal features set that illustrates the data precisely. Subsequently, the WAR Strategy optimization also will be used for hyperparameters tuning of six traditional ML models to obtain developed classifiers that will be utilized in the final phase of the proposed speech recognition system. The subsequent subsections will offer a comprehensive breakdown of each stage, accompanied by an analysis of the rationale and advantages of its implementation.

4.1. Feature extraction

Speech signals are continuous-time waveforms that contain a vast amount of information with high variability, so features extraction is crucial step when recognizing the speech signal using ML since it transforms these raw speech signals into a suitable representation that captures relevant information for the recognition. Choosing a suitable feature

Algorithm 1: The Metaheuristic WAR Strategy Optimizer

Initialize A = 30 // The forces size
Iterations = 1000.
Ki = Zeros (1, dim) // site of the king.
FC = Zeros (1, dim) // site of the forces Commander.
PE_r = 0:5 // the regarded percentage of the war strategies (attack, defense)
RA = zeros (1, A).
WI = 2 × ones (1, A). // Parameter of the weight
V_i = prior site of the FC
$V_i(u+1)$ = The new site of FC
Input: highest bound.
lowest bound.
The diffusion space // the dimension of the war area.
Begin:
Diffuse the forces in the war space randomly for attack.
For i = 1 to A
Determine the offensive capabilities of each combatant.
End For
Arrange the offensive capabilities (fitness) of the combatants.
Select the combatant having highest fitness to be appointed as a King (Ki)
Select the combatant with second-highest fitness to be appointed as a commander (FC)
While j < Iterations
For i = 1 to A
PE = rand.
If PE < PE_r
Update the positions of the combatants based on the positions of Ki, the army head and a random combatant // **Defense**
$V_i(u+1) = V_i(u) + 2 \times PE \times (Ki - V_{rand}(u)) + rand \times WI_i \times (FC - V_i(u))$// Exploration
Else
Update the position of all combatants based on the positions of (Ki), and (FC) // **Attack**
$V_i(u+1) = V_i(u) + 2 \times PE \times (FC - Ki) + rand \times (WI_i \times Ki - V_i(u))$ // Exploitation.
End if
Compute the magnitude of the offensive power for each combatant.
Arrange the offensive capability of each combatant.
If the offensive power in the new site (E_n) < the previous site (E_u)
Update the position of every soldier to be in the previous position.
$V_i(u+1) = (V_i(u+1) \times (E_n \geq E_u) + V_i(u)) \times (E_n < E_u)$
Update both rank and weight of all combatants.
$RA_i = (RA_i + 1) \times (E_n \geq E_u) + RA_i \times (E_n < E_u)$
End if.
End for
Define the combatant having the lowest offensive power as weak combatant.
Transfer the weak combatant.
$V_w(u+1) = -(1 - randn) \times (V_w(t) - median(V)) + Ki$
Update the site of the Ki and FC
u = u + 1
End while.
Return the offensive power and site of the King.
End

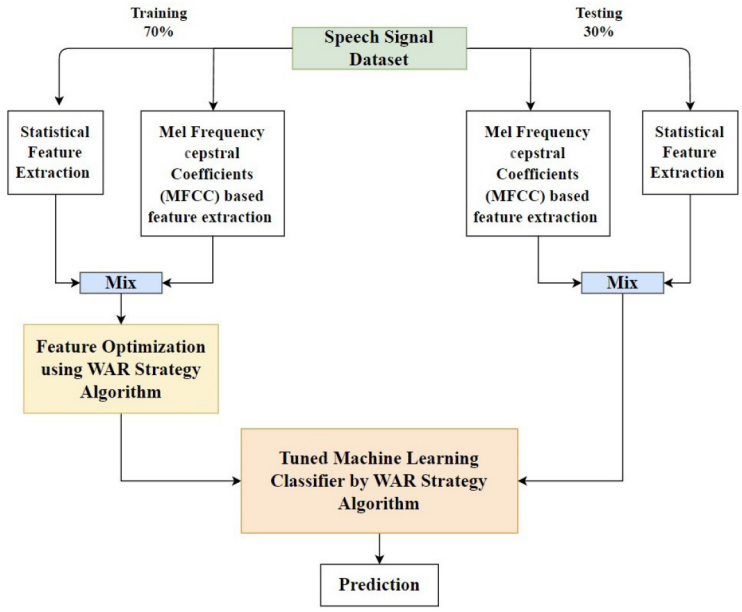


Fig. 1. The speech recognition system block diagram.

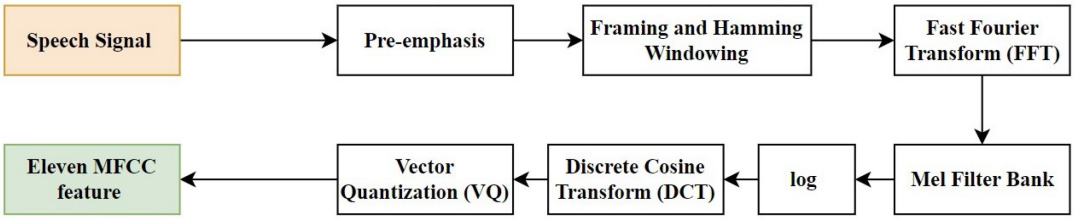


Fig. 2. The MFCC technique main steps followed for features extraction.

extraction technique exerts a substantial influence on the accuracy and the overall performance of speech recognition systems because it effects on the robustness against variability, reducing dimensionality making the model focus only on the relevant information, reducing computational complexity and memory requirements and avoid overfitting, especially when dealing with high-dimensional input data. In this paper two types of features are extracted for the speech signal including nine statistical features and eleven acoustic and spectral features using the Mel Frequency Cepstral Coefficients (MFCC) technique.

4.1.1. Mel frequency cepstral coefficients (MFCC) features

MFCC is a fast, reliable and easy, and widely utilized feature extraction technique in speech and audio signal processing. Its main aim is to capture the essential spectral and perceptual characteristics of a sound signa, in which it mimics the human auditory system utilizing the established variations in the essential bandwidth and frequency of the human ear [25]. The MFCC decreases the frequencies of the entered speech signal and represents them as coefficients depending on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. This feature extraction technique is robust to variations in speaker characteristics, background noise, and other acoustic conditions [26]. A sequence of steps must be followed in order to implement the MFCC technique and obtain the desired feature set as shown in Fig. 2.

Step 1. Pre-emphasis: is a technique of signal processing, utilized to improve spectral characteristics and minimize artifacts. By applying pre-emphasis, the signal's high-frequency content is enhanced, which in turn improve the intelligibility and quality of speech signals, as high-frequency components often carry important information related to consonants and details in the audio. The widely utilized filter for pre-emphasis is a first order type [27].

Step 2. Framing and Hamming Windowing: refers to the process of dividing the speech signal into short, overlapping segments so called frames, that typically contains a small section of the signal

Table 2. Statistical measures descriptions.

Mean	Provides information about sound signal average loudness or intensity over a specific time interval.
SD	Identify abrupt changes or events in the signal. Sudden increases in SD might indicate the presence of an event or a significant change in the sound. Thus, used for quality assessment, noise and events detection in the sound signal
ZC	Used to represent the pattern of speech in sound signal. It identifies the points where the signal changes rapidly and differentiates between voiced and unvoiced speech segments. slowly speech means high ZC value, In contrast low ZC.
Amplitude	The loudness and intensity of speech signal (energy).
Min	The minimum amplitude in the speech signal
Max	The maximum amplitude in the speech signal
Variance	A measure of amplitude variability and used in sound signal analysis to capture information about fluctuations in amplitude values.
Mod	Describe the small differences, variations, or changes that occur, whether it's in amplitude, frequency, phase, or other characteristics of the sound.
Pitch	It is a perceptual attribute that corresponds to the sensation of how high or low a sound is. for example, women's voice sharper than men

typically around 20 to 40 milliseconds in duration. And afterward multiply each frame by specific windowing function. The framing balance of the trade-off between time and frequency resolution in signal analysis, and its overlapping nature helps to reduce artifacts and discontinuities that can occur at the edges [28]. The utilized windowing function in this paper is the Hamming type which is a symmetric windowing designed to minimize the side lobes in the frequency domain. This helps in achieving better frequency resolution and reducing spectral leakage. The formula for the Hamming window is given by [29]:

$$D(g) = 0.54 - 0.46 * \cos\left((2\pi g) / (Z - 1)\right) \quad (1)$$

where D(g) refers to the window's value at sample index g, while Z is the overall count of samples in the window.

Step 3. Fast Fourier Transform (FFT): is an efficient, and rapid method used to transfer the domain of the speech signal from time to frequency, by computing the frequency spectrum and making it suitable for real-time applications [30]. The FFT can provide information about the signal's spectral content, including the magnitudes and phases of different frequency components [31].

Step 4. Mel Filter Bank: it is composed of filters type triangular utilized to calculate the sum weights related to the filter spectral components resulting in an output that closely resembles the Mel scale. It mimics the auditory system of the human being's frequency resolution, which is more perceptually relevant than the linear frequency scale [32].

Step 5. log Mel spectrum: The logarithm of the Mel filter bank outcomes is taken to compress the dynamic range and approximate the non-linear human perception of loudness [33].

Step 6. Discrete Cosine Transform (DCT): it is implemented to return the log Mel spectrum back to the spatial domain by dividing the data of a sequence of finite length into discrete vector. The DCT signal requires less memory to illustrate the Mel spectrum in a relatively small count of coefficients due to it having more information concentrated in a small number of coefficients [34]. The final outcome of the DCT is the MFCC (Mel Frequency Cepstral Coefficient).

Step 7. Vector Quantization (VQ): it is a quantization technique that is used to divide a large count of points into smaller groups that contain approximately the same count of points. The representation of each group of points is done by a centroid point [35]. After acquiring the MFCCs, the VQ is applied to quantize the MFCC feature vectors. VQ uses the clustering algorithm (k-means) depending on some distance metric, often Euclidean distance to finally generate a vector of efficient features [36].

4.1.2. Statistical features

The second set of features that have been extracted from the speech signal are acquired using nine widely known statistical equations which clearly describe the main characteristics of the speech signal and mainly related to the tone of the speakers' voice [37]. These features provide quantitative information about the distribution, variability, and other statistical properties of the speech signal. Each of the implemented statistical equations will give one feature, as a result nine features are acquired from these statical equations. The utilized statistical measures in this paper included, Mean, Standard Deviation (SD), Zero Crossing (ZC), Amplitude, Min, Max, Variance, Mod, and Pitch [38–40]. Table 2 illustrates a brief description about these statistical measures and what kind of information they represent.

4.2. Feature optimization using WAR strategy algorithm

The total count of the previously applied feature extraction methods are twenty-one features which

Table 3. The WAR strategy parameters utilized for feature optimization.

Name	Representation
Iteration	1000
Lowest bound	A minimal features' Count
Highest bound	A maximal features' Count
Initial Population	500 agents
New Population	The new population of agents in each iteration
function	"Easom"

represent the speech signal precisely, these features will be optimized, and a feature optimization operation will be performed in order to acquire the best feature set that represent the speech signal to be used as an input to the hyperparameter tuned classifiers as the final phase for recognizing the speech signal.

It is essential to get the set of features that which are optimal or nearly optimal that accurately capture the most relevant information and all the important characteristics of the speech signal, which can have numerous variations. The metaheuristic algorithm (WAR Strategy) has been implemented in this paper as a feature optimizer to optimize the extracted feature set and obtain the best features that clearly represent the speech signal. The parameters utilized for optimizing the features related to the metaheuristic algorithm (WAR Strategy) are presented in Table 3 [41] which are taken from our previous work in recognizing the sign language.

The Easom function is a type of unimodal test function and can be calculated as follow [42]:

$$f(z_1, z_2) = -\cos(z_1)\cos(z_2) \times \exp(-(z_1-\pi)^2 - (z_2-\pi)^2) \quad (2)$$

Test area is often restricted to square $-100 \le z_1 \le 100$, $-100 \le z_2 \le 100$, and its global minimum is equal to $f(x) = -1$ that is attainable for $(z_1, z_2) = (\pi, \pi)$.

4.3. Classification using optimized machine learning classifiers with the WAR strategy algorithm

The bio-inspired metaheuristics are acknowledged for their efficiency in hybrid procedures for adjusting the hyperparameters of the ML. The procedure of hyperparameter tuning refers to optimizing the hyperparameters of the ML model for enhancing and raising its performance in handling the vast number of challenges that exist in the data when classifying them [43]. The metaheuristic algorithm (WAR Strategy) has been employed to optimize the hyperparameters of well-known classical ML models in this paper. The optimized ML models include: Naïve Bayes (NB) [44], Logistic Regression (LR) [45], Random Forest (RF) [46], K-Nearest-Neighbors (KNN) [47], Support Vector Machine (SVM) [48], and Decision Tree (DT) [49], The WAR Strategy optimization technique is employed to optimize the most influenced hyperparameters in these ML models, and employ them as developed classifiers in the last phase of the proposed speech recognition system. The type of the hyperparameter that is optimized and belongs to each of the earlier mentioned ML models are illustrated in Table 4, and the utilized WAR Strategy parameters for optimizing the parameters are shown in Table 5.

Ackley's is a type of multimodal test function with the following calculation [50]:

$$\text{Ackley's} = -20\exp\left(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20 + \exp(1) \quad (3)$$

Mainly the test area is limited to the hypercube $-32.768 \le x_i \le$ 32:768, i = 1, …, n and the global minimum f(x) = 0, that acquired for $x_i = 0$, i = 1, …, n.

The default values of the hyperparameter of the classical ML models mentioned earlier are given

Table 4. The parameters description of the ML models.

ML Algorithm	Parameter	Illustration
NB	var-smoothing	Add a small amount (smoothing factor) to all feature values to keep them from being completely left out if they don't show up in the training data.
LR	C	Manages the level of regularization and the degree of complexity which help to find an ideal balance among overfitting and underfitting.
RF	n-estimators	Sets how many decision trees will be used in the forest. This number affects how complicated the model is, how much it costs to run, how accurate it is, and how well it handles noisy data
KNN	n-neighbors	Finds the class or value of the question point to use for classification or regression by counting the number of points of data that are closest to it.
SVM	Degree	Choose the SVM's polynomial kernel, which impacts the decision limit difficulty, separation, training time, and achievement of the model.
Decision Tree	max features	Sets the exact set of features that will be considered at each node; this affects how complicated, random, and fast the decision tree is to compute.

Table 5. The parameters and variables utilized for hyperparameters optimization of the metaheuristic algorithm (WAR Strategy).

Parameter	Representation
Iteration	20
Lowest-bound	The lowest parameter's value
Highest-bound	The highest parameter's value
Initial set of population	500 Agent
New set of population	New optimized parameter's value in each iteration
Function of Evaluation	"Ackley"

Table 6. The default and optimized values of the hyperparameters of ML models.

ML Algorithm	Parameter	Default	Tuned
NB	var-smoothing	1.00E-09	0.047619048
LR	C	0.166666667	0.142857143
RF	n-estimators	100	60
KNN	n-neighbors	5	31
SVM	Degree	3	5
DT	max-features	Features-count	25

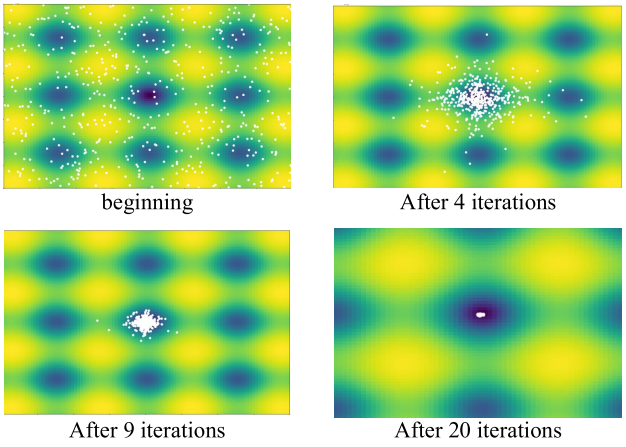


Fig. 3. The population of the WAR strategy algorithm through hyperparameters optimization [41].

in Table 6, in addition to the resulted optimized hyperparameters values after employing the WAR algorithm [41]. The tuned hyperparameters will be adopted in the ML for speech recognition in the proposed system. Moreover, Fig. 3 shows the population of the WAR Strategy optimization algorithm during tuning the hyperparameters.

5. Experimental results

The speech recognition system provided in this study has undergone evaluation for the purpose of recognizing speech signals in three public datasets in three variant languages including English, Arabic, and Malaysian. The efficacy of the speech recognition system, which adopted the metaheuristic algorithm (WAR Strategy) for both feature optimization and

Table 7. Words samples of the English speech dataset.

Bed	Bird	Cat	Eight	Five
Eight	Happy	House	Three	Tree

ML hyperparameters tuning has been demonstrated. Whereas this system has effectively handled the numerous variances seen in the three tested datasets. Moreover, it exhibits exceptional recognition accuracy while reducing the necessary training time, particularly when dealing with several speech signals with diverse properties. The impact of the utilized optimization algorithm has been assessed for both feature optimization and hyperparameter tuning. In which the presented speech recognition system has been implemented and validated in three different scenarios as follow:

- First scenario, present the speech recognition without the use of metaheuristic algorithm (WAR Strategy).
- The second scenario includes features optimization by the WAR Strategy algorithm.
- The third scenario is the proposed system implementation.

A set of widely utilized statistical measures are adopted to assess the performance of each one of the aforementioned scenarios in addition to the proposed system including accuracy, precision, F-measure, and recall. Moreover, the required training time of the six classifiers are measured in recognizing the speech signal with the use of the WAR Strategy Algorithm and without using it.

The system's implementation environment comprises an ASUS laptop equipped with an AMD Ryzen 9 5900HS processor including Radeon Graphics at 3.30 GHz, 16 GB of RAM, and an NVIDIA GeForce RTX graphics card, working on a 64-bit Windows 10 operating system.

5.1. Experiments on English speech dataset

This dataset contains 64721 speech audio files saved in a wav format for English words of count thirty word which means this dataset has thirty class. All these words are from a small set of commands that are spoken by a variant speakers with a background noise including doing the dishes, cat sound, bike exercise, and other confusion noise types. Thus, the dataset implies a high variation in the speech signal [51]. Table 7, illustrate examples of the words spoken in this dataset. Fig. 4, exhibit the WAR Strategy algorithm population on the features of English speech dataset. While Tables 8 to 10, illustrate the acquired results from recognizing the speech signals of this dataset without using the WAR algorithm, after

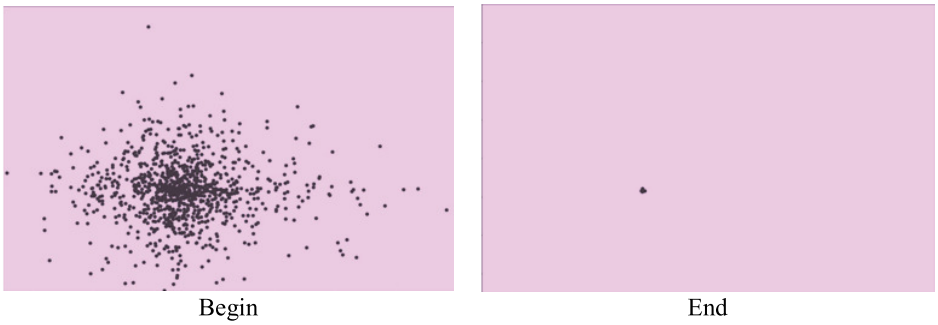


Fig. 4. The population of the utilized optimization algorithm (WAR Strategy) during features optimization of English speech signal dataset.

implementing it for feature optimization, and finally the proposed system in sequence.

5.2. Experiments on Arabic speech dataset

This dataset consists of a list of 12000 pairs (x, y), in which x represents the input speech signal, and y shows the corresponding keyword. The total count of the spoken words is 40, and each audio file is one-second in length saved in a wav format. There are a total of 30 participants, and each participant recorded 10 utterances for every keyword. There are 40 folders, each of which represents one keyword and contains 300 files so in total (30 * 10 * 40 = 12000). The dataset also includes several background noise recordings that have been obtained from various natural sources of noise [52]. Table 11 exhibits samples of these words in Arabic and its corresponding meaning in English. On the other hand, the obtained results from implementing the previously explained scenarios and the proposed system are presented in Tables 12 to 14 in sequence. While Fig. 5 shows the population behavior of the WAR algorithm on the features of this dataset.

5.3. Experiments on Malaysian speech dataset

This dataset is composed of 20 words in Malaysian language in total 1600 clips saved in wav format recorded using the mobile phone by WhatsApp

Table 8. Speech recognition results without employing metaheuristic algorithm on speech signals dataset in English language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training-Time (s)	Hyperparameter-Parameter
NB	68.55%	68.55%	68.55%	68.55%	109.6795	1.00E-09
LR	56.86%	56.86%	56.86%	56.86%	3601.971	0.166666667
RF	63.68%	63.68%	63.68%	63.68%	26712.52	100
KNN	76.59%	76.59%	76.59%	76.59%	13472.1	5
SVM	82.38%	82.38%	82.38%	82.38%	22853.45	3
DT	77%	77%	77%	77%	1425.197	Features' Count

Table 9. Speech recognition results after optimizing the features using WAR algorithm on speech signals dataset in English language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training-Time (s)	Hyperparameter-Parameter
NB	88.17%	88.17%	88.17%	88.17%	106.7891	1.00E-09
LR	87.19%	87.19%	87.19%	87.19%	3178.258	0.166666667
RF	93.10%	93.10%	93.10%	93.10%	23426.28	100
KNN	96.6%	96.6%	96.6%	96.6%	12291.73	5
SVM	91.39%	91.39%	91.39%	91.39%	19976.21	3
DT	96.78%	96.78%	96.78%	96.78%	1380.88	Features' Count

Table 10. The results of the proposed speech recognition system on speech signals dataset in English language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training Time (s)	Tuned-Parameters
NB	99.27%	99.27%	99.27%	99.27%	0.082	0.047619048
LR	99.49%	99.49%	99.49%	99.49%	1.468	0.142857143
RF	100%	100%	100%	100%	19.844	60
KNN	100%	100%	100%	100%	3.872	31
SVM	100%	100%	100%	100%	9.354	5
DT	100%	100%	100%	100%	0.825	25

Table 11. Sample of arabic words in the Arabic speech dataset.

Keyword	Translation
افتح	Open
توقف	Stop
التالي	Next
واحد	one

Table 12. Speech recognition results without employing Metaheuristic algorithm on speech signals dataset in Arabic language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training-Time (s)	Hyperparameter-Parameter
NB	31%	31%	31%	31%	21.94142	1.00E-09
LR	40%	40%	40%	40%	1478.06	0.166666667
RF	50.11%	50.11%	50.11%	50.11%	5188.064	100
KNN	48.63%	48.63%	48.63%	48.63%	641.2861	5
SVM	51.16%	51.16%	51.16%	51.16%	8404.33	3
DT	25.55%	25.55%	25.55%	25.55%	272.2702	Features' Count

Table 13. Speech recognition results after optimizing the features using WAR algorithm on speech signals dataset in Arabic language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training-Time (s)	Hyperparameter-Parameter
NB	81.38%	81.38%	81.38%	81.38%	17.89546	1.00E-09
LR	79.77%	79.77%	79.77%	79.77%	1314.035	0.166666667
RF	92.52%	92.52%	92.52%	92.52%	4741.333	100
KNN	98.83%	98.83%	98.83%	98.83%	600.6465	5
SVM	91.94%	91.94%	91.94%	91.94%	7343.085	3
DT	85.36%	85.36%	85.36%	85.36%	244.3659	Features' Count

Table 14. The results of the proposed speech recognition system on speech signals dataset in Arabic language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training-Time (s)	Tuned-Parameters
NB	99.74%	99.74%	99.74%	99.74%	0.013	0.047619048
LR	97.61%	97.61%	97.61%	97.61%	0.925	0.142857143
RF	100%	100%	100%	100%	1.376	60
KNN	100%	100%	100%	100%	0.33	31
SVM	100%	100%	100%	100%	3.682	5
DT	100%	100%	100%	100%	0.172	25

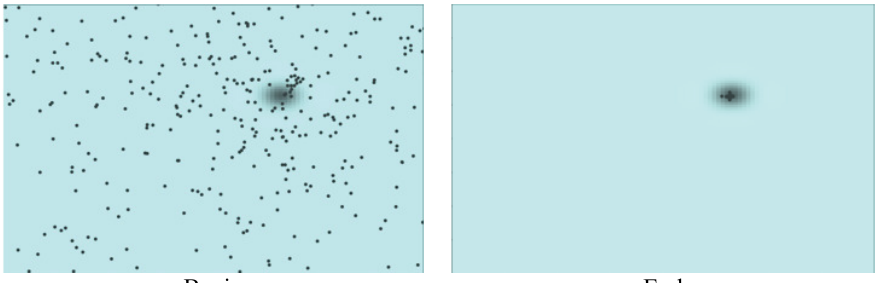


Fig. 5. The population of the WAR algorithm during features optimization for the Arabic speech signal dataset.

application with one second length. Table 15 shows samples of those words from the Malaysian speech dataset [53]. The results of this dataset are illustrated in Tables 16 to 18, and the population of the WAR strategy algorithm on its features is presented in Fig. 6.

5.4. Discussion

The result analysis clearly demonstrates that the combination of the optimization metaheuristic algorithm (i.e. WAR Strategy) and classical ML models has reached a remarkable degree of accuracy in

Table 15. Samples of the words recorded in the Malaysian speech dataset.

Keyword	Translation
aku	me
belik	buy
dekat mane	Look at me
harge	price

Table 16. Speech recognition results without employing metaheuristic algorithm on speech signal dataset in Malaysian language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training-Time (s)	Hyperparameter-Parameter
NB	55.83%	55.83%	55.83%	55.83%	3.98922	1.00E-09
LR	89.58%	89.58%	89.58%	89.58%	149.682	0.166666667
RF	98%	98%	98%	98%	314.1589	100
KNN	89.58%	89.58%	89.58%	89.58%	24.9331	5
SVM	96.7%	96.7%	96.7%	96.7%	109.7059	3
DT	96.66%	96.66%	96.66%	96.66%	7.978678	Features' Count

Table 17. Speech recognition results after optimizing the features using WAR algorithm on speech signal dataset in Malaysian language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training-Time (s)	Hyperparameter-Parameter
NB	87.58%	87.58%	87.58%	87.58%	2.984285	1.00E-09
LR	97.79%	97.79%	97.79%	97.79%	131.0821	0.166666667
RF	99.16%	99.16%	99.16%	99.16%	310.0608	100
KNN	91.04%	91.04%	91.04%	91.04%	23.93913	5
SVM	98.1%	98.1%	98.1%	98.1%	90.81841	3
DT	97.68%	97.68%	97.68%	97.68%	6.977413	Features' Count

Table 18. The results of the proposed speech recognition system on speech signal dataset in Malaysian language.

ML Algorithm	Acc.%	Prec.%	Rec.%	F1-Score%	Training-Time (s)	Tuned-Parameters
NB	98.38%	98.38%	98.38%	98.38%	0.001	0.047619048
LR	100%	100%	100%	100%	0.083	0.142857143
RF	100%	100%	100%	100%	0.139	60
KNN	100%	100%	100%	100%	0.01	31
SVM	100%	100%	100%	100%	0.049	5
DT	100%	100%	100%	100%	0.004	25

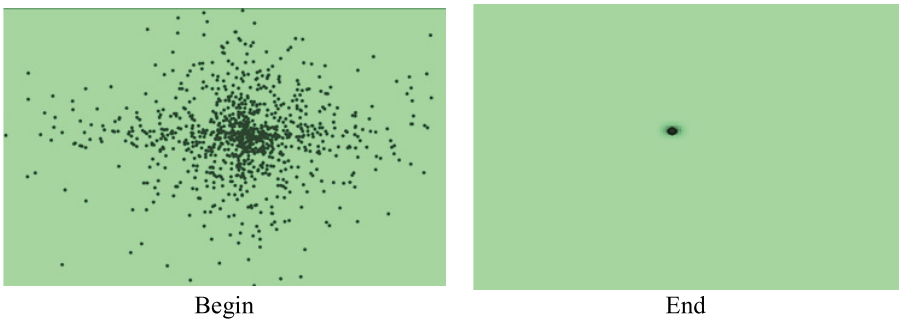


Fig. 6. The population of the WAR algorithm during features optimization of Malaysian speech signal dataset.

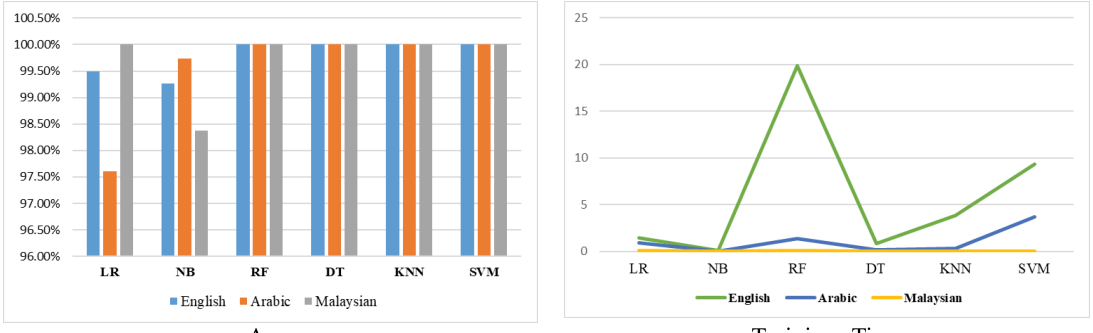
speech recognition for English, Arabic, and Malaysian languages. The superior results acquired from the proposed system has proven the efficiency of the features optimization using the WAR algorithm in handling the challenges associated with speech recognition, such as background noise, speaker variability, ambiguity in words, variations in speaking style and accent and the emotional and disfluent speech. In which,

Table 19. A comparison with other speech recognition related works.

Study	Optimization algorithm	Method	Classifier	Dataset	Acc.%	Prec.%	Rec.%	F1-Score
[14]	ABC and PSO	Feature selection	SVM	Indian	94.47%	-	-	-
[16]	ABC	Layers optimization	ANN	English	99.36%	-	-	-
[17]	PSO	Feature optimization	SVM	Brazilian	92%	-	-	-
[18]	BBA and LAHC	Feature selection	RF	Indian	100%	-	-	-
[20]	BFOA	Parameters optimization	HMM	Arabic	92%	-	-	-
[21]	WOA	Layers optimization	DNN	English	99.6%	-	-	-
[22]	GOA	Parameters optimization	RNN	Indian	96%12.	-	-	-
Proposed	WSO	Features and parameters optimization	DT DT DT	English Arabic Malaysian	**100%** **100%** **100%**	**100%** **100%** **100%**	**100%** **100%** **100%**	**100%** **100%** **100%**

the speech recognition accuracy has achieved a maximum of 100% by employing many fine-tuned ML algorithms. Moreover, the rapid convergence rate, and the ability of the WAR Strategy algorithm in avoiding the local optima has been proved, due to its optimal selection of parameters related to ML classifiers that makes them achieved a remarkable performance. In addition to the enhanced recognition accuracy, the required training time also decreased and reached less than one second in almost all the fine-tuned classifiers. The generalization problem as one of most unsolved problem in speech recognition systems has been tackled by implementing the suggested system on three different speech datasets with diverse characteristics, resulting in a commendable level of accuracy and efficiency. A noticeable disparity is observed when comparing the outcomes of employing the ML models for speech recognition without using the metaheuristic algorithm (WAR Strategy), with the developed ML models in this paper. The efficiency after using the WAR Strategy algorithm was initially observed for feature optimization, based on the findings obtained from the second implementation scenario mentioned earlier, and from the proposed system after tuning the ML parameters. when comparing the acquired results from the proposed speech recognition system with the results obtained from implementing the speech recognition system without using the WAR algorithm, it is obvious, the proposed systems has achieved a remarkable performance for precise and fast speech recognition and avoids the need for using deep learning approaches that require substantial effort and complexity.

In order to highlight the advantages of the proposed speech recognition system, a comparative study has been carried out in comparison to previous studies. The results of this comparison are illustrated in Table 19. The proposed system has achieved a recognition accuracy equal to 100% with a training time of 0.004 second using the optimized decision tree on Malaysian speech dataset. Fig. 7 exhibits the accuracy rate and the training time of all the optimized classifiers on the three speech datasets.



Acc. Training -Time

Fig. 7. The statistical results of the proposed speech recognition system in three different language datasets.

6. Conclusions and future work

This paper presents a highly efficient speech recognition system that takes into account variant speech datasets in different languages including, English, Arabic, and Malaysian. The recently developed metaheuristic optimization algorithm (WAR Strategy) has been utilized to as an optimizer for both features and the hyperparameters of six traditional ML algorithm in order to achieve a precise speech recognition. First, two types of features are extracted and mixed which are the statistical features that describe the general properties of the speech signal, and the MFCC features that capture the spectral characteristics of the speech signal. Afterward, the extracted features are optimized using the WAR Strategy algorithm to acquire optimum features set of the speech signal. At the end, the WAR Strategy algorithm optimizes the hyperparameters of six ML algorithm and uses this optimized version as a classifier in the proposed speech recognition system. The proposed system has successfully addressed various challenges in speech signal recognition, including noisy environments, speaker variability, word ambiguity, variations in style of speaking and accent, as well as sentimental and disfluent speech variance. This has been achieved by utilizing two types of features and optimizing them through the implementation of the WAR Strategy algorithm. The implementation of the optimization algorithm for fine-tuning the hyperparameters has greatly improved the overall performance, as evidenced by the outstanding recognition accuracy of 100% and the minimal training time of less than 0.5 seconds achieved when recognizing three distinct speech datasets. The findings demonstrate that the proposed system is highly efficient in speech signal recognition, even when dealing with datasets that have substantial variations and a large number of samples. In which, it reduces the training time and complexity and preserving exceptional generalization. As future work, the hyperparameters of the deep learning models will be optimized to achieve better recognition accuracy and eliminate the high complexity in deep models.

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.

Acknowledgments

Not Applicable.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Ethical approval

Not Applicable.

Consent to participate

The authors provide the appropriate consent to participate.

Consent for publication

The authors provide the consent to publish the images in the manuscript. The data used in the publication is publicly available. We provide respective citations for each of the data sources.

Code availability

Not Applicable.

Authors' contributions

Shahad Th. has written the whole manuscript, and suggested the system and implemented it, Salman y. and Azhana A, responsible of the supervision of this manuscript, Saif M and Ahmed A. reviewed the manuscript and give advice and adjustments.

References

1. Björn Lindblom and Johan Sundberg, "The human voice in speech and singing," *Springer Handbook of Acoustics*, pp. 703–746, 2014.
2. Timothée Proix, Jaime Delgado Saa, Andy Christen, Stephanie Martin, Brian N. Pasley, Robert T. Knight, Xing Tian et al. "Imagined speech can be decoded from low-and cross-frequency intracranial EEG features," *Nature Communications*, vol. 13, no. 1, p. 48, 2022.
3. Jinyu Li, "Recent advances in end-to-end automatic speech recognition," *APSIPA Transactions on Signal and Information Processing*, vol. 11, no. 1, 2022.
4. Sadeen Alharbi, Muna Alrazgan, Alanoud Alrashed, Turkiayh Alnomasi, Raghad Almojel, Rimah Alharbi, Saja Alharbi, Sahar Alturki, Fatimah Alshehri, and Maha Almojil, "Automatic speech recognition: Systematic literature review," *IEEE Access*, vol. 9, pp. 131858–131876, 2021.
5. Habib Ibrahim and Asaf Varol, "A study on automatic speech recognition systems," In *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*, IEEE, pp. 1–5, 2020.

6. Mishaim Malik, Muhammad Kamran Malik, Khawar Mehmood, and Imran Makhdoom, "Automatic speech recognition: A survey," *Multimedia Tools and Applications*, vol. 80, pp. 9411–9457, 2021.
7. Laurent Besacier, Etienne Barnard, Alexey Karpov, and Tanja Schultz, "Automatic speech recognition for under-resourced languages: A survey," *Speech Communication*, vol. 56, pp. 85–100, 2014.
8. Raed Abdulkareem Abdulhasan, Shahad Thamear Abd Allatief, and Saif Mohanad Kadhim, "Instant learning based on deep neural network with linear discriminant analysis features extraction for accurate iris recognition system," *Multimedia Tools and Applications*, vol. 83, no. 11, pp. 32099–32122, 2024.
9. Jayashree Padmanabhan and Melvin Jose Johnson Premkumar, "Machine learning in automatic speech recognition: A survey," *IETE Technical Review*, vol. 32, no. 4, pp. 240–251, 2015.
10. R. Arun Kumar, J. Vijay Franklin, and Neeraja Koppula, "A comprehensive survey on metaheuristic algorithm for feature selection techniques," *Materials Today: Proceedings*, vol. 64, pp. 435–441, 2022.
11. Mohammed H. Al-Farouni, "Enhanced bird swarm algorithm with deep learning based electroencephalography signal analysis for emotion recognition," *Journal of Smart Internet of Things* 2022, no. 1, pp. 33–52, 2023.
12. Mainak Biswas, Saif Rahaman, Ali Ahmadian, Kamalularifin Subari, and Pawan Kumar Singh, "Automatic spoken language identification using MFCC based time series features," *Multimedia Tools and Applications*, vol. 82, no. 7, pp. 9565–9595, 2023.
13. Li Yang and Abdallah Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.
14. Sunanda Mendiratta, Neelam Turk, and Dipali Bansal, "Automatic speech recognition using optimal selection of features based on hybrid ABC-PSO," In *2016 International Conference on Inventive Computation Technologies (ICICT)*, IEEE, vol. 2, pp. 1–7, 2016.
15. Teena Mittal and Rajendra Kumar Sharma, "Speech recognition using ANN and predator-influenced civilized swarm optimization algorithm," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 24, no. 6, pp. 4790–4803, 2016.
16. Shilpi Shukla and Madhu Jain, "A novel system for effective speech recognition based on artificial neural network and opposition artificial bee colony algorithm," *International Journal of Speech Technology*, vol. 22, no. 4, pp. 959–969, 2019.
17. Gracieth Cavalcanti Batista, Washington Luis Santos Silva, Duarte Lopes de Oliveira, and Osamu Saotome, "Automatic speech patterns recognition of commands using SVM and PSO," *Multimedia Tools and Applications* vol. 78, no. 22, pp. 31709–31731, 2019.
18. Aankit Das, Samarpan Guha, Pawan Kumar Singh, Ali Ahmadian, Norazak Senu, and Ram Sarkar, "A hybrid meta-heuristic feature selection method for identification of Indian spoken languages from audio signals," *IEEE Access*, vol. 8, pp. 181432–181449, 2020
19. Samarpan Guha, Aankit Das, Pawan Kumar Singh, Ali Ahmadian, Norazak Senu, and Ram Sarkar, "Hybrid feature selection method based on harmony search and naked mole-rat algorithms for spoken language identification from audio signals," *IEEE Access*, vol. 8, pp. 182868–182887, 2020.
20. Abdelmadjid Benmachiche, Amina Makhlouf, and Tahar Bouhadada, "Optimization learning of hidden markov model using the bacterial foraging optimization algorithm for speech recognition," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, vol. 24, no. 3, pp. 171–181, 2020.
21. Shilpi Shukla and Madhu Jain, "A novel stochastic deep resilient network for effective speech recognition," *International Journal of Speech Technology*, vol. 24, no. 3, pp. 797–806, 2021.
22. Ravindra Parshuram Bachate, Ashok Sharma, Amar Singh, Ayman A. Aly, Abdulaziz H. Alghtani, and Dac-Nhuong Le, "Enhanced marathi speech recognition facilitated by grasshopper optimisation-based recurrent neural network," *Comput. Syst. Sci. Eng*, vol. 43, no. 2, pp. 439–454, 2022.
23. Padam Singh and Sushil Kumar Choudhary, "Introduction: Optimization and metaheuristics algorithms," *Metaheuristic and Evolutionary Computation: Algorithms and Applications*, pp. 3–33, 2021.
24. Tummala SLV Ayyarao, N. S. S. Ramakrishna, Rajvikram Madurai Elavarasan, Nishanth Polumahanthi, M. Rambabu, Gaurav Saini, Baseem Khan, and Bilal Alatas, "War strategy optimization algorithm: A new effective metaheuristic algorithm for global optimization," *IEEE Access*, vol. 10, pp. 25073–25105, 2022.
25. Zrar Kh Abdul and Abdulbasit K. Al-Talabani, "Mel frequency cepstral coefficient and its applications: A review," *IEEE Access*, vol. 10, pp. 122136–122158, 2022.
26. Sreenivas Sremath Tirumala, Seyed Reza Shahamiri, Abhimanyu Singh Garhwal, and Ruili Wang, "Speaker identification features extraction methods: A systematic review," *Expert Systems with Applications*, vol. 90, pp. 250–271, 2017.
27. Erfan Loweimi, Seyed Mohammad Ahadi, Thomas Drugman, and Samira Loveymi, "On the importance of pre-emphasis and window shape in phase-based speech recognition," In *Advances in Nonlinear Speech Processing: 6th International Conference, NOLISP 2013, Mons, Belgium, June 19-21, 2013*. Proceedings 6, Springer Berlin Heidelberg, pp. 160–167, 2013.
28. F. L. Teixeira, S. P. Soares, J. L. P. Abreu, P. M. Oliveira, and J. P. Teixeira, "Comparative analysis of windows for speech emotion recognition using CNN," *International Conference on Optimization, Learning Algorithms and Applications*, Springer Nature Switzerland, pp. 233–248, 2023.
29. Prajoy Podder, Tanvir Zaman Khan, Mamdudul Haque Khan, and M. Muktadir Rahman, "Comparative performance analysis of hamming, hanning and blackman window," *International Journal of Computer Applications*, vol. 96, no. 18, pp. 1–7, 2014.
30. Weiqiang Liu, Qicong Liao, Fei Qiao, Weijie Xia, Chenghua Wang, and Fabrizio Lombardi, "Approximate designs for fast Fourier transform (FFT) with application to speech recognition," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 12, pp. 4727–4739, 2019.
31. Xavier Amatriain, Jordi Bonada, Alex Loscos, and Xavier Serra, "Spectral processing," *DAFX: Digital Audio Effects*, pp. 373–438, 2002.
32. M. S. Likitha, Sri Raksha R. Gupta, K. Hasitha, and A. Upendra Raju, "Speech based human emotion recognition using MFCC," In *2017 International Conference On Wireless Communications, Signal Processing and Networking (WiSPNET)*, IEEE, pp. 2257–2260, 2017.
33. Shashidhar G. Koolagudi, Deepika Rastogi, and K. Sreenivasa Rao, "Identification of language using mel-frequency cepstral coefficients (MFCC)," *Procedia Engineering*, vol. 38, pp. 3391–3398, 2012.
34. Md Afzal Hossan, Sheeraz Memon, and Mark A. Gregory, "A novel approach for MFCC feature extraction," In *2010 4th International Conference on Signal Processing and Communication Systems*, pp. 1–5, IEEE, 2010.
35. Saswati Debnath and Pinki Roy, "Automatic speech recognition based on clustering technique," In *Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph* 2018, Springer Singapore, pp. 679–688, 2020.

36. Erhan Akbal, "An automated environmental sound classification methods based on statistical and textural feature," *Applied Acoustics* vol. 167, pp. 107413, 2020.
37. Dong Kyu Lee, Junyong In, and Sangseok Lee, "Standard deviation and standard error of the mean," *Korean Journal of Anesthesiology*, vol. 68, no. 3, pp. 220, 2015.
38. Yuichi Goto and Masanobu Taniguchi, "Robustness of zero crossing estimator," *Journal of Time Series Analysis*, vol. 40, no. 5, pp. 815–830, 2019.
39. Takumi Abe, Shoichi Koyama, Natsuki Ueno, and Hiroshi Saruwatari, "Amplitude matching for multizone sound field control," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 656–669, 2022.
40. Michael J. Carey, Eluned S. Parris, Harvey Lloyd-Thomas, and Stephen Bennett, "Robust prosodic features for speaker identification," In *Proceeding of Fourth International Conference on Spoken Language Processing*. ICSLP'96, IEEE, vol. 3, pp. 1800–1803, 1996.
41. Shahad Thamear Abd Al-Latief, Salman Yussof, Azhana Ahmad, Saif Mohanad Khadim, and Raed Abdulkareem Abdulhasan, "Instant sign language recognition by WAR strategy algorithm based tuned machine learning," *International Journal of Networked and Distributed Computing*, pp. 1–18, 2024.
42. E. J. Solteiro Pires, J. A. Tenreiro Machado, P. B. de Moura Oliveira, J. Boaventura Cunha, and Luís Mendes, "Particle swarm optimization with fractional-order velocity," *Nonlinear Dynamics* vol. 61, pp. 295–301, 2010.
43. Răzvan Andonie, "Hyperparameter optimization in learning systems," *Journal of Membrane Computing*, vol. 1, no. 4, pp. 279–291, 2019.
44. Marlis Ontivero-Ortega, Agustin Lage-Castellanos, Giancarlo Valente, Rainer Goebel, and Mitchell Valdes-Sosa, "Fast gaussian naïve bayes for searchlight classification analysis," *Neuroimage*, vol. 163, pp. 471–479, 2017.
45. Stephan Dreiseitl and Lucila Ohno-Machado, "Logistic regression and artificial neural network classification models: A methodology review," *Journal of Biomedical Informatics*, vol. 35, no. 5–6, pp. 352–359, 2002.
46. Mahesh Pal, "Random forest classifier for remote sensing classification," *International Journal of Remote Sensing*, vol. 26, no. 1, pp. 217–222, 2005.
47. Padraig Cunningham and Sarah Jane Delany, "K-nearest neighbour classifiers-a tutorial," *ACM Computing Surveys (CSUR)*, vol. 54, no. 6, pp. 1–25, 2021.
48. Derek A. Pisner and David M. Schnyer, "Support vector machine," In *Machine Learning*, Academic Press, pp. 101–121, 2020.
49. Arundhati Navada, Aamir Nizam Ansari, Siddharth Patil, and Balwant A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," In *2011 IEEE Control and System Graduate Research Colloquium*, IEEE, pp. 37–42, 2011.
50. Jing J. Liang, A. Kai Qin, Ponnuthurai N. Suganthan, and S. Baskar, "Comprehensive learning particle swarm optimizer for global optimization of multimodal functions," *IEEE Transactions on Evolutionary Computation*, vol. 10, no. 3, pp. 281–295, 2006.
51. Bharat sahu. (2018). Speech commands classification dataset. *Kaggle*. https://www.kaggle.com/datasets/bharatsahu/speech-commands-classification-dataset?select=sheila. (Accessed 12 December 2023).
52. Abdulkader Ghandoura, Farouk Hjabo, and Oumayma Al Dakkak, "Building and benchmarking an arabic speech commands dataset for small-footprint keyword spotting," *Engineering Applications of Artificial Intelligence* 102, pp. 104267, 2021.
53. Wicara Monsoleil. 2022. Dataset bahasa daerah Indonesia. *Kaggle*. https://www.kaggle.com/datasets/wicaramonsoleil/dataset-bahasa-daerah-indonesia. (Accessed 2 January 2024).