Prediction of neuro-degenerative disorders using sunflower optimisation algorithm and Kernel extreme learning machine: A case-study with Parkinson's and Alzheimer's disease

Kishore Balasubramanian[1] ⓘ, Ananthamoorthy NP[2]
and Ramya K[3]

Abstract

Parkinson's and Alzheimer's Disease are believed to be most prevalent and common in older people. Several data-mining approaches are employed on the neuro-degenerative data in predicting the disease. A novel method has been built and developed to diagnose Alzheimer's (AD) and Parkinson's (PD) in early stages, which includes image acquisition, pre-processing, feature extraction and selection, followed by classification. The challenge lies in selecting the optimal feature subset for classification. In this work, the Sunflower Optimisation Algorithm (SFO) is employed to select the optimal feature set, which is then fed to the Kernel Extreme Learning Machine (KELM) for classification. The method is tested on the Alzheimer's Disease Neuroimaging Initiative (ADNI) and local dataset for AD, the University of California, Irvine (UCI) machine learning repository and the Istanbul dataset for PD. Experimental outcomes have demonstrated a high accuracy level in both AD and PD diagnosis. For AD diagnosis, the highest classification rate is obtained for the AD versus NC classification using the ADNI dataset (99.32%) and local dataset (98.65%). For PD diagnosis, the highest accuracy of 99.52% and 99.45% is achieved on the UCI and Istanbul datasets, respectively. To show the robustness of the method, the method is compared with other similar methods of feature selection and classification with 10-fold cross-validation (CV) and with unseen data. The method proposed has an excellent prospect, bringing greater convenience to clinicians in making a better solid decision in clinical diagnosis of neuro-degenerative diseases.

Introduction

Alzheimer's disease and Parkinson's disease are distinct diseases or ends of a neurodegenerative continuum. The World Health Organization (WHO)[1] recognised dementia as a major health issue, in particular Parkinson's Disease (PD). Parkinson's disease (PD) is a leading neuro-degenerative disease in people over 65, affecting nearly 10 million people worldwide and on the rise.[2] It is projected to touch 131.5 M by around 2050.[3] PD is based on the dopamine receptors affecting the subject's mobility. The disease progresses with motor as well as non-motor symptoms. Patients diagnosed with PD show signs and symptoms related to Parkinsonism. There are other causes due to drugs and rare conditions like multiple cerebral infractions and degenerative conditions.[4] Research studies have identified that genetic mutations, pathogenic inflammation and misfolded proteins also contribute to the death of brain cells in PD patients.[5,6] PD is a complicated neuro-pathological disease containing molecular pathway

[1]Dr Mahalingam College of Engineering and Technology, Pollachi, Tamil Nadu, India
[2]Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu, India
[3]P A College of Engineering and Technology, Pollachi, Tamil Nadu, India

Corresponding author:
Kishore Balasubramanian, Dr Mahalingam College of Engineering and Technology, Udumalai Road, M.K. Patti (PO), Pollachi, Tamil Nadu 642003, India.
Email: kishore_iyer7@rediffmail.com

arrays implicated in the disease physiology.[7] Noticeable onset of symptoms or factors like motor factors is diagnosed and thereafter treatment is started. Gene factors also contribute to the disease, which often cannot be diagnosed easily. Hence, gene factors can be used as bio-markers for expeditious screening and diagnosis.[8]

Alzheimer's disease (AD) produces progressive loss of memory. Symptoms start to appear in older adult stages, normally and their prevalence rises with age sharply.[9] Genes play a vital impact in the disease onset, and certain specific genes may increase the burden by aggravating the disease. Age, habits like alcohol consumption, smoking, etc. are also some of the causes.[10] AD worsens as time progresses.[11] Nearly 70% of the risk is genetic, with depression, stress and hypertension also playing a significant role.[12] There are no supplements or medications to reduce the risk.[13] AD people definitely have to rely on other people for assistance. They may be subjected to more stress or psychophysical and economic pressures.[14] Magnetic Resonance Imaging (MRI) Scans and neuropsychological tests are conducted to diagnose the disease if early symptoms are noticed.[15] These diseases are required to be diagnosed at an early stage, that paves way for faster treatment and a significant reduction in symptoms for which various Machine Learning (ML) approaches were proposed.[16] Current clinical diagnosis of AD and PD relies on the skills of the professionals, which necessitates studying hundreds of brain tissue slides, which is time-consuming and costly. Learning-based strategies can be deployed to fasten the process and lower the cost of diagnosis. Moreover, early diagnosis using CAD systems is required for integrative analysis that combines a number of probabilistic and optimisation techniques to enable computers to extract information from large, complicated datasets. To better identify diseases, the machine learning platform combines several data types into a single model. As a result, researchers are increasingly relying on machine learning to detect neurodegenerative disorders in their early stages.

Computer Aided Systems (CAD) have been developed in recent years to aid in disease and management, as the ML models have shown immense strength in the interpretation of data, aiding in diagnosis and sound clinical decision making.[17] Various computational intelligence approaches have been developed these days which involve multiple procedures mainly devoted to disease prediction, classification and clustering issues. These methods are mostly inspired by nature,[18,19] evolution phenomenon,[20] swarm intelligence,[21] evolutionary algorithms[22] and many more.[23] Artificial Neural Networks (ANN) have shown some promising results as reported in various disciplines, but designing such a network is not an easy task. ANNs have regulated procedures defined by a set of parameters. Setting of such parameters is a vital and crucial task as they determine or regularise the ANNs in obtaining better performance. As many studies have indicated, there is no optimal procedure for this and it also varies from problem to problem. In a research to construct a smaller number of parameters for setting optimal procedure, Extreme Learning Machine (ELM) was proposed.[24] But the ELM had its own demerit in obtaining the most accurate result. Hence, a variant of ELM called Kernel ELM was proposed.[25] With technical advancements blooming, literature has shown different machine learning approaches towards detecting neuro degenerative diseases.

In the next section, a brief literature study has been outlined with respect to PD and AD diagnosis.

Literature review

Little et al.[26] demonstrated a method that identifies PD in patients having dysphonic indications employing Support Vector Machines (SVM), Efficient Learning Machines (ELM) with feature selection. The experimental outcome showed that the model was able to identify efficiently, PD patients having four such features only. Das[27] compared the classification scores of ANN, DMneural, Regression and Decision Trees in diagnosing PD, where ANN obtained 92.9% results. Sakar et al.[28] developed a method combining SVM and mutual information feature selection in detecting PD. 92.75% classification accuracy was reported. An improved accuracy of 93.47% was reported by Li et al[29] employing SVM and Fuzzy Nonlinear Transformation (FNLT) method in diagnosing PD. Chen et al.[30] proposed a method combining Fuzzy k-nearest neighbour classifier (FKNN) and Principal Component Analysis (PCA-FKNN) to diagnose PD, yielding 96% accuracy. Ozcift and Gulten[31] reported ensemble classifiers based on rotation forest involving feature selection by correlation-based method in identifying PD in patients and observed 87.13%. accuracy. Shen et al.[32] presented a method based on enhanced SVM and fruit fly optimisation algorithm, attaining 96.67% accuracy. Authors in Cai et al.,[33] reported an optimal SVM using Bacterial Foraging Optimisation (BFO) in combination with relief selection of features in predicting PD, exhibiting 97.42% accuracy of classification. The Feed Forward Parallel NN was developed by the authors in[34] to diagnose PD and attained 84% accuracy. Spadoto et al.[35] presented a method combining evolutionary techniques with an Optimum-Path Forest (OPF) classifier to diagnose PD, yielding 84.01% accuracy. Integrated Fuzzy C-Means clustering with feature weighting (FCMFW) along with KNN was reported by Polat[36] in classifying PD, achieving 97.93% accuracy. Cai et al.[37] proposed an intelligent technique to identify PD employing Chaotic Bacterial Foraging Optimisation with Enhanced fuzzy KNN classifier, obtaining 97.89% accuracy on the basis of vocal measurements.

Authors in[38] proposed a novel multivariate technique to detect Alzheimer's Disease (AD) utilising Stationary Wavelet Entropy (SWE) and Predator-Prey Particle Swarm Optimisation (PPPSO). Brain imaging,

axial scale selection, feature extraction by SWE and classification by NN were steps involved. PPPSO was employed to adjust the bias and weights of the NN. Thirteen features were selected and the method yielded 92.73% ± 1.03% accuracy. Dessouky and Elrashidy[39] reported feature extraction techniques of AD images employing various optimisation algorithms and compared evolutionary algorithms such as PSO, BA, GA, pattern search, etc., in detecting AD. Pattern search was recorded as the best with the highest accuracy rate. The authors of[40] presented a comprehensive review of 165 AD papers involving feature extraction and Machine Learning (ML) approaches from 1995 to 2019, based on SVM, ANN and Deep Learning (DL).

Andres Ortiz et al,[41] reported an ensemble of DL architectures to diagnose AD at an early stage. Peng et al[42] employed genetic information and MRI data for feature extraction and utilised SVM multiple kernel learning. Leave one-out cross validation (LOOCV) was utilised for classifying Alzheimer's. Shree and Sheshadri[43] provided a detailed investigation into AD diagnosis using various classification methods. A review on the early diagnosis of AD was presented by Khan and Usman.[44] The instance-based learning method was adopted in Khan et al.[45] to diagnose AD. Shankar et al.[46] proposed AD detection employing an optimised feature set selected by Group Grey Wolf Optimisation (GGWO) with a Convolutional Classifier yielding 96.23% classification accuracy. Islam and Zhang[47] investigated a DL technique to identify AD from the MRI data. Xiao et al.[48] proposed an experimental method to identify AD involving textural features and voxel-based morpho-metric neuro-imaging. SVM was used as a classifier. Entropy feature selection was used in reducing issues related to dimensionality, achieving 88% accuracy. Silva et al.[49] attempted to predict AD employing Haralick features by a feature extraction process. GA and PSO were employed for feature selection with the RF classifier producing 78.9% and 77.66% accuracy while using GA and PSO, respectively. Acharya et al.,[50] the authors developed a Computer-Aided-Brain-Diagnosis (CABD) system to identify if a brain scan shows AD signs using feature extraction methods. Authors proved that the Shearlet Transform (ST) method of feature extortion offered better outcomes in AD diagnosis, compared to other works. The KNN classifier was used to achieve 94.54% accuracy, 96% sensitivity and 93.64 % specificity. Feature ranking and GA was employed in Beheshti et al.[51] to diagnose AD by predicting mild cognitive impairment-to-Alzheimer's conversion obtained from structural MRI (sMRI). State-of-the-art approaches reviewed in the literature are listed in Table 1. The table summarises certain important approaches adopted to diagnose AD and PD, with their contributions and implications.

To put it in a nutshell, there is research going on to detect neuro-degenerative disorders like PD and AD, employing various hybrid and DL techniques. Many optimisation algorithms have been evolved and employed in various aspects, like feature selection, optimising the weights and bias of the classifier, etc., in delineating the diseased condition. In this study, the Sunflower Optimisation algorithm (SFO) is deployed to select optimal features and fed to KELM for classification. Contributions to the work include:

(i) The newly developed Sunflower Optimisation algorithm (SFO) is implemented for selecting relevant features in neurodegenerative disease classification employing image and vocal measurement properties.
(ii) SFO performed well in selecting the optimal feature set with a high rate of convergence when compared with other similar algorithms.
(iii) Kernel ELM is employed for classification which outperformed other classifiers.
(iv) To obtain a better estimation of the final model's error, the model is tested with unseen data and 10-fold cross validation.
(v) A balance between maximum classification accuracy and minimum error is achieved through SFO-KELM.

Rest of the paper is structured as: Section 2 outlines the proposed framework consisting of dataset preparation, feature extortion, optimal feature selection using SFO and classification using KELM. Section 3 discusses the experiment conducted and illustrates the results. The conclusion is provided in Section 4.

Proposed frame work

This section outlines the framework developed for investigating neuro-degenerative disorders. Initially, the dataset is prepared from the ADNI and local datasets for AD and from the UCI repository and Istanbul for PD, followed by feature selection using the Sunflower Algorithm (SFO). The SFO algorithm is inspired by the natural behaviour of sunflowers and can be classified as an iterative metaheuristic optimisation algorithm based on population for solving multi-dimensional problems. SFO has proved to be efficient in finding global optimal solution.[52] SFO avoids getting stuck at local optima and doesn't require derivatives while the objective function is evaluated. SFO was employed successfully in solving benchmark problems with better convergence in spite of the non-refined parameters. This proves SFO's healthier performance when compared to the well-known techniques like GA and PSO.[62]

Dataset collected from the AD database is subjected to pre-processing before being applied in the process. The optimal feature subset is generated from the feature selection process and applied to the Kernel Extreme Learning Machine (KELM) classifier. ELM proved to be better at learning the forward neural networks and avoided gradient descent training methods as encountered in back propagation networks. ELM has been better at classification for medical diagnosis. KELM

Table 1. State of art approaches towards AD and PD disease diagnosis.

Author	Objective	Implications
Little et al.[26]	Identifies PD in patients having dysphonic indications employing SVM, ELM with feature selection	Non-standard methods and traditional harmonics-to-noise ratios demarcate normal from PD.
Das et al.[27]	Classification methods such as Neural Networks, DMneural, Regression and Decision Tree are compared for PD diagnosis.	Reliable PD diagnosis is notoriously hard to attain with mis-diagnosis accounting to be as high as 25% of the cases.
Sakar et al.[28]	Selects a minimal feature subset with maximal joint relevance to the PD-score and to introduce a predictive model with minimal bias	Mutual information score with permutation for relevance assessment and its statistical significance in relation to the features and PD-score
Li et al.[29]	Extraction of optimal feature subset to raise the analytical performance when small data set is available.	Fuzzy- transformation method is better to PCA and KPCA methods.
Chen et al.[30]	Employs PCA to generate the most discriminative feature sets, FKNN is applied to detect PD	Better accuracy compared to SVM
Shen et al.[32]	Proposes a new SVM parameter tuning method using fruit fly optimisation algorithm (FOA) to identify PD	Compared with PSO-SVM, GA-SVM, the method attained better accuracy and reduced computational time
Cai et al.[33]	Develops an optimal SVM with bacterial foraging optimisation (BFO) in predicting PD	Relief feature selection increased the accuracy to 97.42%
Cai et al.[37]	Couples the chaotic bacterial foraging optimisation with Gauss mutation (CBFO) approach with FKNN to detect PD	Better performance when compared to PSO, GA and FOA
Zhang et al.[38]	Detects AD using multivariate approach Using Stationary Wavelet Entropy and PP-PSO	Four level decomposition method identified the subject in 0.88 s after volumetric pre-processing
Shankar et al.[46]	Employs Group GWO (GGWO) technique to improve AD detection using KNN, DT and CNN	Reduced feature sets obtained with 96% classification accuracy.
Xiao et al.[48]	Combines different features from sMRI: grey-matter volume, grey-level cooccurrence matrix and Gabor feature to detect AD.	Multi-feature combination correlation is analysed to enhance SVM-RFE algorithm by covariance method
Silva et al.[49]	Combines textural features from GLCM with morphometry neuroimaging analysis based on voxels	Certain anatomical brain regions are targeted and classifiers are compared
Beheshti et al.[51]	Predicts conversion of MCI to AD between first and third year ahead of clinical diagnosis.	Feature-ranking with GA employed to analyse structural magnetic resonance imaging data.

has multi-dominance over ELM by combining the kernel function. There is a non-linear mapping of linear non-separable mode onto a high dimensional (HD) feature space. In this way, improved accuracy and linear separability can be attained. If there are large training samples and where the hidden feature map is not known, KELM has been beneficial. KELM has proved to be better at tackling the issue of large variations in the classification accuracy which are encountered in ELM as reported in different studies. Further, it does not need any randomness in the process of assigning the connection weights between the input and hidden layers. KELM has proved its efficiency in disease diagnosis, hyperspectral remote-sensing image classification, activity recognition, etc. In KELM, hidden layer node optimisation is avoided and hence, possible to obtain the least square optimal solution. SFO and KELM are described in subsequent sections. KELM is used in a wide variety of classification problems as compared to SVM and ELM.[25,53] More information on KELM is established in section 2.4. The overall framework is represented in Figure 1.

Dataset preparation

Alzheimer's Disease (AD) dataset. The Alzheimer's Disease Neuroimaging Initiative (ADNI) database[54] is used (http://adni.loni.usc.edu). The database was created as a part of an initiative to develop biomarkers for AD detection. Samples of 1409 subjects obtained from ADNI's MRI investigations were collected using a 1.5T MRI scanner. The subjects were analysed using the MP-RAGE sequence (T1-weighted) that included 294 subjects with AD, 763 with MCI and 352 normal controls). The database contains images of normal person denoted as NC and disease conformed patients denoted by AD. Persons with a slight, noticeable and measurable decline in their cognitive skills like memory, thinking, etc. are at an elevated risk of developing AD or another dementia which is denoted as MCI. ADNI-1, ADNI-2 and ADNI-GO subjects were included which comprise baseline 3D T1-weighted scans. Those images are also included in the dataset. Table 2 shows the demographic information collected from the ADNI.

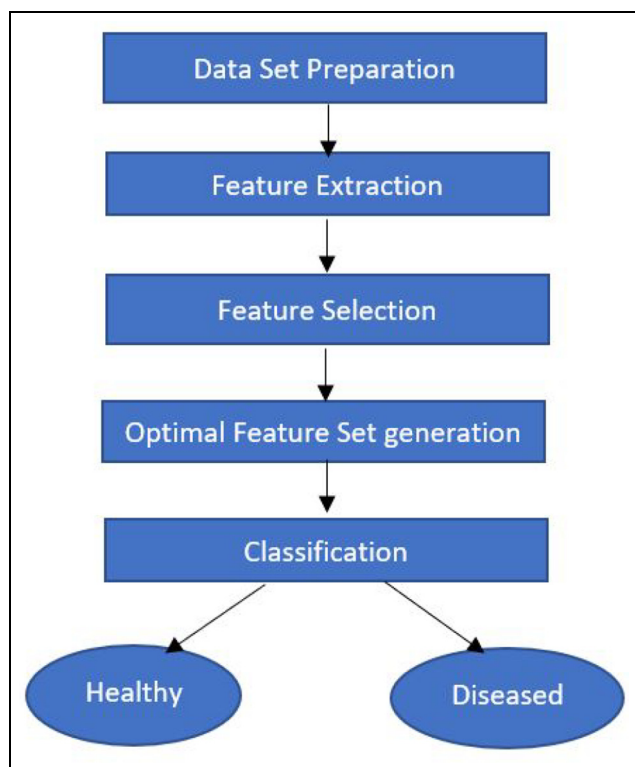Sample images taken from the ADNI dataset is illustrated in Figure 2.

Figure 1. Framework of the method.

Figure 2. Sample images from ADNI dataset: (a) normal controls, (b) Alzheimer disease and (c) mild cognitive impairment.

Table 2. Demographics from ADNI set.

Parameter	NC	AD	MCI
Sample size	352	294	763
Men: Women	167:185	158:136	438:325
Age	74.53 ± 6.16	75.13 ± 7.75	73.80 ± 7.35
CDR	0.03 ± 0.12	4.46 ± 1.61	1.95 ± 0.97
MMSE	29.07 ± 1.16	23.12 ± 2.1	26.91 ± 1.78

CDR: clinical dementia rating scale; NC: normal controls; MCI: mild cognitive impairment; MMSE: mini mental state examination.

Table 3. Demographics from local hospital.

Parameter	NC	AD	MCI
Sample size	40	90	45
Men: Women	22:18	50:40	20:25
Age	67.1 ± 6.2	68.43 ± 8.4	70.6 ± 7.1
CDR	0.04 ± 0.2	1.2 ± 0.6	0.5 ± 0.1
MMSE	29.4 ± 1.2	18.4 ± 4.5	26.4 ± 2.0

CDR: clinical dementia rating scale; NC: normal controls; MCI: mild cognitive impairment; MMSE: mini mental state examination.

An independent dataset is obtained from the local clinical diagnostic centre. A 1.5T MRI scanner was used to obtain MRI brain examinations from which HR T1-weighted MP-RAGE 3D-sequence were acquired. Table 3 depicts the demographics of the set collected. On comparing with the patients in the ADNI dataset, the figures show that the AD patients at the local clinic had acute symptoms. Clinical assessments were completed by an experienced neurologist who was blinded to the MRI results.

performed by pixel replacement with the median score of the raw input image. The procedure followed is the same as that used in.[55] The median filter is preferred owing to its excellent noise reduction with limited blurring, specifically with large magnitudes. A sample pre-processed image is shown in Figure 3.

MRI pre-processing. Brain MR images obtained from the ADNI database are subjected to pre-processing in order to eliminate noise in the MR images. The images are processed to reduce noise using a median filter. Median Filtering is applied on the collected dataset which preserves the edges of the image, rejecting noise and enhances the image quality. Computation is
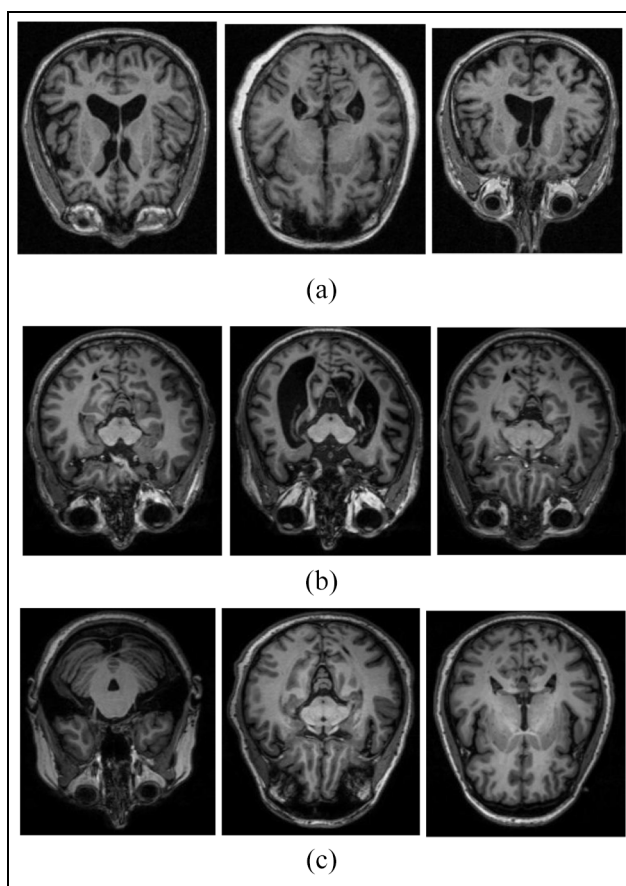
Parkinson's disease (PD) dataset. The PD dataset acquired from the UCI Machine Learning Repository[56] was utilised for evaluating the proposed work. Bio-medical voice measurements are obtained from 31 subjects, out of which 23 are diseased. Each of the columns provided denotes a particular vocal measure, and each of the rows represents one of the 195 voice recordings. The
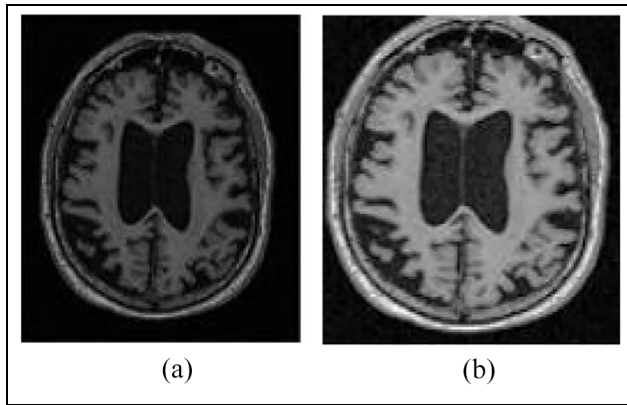
Figure 3. Sample images of original and pre-processed MRI: (a) original MRI and (b) filtered MRI.

Table 4. Demographics from UCI dataset.

Parameters	Values
No of subjects	23 PD and 8 normal
Time line period for recording voice	0–28 years
Age	46–85
Average phonations of the alphabetic vowel	6
Duration	36 s

time line for the diagnosis is from 0 to 28 years with subjects aged between 46 and 85. Each and every subject contributed six phonations on an average, the alphabetic vowel with a duration of 36 s. Table 4 depicts the demographics from the UCI dataset.

Description of the same is provided in Table 5.

Sakar et al.[57] deposited the second data set used in this investigation, known as the Istanbul dataset for PD diagnosis. It included a variety of recordings of the sound, such as sustained vowels, numbers, words and sentences in short collected from 68 persons. A training dataset was obtained from 40 people, which included 20 diagnosed with Parkinson's in the age group of 43–77 and 20 healthy people in the age group of 45–83. The data for testing was acquired from 28 subjects with PD in the range 39–79. Three types of sustained vowel recordings are utilised in this study: /a/, /o/ and /u/, all of which had same data types as the UCI dataset. All are combined and created a database with 288 sustained vowel samples in total, on which the analyses were performed. For each speech sample, a set of 26 linear features and time and frequency based features is extorted, as shown in Table 6.

Feature extraction

AD dataset. Extracting relevant features from the input image is a very specialised task as it increases the system efficiency. Features explain certain computational properties of the input image given. After pre-processing, promising features were identified from the images

Table 5. Parkinson's disease dataset description from UCI.

Features	Attribute	Description
F1	MVDP : Fo (Hz)	Average vocal fundamental frequency
F2	MVDP : Fhi (Hz)	Maximum vocal fundamental frequency
F3	MVDP : Flo (Hz)	Minimum vocal fundamental frequency
F4	MVDP : Jitter (%)	Several measures of variation in fundamental frequency
F5	MVDP : Jitter (Abs)	
F6	MVDP : RAP	
F7	MVDP : PPQ	
F8	Jitter : DDP	
F9	MVDP : Shimmer	Several measures of variation in amplitude
F10	MVDP : Shimmer (dB)	
F11	Shimmer : APQ3	
F12	Shimmer : APQ5	
F13	MVDP : APQ	
F14	Shimmer : DDA	
F15	NHR	Two measures of ratio of noise to tonal components in the voice
F16	HNR	
F17	RPDE	Two non-linear dynamical complexity measures
F18	D2	
F19	DFA	Signal fractal scaling exponent
F20	Spread 1	Three non-linear measures of fundamental frequency variation
F21	Spread 2	
F22	PPE	

that serve as a distinguishing feature for AD diagnosis. From the images that are pre-processed, the following features are extorted as they represent various characteristics of the image which are elucidated below.

Textural features (TF). Texture is characterised by grey scale variation in spatial position. The Grey Level Co-occurrence Matrix (GLCM) and Grey-Level Run Length Matrix (GLRLM) were considered. GLCM represents the intensity correlation between the image pixels at a certain distance in the designated direction.[58] Periodicity and spatial grey level dependencies are depicted by these texture features. GLRLM is generated from surface features. Run length is the count of neighbouring pixels having similar grey power slated under a particular header.

Histogram features (HF). These demonstrate the pixels at every intensity value in the image.

SURF, SIFT and oriented FAST and rotated BRIEF (ORB) Features. These features are object features that are not affected by complications like object scaling, rotation, etc. and are resilient to noise effects.

Table 6. Parkinson's disease dataset description from Istanbul dataset.

Label	Feature
S1	Jitter (local)
S2	Jitter (local, absolute)
S3	Jitter (rap)
S4	Jitter (ppq5)
S5	Jitter (ddp)
S6	Number of pulses
S7	Number of periods
S8	Mean period
S9	Standard dev. of period
S10	Shimmer (local)
S11	Shimmer (local, dB)
S12	Shimmer (apq3)
S13	Shimmer (apq5)
S14	Shimmer (apq11)
S15	Shimmer (dda)
S16	Fraction of locally unvoiced frames
S17	Number of voice breaks
S18	Period of voice breaks
S19	Median pitch
S20	Mean pitch
S21	Standard deviation
S22	Minimum pitch
S23	Maximum pitch
S24	Autocorrelation
S25	Noise-to-Harmonic
S26	Harmonic-to-Noise

Table 7. Features extracted from AD set.

S.No	Feature set	Features extracted
1	GLCM (f1–f22)	Autocorrelation, Contrast, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Entropy, Homogeneity, Maximum Probability, Sum Of Squares, Sum Average, Sum Variance, Difference Variance, Difference Entropy, Information Measure of Correlation 1 And 2, Inverse Difference Normalised, Inverse Difference Moment Normalised, Angular Second Moment, Inertia.
2	GLRLM (f23–f29)	Long Runs Emphasis, Short Runs Emphasis, Grey-Level Non-Uniformity, Run Length Non-Uniformity, Run Percentage, Low Grey-Level Emphasis, High Grey-Level Run Emphasis
3	Histogram (f30–f34)	Mean, Variance, Standard Deviation, Kurtosis, Skewness
4	SURF, SIFT, ORB, LBP (f35–f38)	
5	Wavelet (f39–f42)	Mean, Variance, Homogeneity, Energy
6	Fractal (f43–45)	Fractal Dimension, Lacunarity, Correlation co-efficient

Local binary pattern features (LBP). Binary pattern features represent the surroundings of pixels in the regions.

Wavelet features (WF). Spatial and frequency domain information from the anatomical structures can provide more information by releasing hidden frequency selective information. Anisotropic Dual Tree Complex Wavelet transform (ADT-CWT) was performed on MRI to extract wavelet features. Horizontal or Vertical decomposition of sub bands was performed as in.[59] Anisotropic and directional basis features were extracted that resulted in 10 sub band information. Decomposing sub bands resulted in extraction of first and second order textural features such as mean, variance, homogeneity and energy.

Fractal dimension (FD) features. These features illustrate the appearance and shape of the objects having properties of scale-invariance and self-similarity. Fractal dimension is a measure of surface roughness or boundary irregularity.[60] Box counting method[61] was employed to extract the features such as FD, lacunarity, correlation coefficient.

The list of features extracted is shown in Table 7.

UCI dataset. The list of standard features taken from the UCI is illustrated in Tables 5 and 6 in the previous section.

Feature selection

This step is most significant, as computing all the obtained features is highly complex. There may be some irrelevant/redundant features that do not contribute in identifying the disease, which may have to be discarded. Hence, this step is highly essential. In short, it is necessary to identify the features that have a high potential for influencing or triggering the disease, which aids in improving classification or prediction accuracy. The extracted features are subjected to feature selection by the Sunflower Algorithm (SFO). The next section explains the SFO with initial background information and its applicability in the feature selection process for the datasets collected.

Sunflower optimisation algorithm (SFO). The SFO algorithm, inspired by nature, is based on population proposed by Gomes et al.[62] The algorithm emulates the orientation of sunflowers towards the sun. Sunflowers search to orient themselves in a better way to capture the radiation from the sun. During such a movement, fertilisation happens between the neighbouring flowers in the close vicinity. The aggregate of radiation absorbed depends on the distance between the flowers and the sun in relation to the inverse square law of radiation.[63] If the distance is less between the sun and the plant, more will be radiation absorbed by the

flower, else otherwise. The amount of radiation received by the plant will be

$$I = \frac{P}{4\pi d^2} \qquad (1)$$

Where I denote solar radiation intensity, *P* denotes the source power and d represents inter flower distance and the sun. Each sunflower orients itself to the direction of the sun as shown in equation (2)

$$\overrightarrow{S_i} = \frac{X^* - X_i}{\|X^* - X_i\|}, i = 1, 2, 3, ...n \qquad (2)$$

Where *X** represents the overall best solution, X_i denotes the current solution and n the population size. The step of individuals (sunflowers) towards the sun is expressed as in equation (3)

$$d_i = \gamma P_i (\|X_i + X_{i-1}\|) \times \|X_i + X_{i-1}\| \qquad (3)$$

In the above equation, γ denotes the inertial displacement of the sunflowers. The probability of pollination that each sunflower *i*, fertilises with the closest neighbour *i*–1 producing new individuals in a new position that changes with respect to the distance between the flowers is represented by *PiXi* + X_{i-1}. Each individual that is close to the sun takes small steps in searching for a local refinement while others move normally. Individuals are seen that don't skip regions conforming to the global minimum. Hence, there is a restriction on the maximum step defined for each individual as given by equation (4).

$$d_{max} = \frac{\|X_{max} - X_{min}\|}{2n} \qquad (4)$$

Where X_{min} and X_{max} denote the lower and upper bounds. n represents the population.

Each sunflower *i* updates its position to generate new generation, on the basis of orientation of sunflowers towards the sun as shown in equation (5).

$$\overrightarrow{X_{i+1}} = \overrightarrow{X_i} + d_i \times \overrightarrow{S_i} \qquad (5)$$

Where X_{i+1} represents the newly generated individual (sunflower) position.

SFO starts with the population generation and may be even or randomised. The individual with the best evaluation will be converted to the sun. In the proposed work, the number of suns is restricted to one. All the other sunflowers will then align themselves as sunflowers move towards the sun, randomly moving in a specified direction. Figure 4 represents the SFO algorithm.

Classification

The selected feature set from SFO is fed to the Kernel Extreme Learning Machine (KELM) for classification. KELM, developed by Huang,[24] is an enhanced version of the original ELM. KELM has dominance over the original ELM by combining kernel function onto the ELM. KELM has the property to map non-linearly, the linear non-separable model onto a high dimensional (HD) feature space, ensuring linear separability and a higher accuracy rate. As the ELM learning algorithm is for Single Hidden Layer Feed Forward Neural

---

Algorithm: Sunflower Optimization

1. Set the initial values such as mortality rate *m*, population size *n*, pollination rate *P*, maximum number of iterations *max_it*
2. Set the iteration counter t = 0
3. Generate the population randomly
4. Evaluate the fitness of the individuals (sunflowers) in the population
5. Assign the overall best solution in population X*
6. Adjust all the individuals (sunflowers) towards the best solution (sun) as in Eq (2)
7. Repeat
8.   Compute the fitness (direction vector) for each individual
9.   Remove the worst m% individuals
10.   Calculate the step of the individual towards the sun as in Eq (3)
11.   Fertilise the best sunflowers around the sun
12.   Check the maximum step of the individuals as in Eq (4)
13.   Update the individuals as in Eq (5)
14.   Evaluate the new individuals.
15.   Accept the new individuals, if their fitness is better than the current values.
16.   Set t = t+1
17. Until (t > *max_it*)
18. Produce the best individual.

Figure 4. Pseudocode of SFO.

Network (SLFN), it is possible to represent SLFN as in Huang et al.[64]

$$f(x) = h(x) = H\beta \qquad (6)$$

Where x denotes sample, $f(x)$ denotes the NN output, $h(x)$ is the class vector, H specifies the matrix of hidden layer feature mapping, β represents the weight of the hidden layer output link layer. For minimising training error and enhancing the generalised performance, β is represented as

$$\beta = H^T \left(\frac{1}{C} + HH^T\right)^{-1} T \qquad (7)$$

Where T represents the output matrix of samples expected and C is the regularisation co-efficient. The output function of ELM is

$$f(x) = H\beta = HH^T \left(\frac{1}{C} + HH^T\right)^{-1} T \qquad (8)$$

If $h(x)$ is unknown, the kernel matrix of ELM satisfying Mercer's conditions is stated as

$$m = HH^T; m_{ij} = h(x_i) \cdot h(x_j) = K(x_i, x_j) \qquad (9)$$

Hence, the KELM output function could be expressed by

$$f(x) = [K(x, x_1), \dots K(x, x_N)] \left(\frac{1}{C} + m\right)^{-1} T \qquad (10)$$

$K(x, y)$ represents the kernel function of hidden neurons of SLFN

Many types of kernels such as linear, exponential, wavelets, polynomial, etc. have extensively used in research works. In this work, the kernel used is Gaussian[57] denoted as

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\gamma^2}\right) \qquad (11)$$

For KELM, the vital factors that have to be tuned properly are the regularisation co-efficient C and Kernel parameter γ.[65] They decide the KELM performance.

Experiment and results

In this section, the experiment carried out is investigated and the outcomes are recorded. The experiments are investigated in MATLAB run on Windows 7 ultimate OS, Intel Core i3-7100U CPU @ 2.4 GHz and 8 GB RAM. All the algorithms are implemented from scratch. From the final set of features, 45 from ADNI

Table 8. Features selected from AD dataset.

Feature set	Features selected	Number of features
GLCM	Contrast, Correlation, Dissimilarity, Energy, Entropy, Difference Entropy, Difference Variance, Angular Second Moment	8
GLRLM	Long Run Emphasis, Grey-Level Non uniformity, Run Percentage, High Grey-Level Run Emphasis	4
Histogram	Skewness, Kurtosis	2
	SURF, ORB, LBP	3
Wavelet	Homogeneity, Energy	2
Fractal	Fractal Dimension, Lacunarity	2

Table 9. SFO control parameters.

Parameters	Value
Number of sunflowers	100
Pollination	0.6
Sun	1
Days/Generations	100

and 22 from the PD dataset, the SFO algorithm selected only the top ranked features as provided in Table 8 for the AD set. Totally, 21 features were selected from the AD set. For the PD dataset, six (F1, F16, F17, F18, F19, F22) features from the UCI dataset and eight (S2, S9, S16, S18, S21, S24, S25, S26) features from the Istanbul dataset are selected by the SFO. The selected features are fed separately to the KELM for classification. The efficiency of the SFO depends on maximum number of iterations, (generation, days), which will be the stopping criterion. In the proposed work, SFO is employed to choose the best minimal set of features that enhance the classification rate. The fitness function in this work is set by

$$Fitness = \mu P + (1 - \mu)\frac{N - L}{N}$$

Where P denotes the classification accuracy, L represents the no. of features selected and the total number of features is shown by N. μ denotes the accuracy weight usually selected between 0 and 1. 1–μ is the feature selection quality weight.

Control Parameters set for the SFO is shown in Table 9.

Aside from the standard control parameters described in Table 9, the mortality rate is set as 0.1 and the survival rate as 1 (pollinate rate + mortality rate). Lower bound and upper bound values range from (-5 to $+5$). SFO is compared with Genetic Algorithm (GA) and Grey Wolf Optimisation Algorithm (GWO) on the selection of features as they are prominent
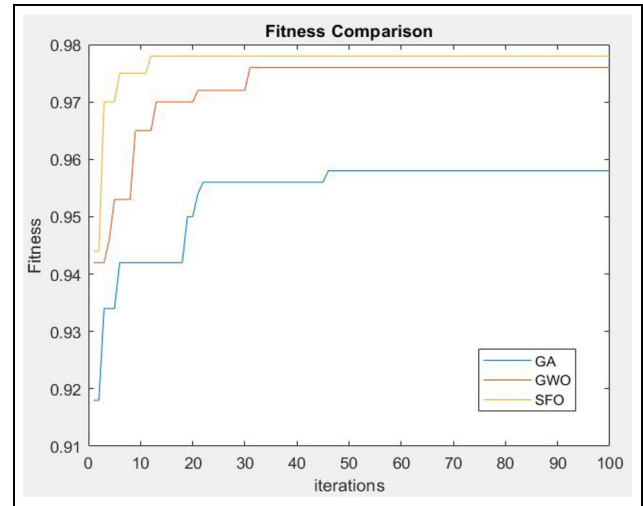
Table 10. Control parameters for GA and GWO.

Parameters	Value
Population size	100
Maximum iteration	100
Problem dimension	N (feature size)
Crossover probability – GA	0.8
Mutation probability – GA	0.01
α and β– GWO fitness function	0.99 and 0.01
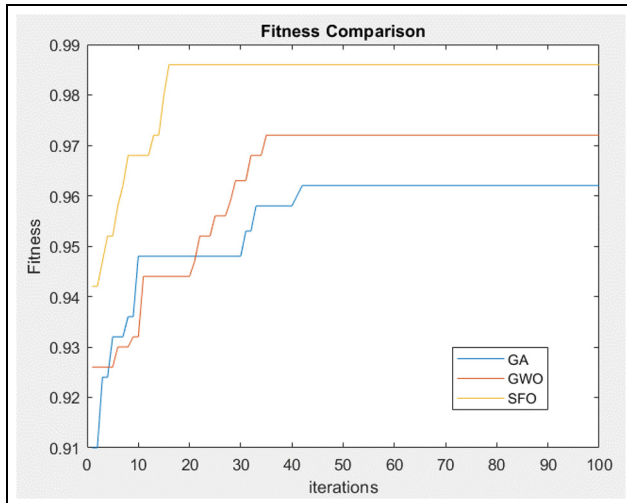
Figure 5. Average Fitness comparison between the algorithms for AD dataset.

Figure 6. Average Fitness comparison between the algorithms for PD dataset.

Table 11. Performance metrics.

Parameter	Expression
Sensitivity	$\dfrac{TP}{TP+FN}$
Specificity	$\dfrac{TN}{TN+FP}$
Accuracy	$\dfrac{TP+TN}{TP+FN+TN+FP}$
Precision/Positive predictive value	$\dfrac{TP}{TP+FP}$
Recall	$\dfrac{TP}{TP+FN}$
F-1 Score	$2*\dfrac{(Precision)(Recall)}{Precision+Recall}$
G-Mean	$\sqrt{sensitivity \times specificity}$

Where
TP: Number of correct predictions that an instance is positive
FP: Number of incorrect predictions that an instance is positive
FN: Number of incorrect predictions that an instance is negative
TN: Number of correct predictions that an instance is negative

optimisation algorithms widely used. Control Parameters for GA and GWO for comparison are given in Table 10. To obtain classification results unbiased, 10-fold cross validation and unseen data validation are performed.[66,67] Most of the studies have preferred k to be 10 for cross fold validation. ninety percent of the available samples were utilised for training purposes, the rest for testing. The average result of all the 10 trials is calculated. In this case, the testing sets stay independent, thus promising reliable and precise results

The optimisation process of GA, GWO and SFO is observed through an iteration process as shown in Figures 5 and 6 for AD and PD, respectively.

It is seen from Figure 5 that SFO converges completely after the 16th iteration while GA and GWO, converge after the 42nd and 35th iterations respectively, on the AD dataset. Similarly, from Figure 6, it can be seen that SFO converges after the 11th iteration compared to GA and GWO which converged after 45th and 30th, respectively on the PD dataset. It clearly indicated that SFO is effective in quickly finding the best solution. The fitness value is also higher compared to GA and GWO. GA and GWO have mixed responses in the initial iterations, while SFO has a clear way to converge.

The features optimised by SFO are fed to KELM for classification. Regularisation co-efficient C is chosen as 32 and kernel parameter γ is selected as 1. It is experimented with different combinations using a grid search technique, where the results recommended the best performance of KELM with (C, γ) as (32, 1). The search range for C is between $\{2^{-5}, 2^{3-}, ...2^5\}$ and that of γ between $\{2^{5-}, 2^{3-}, ...2^5\}$. The classifier produced output that is either diseased or normal. Performance evaluation metrics adopted are tabulated in Table 11.

The performance of the method proposed is validated and tested on patients and controls, with the following classifications such as: AD versus NC, MCI versus NC, AD versus MCI. The model is experimented on the ADNI dataset first, and subsequently on the local dataset for each classification. The dataset is divided into training set and validation set (70% of the images) and

Table 12. Binary classification results on the AD datasets.

Class	Dataset	Acc (%)	Sen (%)	Spec (%)	G-Mean (%)	F-1 score (%)
AD versus NC	ADNI	99.32	98.9	99.66	99.28	99.6
	Local	98.65	98.4	98.5	98.45	98.5
MCI versus NC	ADNI	97.85	96.56	96.67	96.61	96.3
	Local	97.38	97.01	97.1	97.05	95.6
AD versus MCI	ADNI	97.62	95.81	96.87	96.34	97.7
	Local	95.56	94.89	94.76	94.82	96.4

a test set (30% of the images). In addition to accuracy, sensitivity and specificity, G-mean and F1-score measures are calculated. The sensitivity of an imbalanced categorisation may be more significant than the specificity. The geometric mean, or G-Mean, is a measure that combines sensitivity and specificity into a single value that balances both objectives. Even if the negative cases are accurately categorised, a low G-Mean indicates poor performance in the categorisation of positive cases. This criterion is crucial for avoiding overfitting the negative class and underfitting the positive class. Table 12 reports the proposed method's binary classification rates on the test dataset. In all of the comparisons, the outcome showed high classification accuracy levels. The highest accuracy, sensitivity and specificity are obtained in the AD versus NC classification tests employed on both the ADNI and the local dataset. The model provided an appreciable discrimination between MCI and NC with an appreciable performance.

A high F-1 score indicates that the classifier performs well in the case of minority classes. A low value of G-mean show poor classification in categorising positive cases. This criterion is crucial for avoiding overfitting the negative class and underfitting the positive class. The proposed method attains a high value of G-Mean and an F-1 score, thus overcoming the aforementioned issue. Comparative investigations are done between SFO-KELM and other competitive approaches, including GA-KELM, GWO-KELM, to prove the efficacy of the method. Competitive algorithms are used for comparison. Ten-fold cross validation and unseen data validation is performed to avoid bias during the process. Table 13 shows the average classification outcomes of the three optimisation approaches with KELM in terms of selected attributes or features and performance evaluation metrics on the ADNI and local dataset.

From Table 13, it can be inferred that among the three approaches, SFO-KELM has performed better with the least number of features selected with 99.32% accuracy on the ADNI and 98.65% on the local dataset. Figure 7 shows the computational time in seconds during 10-fold CV runs on the ADNI dataset as it contained a greater number of images. It can be inferred that the SFO-KELM requires almost 45sec to finish the training and prediction process in each of the folds for the AD dataset. SFO-KELM on the ADNI dataset

Table 13. Experimental outcomes of the feature selection approaches on AD dataset.

Method	Features selected	Class	Accuracy (%)	
			ADNI	Local
SFO-KELM	21	AD versus NC	99.32	98.65
		MCI versus NC	97.85	97.38
		AD versus MCI	97.62	95.56
GA-KELM	30	AD versus NC	92.47	93.42
		MCI versus NC	93.38	94.18
		AD versus MCI	91.09	92.45
GWO-KELM	25	AD versus NC	95.49	96.67
		MCI versus NC	94.26	95.21
		AD versus MCI	93.22	92.16

requires significantly less computational effort. Figure 8 shows the average computational time in seconds during 10-fold CV runs on the UCI dataset.

As seen from Figure 8, SFO-KELM needs an average of almost 16sec to finish the training and prediction process in each of the folds for the PD dataset. Using SFO-KELM on the PD dataset requires less computational time. To evaluate the KELM classifier, KELM and other classifiers like Random Forest (RF), Support Vector Machine (SVM) and K-NN are compared. Table 14 shows the comparative performance of SFO-KELM with other classifiers on the ADNI and local datasets. From Table 14, it can be inferred that SFO-KELM outperformed other classifiers in predicting the class label. It can also be seen that SFO when used with other classifiers had a variation of around 5% which showed the efficacy of SFO.

Tables 15 and 16 show the average classification outcomes of the three optimisation approaches with KELM in terms of selected attributes or features and performance evaluation metrics on the UCI dataset and Istanbul dataset respectively.

From Table 15, it can be inferred that among the three approaches, SFO-KELM has performed better with the least number of features selected with 99.52% accuracy, 99.11% sensitivity, 98.66% specificity, 98.8% F-1 Score and 98.88% G-Mean. Similar performance is obtained for the Istanbul dataset as well, as observed from Table 16. Though the number of features selected differs from GWO-KELM for the second dataset, the method attained the highest accuracy with 100%
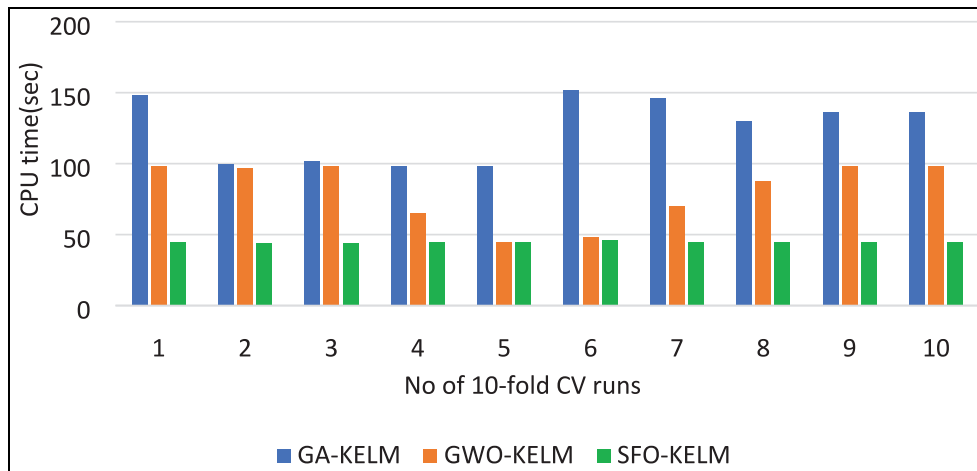
Figure 7. Comparison of the methods in terms of CPU time on the ADNI dataset.
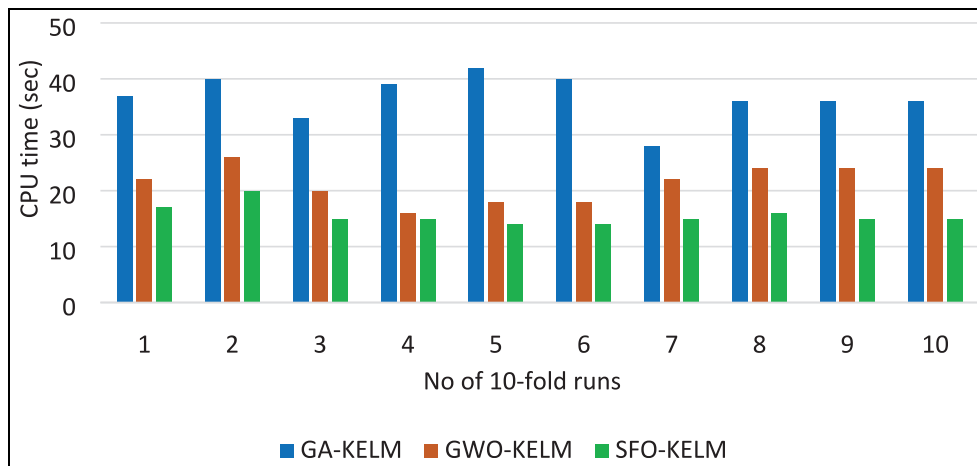
Figure 8. Comparison of the methods in terms of CPU time on the UCI dataset.

Table 14. Comparison of proposed SFO-KELM with other classifiers in AD diagnosis.

Method	Accuracy (%)	
	ADNI	Local
SFO-RF	93.67	92.46
SFO-KNN	95.16	94.28
SFO-SVM	97.45	96.82
SFO- KELM	99.32	98.65

specificity and 99.01% sensitivity. A high F-1 score indicates that the classifier performs well in the case of minority classes. A low value of G-mean show poor classification in categorising positive cases. This criterion is crucial for avoiding overfitting the negative class and underfitting the positive class. The proposed method attains a high value of G-Mean, thus overcoming the afore mentioned issue. To estimate KELM classifier performance, KELM and other classifiers like Random Forest (RF), Support Vector Machine (SVM) and K-NN are compared. Table 17 shows the comparative performance of SFO-KELM with other classifiers. From Table 17, it can be inferred that SFO-KELM

Table 15. Experimental outcomes of the feature selection approaches on UCI dataset.

Method	Features selected	Acc (%)	Sen (%)	Spec (%)	G-Mean (%)	F-1 score (%)
SFO-KELM	6	99.52	99.11	98.66	98.88	98.8
GA-KELM	10	91.69	95.32	94.45	94.90	93.5
GWO-KELM	7	95.66	97.99	94.57	96.26	96.4

Table 16. Experimental outcomes of the feature selection approaches on Istanbul dataset.

Method	Features selected	Acc (%)	Sen (%)	Spec (%)	G-mean (%)	F-1 score (%)
SFO-KELM	8	99.45	99.01	100	99.50	98.9
GA-KELM	10	92.61	94.32	95.55	94.93	94.8
GWO-KELM	6	95.94	97.86	96.74	97.29	96.46

Table 17. Comparison of proposed SFO-KELM with other classifiers in PD diagnosis.

Method	Accuracy (%)	
	UCI	Istanbul
SFO-RF	94.70	95.88
SFO-KNN	96.24	95.76
SFO-SVM	98.12	97.87
SFO- KELM	99.52	99.45

outperformed other classifiers in predicting the class label. It can also be seen that SFO when used with other classifiers had a variation of around 4%–5% which showed the efficacy of SFO.

To accomplish better estimation of the final model's error, the model was tested with unseen data, where the data is randomly split into two sets, one each for training and testing. Table 18 shows the overall experimental results using unseen datasets.

Further, the proposed framework is compared with approaches illustrated in the related literature towards PD and AD as depicted in Table 19.

Conclusion

Data mining techniques have taken a tremendous path in medical diagnosis after the advancement in computer aided technologies. In this work, SFO-KELM was employed in detecting AD and PD. The methodology encompassed the main steps of feature selection and final classification. Distinct datasets were taken for the study: ADNI and local for AD, UCI and Istanbul for PD. The AD dataset was subjected to pre-processing and features were extracted. The PD dataset contained distinct features and were directly subjected to feature extraction. SFO was proposed to select the most informative and distinct features from both the datasets. Lastly, the KELM classifier classified them as either diseased or normal. The feature selection process was compared with other similar algorithms like Genetic Algorithm (GA) and Grey Wolf Optimisation (GWO). Similarly, the effectiveness of the method proposed was also assessed with other classifiers like Random Forest (RF), K-NN and Support Vector Machine (SVM). Experimental outcomes have demonstrated a high accuracy level in both AD and PD diagnosis with appreciable sensitivity and specificity. For AD diagnosis, the

Table 18. Overall results obtained with unseen data.

Dataset	Acc (%)	Sen (%)	Spec (%)	G-Mean (%)	F-1 score (%)
ADNI	97.64	94.55	96.60	95.81	95.06
UCI	96.97	98.16	95.73	96.62	98.16
Istanbul	98.24	97.60	98.58	98.09	99.10

Table 19. Comparison of the proposed method with other state of art methods.

Study	Method	Disease	Acc (%)
Spadoto et al.[35]	Evolutionary based – Harmony Search, gravitational search and OPF	PD	84.01
Astrom et al.[34]	Parallel NN	PD	91.02
Chen et al.[30]	Fuzzy KNN	PD	96.07
Cai et al.[33]	SVM and Bacterial Foraging Optimisation	PD	97.42
Cai et al.[37]	Chaotic Bacterial Foraging Optimisation Enhanced Fuzzy KNN Approach,	PD	97.89
Zhang et al.[38]	Stationary Wavelet Entropy and Predator-Prey Particle Swarm OptimisationOptimization'	AD	92.73
Jialin Peng et al.[42]	Structured Sparse Kernel Learning	AD	96.01
Khan et al.[45]	Instance-Based Learning Techniques	AD	96
Shankar et al.[46]	Group Grey Wolf Optimisation Based Features with Convolutional Classifier	AD	96.23
Islam and Zhang[47]	Ensemble of Deep CNN	AD	93.18
Silva et al.[49]	Haralick Texture Features, Feature Selection using GA and PSO	AD	78.9 and 77.66
Proposed method	SFO and KELM	AD/PD	99.3/99.5

highest classification rate was attained in the AD versus NC classification using the ADNI dataset (99.32%) and local dataset (98.65%). For PD diagnosis, the highest accuracy of 99.52% and 99.45% was achieved on the UCI and Istanbul datasets respectively. It was also observed that running the algorithm 10 times resulted in better selection strategy and classification accuracy. Experiment with unseen data was also adopted for a better estimation of the final model's error. The SFO algorithm was quick to converge, yielding the highest fitness among the other optimisation algorithms. Trained with AD and controls, and on PD, the model could be specifically be efficient when trying to differentiate between MCI versus AD versus NC and PD versus non-PD. Therefore, the model is an ideal tool effective in clinical practise as it consumes less time and provide significant performance in distinguishing only slightly different images. This method has proved better at predicting the class and can act as a second expert in diagnosis at an early stage. It can be concluded that the method proposed can be extended to other diseases like glaucoma, age-related macular degeneration, breast cancer, hepatitis and so on in the future. The results can help physicians start the treatment early and have control over the disease progression. Future research is needed to determine the procedure's accuracy in diagnosing persons in pre-clinical stage of the disease and, perhaps, as a diagnostic tool for people at an elevated risk of acquiring dementia. Secondly, the model must be investigated combining clinical, genetic, cognitive and other biomarkers to enhance the prediction rate of complete-blown development of dementia in MCI patients and near PD subjects. Thirdly, hybridisation of algorithms may yield better accuracy with even lesser number of features being selected.

References

1. World Health Organization. *Report of the WHO independent high-level commission on non-communicable diseases.* 4th ed. Geneva, Switzerland: WHO, 2018. http://apps.who.int/iris/bitstream/handle/10665/272710/97892415141 63-eng.pdf?ua = 1

2. Wirdefeldt K, Adami HO, Cole P, et al. Epidemiology and etiology of Parkinson's disease: a review of the evidence. *Eur J Epidemiol* 2011; 26 (Suppl 1): S1–58.

3. Prince M, Ali GC, Guerchet M, et al. Recent global trends in the prevalence and incidence of dementia, and survival with dementia. *Alzheimers Res Ther* 2016; 8(1): 23.

4. Shubham B, Arvind Kumar T and Kumar Sahani A. A survey of machine learning based approaches for Parkinson disease prediction. *IAARD Int J Comput Sci Inform Technol* 2015; 6(2): 1648–1655.

5. Lim KL and Tan JM. Role of the ubiquitin proteasome system in Parkinson's disease. *BMC Biochem* 2007; 8(Suppl 1): S13.

6. Singleton AB, Farrer MJ and Bonifati V. The genetics of Parkinson's disease: progress and therapeutic implications. *Mov Disord* 2013; 28(1): 14–23.

7. Moore DJ, West AB, Dawson VL, et al. Molecular pathophysiology of Parkinson's disease. *Annu Rev Neurosci* 2005; 28(57–87): 57–87.

8. Sharma S, Moon CS, Khogali A, et al. Biomarkers in Parkinson's disease (recent update). *Neurochem Int* 2013; 63(3): 201–229. Epub. DOI: 10.1016/j.neuint.2013.06.005

9. Amira Ben R, et al. Diagnosis of Alzheimer diseases in early step using SVM (support vector machine). In: *2016 13th International conference on computer graphics, imaging and visualization (CGiV)*, Beni Mellal, Morocco, 29 March–1 April 2016, pp. 364–367. New York: IEEE.

10. Bhagyashree SR and Sheshadri HS. An approach in the diagnosis of Alzheimer disease – a survey. *Int J Eng Trends Technol* 2014; 7(1): 41–43.

11. Joshi SD and Mandal A. P1-190: Alzheimer's disease and how the community responses in developing country. *Alzheimer Dement* 2009; 5(4S_Part_8): 232–233.

12. Parry J and Weiyuan C. Looming dementia epidemic in Asia. *Bull World Health Organ* 2011; 89(3): 166–167.

13. Maharaj T. Live-model simulation: improving nursing students' attitudes and knowledge of Alzheimer's disease. *Clin Simul Nurs* 2017; 13(9): 446–451.

14. Anderson LA and Egge R. Expanding efforts to address Alzheimer's disease: the healthy brain initiative. *Alzheimers Dement* 2014; 10(5 Suppl): S453–S456.

15. Dementia Australia. https://www.dementia.org.au/information/diagnosing-dementia

16. Joshi S, Shenoy D, Vibhudendra Simha G, et al. Classification of Alzheimer's disease and Parkinson's disease by Using Machine Learning and Neural Network Methods. In: *Proceedings of the 2010 second international conference on machine learning and computing (ICMLC 10)*, Bangalore, India, 9–11 February 2010, pp. 218–222. New York: IEEE. . DOI: 10.1109/ICMLC.2010.45

17. Pagan FL. Improving outcomes through early diagnosis of Parkinson's disease. *Am J Manag Care* 2012; 18(7 Suppl): S176–S182.

18. Mirjalili S and Lewis A. The whale optimization algorithm. *Adv Eng Softw* 2016; 95: 51–67.

19. Wu T, Yao M and Yang J. Dolphin swarm algorithm. *Front Inf Technol Electron Eng* 2016; 17(8): 717–729.

20. Eltaeib T and Mahmood A. Differential evolution: a survey and analysis. *Appl Sci* 2018; 8(10): 1945.

21. Chakraborty A and Kar AK. Swarm intelligence: a review of algorithms. In: Patnaik S, Yang XS and Nakamatsu K (eds) *Nature-Inspired computing and optimization. Modeling and optimization in science and technologies*. Vol. 10. Cham: Springer, 2017, pp.475–494.

22. Câmara D. Evolution and evolutionary algorithms. In: Câmara D (ed.) *Bio-inspired networking*. Amsterdam, The Netherlands: Elsevier, 2015, pp.1–30.

23. Darwish A. Bio-inspired computing: algorithms review, deep analysis, and the scope of applications. *Future Comput Inform J* 2018; 3(2): 231–246.

24. Huang GB, Zhu QY and Siew CK. Extreme learning machine: theory and applications. *Neurocomputing* 2006; 70: 489–501.

25. Huang GB, Zhou H, Ding X, et al. Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern B Cybern* 2012; 42(2): 513–529.

26. Little MA, McSharry PE, Hunter EJ, et al. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans Biomed Eng* 2009; 56(4): 1015.

27. Das R. A comparison of multiple classification methods for diagnosis of Parkinson disease. *Expert Syst Appl* 2010; 37(2): 1568–1572.

28. Sakar CO and Kursun O. Telediagnosis of Parkinson's disease using measurements of dysphonia. *J Med Syst* 2010; 34(4): 591–599.

29. Li DC, Liu CW and Hu SC. A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. *Artif Intell Med* 2011; 52(1): 45–52.

30. Chen HL, Huang CC, Yu XG, et al. An efficient diagnosis system for detection of Parkinsons disease using fuzzy k-nearest neighbor approach. *Expert Syst Appl* 2013; 40(1): 263–271.

31. Ozcift A and Gulten A. Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. *Comput Methods Programs Biomed* 2011; 104(3): 443–451. Epub. DOI: 10.1016/j.cmpb.2011.03.018

32. Shen L, Chen H, Yu Z, et al. Evolving support vector machines using fruit fly optimization for medical data classification. *Knowl Based Syst* 2016; 96: 61–75.

33. Cai Z, Gu J and Chen HL. A new hybrid intelligent framework for predicting Parkinson's disease. *IEEE Access* 2017; 5: 17188–17200.

34. Åström F and Koker R. A parallel neural network approach to prediction of Parkinsons disease. *Expert Syst Appl* 2011; 38(10): 12470–12474.

35. Spadoto AA, Guido RC, Carnevali FL, et al. Improving Parkinsons disease identification through evolutionary-based feature selection. *Annu Int Conf IEEE Eng Med Biol Soc* 2011; 2011: 7857–7860.

36. Polat K. Classification of Parkinson's disease using feature weighting method on the basis of fuzzy C-means clustering. *Int J Syst Sci* 2012; 43(4): 597–609.

37. Cai Z, Gu J, Wen C, et al. An intelligent Parkinson's disease diagnostic system based on a chaotic bacterial foraging optimization enhanced fuzzy KNN approach. *Comput Math Methods Med* 2018; 2018: 2396952.

38. Zhang Y, Wang S, Sui Y, et al. Multivariate approach for Alzheimers disease detection using stationary wavelet entropy and predator-prey particle swarm optimization. *J Alzheimers Dis* 2018; 65(3): 855–869.

39. Dessouky MA and Elrashidy M. Feature extraction of the Alzheimer's disease images using different optimization algorithms. *J Alzheimers Dis Parkinsonism* 2016; 06: 230.

40. Tanveer M, Richhariya B, Khan RU, et al. Machine learning techniques for the diagnosis of Alzheimer's

disease: a review. *ACM Trans Multimed Comput Commun Appl* 2020; 16(1s): 35.

41. Ortiz A, Munilla J, Górriz JM, et al. Ensembles of deep learning architectures for the early diagnosis of the Alzheimer's disease. *Int J Neural Syst* 2016; 26(7): 1650025.

42. Peng J, Zhu X, Wang Y, et al. Structured sparsity regularized multiple kernel learning for Alzheimers disease diagnosis. *Pattern Recognit* 2019; 88: 370–382.

43. Shree SR and Sheshadri HS. An initial investigation in the diagnosis of Alzheimers disease using various classification techniques. In: *2014 IEEE International conference on computational intelligence and computing research*, Coimbatore, India, 18–20 December 2014, pp. 1–5. New York: IEEE.

44. Khan A and Usman M. Early diagnosis of Alzheimer's disease using machine learning techniques – a review paper. *In Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K)* 2015; 1: 380–387.

45. Khan A, Liu LS, Usman M, et al. Early diagnosis of Alzheimer disease using instance-based learning techniques. *J Med Imaging Health Informat* 2016; 6(4): 1111–1118.

46. Shankar K, Lakshmanaprabu S, Khanna A, et al. Alzheimer detection using group grey Wolf optimization based features with convolutional classifier. *Comput Electr Eng* 2019; 77: 230–243.

47. Islam J and Zhang Y. 2017. An ensemble of deep convolutional neural networks for Alzheimer's disease detection and classification. *arXiv preprint arXiv:* 1712.01675

48. Xiao Z, Ding Y, Lan T, et al. Brain MR image classification for Alzheimer's disease diagnosis based on multifeature fusion. *Comput Math Methods Med* 2017; 2017: 1952373.

49. Silva IR, N. N. Lima M, Santos W, et al. A model for diagnosis of Alzheimer's disease using Haralick texture features and meta feature selection. *MLDM* 2019; 1: 132–145.

50. Acharya UR, Fernandes SL, WeiKoh JE, et al. Automated detection of Alzheimer's disease using brain MRI images – a study with various feature extraction techniques.. *J Med Syst* 2019; 43: 302.

51. Beheshti I, Demirel H and Matsuda H. Classification of Alzheimers disease and prediction of mild cognitive Impairment-to-Alzheimers conversion from structural magnetic resource imaging using feature ranking and a genetic algorithm. *Comput Biol Med* 2017; 83: 109–119.

52. Qais MH, Hasanien HM and Alghuwainem S. Identification of electrical parameters for three-diode photovoltaic model using analytical and sunflower optimization algorithm. *Appl Energy* 2019; 250: 109–117.

53. Iosifidis A, Tefas A and Pitas I. Large-scale nonlinear facial image classification based on approximate kernel Extreme Learning Machine. In: *Proceedings of the IEEE international conference on image processing (ICIP 15)*, Quebec, Canada, 2015, pp. 2449–2453. New York: IEEE.

54. Mueller SG, Weiner MW, Thal LJ, et al. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin N Am* 2005; 15(4): 869–877, xi–xii.

55. Sun T and Neuvo Y. Detail-preserving median based filters in image processing. *Pattern Recognit Lett* 1994; 15(4): 341–347.

56. Dua D and Graff C. *UCI Machine learning repository* [http://archive.ics.uci.edu/ml]. Irvine, CA: University of

California, School of Information and Computer Science, 2019.

57. Sakar BE, Isenkul ME, Sakar CO, et al. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J Biomed Health Inform* 2013; 17(4): 828–834.

58. Haralick RM, Shanmugam K and Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973; SMC-3(6): 610–621.

59. Kausu TR, Gopi VP, Wahid KA, et al. Combination of clinical and multiresolution features for glaucoma detection and its classification using fundus images, Biocybern. *Biomed Eng* 2018; 38(2): 329341.

60. Chaudhuri BB and Sarkar N. Texture segmentation using fractal dimension. *IEEE Trans Pattern Anal Mach Intell* 1995; 17: 72–77.

61. Buczkowski S, Hildgen P and Cartilier L. Measurements of fractal dimension by box-counting: a critical analysis of data scatter. *Physica A: Statistical Mechanics and its Applications* 1998; 252(1-2): 23–34.

62. Gomes GF, da Cunha SS and Ancelotti AC. A sunflower optimization (SFO) algorithm applied to damage identification on laminated composite plates. *Eng Comput* 2019; 35(2): 619–626.

63. McInnes CR. Solar radiation pressure. In: *Solar sailing. Astronomy and planetary sciences*. London: Springer, 1999.https://doi.org/10.1007/978-1-4471-3992-8_2

64. Huang GB, Wang DH and Lan Y. Extreme learning machines: a survey. *Int J Mach Learn Cybern* 2011; 2: 107–122.

65. Duan L, Dong S, Cui S, et al. 2016) Extreme learning machine with Gaussian Kernel based relevance feedback scheme for image retrieval. In: Cao J, Mao K, Wu J, et al. (eds) *Proceedings of ELM-2015 Volume 1. Proceedings in adaptation, learning and optimization*, vol 6. Springer, Cham. DOI: 10.1007/978-3-319-28397-5_31.

66. Li B, Rong X and Li Y. An improved kernel based extreme learning machine for robot execution failures. *Sci World J* 2014; 2014: 1–7.

67. Kohavi RON. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th international joint conference on artificial intelligence –Volume 2 (IJCAI95)*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1995, pp.1137–1143.