

Effective Prediction of Rheumatoid Arthritis (Ra) Diagnosis Using Hybrid Harmony Search with Adaptive Neuro Fuzzy Inference System

S SHANMUGAM (✉ shanmugam.network13@gmail.com)

Kongunadu College of Engineering and Technology <https://orcid.org/0000-0001-6935-4980>

C Saravanabhavan

Kongunadu College of Engineering and Technology

T Arunkumar

Vellore Institute of Technology: VIT University

Research Article

Keywords: Rheumatoid Arthritis, Diagnosis Model, Harmony Search, ANFIS, Classification

Posted Date: March 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-252272/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

EFFECTIVE PREDICTION OF RHEUMATOID ARTHRITIS (RA) DIAGNOSIS USING HYBRID HARMONY SEARCH WITH ADAPTIVE NEURO FUZZY INFERENCE SYSTEM

S Shanmugam¹, C Saravanabhavan², T Arunkumar³

¹Assistant Professor, Department of Computer Science and Engineering, Kongunadu College of Engineering, Thottiyam, Trichy, India.

² Professor, Department of Computer Science and Engineering, Kongunadu College of Engineering, Thottiyam, Trichy, India.

³Senior Professor, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India.

E-Mail CA: shanmugams@kongunadu.ac.in

ABSTRACT

One of the severe auto immune diseases that affects the entire human body is Rheumatoid Arthritis (RA), the disease triggers one's immune system to attack the inner linings of bones and causes severe inflammation of the synovium. The continuous erosion of bone lining leads to permanent loss of the joint, accounting this severity the early prognosis of the disease is a significant and inevitable process. But, the sign and symptoms of the disease are always uncertain. The symptom of RA disease is similar to other inflammatory diseases, so highly experienced experts can identify the disease in its early stage. To support the clinicians and technicians for early prognosis of the disease, a computer-aided decision support model based on Harmony Search –Adaptive Neuro Fuzzy Inference System is presented in this study. The Harmony search algorithm is employed to select the optimal features, and ANFIS is adopted to perform classification. To demonstrate the effectiveness of the model, metrics such as Accuracy, Sensitivity, Specificity, Precision, Recall, F-measure, Positive Predictive Value, Negative Predictive Value, Root Mean Square Error, and Mean Absolute Error are employed and evaluated in MATLAB simulation environment. The proposed HS-ANFIS outperformed other models developed in this research and existing works of literature.

Index terms- Rheumatoid Arthritis; Diagnosis Model; Harmony Search; ANFIS; Classification.

1. INTRODUCTION

Rheumatoid Arthritis (RA) is a chronic systemic inflammatory sickness [1&2] that occurs on joints and muscles, resulting in apparent interference of joint structure and capacity. According to the World Health Organization (WHO) report, the prevalence rate of RA is 0.3%

to 1% worldwide, and in developed countries the disease is becoming a major threat. In India, RA's prevalence is estimated to be 0.92%, and women are highly vulnerable to the disease than men [3]. According to the WHO report, it was estimated that there are more than 100 types of arthritis [4], arthritis affects more than one among four adults in the population [5]. According to arthritis foundation America, one in three adults have arthritis. In spite of encountering greater challenges in treating RA disease, clinicians have a high degree of challenges to diagnose the RA disease.

The RA disease is basically diagnosed by satisfying the American College of Rheumatology (ACR)/European League Against Rheumatism (EULAR) classification criteria for RA [6]. The diagnosis of RA over suspected patients is carried out by undertaking a series of laboratory tests and analysing the brief case history. Initially, the RA started with severe pain without any swelling, the symptom is a non-specific and also appears to be in conjunction with other immune diseases. The major clue for suspecting RA disease is a prolonged period of morning stiffness in limited number of joints. The laboratory test results can improve the diagnostic sensitivity, and they are carried out by examining the level of ESR (Erythrocyte sedimentation rate), CRP (C-reactive protein), RF (Rheumatoid factor), anti-CCP (Anti-cyclic citrullinated protein), ANA (Antinuclear antibody), blood count rate, and other imaging tests. The ESR presents the blood settlement rate through a liquid column, and the ESR rate helps to distinguish the inflammatory and non-inflammatory diseases. Still ESR level varies with age [7] and also it may include other diseases like malignancy, pathogenic infections, and other related conditions. The RA factor represents auto immune proteins present in our body, about 5% to 10% of peoples have a higher level of RA factor, which get rises with age [8]. The most effective laboratory test for RA diagnosis is counting the anti-CCP anti-bodies. ELISA is the test used to calculate the anti-CCP rate, and it is more specific test to report positive RA than Rheumatoid factor [9]. The test based on estimating the C-reactive protein, the protein which is released by the liver during inflammatory diseases. With this test, the patient may reported to be affected with Systemic Lupus Erythematosus (SLE) or RA. The ANA has a specific disease association with RA, and about 98% of patients affected with SLE reported with positive ANA, and those who affected with other connective tissue disease show 40% to 70%. Still, it was reported that ANA might present about 5% in many healthy individuals [10]. So, with ANA test the patient health history also should be analysed to report positive RA. Thus, it is clearly understandable that the single laboratory test is not enough to diagnose the RA disease, the patient is supposed to undergo series of various laboratory tests. Based on all the

test results investigated with a patient health history, age, genetic factors and so on, and the clinician can come to a valid conclusion of reporting positive RA or Non-RA. Meanwhile, if the diagnosis procedure is delayed, the course of the disease may become severe, which may pay the way for other risk factors and even permanent loss of life.

The designing of intelligent -making model for early diagnosis of RA is always an open field of research. Numerous research studies have been presented in existing literature. Still, the thirst of finding the best model has not elapsed because of its challenge to meet better classification accuracy. In this research article, a computer-aided decision support model based on Harmony Search –Adaptive Neuro Fuzzy Inference System is developed, and the performance is evaluated by employing metrics such as classification accuracy, sensitivity, specificity, precision, recall, F-measure, PPV, NPV, RMSE, and MAE.

1.1 Research Contribution

The contributions made in the proposed work are as follows:

- To collect and preprocess the real-time dataset includes both RA and Non-RA affected patients.
- Feature selction is performed by employing Harmony search strategy.
- Classification is performed by employing Adaptive Neuro Fuzzy Inference approach.
- To evaluate the performance of proposed HS-ANFIS approach is demonstrated by comparing with state of the art approaches.

The remaining section of this paper is organized as follows: Related works are presented in section 2, the methodology is discussed in section 3, the experimental modelling is discussed in section 4, the performance metrics are presented in section 5, the simulation results are briefed in section 6 and based on the investigation of results, inferences and conclusion are presented in section 7.

2. RELATED WORKS

The RA is an auto-immune disease which affects the joints of hands, legs, wrist, knees, angles etc. The disease is also termed as systematic disease as its nature of affecting is not only the joints but also the other organs of the body such as lungs, heart etc. The essential domains and tests of clinical examinations in rheumatology to report a patient with RA, numerous clinical examinations are needed. Indeed the experience and more skilled man power is needed

to avoid late and wrong diagnosis [11]. The mortality risk associated with ILD and RA has been made and it was identified that the patients affected by ILD and RA reported high mortality than the patients affected by RA without ILD [12]. The risk factors associated with RA include, development of malignant lymphoma, cardio vascular disease, atherosclerosis, Temporomandibular disorders, depression among women's voice disorder [13-24].

Considering the major stress patients undergo, researchers have proposed many intelligent decision support models for the past few decades for prognosis of the disease. Machine learning methods for early prediction of RA based on electronic health records [25-29], deep learning strategy on X-ray images [30], an ensemble approach for disease gene identification, where EPU achieved an accuracy of 84.8% [31]. The Decision Stump as weak Learner, and Cuckoo search named CS-Boost for early prognosis of the disease [32]. Adaboost based classifier model for early diagnosis of fibromyalgia and arthritis [33], Numerous Neural Network based diagnosis model for arthritis diagnosis. Neural network based RA diagnosis is investigated in [34]. A Java based software tool has been presented which performed pairwise alignment and analysis based on mutability score [35].

The risk factors associated with the RA diseases and the existing models on clinical decision support system for RA disease diagnosis are presented in the survey. The presented models are mainly concentrated on classification accuracy as performance metric to evaluate the model that is inadequate to validate a medical diagnosis procedure. In this article, numerous metrics are employed to validate the model. Further, there is no much studies available in this field of RA diagnosis research because of the complex nature of disease symptoms which imposes a huge burden to attain the outcome. In this article ANFIS model is developed to perform classification because of its adaptive nature to handle the high level of uncertainty in disease diagnosis [36-44] , and feature selection is a significant step in the design of classification model. Here, Harmony search algorithm is employed to select the necessary features, due to its advantage of simple implementation, better learning ability and improved convergence [45-49].

3. PROPOSED METHODOLOGY FOR RA DIAGNOSIS

3.1 Harmony Search algorithm

Harmony Search (HS) algorithm is a meta-heuristic calculation that emulates the ad-lib procedure of a music player. Every performer plays a note while discovering the best notes of Harmony from end to end. The fundamental goal of the calculation is to eliminate the overall

complexity occur during the search process. The HS method is inspired by the underlying principles of the harmony improvisation [50]. The calculation stream demonstrates parameter HMCR called concordance memory tolerating or thinking about rate. If the HMC rate is excessively low, just a couple of best harmonies are chosen and they lead combine too gradually. If this rate is to a great degree high (close to 1), every one of the harmonies are utilised in the concordance memory, consequently prompting different harmonies are not investigated well conceivably wrong arrangements acquired. Subsequently, ordinarily, we use HMCR=0.9. In standard, the pitch can be balanced straight or nonlinearly, and direct alteration is utilised.

$$X_{new} = X_{old} + PB * \varepsilon \quad (1)$$

Where X_{old} is the current pitch or arrangement from the concordance memory, pitch adjustment determined by a pitch bandwidth PB and a pitch adjusting rate PAR. Here ε is a random number generator in the range of [-1, 1]. We can relegate a pitch-changing rate PAR to control the level of the alteration. Along these lines, we generally use PAR =0.1 ~0.5 in many applications. Three parts in harmony search can be outlined as the pseudo code appeared in Algorithm 1 and flow diagram in Figure 1. We can see that the likelihood of randomisation is

$$P_{random} = 1 - HMCR \quad (2)$$

Authentic likelihood indicates adjusting pitches

$$P_{pitch} = HMCR * PAR \quad (3)$$

Algorithm 1: Pseudo code of Harmonics Search Algorithm

Begin

Objective utility $\Delta N_A = \{N_1, N_2, N_3, \dots, N_n\}$

Produce primary harmonics (real number arrays)

Characterise Pitch Adjusting Rate (PAR)

Pitch limits and transmission capacity

Classify Harmony Memory Accepting Rate (HMCR)

While($t < \text{Max num of repetition}$)

Create innovative harmonics by receiving finest harmonics

Regulate pitch to get innovative harmonics (resultant)fitness value

If($\text{rand} > \text{HMCR}$),

Select a staying harmonic irregularly

Else if ($rand > PAR$),

Fine-tune the pitch irregularly within limits

Else

Make innovative harmonics via randomised algorithms

End if

Admit the innovative harmonics (resultant) if improved

End while

Calculate the fitness value

*Find the present finest conclusion **End***

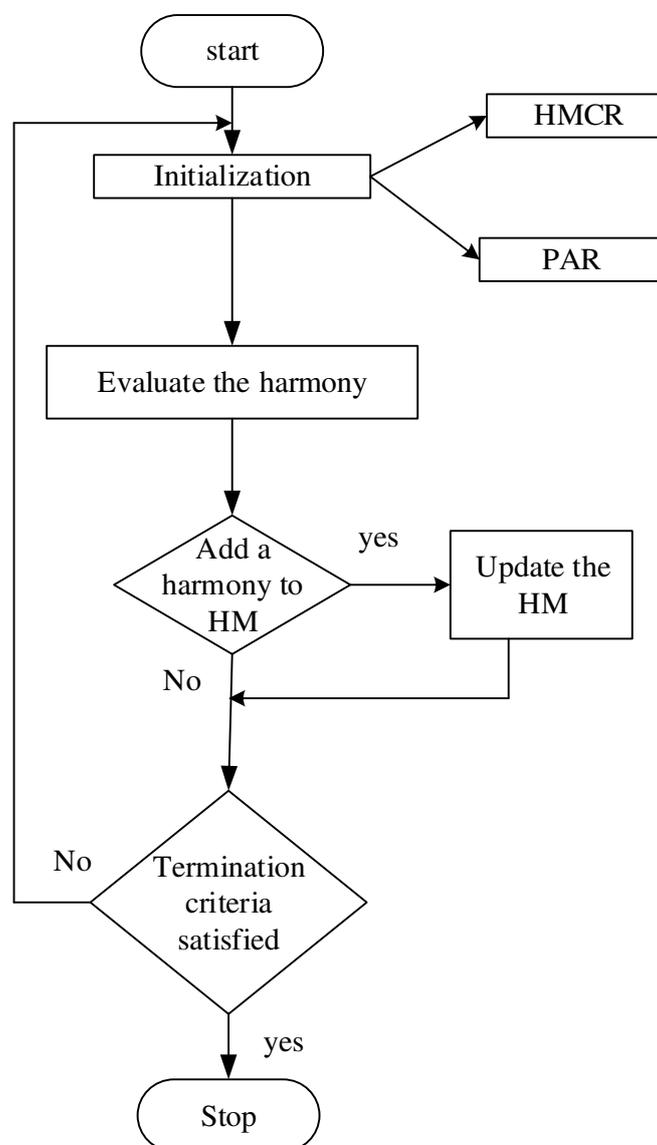


Figure 1 Harmony search flow diagram

3.2 Adaptive Neuro-Fuzzy Inference System (ANFIS)

Adaptive Neuro-Fuzzy Inference System (ANFIS) is a hybrid structure containing the neural framework and the fuzzy logic [51]. The standard base contains the fuzzy on the chance that guidelines of Takagi and Sugeno's sort as seek after: If x is A and y is B then z is f(x, y) where A and B is the fluffy sets in the antecedents and z = f(x, y) is a fresh capacity. For the most part f(x, y) is a polynomial for the data factors x and y. Regardless, it can be whatever other limits that can portray the yield of the system inside the fuzzy. At whatever point f(x, y) is predictable, a zero demand Sugeno-fuzzy model is formed which may be seen as a remarkable example of Mamdanifuzzy reasoning system where every standard coming about is controlled by a fuzzy singleton. If f(x, y) is taken to be a first demand polynomial, Sugeno fuzzy model is encircled. For the first demand two guidelines Sugeno fuzzy derivation structure is given in Figure 2, the two precepts may be communicated as:

$$\textbf{Rule 1:}$$
 If x is A1 and y is B1 then $f_1 = p_1x + q_1y + r_1$ (4)

$$\textbf{Rule 2:}$$
 If x is A2 and y is B2 then $f_2 = p_2x + q_2y + r_2$ (5)

In this above derivation system, the yield of every standard is a direct blend of the information factors included by a steady term. The last yield is the weighted normal of each standard's yield.

Layer 1: All the hubs are versatile hubs. Compute Membership Value for Premise Parameter Output, I for hub i=1, 2

$$O_{1,i} = \mu_{A_i}(x_1) \quad (6)$$

Output O_i , i for node i=3, 4

$$O_{1,i} = \mu_{B_{i-2}}(x_2) \text{ for } (i=1, 2) \quad (7)$$

Layer 2: Firing Strength of Rule Use T-norm (min, product, fuzzy AND ...)

$$O_{2,i} = w_i = \mu_{A_i}(x_1)\mu_{B_i}(x_2) \text{ for } (i=1, 2) \quad (8)$$

Layer 3: Normalize Firing Strength Ratio of ith rule's firing strength vs. all Rules' firing strength

$$O_{3,i} = \overline{W}_i = \frac{w_i}{w_1 + w_2} \quad (9)$$

Layer 4: Consequent Parameters Takagi-Sugeno type output

$$O_{4,i} = \overline{W}_i f_i = \overline{w}_i (p_i x_i + q_i x_2 + r_i) \quad (10)$$

Consequent parameters $\{p_i, q_i, r_i\}$

Layer 5: Overall Output

$$O_{5,1} = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (11)$$

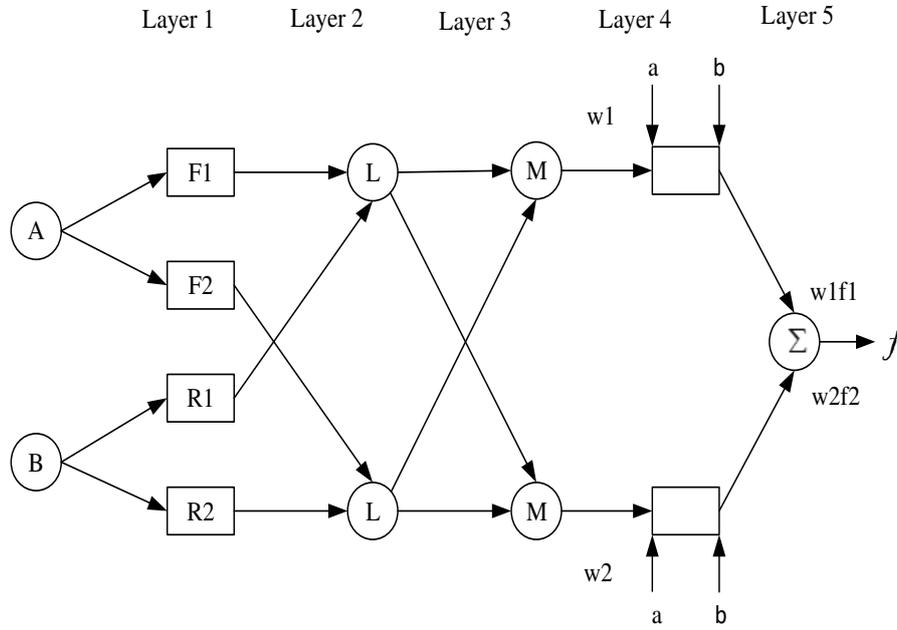


Figure 2 Adaptive Neuro Fuzzy Inference System Structure

3.3 Training Phase Using Hybrid HS-ANFIS

The RA prediction in the data can be considered as a classification problem. It comprises the feature extraction and classification procedure. With the consideration of the property of the RA data with the necessary features identified by HS algorithm, the ANFIS is used as a classifier to differentiate the normal and RA. The technique of the proposed method can be concisely given as follows:

Stage 1: Initially instate the HS Memory (HM). It contains an improvement issue under thought. For an n-measurement issue, a HM with the extent of HMS can be spoken to as pursues:

$$HM = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_n^1 \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_1^{HMS} & x_2^{HMS} & \dots & x_n^{HMS} \end{bmatrix} \quad (12)$$

Where n is the dimension of the problem.

Stage 2: Improvise another arrangement from the HM. Relies upon the Harmony Memory Considering Rate (HMCR) Each segment of this arrangement is gotten.

Stage 3: Update the HS Memory.

Stage 4: Repeat Stage 2 to Stage 3 until the maximal number of emphasis is met.

The ANFIS applies a hybrid learning strategy for advancing the ensuing parameters of the ANFIS which consolidate the inclination plunge method for improving the forerunner parameters and the Least Square Estimator (LSE). Then again, escape from the nearby optima utilising the HS strategy. In this paper, the HS is connected to streamline the initiate parameters of the ANFIS, in particular HYBRID HS-ANFIS. The learning system of the HYBRID HS-ANFIS is abridged as pursues:

Stage 1: Set the HS parameters including HMS, (HMCR) harmony memory consideration rate, and (PAR) pitch adjustment rate.

Stage 2: Set the HM by arbitrarily producing HMS arrangement vectors inside specific limits. Every one of these arrangement vectors speaks to the predecessor parameters of the ANFIS.

Stage 3: Evaluate the Rheumatoid Arthritis (RA) Diagnosis for every arrangement vector in the HM.

Stage 4: Search for the record of the Best Feature Selection in the HM.

Stage 5: Improvise a selected feature from the HM by using three principles, the congruity memory thought, the pitch alteration, and the arbitrary determination.

Stage 6: Update the HM.

Stage 7: Categorise the Selected Feature and compare it with the best features using ANFIS Training-diagnosis procedure of RA else, return to Step 4.

Training phase of HS-ANFIS classifier is pretty fast and the training samples from the HS-ANFIS classifier is stored in the database, which is needed during the testing phase of HS-ANFIS classifier. HS-ANFIS keeps all the training data which leads to a lack in simplification.

4. EXPERIMENTAL MODELLING OF THE RA DIAGNOSIS MODEL

4.1 Dataset Construction:

The reliable dataset construction is an essential part of developing a disease diagnose model using artificial intelligence techniques. Now, there is no existing dataset available for Rheumatoid Arthritis, so data collection is the first set of the proposed study. In this work, the clinical data is collected from the various outpatient units in Coimbatore, India. The dataset is labelled into RA and non RA with the help of clinical experts, about 1000 data samples are

collected among them 625 participants are female, and 375 are male. The classification is made based on the American College of Rheumatology (ACR) and European League against Rheumatism (EULAR) strategy, about 927 are RA patients and 73 are non-RA participants as shown in Table 1. The collected data consist of attributes such as the participant's Age, Anti-CCP (-/ml), Body Pain, Erythrocyte Sedimentation Rate (ESR) mm/hr, Family history, Location of pain, Morning Stiffness, Myalgia, Rest Pain, Rheumatoid Factor, Sex, Swelling, Symmetry Joint Inflammation, Symptom Duration, and Weight Loss is shown in Table 2. The dataset has 20 features, whereas the proposed strategy has optimally selected six features.

Table 1: The number of the training samples and test samples

Class	Training set	Testing set	Total
Normal	52	21	73
RA	748	179	927
Total	800	200	1000

Table 2: RA Dataset Features

Feature Number	Total number of features in dataset(20)	Selected Features
F1	Age	0
F2	Anemia	0
F3	Anti-CCP	1
F4	Body Pain	0
F5	C-Reactive Protein	1
F6	Depression	0
F7	ESR	0
F8	Environmental Issue	0
F9	Family history	0
F10	Location	0
F11	Morning stiffness	1
F12	Myalgia	0
F13	Numbness	0
F14	Rheumatoid Factor	0
F15	Rest pain	1

F16	Sex	0
F17	Swelling	0
F18	Symmetry Joint Inflammation	1
F19	Symptom duration	1
F20	Weight loss	0
0- Features not selected, 1 – Features selected		

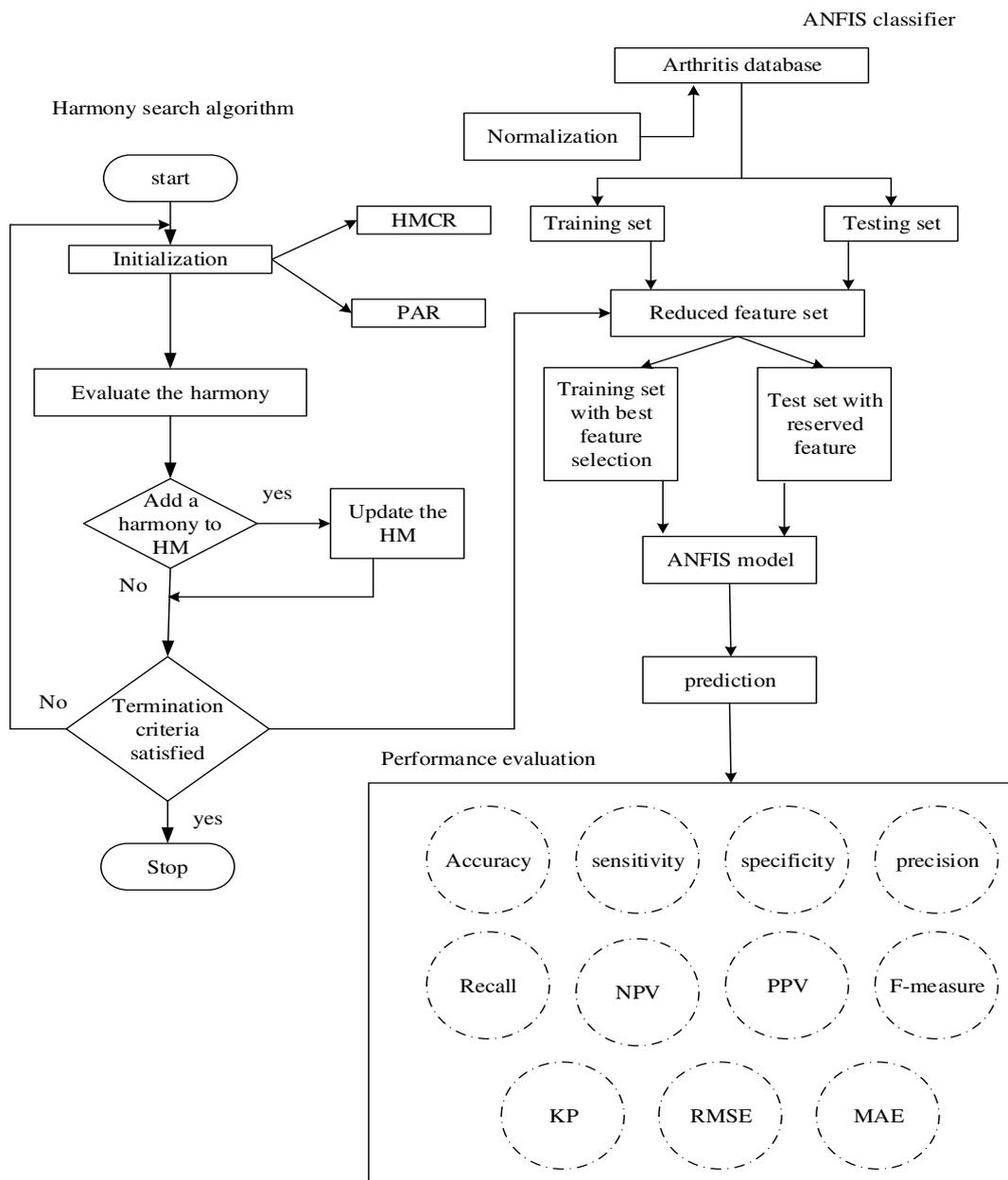


Figure 3 Architecture of Rheumatoid Arthritis diagnosis model

4.2 Normalisation

The purpose of Normalisation techniques is used to map the data to a diverse scale. In this work using Z score normalisation, also called as Zero mean normalisation. Here, the RA data is normalised based on the mean and standard deviation.

$$\Delta N = \frac{\text{mean}(\Delta Ra)}{\text{std}(\Delta Ra)} \quad (13)$$

4.3 Data Segregation

After consummating the Normalisation process, the dataset is divided into two portions using the percentage (%). Dataset splitting is the procedure of dividing the entire dataset into two portions. The first portion is training data and it contains 80% of ΔN is represented as $\Delta N_A = \{N_1, N_2, N_3, \dots, N_n\}$ and the second portion is testing data it contains 20% of ΔN is representation $\Delta D_B = \{D_1, D_2, D_3, \dots, D_n\}$.

4.4 Training and Testing phase

The Pre-processed data is feed into Harmony Search Algorithm, the algorithm tends to select the optimal features. The selected features should follow the constrain such that, the selected number of features should be minimum. Meanwhile, maximising the accuracy, the cost function is presented as follows:

$$\text{cost function} = \mu \text{Acc} + \mu' \frac{N - L}{N} \quad (14)$$

Where 'Acc' is the accuracy of the classification model, 'L' is the selected attribute length, N represents the number of features, μ and μ' indicates the weight of classification accuracy and feature selection quality, $\mu \in [0, 1]$ and $\mu' = 1 - \mu$.

In the testing phase, each new RA data is analysed and its principal features are located and compared with the principal features of trained RA data. If some matches are found, the data is classified by the HS-ANFIS according to the previously defined rules. Initially, the test query RA data is to be received from the user, and then feature extraction is done. Next, the proposed HS-ANFIS classifier approach is applied on the given query data to flag the data into 'Normal' or 'RA'. The parameters for proposed HS-ANFIS technique for RA disease classification is presented in Table 3.

Table 3 Parameters of RA diagnosis model

Parameters	Values
Number of iterations	100
Harmony Search Memory (HMS)	10
Numeral of New harmonics (nNew)	10
Pitch Adjustment Rate (PAR)	0.1
Fret width bandwidth FW	0.02*(Varmax-Varmin)
Fret width damp ratio	fw_damp=0.995

5. PERFORMANCE METRICS

The proposed RA Diagnosis model is simulated in the MATLAB version-7 environment using i5 quad core processor with 4GB RAM. The metric values are based on the confusion matrix framed between the true positive and negative predictions. TP signifies the quantity of true positives, TN the quantity of true negatives, FP the quantity of false positives, and FN the quantity of false negatives. Every individual stepping through the exam, either has or has not have the illness. The test result can be sure (ordering the individual as having the illness) or antagonistic (arranging the individual as not having the sickness). The test results for each subject might possibly coordinate the subject's genuine status. In that setting:

- True Positive: Sick individuals accurately recognised as wiped out
- False Positive: Healthy individuals erroneously distinguished as debilitated
- True Negative: Healthy individuals effectively recognised as sound
- False Negative: Sick individuals erroneously recognised as sound

The performance of the model is demonstrated by the performance metrics as follows:

Accuracy: Accuracy is the extent of the aggregate number of forecasts that were right. It is resolved utilising,

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \times 100\% \quad (15)$$

Sensitivity: Sensitivity is the extent of positive occasions that were effectively characterised, as determined utilising,

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \quad (16)$$

Specificity: Specificity is the extent of negative occasions that were effectively characterised, as determined utilising,

$$Specificity = \frac{TN}{TN + FP} \times 100\% \quad (17)$$

Precision: Precision is the extent of the anticipated positive occurrences that were right,

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (18)$$

Recall: The Precision is a proportion of result significance, while recall is a proportion of the number of genuinely important outcomes..

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (19)$$

PPV and NPV: The positive and negative prescient qualities (PPV and NPV separately) are the extents of positive and negative outcomes in insights and symptomatic tests that are genuine positive and genuine negative outcomes individually, the PPV and NPV are presented as follows:

$$PPV = \frac{TP}{TP + FP} \quad (20)$$

$$NPV = \frac{TN}{TN + FN} \quad (21)$$

The PPV and NPV portray the execution of an analytic test.

F-measure: A measure that consolidates precision and recall is the consonant mean of exactness and recall; the conventional F-measure figuring is given underneath,

$$F = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (22)$$

Root Mean Square Error (RMSE): RMSE is the standard deviation of the residuals (expectation mistakes). Residuals are a proportion of how a long way from the relapse line information focuses is;

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (p_i - o_i)^2}}{n} \quad (23)$$

Mean Absolute Error (MAE): MAE is a proportion of contrast between two nonstop factors. The Mean Absolute Error is given by,

$$MAE = \frac{\sum_{i=1}^n |e_i|}{n} \quad (24)$$

Kappa Statistic: Cohen's kappa coefficient (κ) is a measurement which measures between ratter understandings for subjective (specific) things.

$$K = \frac{[K_o - K_e]}{[1 - K_e]} \quad (25)$$

$$\text{Where } K_o = \frac{(TN + TP)}{n}$$

$$K_e = \frac{[(TN + FN) \times (TN + FP) + (FP + TP) \times (FN + TP)]}{n^2} \quad (26)$$

6. DISCUSSION ON OBTAINED RESULTS

The early prognosis of the RA disease using Harmonic Search –Adaptive neuro fuzzy model is trained and tested with real time datasets, and the performance metrics are evaluated to demonstrate the efficiency of the proposed strategy. For the given dataset, the harmony search algorithm is utilised to selected optimal features, about six features are flagged as significant ones to report positive RA. The selected features include Anti-CCP, C-Reactive Protein, Morning Stiffness, Rest pain, Symmetry Joint Inflammation and Symptom Duration.

In disease diagnosis, meeting the better classification accuracy is alone not enough to evaluate the model. It is necessary to analyse the ability of the model to forecast the negative and positive rates because false prediction is dangerous in medical diagnosis, especially in the disease like RA with uncertain symptoms. The performance metrics such as accuracy, sensitivity, specificity, precision, F-measure, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Kappa Statistic results are shown in Figure 4. The adaptive neuro fuzzy model fed with the selected features of Harmonic Search optimiser reported classification accuracy of 0.9453, the sensitivity of the model is 0.9734 and the specificity is 0.5384, recall is 0.9734, the reported F-measure is 0.9708, with reduced RMSE of 0.245 and Kappa statistics of 0.676.

To report the significance of proposed HS-ANFIS the model performance is compared with other techniques such as GWO-ID3, PSO-ID3, GWO-SVM, where the Grey Wolf Optimizer, Particle Swarm Optimization algorithm are employed for feature selection, and RA disease classification is performed by ID3 and SVM as reported in Table 5. The variation in performance of the presented algorithms are shown in Figure 5, depicting the effectiveness of HS-ANFIS for efficient RA disease classification.

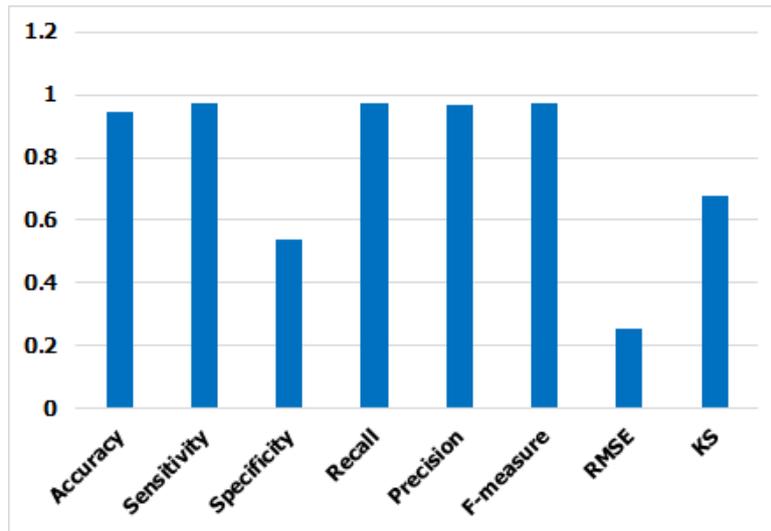
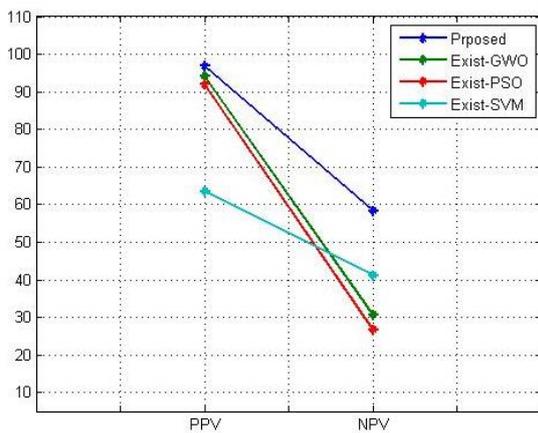
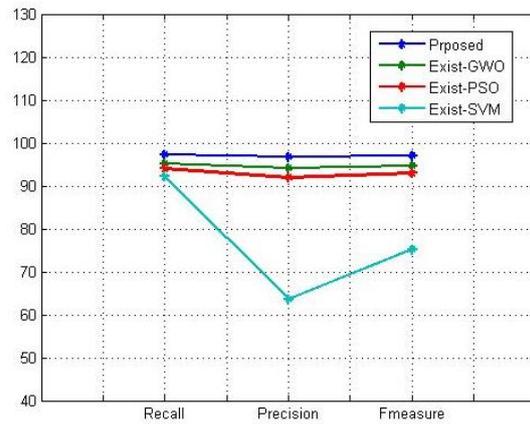


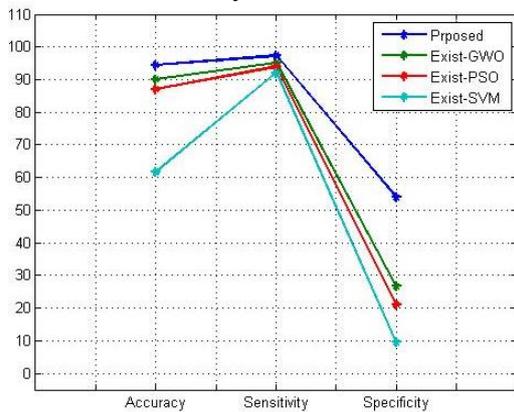
Figure 4 Classification performance results of the HS-ANFIS model



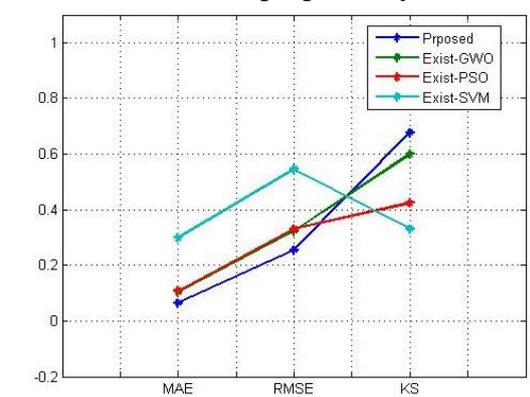
(a) Variations in PPV and NPV of proposed system



(b) Variations in recall, precision and Fmeasure of proposed system



(c) Variations in Accuracy, Sensitivity and Specificity of proposed system

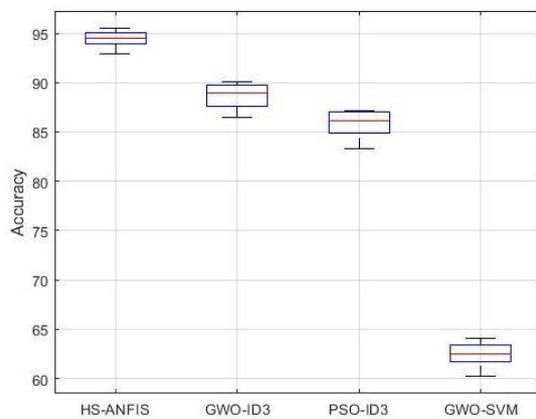


(d) Variations in MAE, RMSE and KS of proposed system

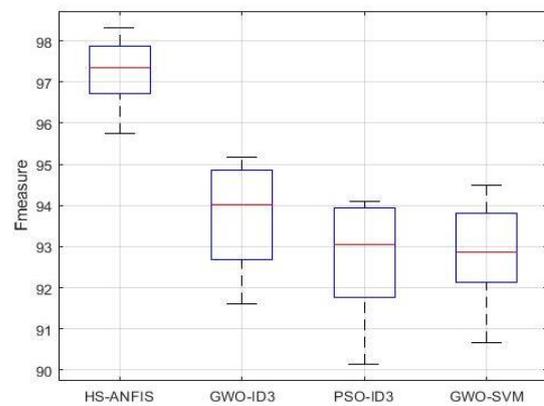
Figure 5 Comparison of performance metric results

Table 5 Performance of the proposed models

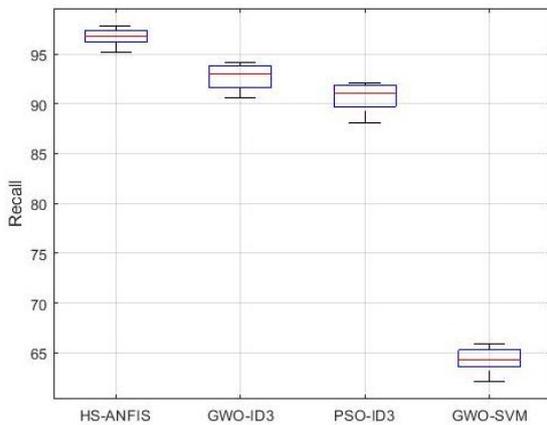
Model	Accuracy	Sensitivity	Specificity	Recall	Precision	F-measure	RMSE	KS
HS-ANFIS	94.527	97.340	53.846	97.340	96.825	97.082	0.254	0.676
GWO-ID3	90.049	95.161	26.667	95.161	94.149	94.652	0.323	0.599
PSO-ID3	87.064	93.956	21.053	93.956	91.935	92.935	0.331	0.424
GWO-SVM	61.691	92.126	9.459	92.126	63.587	75.241	0.546	0.332



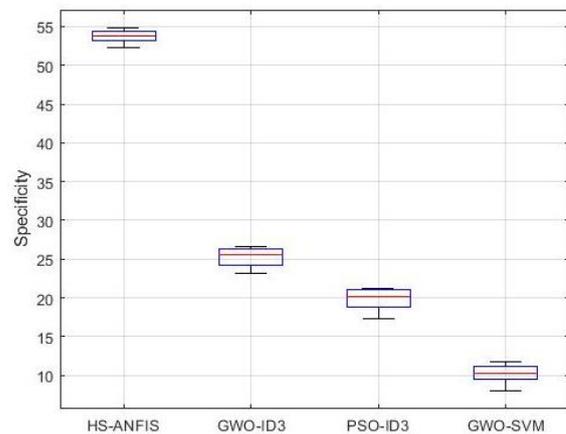
(a) Accuracy



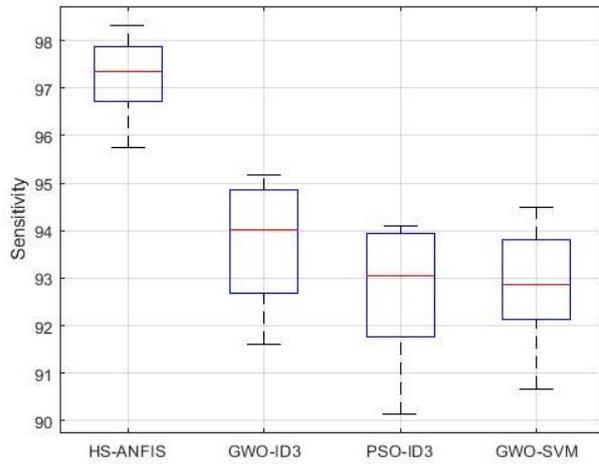
(b) F-measure



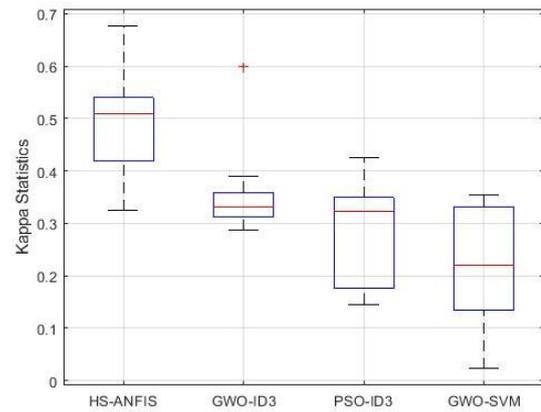
(c) Recall



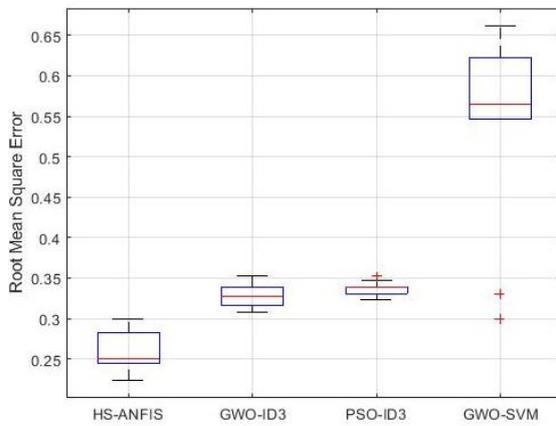
(d) Specificity



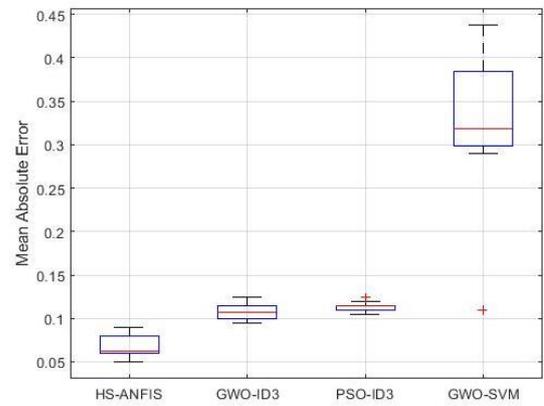
(e) Sensitivity



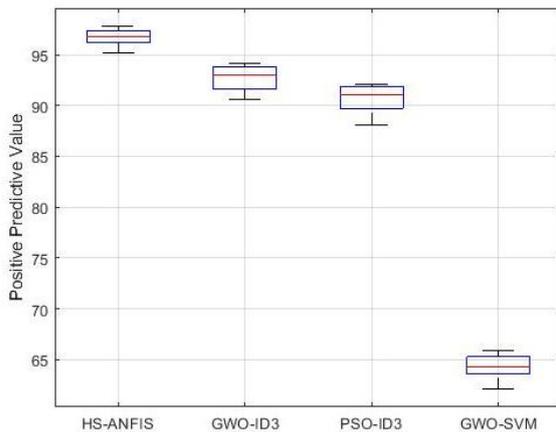
(d) Kappa Statistics



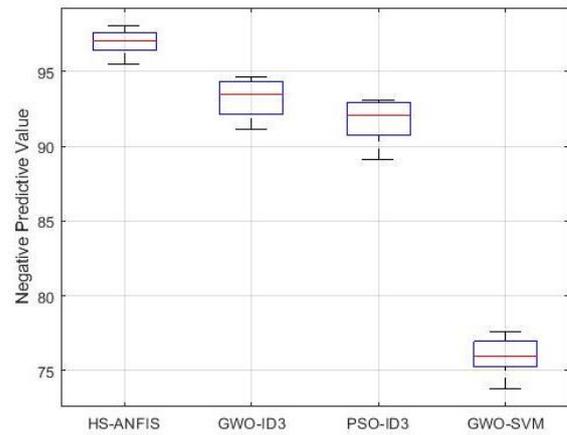
Root Mean Square Error



Mean Absolute Error



Positive Predictive Value



Negative Predictive Value

Figure 6 Box plots for 10 times of trials for each classification method on the **Rheumatoid Arthritis** dataset

The box plots are drawn to exhibit the general values of accuracy, sensitivity, specificity, precision, F-measure, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Kappa Statistic for 10, trial runs are shown in Figure 5. The accuracy of the HS-ANFIS is 4.47% higher than that of GWO-ID3, 7.47% than that of PSO-ID3, 32.84% than that of GWO-SVM. The sensitivity of the HS-ANFIS is 2.18% higher than that of GWO-ID3, 3.38% improved than PSO-ID3 and 5.2%. The specificity is 22.18% improved in HS-ANFIS to GWO-ID3, about 32.79% higher than that of PSO-ID3 and 44.39 % better than that of GWO-SVM. The Recall is improved to the value of 2.18%, 3.38% and 5.214 % than that of GWO-ID3, PSO-ID3 and GWO-SVM. The precision is 2.082% higher than that of GWO-ID3, 4.89% higher than that of PSO-ID3 and 33.34%. The F-measure is 2.43% improved than that of GWO-ID3, 4.14% improved than that of PSO-ID3, 21.84% higher than that of GWO-SVM strategy. Similarly, the RMSE value is significantly decreased than that of other models, with better kappa score.

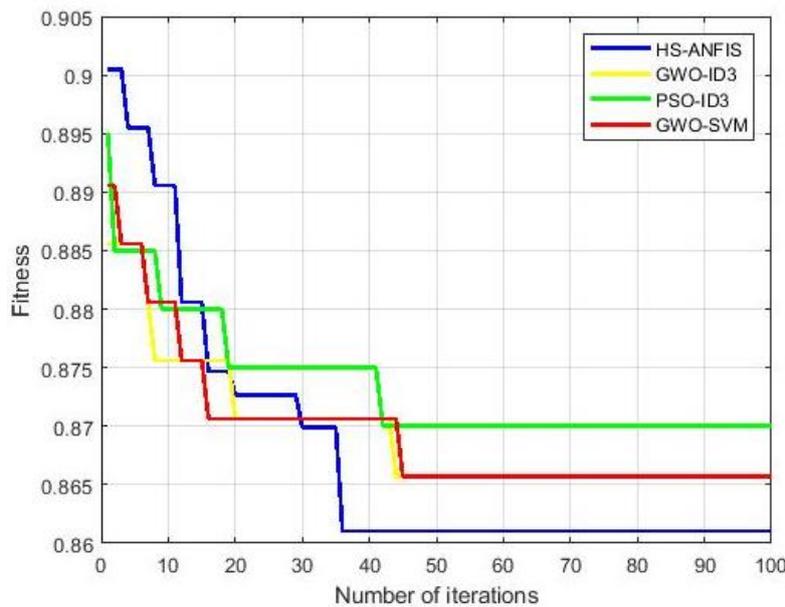


Figure 7 Convergence Graph of proposed RA diagnosis

The Convergence graph showing attainment of fitness value over the iterations as in Figure 7, the feature selection algorithms such as GWO-ID3, PSO-ID3 and GWO-SVM are compared and presented. It is depicted in the convergence plot that the HS-ANFIS converges faster and also the fitness value of HS-ANFIS is minimum representing the minimal number of features selected than that of GWO-ID3, PSO-ID3 and GWO-SVM.

The performance of the HS-ANFIS is compared with the existing works of literature, and the obtained matrix results for the given training and testing dataset is reported in Table 6 and 7. The training accuracy of the proposed HS-ANFIS is 17.1% better than that of C4.5,

8.8% higher than that of PSO-C4.5, 9.3% higher than that of GWO-C4.5 and 3.7 % improved to the HGWO-C4.5 for the same datasets employed [25] 16.9 % higher than that of CSBoost [27]. The sensitivity of the proposed HS-ANFIS is 20% better than that of C4.5, 11.8% higher than that of PSO-C4.5, 9.9% better than that of GWO-C4.5, 3.9% higher than HGWO- C4.5, 10.1% to REACT and 19% higher than that of CS-Boost. The specificity is 6.6% better than that of C4.5, 7.4% to PSO-C4.5, 5.6% better than that of GWO-C4.5, 5.7% higher than HGWO-C4.5 and 4.6% higher than the CS-Boost strategy. The computational time of the proposed HGWO-C4.5 is significantly reduced than that of other models employed in the study.

Table 6 Training Performance of proposed model with existing works of literature

Reference	Method	AUC	Sensitivity	Specificity	Computational time (min)
[25]	C4.5	0.731	0.741	0.512	2.94
	PSO-C4.5	0.813	0.823	0.504	3.22
	GWO-C4.5	0.808	0.842	0.522	3.24
	HGWO-C4.5	0.864	0.902	0.530	3.20
[26]	REACT	0.780	0.840	0.521	3.01
[32]	CS-Boost	0.732	0.751	0.532	3.82
Proposed HS-ANFIS		0.901	0.941	0.578	2.13

Table 7 Testing Performance of proposed model with existing works of literature

Reference	Method	AUC	Sensitivity	Specificity	Computational time (min)
[25]	C4.5	0.703	0.732	0.504	1.20
	PSO-C4.5	0.805	0.804	0.531	2.13
	GWO-C4.5	0.812	0.821	0.534	2.01
	HGWO-C4.5	0.801	0.913	0.530	2.28
[26]	REACT	0.743	0.823	0.504	2.61
[32]	CS-Boost	0.712	0.736	0.521	2.45
Proposed HS-ANFIS		0.945	0.973	0.5446	1.48

The training response of the Proposed HS-ANFIS is presented in Table 6, the testing accuracy is 24% better than that of C4.5, 14% higher than of PSO-C4.5%, 13.3% improved to GWO-C4.5, 14.4% better than that of HGWO-C4.5, 20.2% higher than that of REACT, 23.3% higher than the CS-Boost strategy employed for comparison. The sensitivity is 24.1% higher than that of C4.5, 16.9% better than that of PSO-C4.5, 15.2% improved to GWO-C4.5, 6% higher than that of HGWO-C4.5, 15% higher than that of REACT, 23.7% improved to CS-Boost. The sensitivity of Proposed HS-ANFIS is 4.06% better than that of C4.5, 1.36% higher than that of PSO-C4.5, 1.1% better than that of GWO-C4.5, 1.46% higher than that of HGWO-C4.5, 4.06% higher than that of REACT, 2.36% improved than that of CS Boost. The Testing time is considerably improved for the proposed strategy. Moreover, on comparing the training and testing performance the proposed HS-ANFIS has better generalisation ability than the other models in the existing works of literature.

5. CONCLUSION

Proposed research work focusses on designing an effective Rheumatoid Arthritis diagnosis model by employing hybrid HS-ANFIS strategy. The real-time data is collected and employed to evaluate the performance of the model. The feature selection is performed by Harmony search algorithm, and its' significance is illustrated by comparing the performance with the state of the art methods such as GWO and PSO employed for feature selection. The performance is evaluated using performance metrics such as Accuracy, Specificity, Sensitivity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Recall, Precision, F-measure, Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Kappa Statistics (KS). The proposed model has shown improved classification accuracy with reduced false predictions than existing literature works.

AUTHOR CONTRIBUTIONS

The authors contributed to each part of this paper equally. The authors read and approved the final manuscript.

CRedit Statement

Conceptualization: S Shanmugam, C Saravanabhavan, T Arunkumar; **Methodology:** S Shanmugam, C Saravanabhavan, T Arunkumar; **Formal analysis and investigation:** S Shanmugam; **Writing - original draft preparation:** S Shanmugam, T Arunkumar; **Writing - review and editing:** S Shanmugam, C Saravanabhavan, T Arunkumar; **Resources:** S Shanmugam, C Saravanabhavan, T Arunkumar; **Supervision:** T Arunkumar.

COMPLIANCE WITH ETHICAL STANDARDS

Funding: No funds, grants, or other support was received.

Conflict of Interest: The authors do not have any Conflict of Interest.

Ethical approval: There is no direct human participation in this study.

REFERENCES

- [1] Pallari, Jari HP, Kenneth W. Dalgarno, and James Woodburn. "Mass customization of foot orthoses for rheumatoid arthritis using selective laser sintering." *IEEE Transactions on Biomedical Engineering* 57, no. 7 (2010): 1750-1756.
- [2] Pani, Danilo, GianlucaBarabino, AlessiaDessì, IostoTradori, MatteoPiga, Alessandro Mathieu, and Luigi Raffo. "A device for local or remote monitoring of hand rehabilitation sessions for rheumatic patients." *IEEE journal of translational engineering in health and medicine* 2 (2014): 1-11.
- [3] A. Gibofsky, "Overview of epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis," *The American Journal of Managed Care*, vol. 18, no. 13, pp. S295–S302, 2012.
- [4] Senthelal, S., Li, J., Goyal, A., Bansal, P. and Thomas, M.A., 2020. Arthritis. *StatPearls [Internet]*.
- [5] Barbour KE, Helmick CG, Boring M, Brady TJ. Vital signs: prevalence of doctor-diagnosed arthritis and arthritis-attributable activity limitation—United States, 2013–2015. *MMWR Morb Mortal Wkly Rep* 2017;66:246–53. 10.15585/mmwr.mm6609e1
- [6] Venables, P. and Maini, R.N., 2014. Diagnosis and differential diagnosis of rheumatoid arthritis. *UpToDate2012*.
- [7] Miller, A., Green, M. and Robinson, D., 1983. Simple rule for calculating normal erythrocyte sedimentation rate. *British medical journal (Clinical research ed.)*, 286(6361), p.266.
- [8] Lucić, N., Mahić, A.Z., Lipa, I. and Seremet, M., 1989. Comparison of the immunoenzyme test (ELISA) with other methods in the detection of rheumatoid factor. *Reumatizam*, 36(1-6), pp.24-30.
- [9] Nishimura, K., Sugiyama, D., Kogata, Y., Tsuji, G., Nakazawa, T., Kawano, S., Saigo, K., Morinobu, A., Koshiha, M., Kuntz, K.M. and Kamae, I., 2007. Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis. *Annals of internal medicine*, 146(11), pp.797-808.

- [10] Tan, E.M., Feltkamp, T.E.W., Smolen, J.S., Butcher, B., Dawkins, R., Fritzler, M.J., Gordon, T., Hardin, J.A., Kalden, J.R., Lahita, R.G. and Maini, R.N., 1997. Range of antinuclear antibodies in “healthy” individuals. *Arthritis & Rheumatism: Official Journal of the American College of Rheumatology*, 40(9), pp.1601-1611.
- [11] Heidari, B., 2011. Rheumatoid Arthritis: Early diagnosis and treatment outcomes. *Caspian journal of internal medicine*, 2(1), p.161.
- [12] Hyldgaard, C., Hilberg, O., Pedersen, A.B., Ulrichsen, S.P., Løkke, A., Bendstrup, E., et al. (2017). A population-based cohort study of rheumatoid arthritis-associated interstitial lung disease: comorbidity and mortality. *Annals of the rheumatic diseases*, 76(10), 1700-1706.
- [13] Crowson, C.S., Rollefstad, S., Ikdahl, E., Kitas, G.D., Van Riel, P.L., Gabriel, S.E., Matteson, E.L., Kvien, T.K., Douglas, K., Sandoo, A. and Arts, E., 2018. Impact of risk factors associated with cardiovascular outcomes in patients with rheumatoid arthritis. *Annals of the rheumatic diseases*, 77(1), pp.48-54.
- [14] van der Woude, D. and van der Helm-van, A.H., 2018. Update on the epidemiology, risk factors, and disease outcomes of rheumatoid arthritis. *Best Practice & Research Clinical Rheumatology*, 32(2), pp.174-187.
- [15] Crowson, C.S., Rollefstad, S., Ikdahl, E., Kitas, G.D., Van Riel, P.L., Gabriel, S.E., Matteson, E.L., Kvien, T.K., Douglas, K., Sandoo, A. and Arts, E., 2018. Impact of risk factors associated with cardiovascular outcomes in patients with rheumatoid arthritis. *Annals of the rheumatic diseases*, 77(1), pp.48-54.
- [16] De Cock, D. and Hyrich, K., 2018. Malignancy and rheumatoid arthritis: Epidemiology, risk factors and management. *Best Practice & Research Clinical Rheumatology*, 32(6), pp.869-886.
- [17] Jagpal, A. and Navarro-Millán, I., 2018. Cardiovascular co-morbidity in patients with rheumatoid arthritis: a narrative review of risk factors, cardiovascular risk assessment and treatment. *BMC rheumatology*, 2(1), p.10.
- [18] England, B.R., Thiele, G.M., Anderson, D.R. and Mikuls, T.R., 2018. Increased cardiovascular risk in rheumatoid arthritis: mechanisms and implications. *Bmj*, 361.
- [19] Klein, A., Polliack, A. and Gafter-Gvili, A., 2018. Rheumatoid arthritis and lymphoma: Incidence, pathogenesis, biology, and outcome. *Hematological Oncology*, 36(5), pp.733-739.
- [20] Pundole, X. and Suarez-Almazor, M.E., 2020. Cancer and rheumatoid arthritis. *Rheumatic Disease Clinics*, 46(3), pp.445-462.
- [21] Inose, R., Hosomi, K., Takahashi, K., Yokoyama, S. and Takada, M., 2019. Risk of malignant lymphoma in patients with rheumatoid arthritis treated with biological disease-modifying antirheumatic drugs and methotrexate. *International journal of clinical pharmacology and therapeutics*, 57(2), p.63.
- [22] Dehghan, M., Ahmadi, A., Yousefghahari, B., Kiakojouri, K. and Gholinia, H., 2020. Effects of Rheumatoid Arthritis on the Larynx. *Iranian Journal of Otorhinolaryngology*, 32(3), pp.147-153.

- [23] Kroese, J.M., Volgenant, C.M., Crielaard, W., Loos, B., van Schaardenburg, D., Visscher, C.M. and Lobbezoo, F., 2021. Temporomandibular disorders in patients with early rheumatoid arthritis and at-risk individuals in the Dutch population: a cross-sectional study. *RMD open*, 7(1), p.e001485.
- [24] Sparks, J.A., Malspeis, S., Hahn, J., Wang, J., Roberts, A.L., Kubzansky, L.D. and Costenbader, K.H., 2021. Depression and subsequent risk for incident rheumatoid arthritis among women. *Arthritis Care & Research*, 73(1), pp.78-89.
- [25] Sundaramurthy, S. and Jayavel, P., 2020. A hybrid Grey Wolf Optimization and Particle Swarm Optimization with C4. 5 approach for prediction of Rheumatoid Arthritis. *Applied Soft Computing*, 94, p.106500.
- [26] Shanmugam, S. and Preethi, J., 2019. Improved feature selection and classification for rheumatoid arthritis disease using weighted decision tree approach (REACT). *The Journal of Supercomputing*, 75(8), pp.5507-5519.
- [27] Salmeron, J.L., Rahimi, S.A., Navali, A.M. and Sadeghpour, A., 2017. Medical diagnosis of Rheumatoid Arthritis using data driven PSO-FCM with scarce datasets. *Neurocomputing*, 232, pp.104-112.
- [28] Sharon, H., Elamvazuthi, I., Lu, C.K., Parasuraman, S. and Natarajan, E., 2019, October. Classification of Rheumatoid Arthritis using Machine Learning Algorithms. In 2019 IEEE Student Conference on Research and Development (SCOReD) (pp. 245-250). IEEE.
- [29] Ho, S., Elamvazuthi, I. and Lu, C.K., 2018, December. Classification of Rheumatoid Arthritis using Machine Learning Algorithms. In 2018 IEEE 4th International Symposium in Robotics and Manufacturing Automation (ROMA) (pp. 1-6). IEEE.
- [30] Hemalatha, R.J., Vijaybaskar, V. and Thamizhvani, T.R., 2019. Automatic localization of anatomical regions in medical ultrasound images of rheumatoid arthritis using deep learning. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 233(6), pp.657-667.
- [31] Sharon, H., Elamvazuthi, I., Lu, C.K., Parasuraman, S. and Natarajan, E., 2020. Development of Rheumatoid Arthritis Classification from Electronic Image Sensor Using Ensemble Method. *Sensors*, 20(1), p.167.
- [32] Shiezadeh, Z., Sajedi, H. and Aflakie, E., 2015. Diagnosis of rheumatoid arthritis using an ensemble learning approach. *Comput. Sci. Inf. Technol.(CS & IT)*, 5(15), pp.139-148.
- [33] Garcia-Zapirain, B., Garcia-Chimeno, Y., & Rogers, H. (2015). Machine learning techniques for automatic classification of patients with fibromyalgia and arthritis. *International Journal of Computer Trends and Technology (IJCTT)*, 25(3), 149-152.
- [34] Ramesh, K. B., Prasad, S. S., Mallikarjunaswamy, B. P., & Puttaiah, E. T. (2014). Prediction and Detection of Rheumatoid Arthritis SNPs Using Neural Networks. *Journal of Pharmacy and Biological Sciences (IOSR-JPBS)*, 9(1), 06-09.

- [35] Mallikarjunaswamy, B.P., Ramesh K.B., And E.T.(2011). Puttaiah. Dna Sequence Analysis Of Rheumatoid Arthritis. *International Journal Of Biomedical Signal Processing*, 2(2), 103-107.
- [36] Das, H., Naik, B. and Behera, H.S., 2020. Medical disease analysis using neuro-fuzzy with feature extraction model for classification. *Informatics in Medicine Unlocked*, 18, p.100288.
- [37] Balasubramanian, K. and Ananthamoorthy, N.P., 2020. Improved adaptive neuro-fuzzy inference system based on modified glowworm swarm and differential evolution optimization algorithm for medical diagnosis. *Neural Computing and Applications*, pp.1-12.
- [38] Bandyopadhyay, S. and Dutta, S., 2020. Adaptive Neuro Fuzzy Inference System Used to classify the Five Major Types of Brain Tumors.
- [39] Das, H., Naik, B. and Behera, H.S., 2020. Disease Classification Using Linguistic Neuro-Fuzzy Model. In *Progress in Computing, Analytics and Networking* (pp. 45-53). Springer, Singapore.
- [40] Nikita, B.K., Sadawarti, H. and Singla, J., 2020. A Neuro-Fuzzy Based Intelligent System For Diagnosis Of Renal Cancer. *International Journal of Scientific and Technology Research*, 9(1).
- [41] Ashokkumar, S.R. and Mohan Babu, G., 2020. Extreme learning adaptive neuro-fuzzy inference system model for classifying the epilepsy using Q-Tuned wavelet transform. *Journal of Intelligent & Fuzzy Systems*, (Preprint), pp.1-16.
- [42] Preetha, R. and Jinny, S.V., 2020. Early diagnose breast cancer with PCA-LDA based FER and neuro-fuzzy classification system. *Journal of Ambient Intelligence and Humanized Computing*, pp.1-10.
- [43] Kour, H., Manhas, J. and Sharma, V., 2020. Usage and implementation of neuro-fuzzy systems for classification and prediction in the diagnosis of different types of medical disorders: a decade review. *Artificial Intelligence Review*, pp.1-56.
- [44] El-Hasnony, I.M., Barakat, S.I. and Mostafa, R.R., 2020. Optimized ANFIS model using hybrid metaheuristic algorithms for Parkinson's disease prediction in IoT environment. *IEEE Access*, 8, pp.119252-119270.
- [45] Saha, S., Ghosh, M., Ghosh, S., Sen, S., Singh, P.K., Geem, Z.W. and Sarkar, R., 2020. Feature Selection for Facial Emotion Recognition Using Cosine Similarity-Based Harmony Search Algorithm. *Applied Sciences*, 10(8), p.2816.
- [46] Rahajoe, A.D., Zainal, R.F., Mulyo, B.M., Plangkang, B. and Tias, R.F., 2020, February. Feature Selection Based on Modified Harmony Search Algorithm. In *2020 International Conference on Smart Technology and Applications (ICoSTA)* (pp. 1-7). IEEE.
- [47] Koti, P., Dhavachelvan, P., Kalaipriyan, T., Arjunan, S., Uthayakumar, J. and Sujatha, P., 2020. An efficient healthcare framework for kidney disease using hybrid harmony search algorithm. *Electronic Government, an International Journal*, 16(1-2), pp.56-68.

- [48] Gholami, J., Ghany, K.K.A. and Zawbaa, H.M., 2020. A novel global harmony search algorithm for solving numerical optimizations. *Soft Computing*, pp.1-13.
- [49] Ala'a, A., Alsewari, A.A., Alamri, H.S. and Zamli, K.Z., 2019. Comprehensive review of the development of the harmony search algorithm and its applications. *IEEE Access*, 7, pp.14233-14245.
- [50] Yang, Xin-She. "Harmony search as a metaheuristic algorithm." *Music-inspired harmony search algorithm*. Springer, Berlin, Heidelberg, (2009). 1-14.

Figures

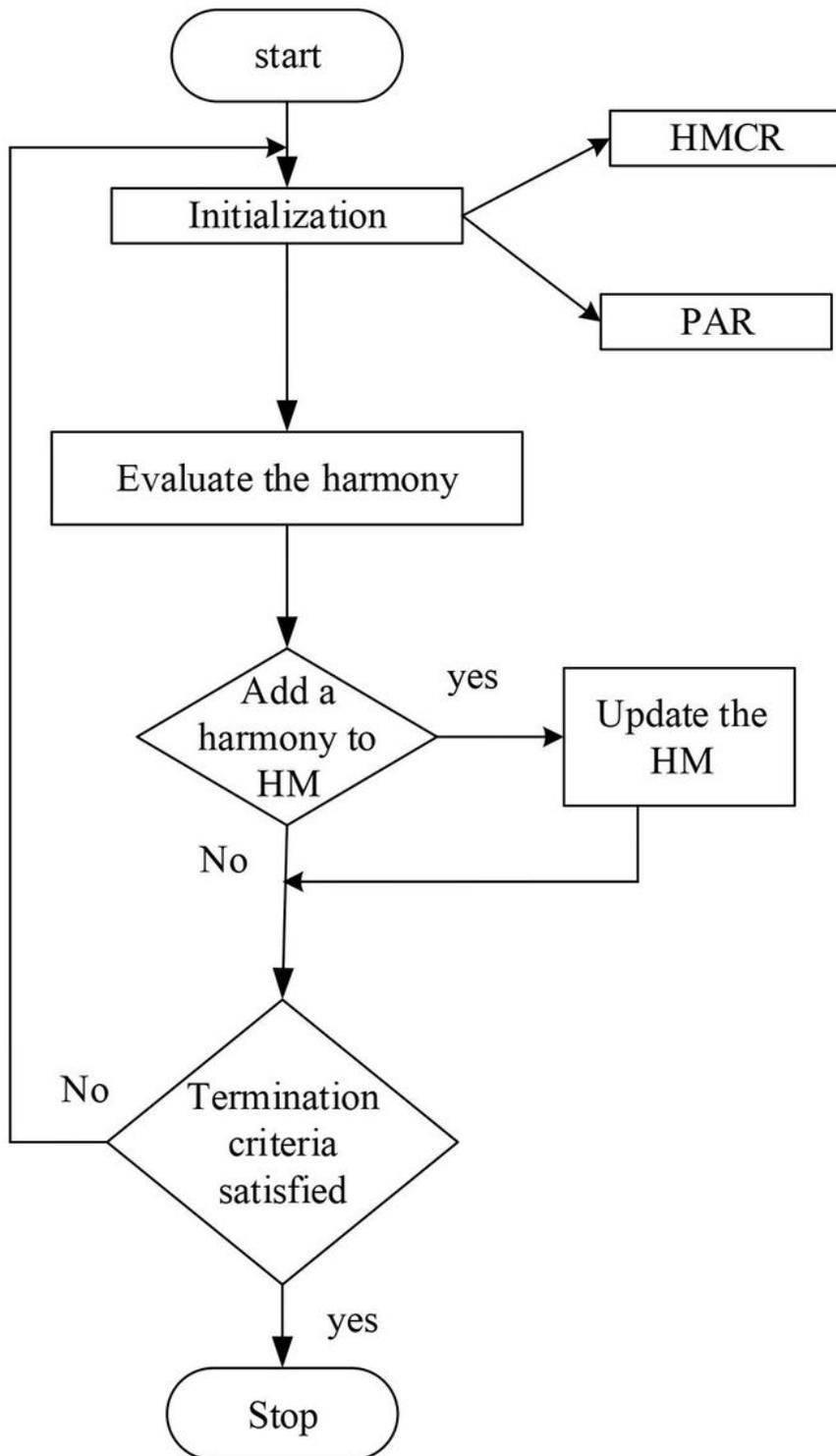


Figure 1

Harmony search flow diagram

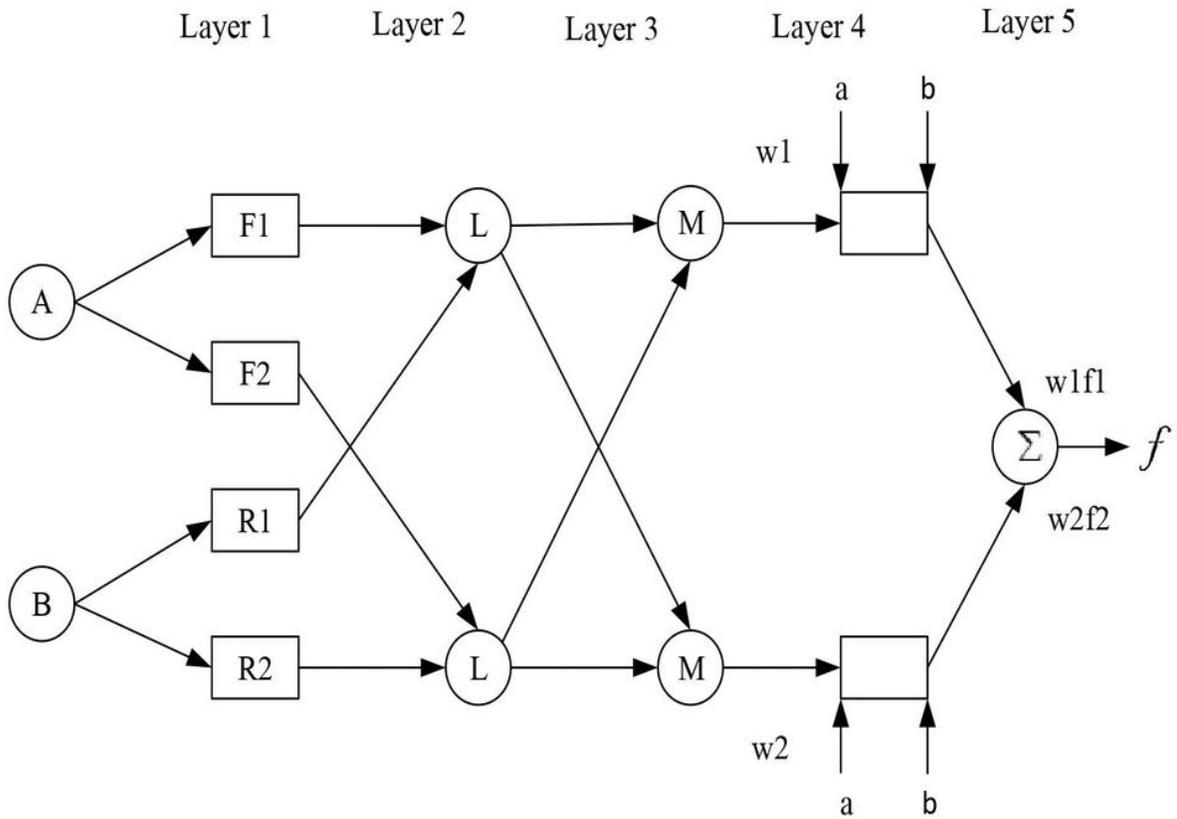


Figure 2

Adaptive Neuro Fuzzy Inference System Structure

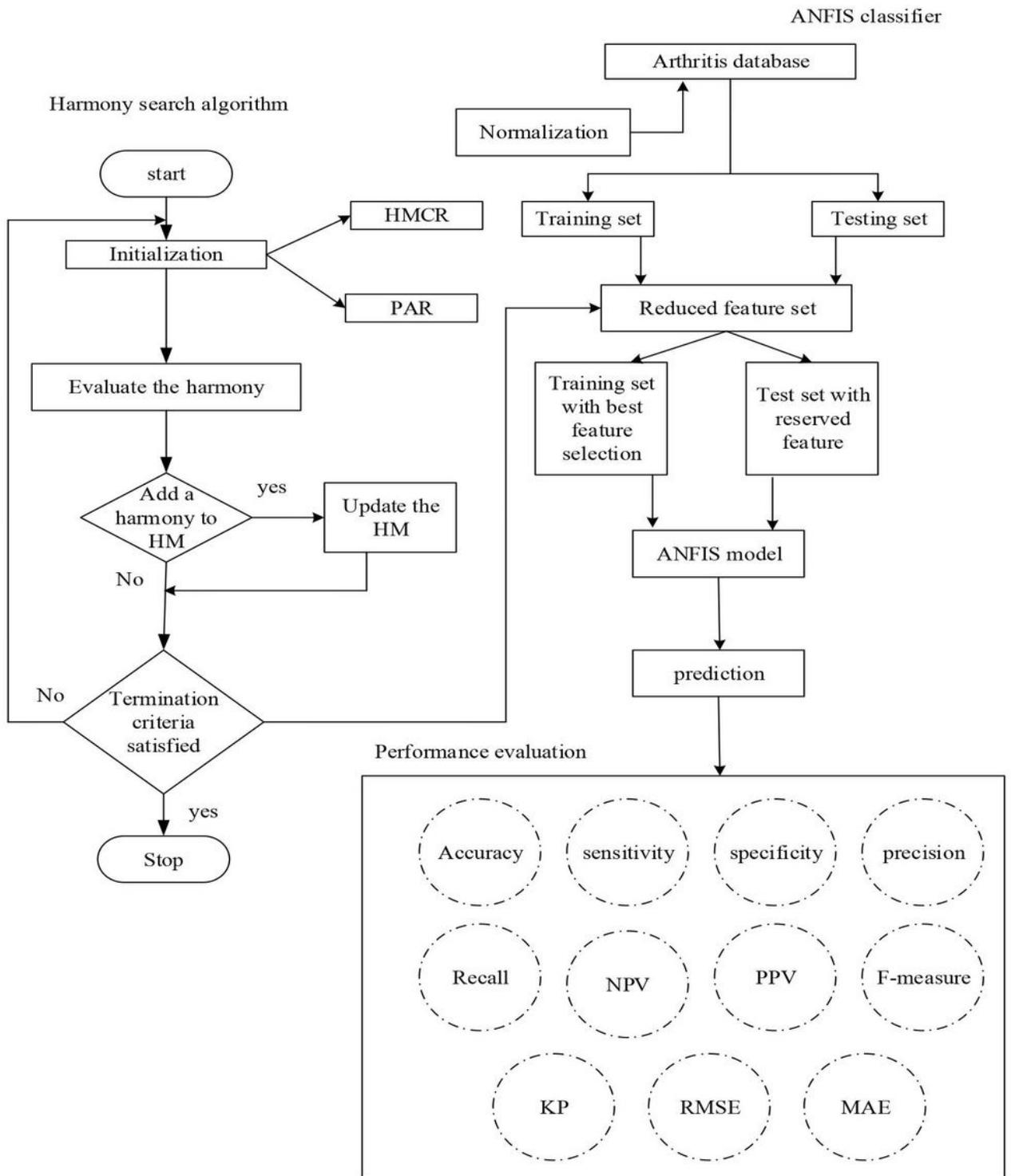


Figure 3

Architecture of Rheumatoid Arthritis diagnosis model

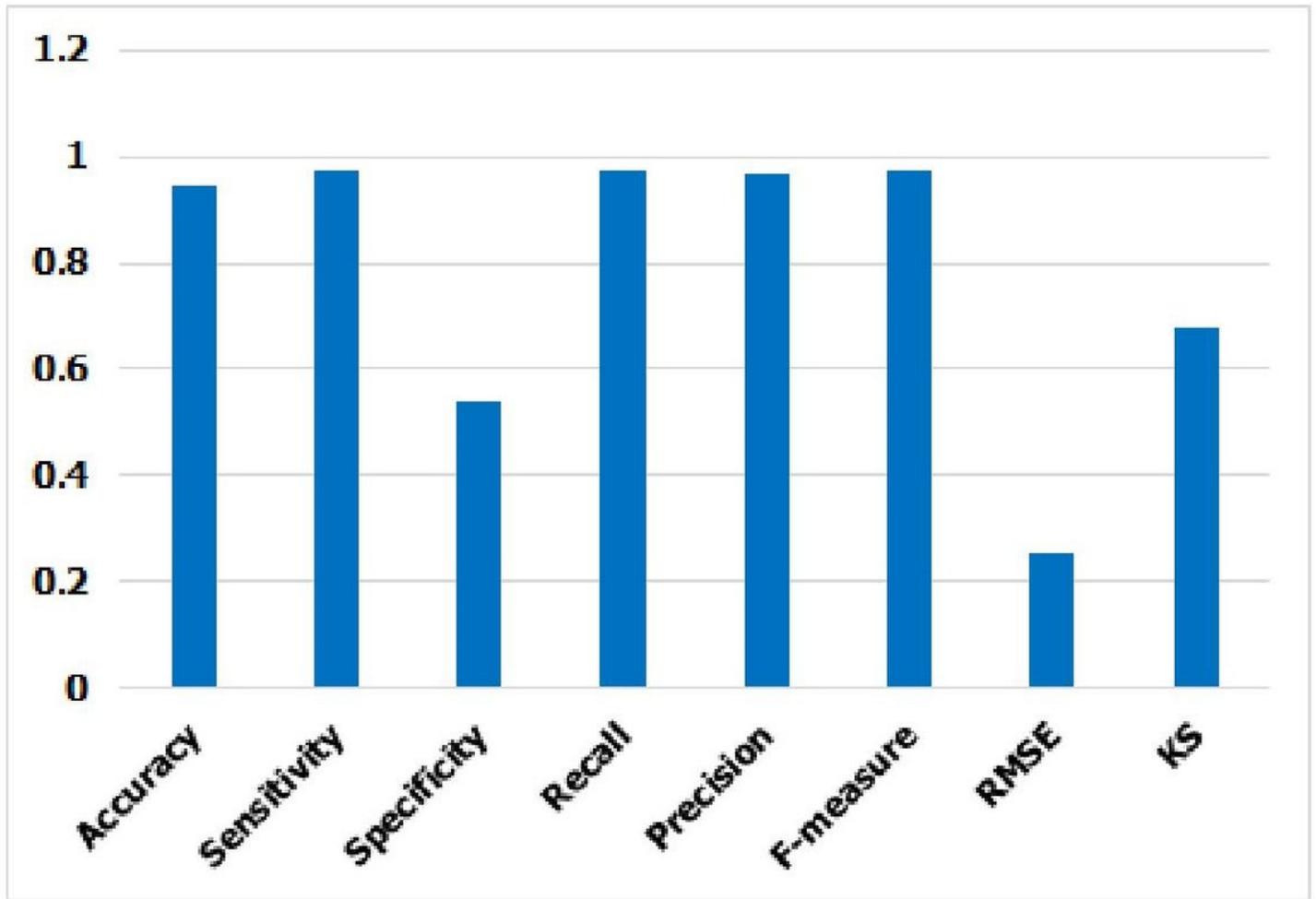
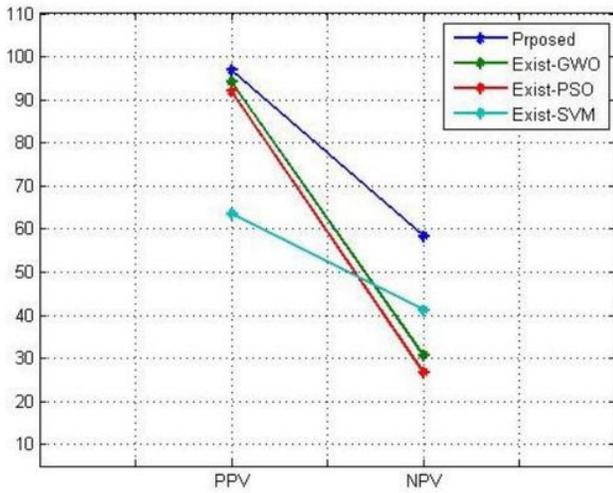
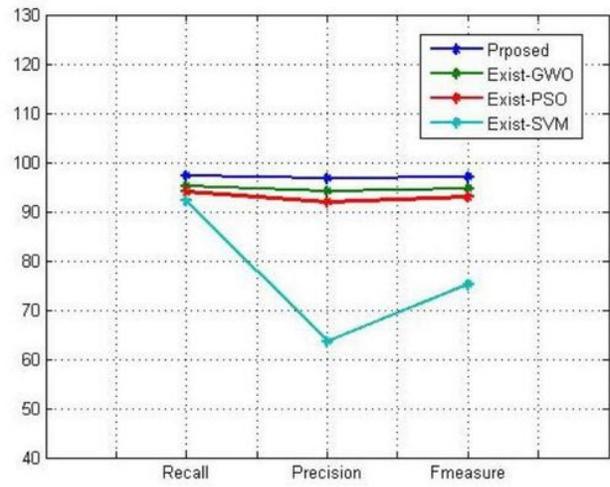


Figure 4

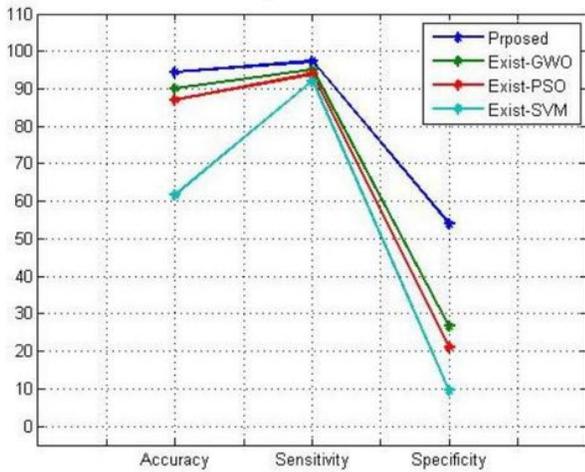
Classification performance results of the HS-ANFIS model



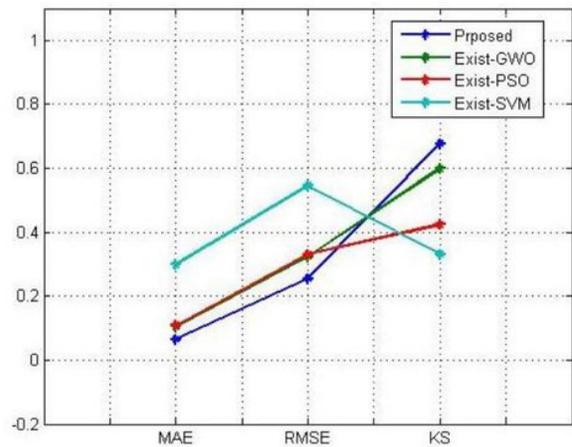
(a) Variations in PPV and NPV of proposed system



(b) Variations in recall, precision and Fmeasure of proposed system



(c) Variations in Accuracy, Sensitivity and Specificity of proposed system



(d) Variations in MAE, RMSE and KS of proposed system

Figure 5

Comparison of performance metric results

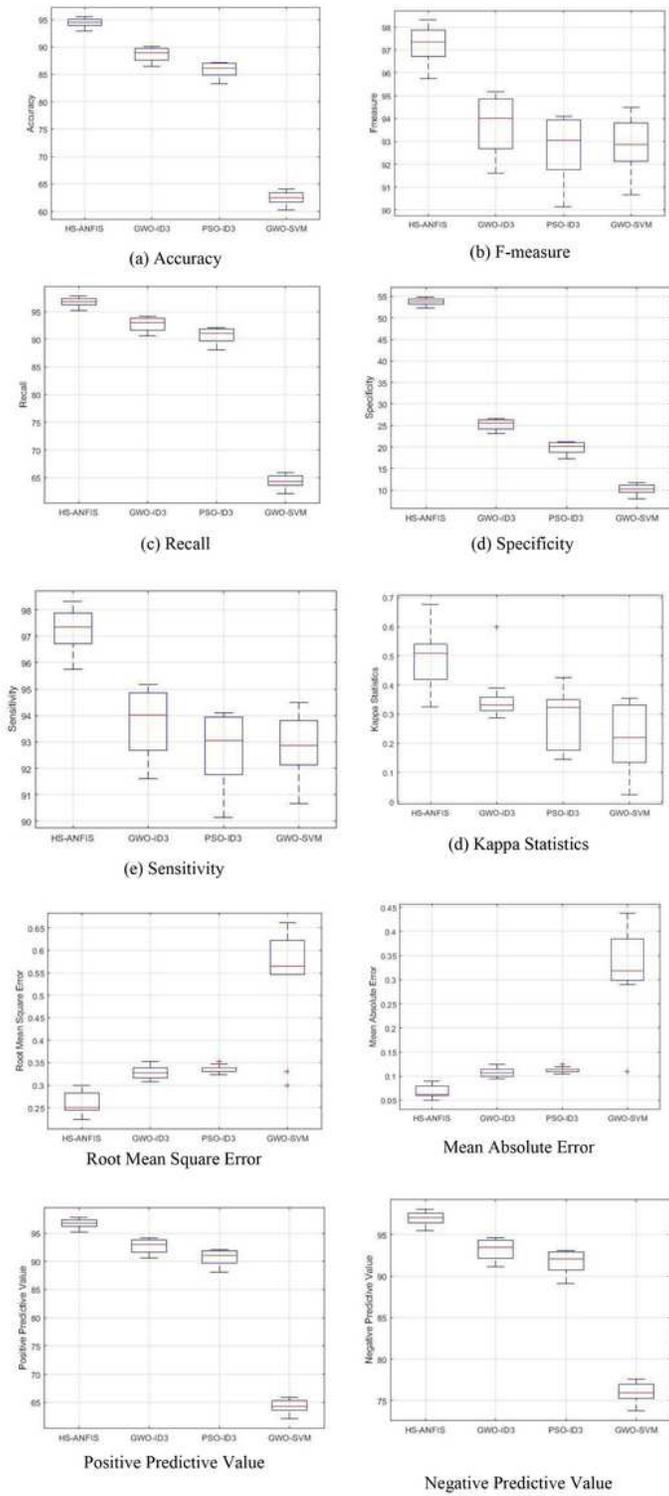


Figure 6

Box plots for 10 times of trials for each classification method on the Rheumatoid Arthritis dataset

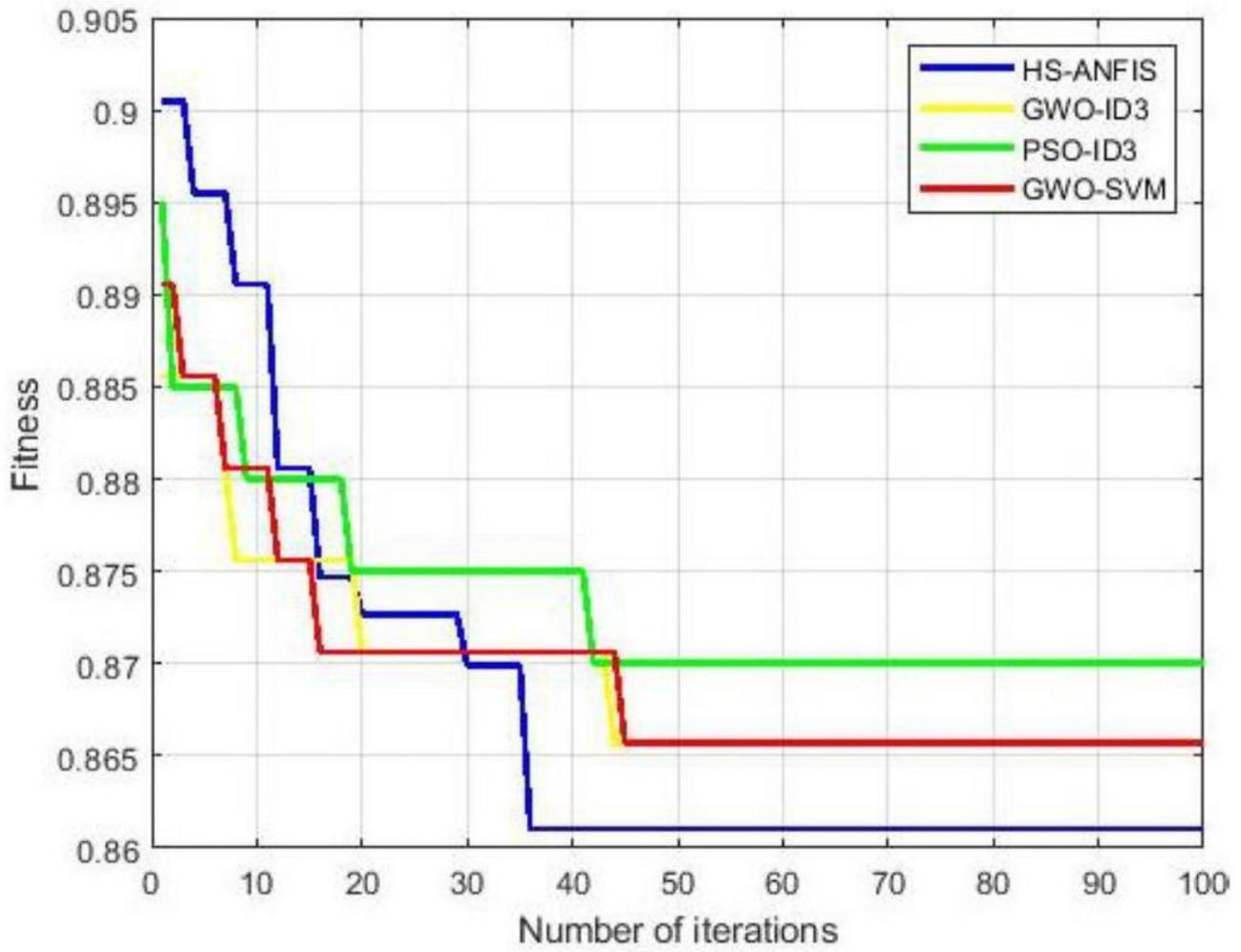


Figure 7

Convergence Graph of proposed RA diagnosis