Applying Artificial Intelligence Techniques to Improve Clinical Diagnosis of Alzheimer's disease

Ahmed Abdullah Farid[1, *], Gamal Ibrahim Selim[1,2], and Hatem Awad A. Khater[3]

[1]Department of Computer Engineering, Faculty of Engineering & Technology, Arab Academy for Science Technology and Maritime Transport(AASTMT), Cairo, Egypt

[2] Professor Dr. Gamal Selim, (email: dgamal55@yahoo.com or dgamalselim@aast.edu)

[3] Associate Professor Dr. Hatem Awad Aly Khater, Faculty of Engineering, Horus University in Egypt (email: hkhater@horus.edu.eg)

*Corresponding author, (email: ahmed.abdullah.farid@gmail.com or portal.ahmed@student.aast.edu)

Abstract.

Alzheimer's disease (AD) is a significant regular type of dementia that causes damage in brain cells. Early detection of AD acting as an essential role in global health care due to misdiagnosis and sharing many clinical sets with other types of dementia, and costly monitoring the progression of the disease over time by magnetic reasoning imaging (MRI) with consideration of human error in manual reading. Our proposed model in the first stage, apply the medical dataset to a composite hybrid feature selection (CHFS), to extract new features for select the best features to improve the performance of the classification process due to eliminating obscures features. In the second stage, we applied a dataset to a stacked hybrid classification system to combine Jrip and random forest classifiers with six model evaluations as meta-classifier individually to improve the prediction of clinical diagnosis. All experiments conducted on a laptop with an Intel Core i7- 8750H CPU at 2.2 GHz and 16 G of ram running on windows 10 (64 bits). The dataset evaluated using an explorer set of weka data mining software for the analysis purpose. The experimental show that the proposed model of (CHFS) feature extraction performs better than principal component analysis (PCA), and lead to effectively reduced the false-negative rate with a relatively high overall accuracy with support vector machine (SVM) as meta-classifier of 96.50% compared to 68.83% which is considerably better than the previous state-of-the-art result. The receiver operating characteristic (ROC) curve was equal to 95.5%. Also, the experiment on MRI images Kaggle dataset of CNN classification process with 80.21% accuracy result. The results of the proposed model show an accurate classify Alzheimer's clinical samples against MRI neuroimaging for diagnoses AD at a low cost.

Keywords: Data Mining, Alzheimer's Dementia, Composite Hybrid Feature Selection, Machine learning, Stack Hybrid Classification, AI Techniques, Classification, AD Diagnose, Clinical AD Dataset.

1. Introduction

Data mining skills involved in biomedical sciences and investigate for providing prediction for help to identify the disease and classify it correctly (Kalló, Gergő & Miklós Emri et al., 2016)(K. Tejeswinee & S.G. Jacob.,2017). (J. Escuderoaccess et al., 2013) (Chi, C. L. & Oh, W. & Borson & S.,2015). AD is a form of dementia that shows for 60-80% of mental disorders (David P Salmon & Mark W. Bondi,2010). AD is the sixth leading cause of death in the united states, according to the national center for health statistics 2019 (CDC)(NCHS,2019). Clinical AD research can create a new challenge for the possibility of effective treatment(j.escudero,2013)(Kloeppel,2008). Alzheimer's Disease is a grave personal, medical, and social issue. Recent research suggests that early and accurate detection is the secret to dealing successfully with it. According to Boise et al., though, diagnosis is incorrect 50 percent of the time, even at the early stages of the disease. (r.chaves&jramiz,2012).Around the world, about 44 million individuals have Alzheimer's or a related type of dementia. Expectations from Ireland show a comparable development design. The Irish National dementia Procedure, Distributed by the Branch of Wellbeing in December 2014, contained evaluations for the rate of Promotion for the years 2011 – 2046 in the Republic of Ireland. The assessments are that the number of sufferers altogether for all age gatherings increments from 47, 829 in 2011 to an aggregate (all age gatherings) of 152,157 of every 2046. In rate terms, this is more prominent than the anticipated development in numbers for the US (bahagyn&shree,2014). We propose in the first stage, a novel composite hybrid feature selection approach based on the optimization of the Genetic Algorithm (CHFS-OGA) to improve the prediction of Alzheimer's disease. In the second stage, we applied the output dataset features from step one to a stacked hybrid classification architecture model to improve the classification accuracy. A proactive evaluation approach that forecasts the future potentials of a novel model of Alzheimer's early diagnosis and hence improves the Feature ranked and classification accuracy, F-measure, true positive prediction of (previous points). The article planned as follows. The next section discusses the literature review of other authors who have used data mining and its relative of machine learning algorithm to analyze and diagnose Alzheimer's disease and various diseases. Section 3 describes the proposed technique used for feature extraction. Section 4 describes the method used for the hybrid classification process, whereas section 5 describes the experiments and discuss the results. Finally, section 6 presents the paper summary and conclusions.

2. Literature Review

Many researchers have used data mining for the diagnosis of various diseases. Some of them are Jyothi that Sony has used classifiers, namely naïve Bayes, k-nearest neighbor, decision tree, to predict heart disease (joyti,2011). Williams ET. Al. (Williams,2013) record clinical dementia rating (CDR) by support vector machine (SVM), decision tree, and neural network and naive Bayes replaced missing value with average one to achieve best accuracy and correlation(chi&borson,2015). Voxel-based morphometry applied to MRI images from an oasis medical dataset(chyzhyk,2010).

S. R. Bhagya Shree et al. compared many classifiers such as naïve Bayes, decision tree algorithm J48, random forest, JRip in the detection of Alzheimer‘s disease(S.rBhayga,2014), the results indicated naïve Bayes, Jrip and random forest perform better, the problem with this paper was the data set was having records of 250 subjects and the data not preprocessed. Tina R. Patil et al., in their paper ‘performance analysis of ayes and J48 classification algorithm for data classification,’ has discussed naive Bayes classification(tina&patil,2013).

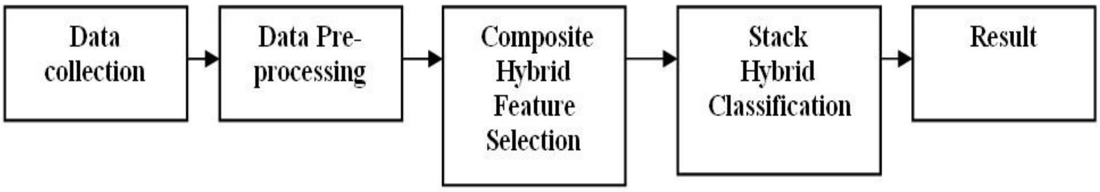
Jehad Ali et al., in their paper, has discussed Random forest and J48 for the classification of data sets(jihad,2012), The prediction of Alzheimer's disease using SVM on the MRI picture implemented by Vemuri, P., Gunter(vemuri&gunter,2008). Dementia can analyze by using various algorithms for enhancement of the accuracy of classification. Analyze MMSE-KC data into a naturalist and unusual CERAD-K used for the classification of mild cognitive impairment and dementia. The results compared to traditional classifiers, i.e., bagging, Bayes network, naive Bayes, logistic regression, SVM, random forest, and MLP(So&hooshyer,2008).

Ramirez. (Ramiz,2013) Carried out a study for finding the ROIs and the most discriminated image parameter for the reduction of the input space dimensionality and enhancement the precision. The data analyzed by using the random forest, Jrip, and naïve Bayes by Sheshadri (she.shadri,2015) and utilized explorer, flow, and API knowledge to evaluate it, an embedded classification model designed to predict AD disease in the future.

3. Proposed Work

In the proposed work, the medical dataset collected from the oasis brain organization (online access Oasis Medical Dataset,2019) . The medical dataset contains several clinical tests of non-demented and demented older adults. The proposed layout in Fig 1.

Fig. 1 The proposed layout



The following steps explain the mechanism of the proposed work:

3.1. Data Collection

The author's collected the dataset form Oasis 3 -a project which contains 1098 subject of longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer's Disease. Our study focuses on clinical data that contains 426 subjects with 1229 records of potential patients, and an oasis medical project is the latest release in the open-access series of AD datasets that aimed at making neuron datasets freely accessible to the scientific society (online access Oasis Medical Dataset,2019) .

3.2. Data Pre-processing

In the real world, data collected tend to be not wholly complete, noisy and conflicting, detection missing of data, data irregularity, prevent the errors and decrease the data to be analyzed lead to massive payouts for decision making(s.r.bahgya,2014). As a primary data has to collect in such a way that the above problems not occur, the missing entries in the collected dataset filled up by using the average values and the author's used unsupervised attribute replace missing values filters in weka data mining tool to solve this, the data is often present in the form of a spreadsheet. However, weka native data storage format is ARFF and transformed from a spreadsheet to CSV format. After that, the CSV file converted to the ARFF file. Thus the data has to be transformed from spreadsheet format to ARFF format (H. Witten & Eibe Frank ,2008) .

doi:10.20944/preprints202003.0297.v1

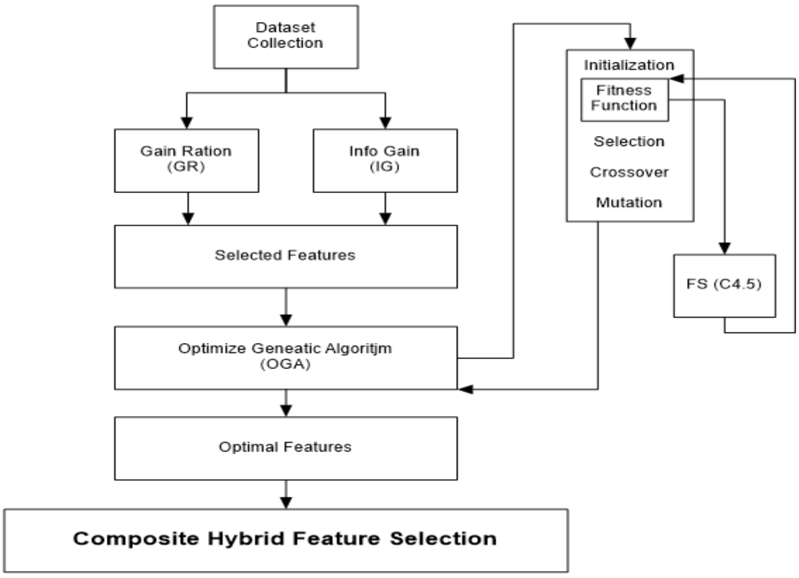
3.3. Proposed Composite Hybrid Feature Selection Model (CHFS)

Feature selection is the approach of taking a subset of relevant features for use in model construction(Chen & ...& F. Li,2010). It combines the advantages of three feature selection approaches (Filter (IG,GR)–Wrapper(improved (Genetic Algorithm)) with Embedded(C4.5)).

3.3.1 Composed Hybrid feature selection architecture

The author combine of three feature extraction technique considered for the optimal selection feature set, and this method is information gain (IG) - gain ratio (GR) and Optimized Genetic Algorithm(Huang & C.,2012) (Yanan Mao & Dingyuan Fan,2016).as shown by fig 2

Fig. 2 The proposed composite hybrid feature selection architecture



In Figure 2, the feature extraction method using gain ratio- information gain to rank the attributes. The superior ranked features taken as the optimized subset, the threshold established on the size of the resumption feature subset and amount of attributes in the dataset that appear in the range (7~10), and using classifier subset with C4.5 decision tree method in early-stage to obtain the optimal subset and vote for best feature selection of attributes in an automated process and produced an optimized feature subset with genetic search elevation planning to recognize the features that give the most predictable for the target class. (Vemuri & et al.Jack & C. R.,2008) (Devi &Cuckoo ,2013).

3.3.2 Information Gain feature selection

The calculation of the information gain for only one attribute according to the algorithm below (Aouatif Amine & ...& Rziza Driss ,2011) : This gain measure gives the effect of the features, and the following algorithm selects features that are larger than the threshold.

Let S be a part of a training set sample with related labels. The m class in training set has si pattern of class I and s have overall patterns in training set predictable information looked-for to grade using (1)(r.chaves&jami,2012)

$$I(s1, s2, \ldots sm) = -\sum_{i=1}^{m} \left(\frac{si}{s}\right) \log 2 \frac{si}{s} \quad (1)$$

Feature F with (f1, f2, and f3 to fv) can separate the medical dataset into v subsets {S1, S2… Sv} where Sj and have value fj for the feature F and Sj include sij samples of class i

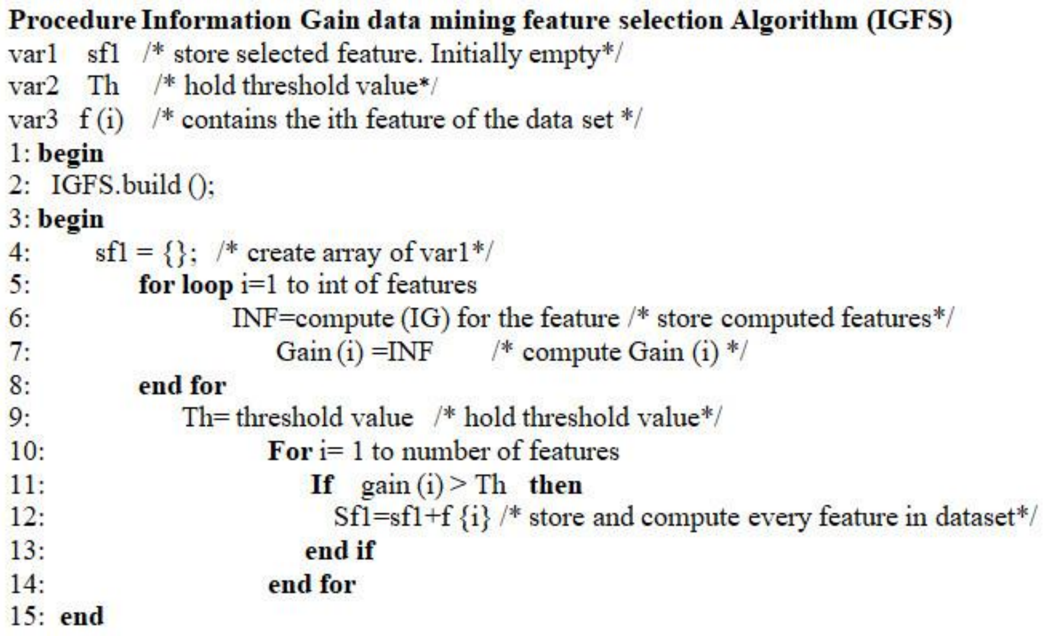
The entropy of F by"(2)"

$$E(f) = -\sum_{i=1}^{m} \frac{sij + s2j + \cdots + smj}{s} I(s1j, s2j, \ldots smj) \qquad (2)$$

And calculate info. Gain by (3)

$$GAIN(F) = I(S1, S2, \ldots . S3) - E(F) \qquad (3)$$

This gain measure gives the effect of the features, and the following algorithm selects features that are larger than the threshold[18,38]and shown in fig 3.



Procedure Information Gain data mining feature selection Algorithm (IGFS)

```
var1   sf1   /* store selected feature. Initially empty*/
var2   Th    /* hold threshold value*/
var3   f (i)  /* contains the ith feature of the data set */
begin
  IGFS.build ();
begin
      sf1 = {};  /* create array of var1*/
          for loop i=1 to int of features
                  INF=compute (IG) for the feature /* store computed features*/
                      Gain (i) =INF       /* compute Gain (i) */
          end for
              Th= threshold value   /* hold threshold value*/
                     For i= 1 to number of features
                         If   gain (i) > Th   then
                           Sf1=sf1+f {i} /* store and compute every feature in dataset*/
                         end if
                     end for
end
```

3.3.3. Gain Ratio Feature Selection

A decision tree can be a simple form when non-terminal nodes perform tests on many attributes to the effect of decision outcomes (J.R. Quinlan,1986), as shown in fig 4.

Let Q set of q data and m is a class and can classify data by

$$I(Q) = -\sum_{i=1}^{m} pi\ log2(pi) \qquad (4)$$

Qij is many samples of class Ci in a subset Qj. Qj contains those samples in Q that have value aj of A. The predictable information(J.R. Quinlan,1986):

$$E(A) = -\sum_{i=1}^{m} \frac{i(q)(q1i + q2i+..qmi)}{q} \tag{5}$$

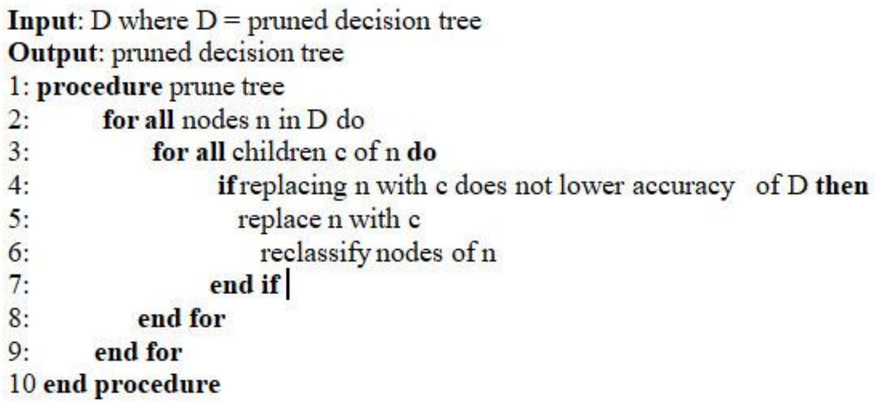
The training information gained by

$$Gain(A) = I(Q) - E(A) \tag{6}$$

$$SplitInfoA(Q) = -\sum_{i=1}^{v} \left(\frac{|Qi|}{|Q|}\right) log2 \left(\frac{|Qi|}{|Q|}\right) \tag{7}$$

The shown value of splitting data into dataset Q into v partitions consequent to v outcomes the test on attribute A(J.R. Quinlan,1986). The gain ratio is

$$Gain\ Ratio(A) = \ Gain\ A/\ QplitInfoA\ (Q) \tag{8}$$



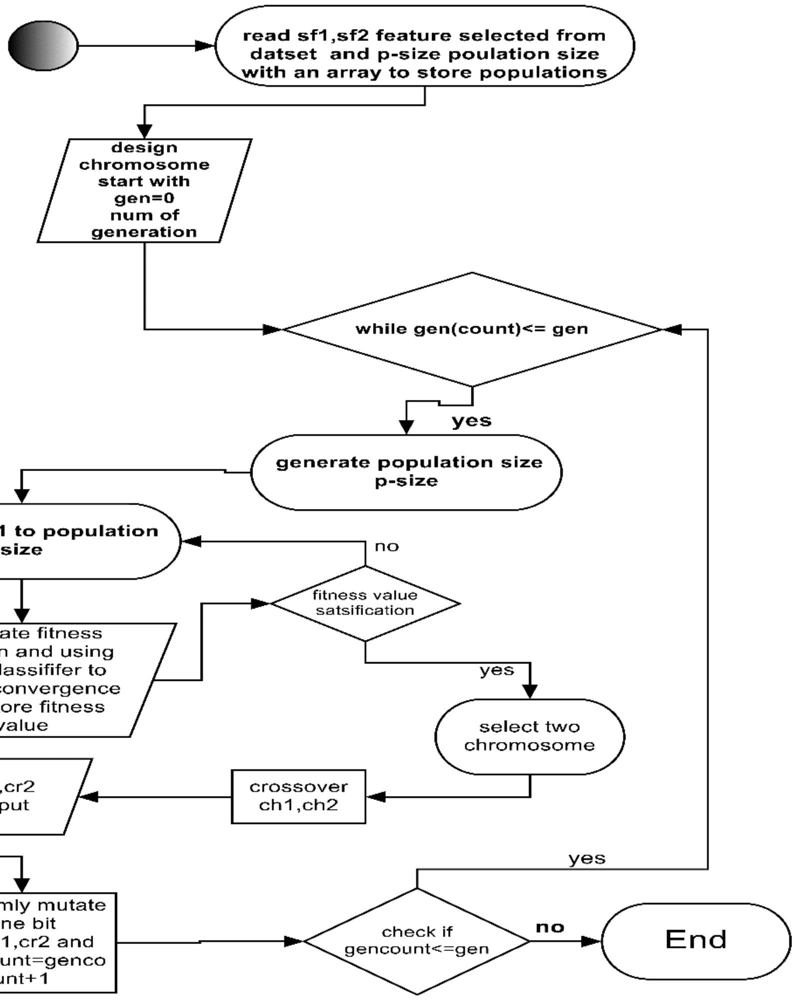
```
Input: D where D = pruned decision tree
Output: pruned decision tree
procedure prune tree
    for all nodes n in D do
        for all children c of n do
            if replacing n with c does not lower accuracy  of D then
                replace n with c
                reclassify nodes of n
            end if
        end for
    end for
end procedure
```

Figure 4. Decision tree algorithm

3.3.4. Optimized Genetic Algorithm (OGA)

The authors propose a method to modify a general genetic algorithm to evaluates specified attributes on training data or a separate testing set and uses a decision tree (J.R. Quinlan,1986) to estimate the 'merit' of a set of attributes to produce an optimized feature subset with genetic search elevation strategy to recognize the features. All feature selection technique should use an evaluation function together with a search strategy to achieve the optimal feature set(Huang & C.,2012) . It is unable to be realized to search all subsets to find out an optimal subset and need much effort to indicate whether a particular feature is present or not in the chromosome, one, and zero used. One in a gene position refers to feature and zero to absent(Yanan Mao & Dingyuan Fan,2016).The number of features and what are the features that are to be present in a chromosome are guided by information gain (IG) and gain ratio (GR). The initial population created using input values of IG and GR of the values present in the chromosome. After Generated the population, the individuals evaluated using a fitness function. There is no general approach to find the fitness function for a genetic algorithm. It is a heuristic approach and depends on the used application. So the authors nominate a C4.5 classifier to be used as a fitness function because C4.5 has some utility of handling both continuous and discrete attributes and training data with missing attribute values, pruning trees after creation - C4.5 goes back through the tree once it has been created and try to eject branches that do not help by replacing them with leaf nodes (Dash & H. Liu, 1997) . (J.R. Quinlan,1986) . The following algorithm selects a feature from the set of features that are gained by OGA, gain ratio, and Information gain, as shown in fig 3.

Fig. 3 Proposed CSFS with a genetic algorithm with the c4.5 decision tree



4. Proposed Stack Hybrid Classification Model Based On Composite Hybrid Feature Selection (CHFS)

A weka software tool (Weka online open-source accessed,2018) shows the list of black-box classifiers. These algorithms, in general, are used to classify the medical dataset.

4.1. Two learning evaluators can be used to evaluate the dataset

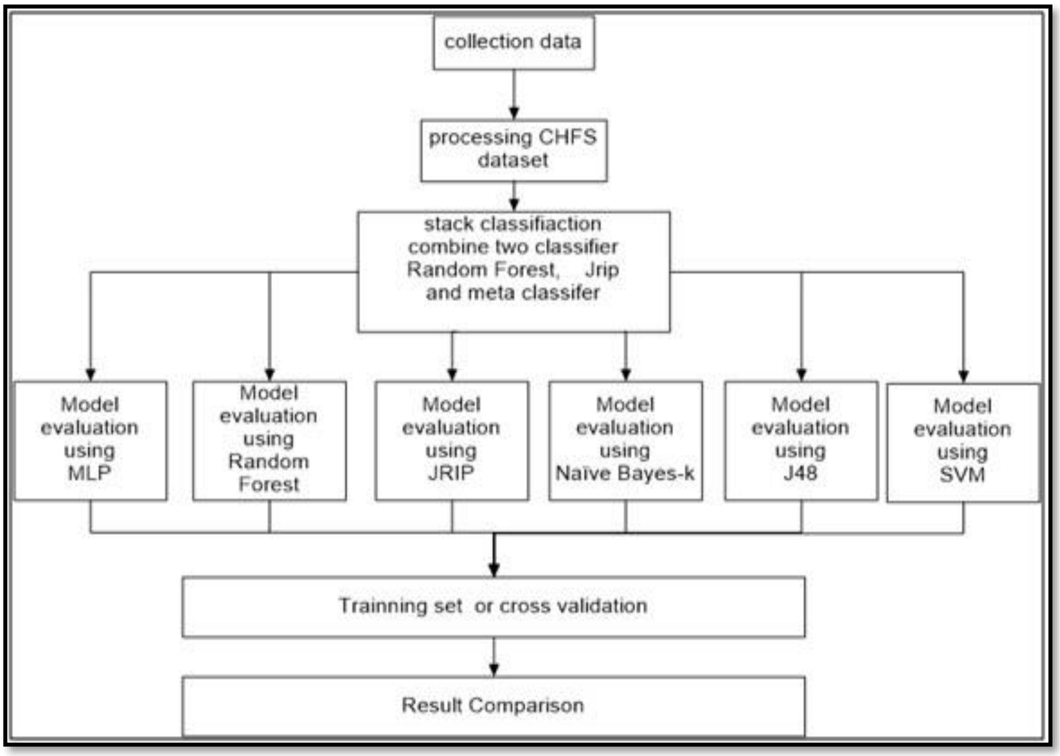
- Training set: the classifier separates a dataset to test and training data. The result of each model can be saved and can visualize.

- Cross-validation: in case of 10 fold cross-validation, weka develops ten models, when it displays the result it uses the average performance of those ten models. It deletes the remaining models. From the observations, the authors conclude that the model saved with cross-validation and the training set is the same. (Divya Jain, Vijendra Singh,2018).

4.2. Stacking technique

Ensemble methods are learning methods that contain a set of classifiers for classifying data by taking a weighted point of their predictions (Leo Breiman,1996) . The authors combine multiple classifiers to get the maximum efficiency of classification accuracy and overcome the weakness of individual classifiers in the classification process on potential patients. Classifiers, as shown in fig 4.

Fig. 4the proposed framework of stack hybrid classification based on the CHFS model



The author chooses permanent (Jrip, RF) based on a result from the table2, which achieves a good indicator of a fitness function problem.

Naive Bayesian classifier is an eclectic classifier and can calculate a group set of probabilities by counting the value and frequency in a given dataset [29]. It assumes that all variables which contribute towards classification are mutually independent [21]. Naive Bayesian classifier based on Baye's notion and theorem of total probabilities. equation 9 is the probability of a document with a vector[7-8]

x= {x1, x2, xn} belongs to hypotheses ‘h’ is given by,

$$P(h1|dx) = \frac{P(xi|h1)P(h1)}{P(xi|h1)P(h1) + P(xi|h2)P(h2)} \quad (9)$$

$$P(h1|xi) = \frac{P(xi|h1)P(h1)}{P(x1)} \qquad (10)$$

$$P(h1|xi) = \sum_{j=1}^{n} P(xi|hj)P(hj) \qquad (11)$$

Decision tree algorithm J48: creates a binary tree to build the model of the classification method[7-8].

Built the tree and applied to the list and results in classification, and J48 ignores the missing values. [8].

Random forests can be implemented to create a group of decision trees at the training period and generate the class. The features randomly selected in each decision split [8].

Rule-based classification: Rules can be representing information if-then rules expressed in the form of if condition, then conclusion a ruler can be assessed by its coverage and accuracy. [22].

$$Coverage(R) = \frac{n\ covers}{|D|} \qquad (12)$$

$$Accuracy\ (R) = \frac{n\ correct}{ion\ covers} \qquad (13)$$

JRip Rules Classifiers[41]: Classes are a measure to rising size and generated a group of rules for the class to reduce the error gradually, JRip (RIPPER) continues by treating every one of the instances of an exacting decision in the training data as a class and discovery group of rules that included in the same class. [23].

Support Vector Machines (SVM)[41]: is a supervised learning algorithm based on statistical learning theory proposed by Vapnik [10]

$$K(Xi,Yi) = (\gamma X_I^T Y_i\ + r)^m, \gamma > 0 \qquad (14)$$

And the Radial basis kernel function (RBF) [26]:

$$K(x_i, y_i) = \exp\left(-\gamma \left\|x_i - y_j\right\|^2\right), \gamma > 0 \qquad (15)$$

Multilayer perceptron (MLP): is a feed-forward network. It utilizes supervised learning, and It contains three layers of nodes apply nonlinear activation functions. [24].

4.3. Metrics used in health check systems for evaluation

The different performance metrics generally used to explore the performance of the various models like sensitivity, accuracy, precision, and f-measure (Divya Jain & Vijendra ,2017) . Accuracy: can be calculated by divide number of accurate predictions by the total number of all predications (Klöppel & ...&Frackowiak & R. S,2008)

$$Accuracy = \frac{TN+TP}{fP+tP+fN+tN}$$

Inside equation

(FN = False Negative, FP = False Positive, TN = True Negative, TP = True Positive)

5. Experimental Results and Evaluation

The CSV file of the medical Dataset (clinical test) to patients of Alzheimer's dementia from oasis.org(online access Oasis Medical Dataset,2019) loaded to the weka tool. All experiments evaluated by the receiver operating characteristic (ROC) curve, accuracy, F-measure (Dina A. Ragab & Omneya Attallah,2019).

5.1.Result from proposed (CHFS) feature selection model

The model implementation from the three feature selection methods,IG, GR, was applied for input to our optimization GA as initialization instead of randomly. Also, The population size is 100, number of generations is 20, the crossover takes place at the middle position and mutation is prepared at one point randomly to reduce features of a dataset and extract the optimal feature subset and result was consisting of 5 features terms as shown in table 1.

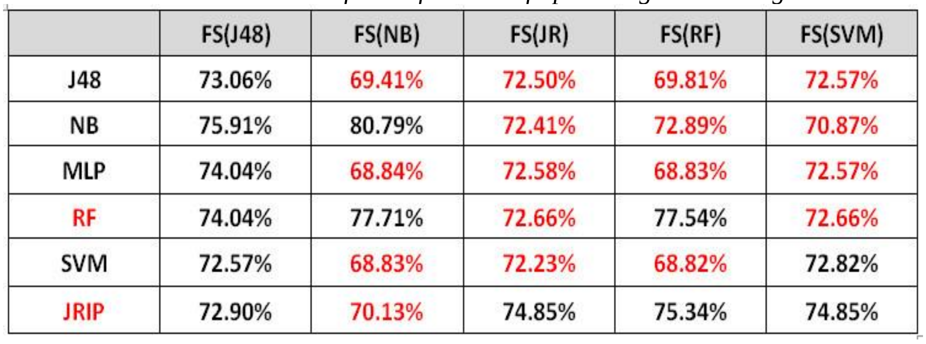
Table 1:Number of feature extraction from three feature selection

	Method	No.of Feature
IG, GR	Filter	6
GA	Wrapper	6
Optimized GA Fitness Fun(C4.5)	Embedded	5

Source: (Weka open source software, 2018)

In table 2 below we calculate the fitness function of j48 and get accuracy with classical classifiers and repeat this procedure with another classifier of calculation a fitness function too and obtain the accuracy for all, summerize result shown that the j48 is perform a good indicator as a fitness function with all classifiers. Furthermore, measure false-positive and false-negative for each chromosome. The chromosome, which has the lowest value, is considered an elite one(Divya Jain, Vijendra Singh,2018).

Table 2 Calculate the fitness function of optimizing Genetic Algorithm



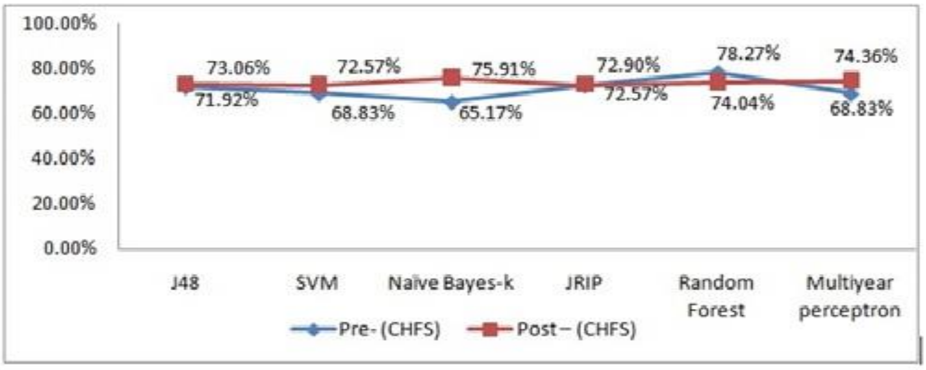
	FS(J48)	FS(NB)	FS(JR)	FS(RF)	FS(SVM)
J48	73.06%	69.41%	72.50%	69.81%	72.57%
NB	75.91%	80.79%	72.41%	72.89%	70.87%
MLP	74.04%	68.84%	72.58%	68.83%	72.57%
RF	74.04%	77.71%	72.66%	77.54%	72.66%
SVM	72.57%	68.83%	72.23%	68.82%	72.82%
JRIP	72.90%	70.13%	74.85%	75.34%	74.85%

The threshold value equal to 0.02 chosen for information gain, as shown in Fig 10, lead to reduce feature term to 5 attributes, and a value greater than or equal to 0.04 was selected for gain ratio method, as shown in Fig.11 to rearrange the feature term as priority and preparation for classification method to obtain the maximum efficiency. Table 1 discusses the results of all the techniques. If 10- fold cross-validation applied. For summarized results in the table (3) from the weka software of the analytics dataset, we used the receiver operating characteristic curve (ROC) to evaluate the output result from our proposed architecture model, as shown in fig 12,13.

Table 3: Experimental result of Accuracy pre-post CHFS model with various classifiers

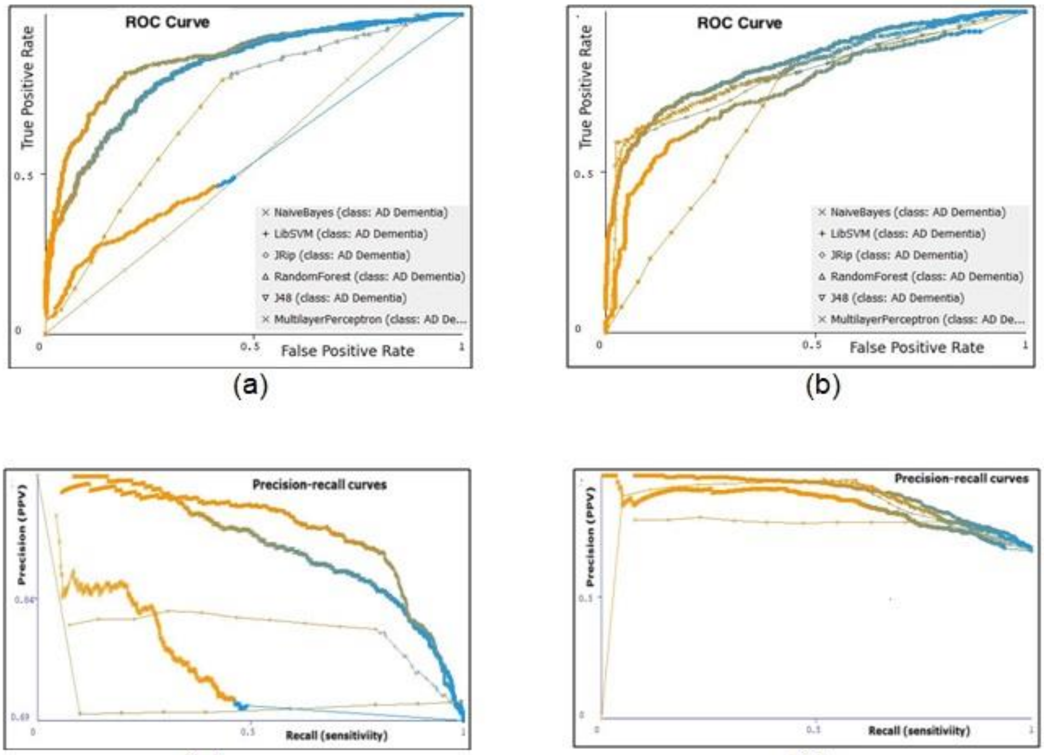
Classifier	Pre- (CHFS) Feature Selection Accuracy	Post – (CHFS) Feature Selection Accuracy
J48	71.92%	73.06%
SVM	68.83%	72.57%
Naïve Bayes-k	65.17%	75.91%
JRIP	72.57%	72.90%
Random Forest	78.27%	74.04%
Multiyear perceptron	68.83%	74.36%

Fig.5 Summarized accuracy results of pre-post (CHFS) feature selection



Source: (Weka open source software, 2018)

Fig.6ROC curve and precision-recall curve for all six classifiers on raw features (a) Pre-CHFS model (b) Post-CHFS model (c) Precision and recall curve pre-CHFS model(d) Precision and recall curve post-CHFS model



(a) (b)

(c) (d)

Source: (Weka open source software, 2018)

As an essential point, the proposed (CHFS) model compared against a popular dimensionality reduction technique, the Principal Component Analysis (PCA) (Joliffe IT,2002) . We have used PCA to reduce the number of variables of our oasis medical dataset(online access Oasis Medical Dataset,2019) , and the result was three features PCA selected at 91.1% accuracy of variance and five features of PCA selected at 94.3% accuracy. The result of this comparison between PCA and our proposed CHFS feature selection model among six traditional classifiers, as shown in Table 4.Furthermore, summarized in fig. 5, Fig 6, explained the ROC curve to evaluate the result of this comparison of our feature selection technique evaluation.

Table 4: Summary of 10-fold Cross-Validation (CHFS)Model Accuracy Compared Against PCA

Classifier	PCA Accuracy		(CHFS) Model Feature Selection Accuracy
	(3) Components (91.1%)	(5) components (94.3%)	
J48	68.72%	68.83%	73.06%
SVM	68.72 %	68.80 %	72.57%
Naïve Bayes	72.78%	72.74%	75.91%
JRIP	69.73%	69.81%	72.90%
Random Forest	72.74%	72.8%	74.04%
Multiyear.perceptron	69.43%	69.83%	74.36%

Fig.7 Summarized accuracy results of PCA versus (CHFS) feature selection

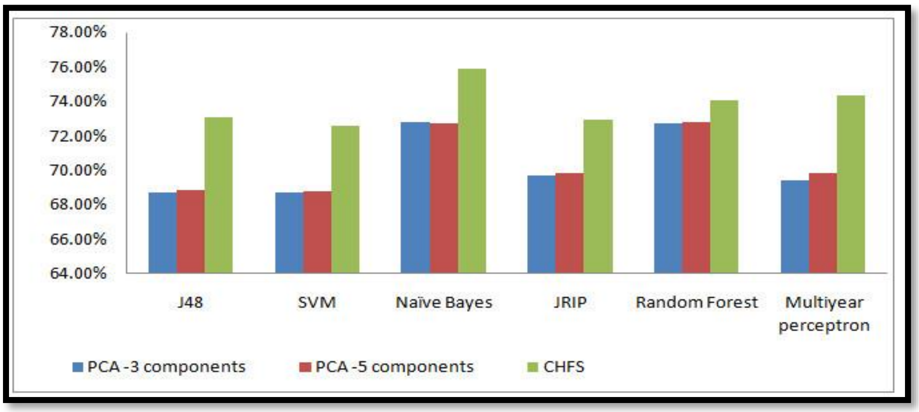
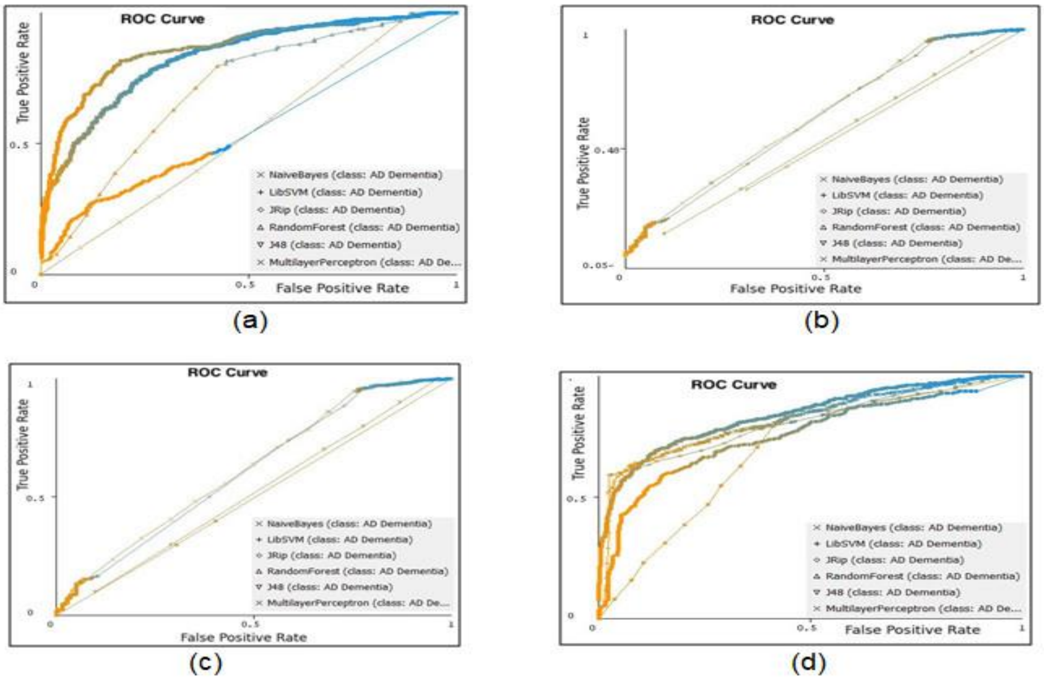


Fig.8 ROC curve for all six classifiers on raw features (a) Roc of a medical dataset on traditional classifiers (b) Of 3- component –PCA(c) Of 5- component -PCA (d) Post-CHFS model(d) Post-CHFS model



Source: (Weka open source software, 2018)

5.2 Result of proposed Stack Hybrid Classification Model Based on (CHFS) Feature Selection

The classification techniques applied to medical dataset. (Yasser Fouad &...& Hatem Khater,2015) . Can be achieved by

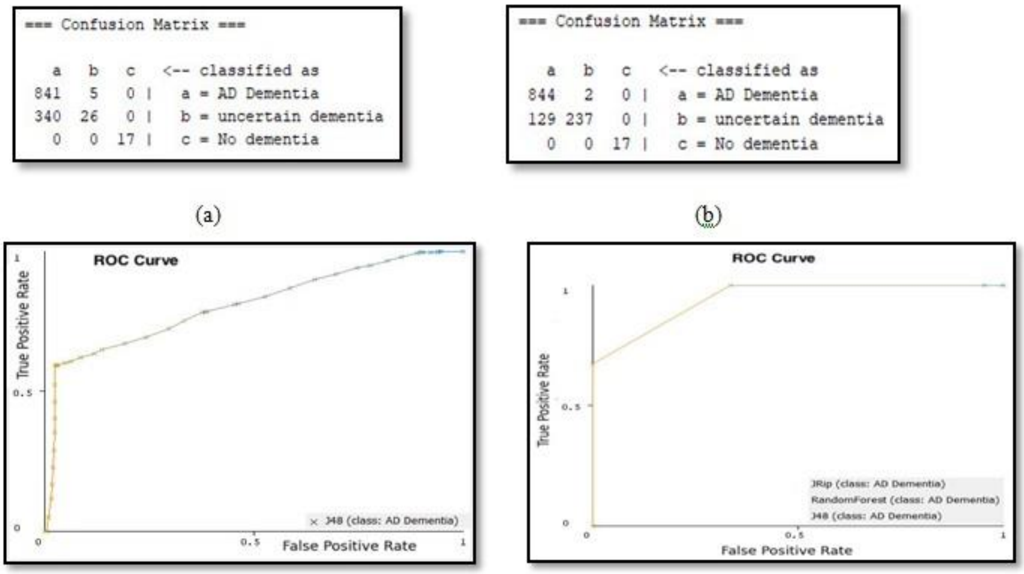
$$ClassificationAccuracy = \frac{CorrectlyClassifiedSamples}{TotalNumberofSamples} * 100 \qquad (3)$$

The authors use the output dataset from CHFS feature selection to run it inside the proposed classification model to flow knowledge in the weka area tool and begin the combination process of random forest and Jrip classifier with six classifiers as a meta-classifier individually.

5.2.1 Hybrid classification combination process of (random forest, Jrip) with j48 stack Meta classifier

The result of non-Alzheimer's samples showed that predicted to be infected with Alzheimer's disease and high overall accuracy (89.34%) compared with the j48 classifier individually (71.92%).

Fig.18Confusion matrix and ROC curve for J48 classifier on raw features (a) Confusion matrix Pre-Hybrid classification model Fig 18. (b) Confusion matrix Post- Hybrid classification modeling 18. (c) ROC curve pre-Hybrid classification modeling 18. (d) ROC curve post- Hybrid classification model



```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 841   5   0 |   a = AD Dementia
 340  26   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(a)

```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 844   2   0 |   a = AD Dementia
 129 237   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(b)

(c)

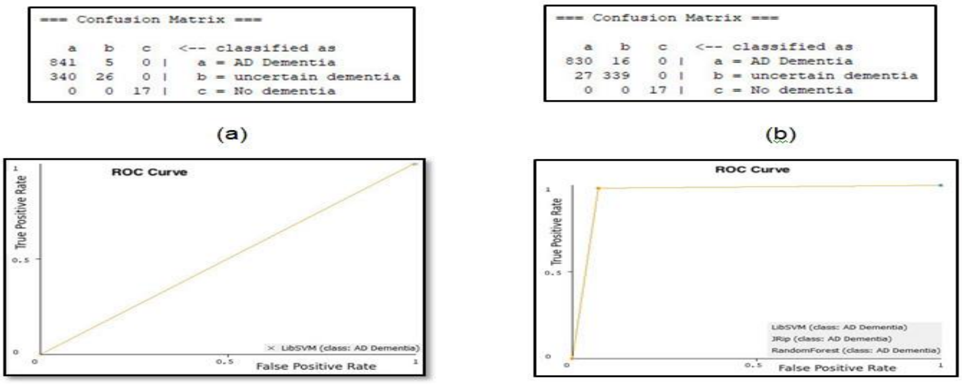
(d)

Source: (Weka open source software, 2018)

5.2.2 Hybrid classification of (random forest, jrip) with SVMmeta-classifier

The result was a high overall accuracy (96.50%) compared with the SVM classifier individually (68.83%), as shown in fig 9.

Fig 9.Confusion matrix and ROC curve for SVM classifier on raw features (a) Confusion matrix Pre-Hybrid classification model (b) Confusion matrix Post- Hybrid classification model(c) ROC curve pre- Hybrid classification model (d) ROC curve post- Hybrid classification model



```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 841   5   0 |   a = AD Dementia
 340  26   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(a)

```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 830  16   0 |   a = AD Dementia
  27 339   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(b)

(c)

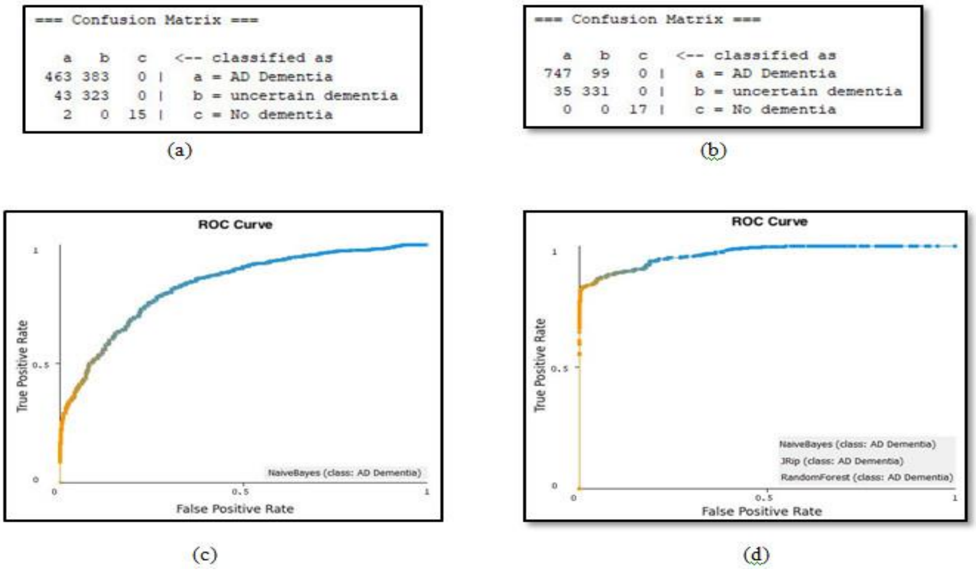
(d)

Source: (Weka open source software, 2018)

5.2.3 Hybrid classification of (random forest, jrip) with Naive Bayesas meta-classifier

The result is a high overall accuracy (89.09%) compared with Naive Bayes classifier individually (65.17%), as shown in fig 20.

Fig.10 Confusion matrix and ROC curve for Naive Bayes classifier on raw features (a) Confusion matrix Pre-Hybrid classification model (b) Confusion matrix Post- Hybrid classification model (c)curve pre- Hybrid classification model(d) ROC curve post- Hybrid classification model



```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 463 383   0 |   a = AD Dementia
  43 323   0 |   b = uncertain dementia
   2   0  15 |   c = No dementia
```

(a)

```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 747  99   0 |   a = AD Dementia
  35 331   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(b)

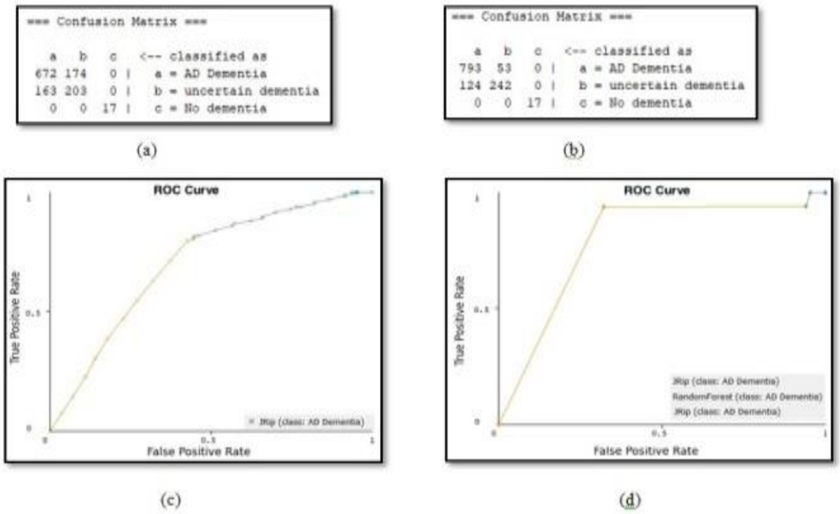
(c) (d)

Source: (Weka open source software, 2018)

5.2.4 Hybrid classification of (random forest, Jrip) with Jripas meta-classifier

the result is a high overall accuracy (85.59%) compared with Naive Bayes classifier individually (72.57%), as shown in fig 11.

Fig.11Confusion matrix and ROC curve for Jrip classifier on raw features (a) Confusion matrix Pre-Hybrid classification model (b) Confusion matrix Post- Hybrid classification model (c) ROC curve pre- Hybrid classification model(d) ROC curve post- Hybrid classification mode



```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 672 174   0 |   a = AD Dementia
 163 203   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(a)

```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 793  53   0 |   a = AD Dementia
 124 242   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(b)

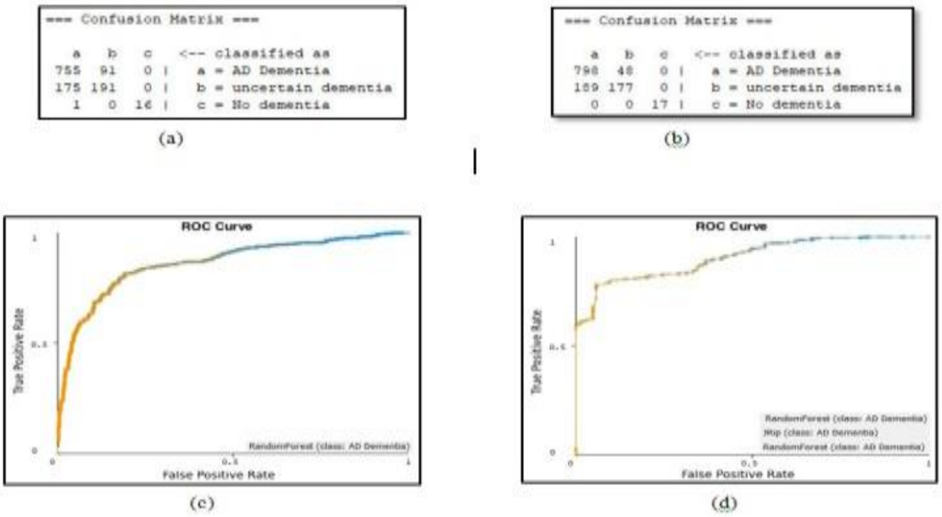
(c) (d)

Source: (Weka open source software, 2018)

5.2.5 Hybrid classification of (random forest, jrip) with a random forest as meta-classifier

The result is a high overall accuracy (80.71%) compared with random forest classifiers individually (78.27%), as shown in fig 12.

Fig.12 Confusion matrix and ROC curve for random forest classifier on raw features (a) Confusion matrix Pre-Hybrid classification model(b) Confusion matrix Post- Hybrid classification model (c) ROC curve pre- Hybrid classification model(d) ROC curve post- Hybrid classification model



```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 755  91   0 |   a = AD Dementia
 175 191   0 |   b = uncertain dementia
   1   0  16 |   c = No dementia
```

(a)

```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 798  48   0 |   a = AD Dementia
 189 177   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(b)

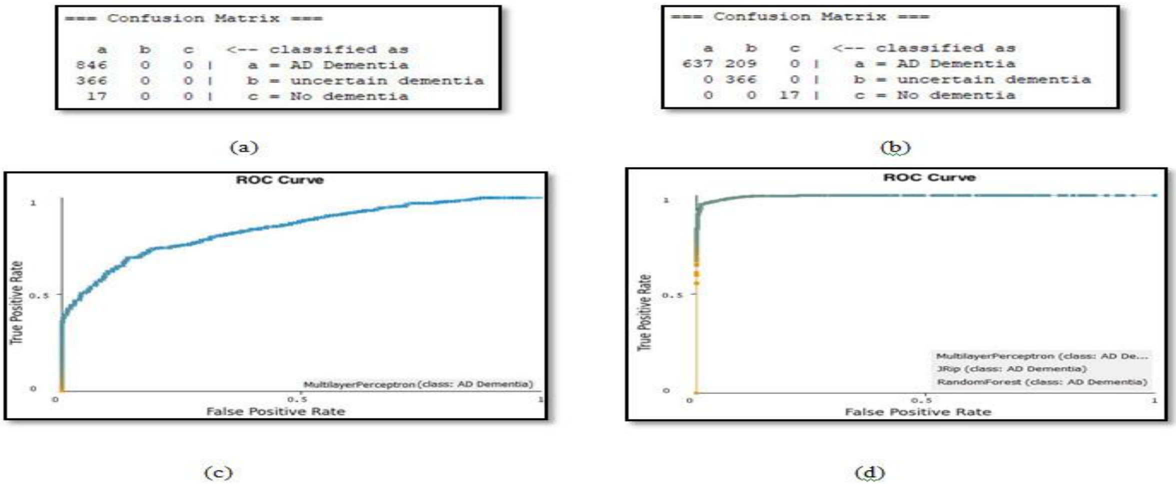
(c) (d)

Source: (Weka open source software, 2018)

5.2.6 Hybrid classification of (random forest, jrip) with Multilayer Perceptron as meta-classifier

The result is a high overall accuracy of (83%) compared with the Multilayer Perceptron classifier individually of (68.83%) in training set mode, as shown in fig 13.

Fig.13 Confusion matrix and ROC curve for Multilayer Perceptron classifier on raw features (a) Confusion matrix Pre-Hybrid classification model (b) Confusion matrix Post- Hybrid classification model(c) ROC curve pre- Hybrid classification model(d) ROC curve post- Hybrid classification model



```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 846   0   0 |   a = AD Dementia
 366   0   0 |   b = uncertain dementia
  17   0   0 |   c = No dementia
```

(a)

```
=== Confusion Matrix ===

   a   b   c   <-- classified as
 637 209   0 |   a = AD Dementia
   0 366   0 |   b = uncertain dementia
   0   0  17 |   c = No dementia
```

(b)

(c) (d)

Source: (Weka open source software, 2018)

6. Discussion

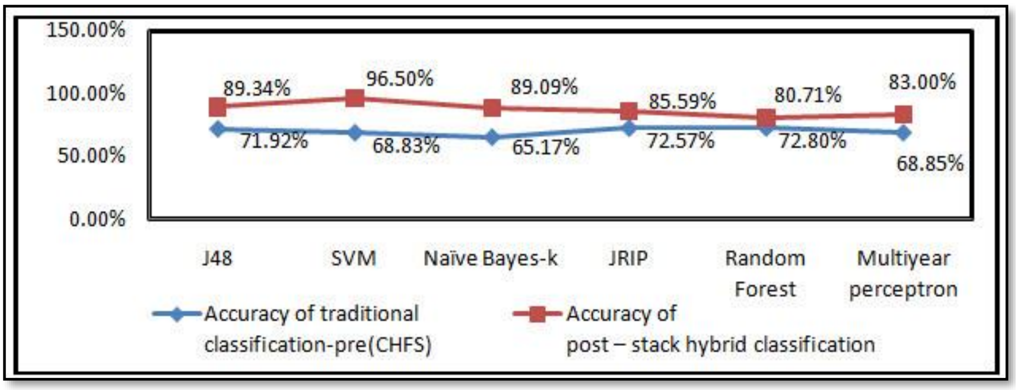
The authors compare the results of different classification methods obtained with the results of our proposed stack hybrid classification based on (CHFS) feature selection on the current medical dataset of 1229 potential patient samples. The neural network classifier (Hagan, Demuth, and Beale1996) used that produced true-positive rate (TPR) of 84.0%, and false-positive rate (FPR) of 85.0%, Multilayer perceptron classifier (Yan et al. 2006) produced a TPR of 74.9% and FPR of 79.3%. Furthermore, the linear regression classifier (KO and Barkana 2014) presented a TPR of 67.7% and FPR of 68.7%. Naive Bayesian network (John and Langley 1995) showed a TPR of 89.2% and an FPR of 89.5%. The proposed stack hybrid classification based on (CHFS) feature selection when applied on an oasis medical dataset(online access Oasis Medical Dataset,2019).

The sensitivity of 96.50% when combining random forest, Jrip classifiers with SVM classifier as meta-classifier, and resulted in a sensitivity of 85.59%with Jrip as meta-classifier, and resulted in a sensitivity of 83%with Multiyear perceptron classifier as meta-classifier, and resulted of 89.09%with Naive Bayes-k classifier as meta-classifier, and resulted of 89.34%with J48classifier as meta-classifier, and resulted of 80.71%with random forest classifier as meta-classifier. These comparisons, according to our proposed model presented in this study, were reduced a false negative rate and showed a relatively high overall accuracy with more accurate results, as shown in fig 14, 15, and table 5.

Table 5: Summary of 10-fold Cross-validation of Stack Hybrid Classification Based on (CHFS) In Weka3.8

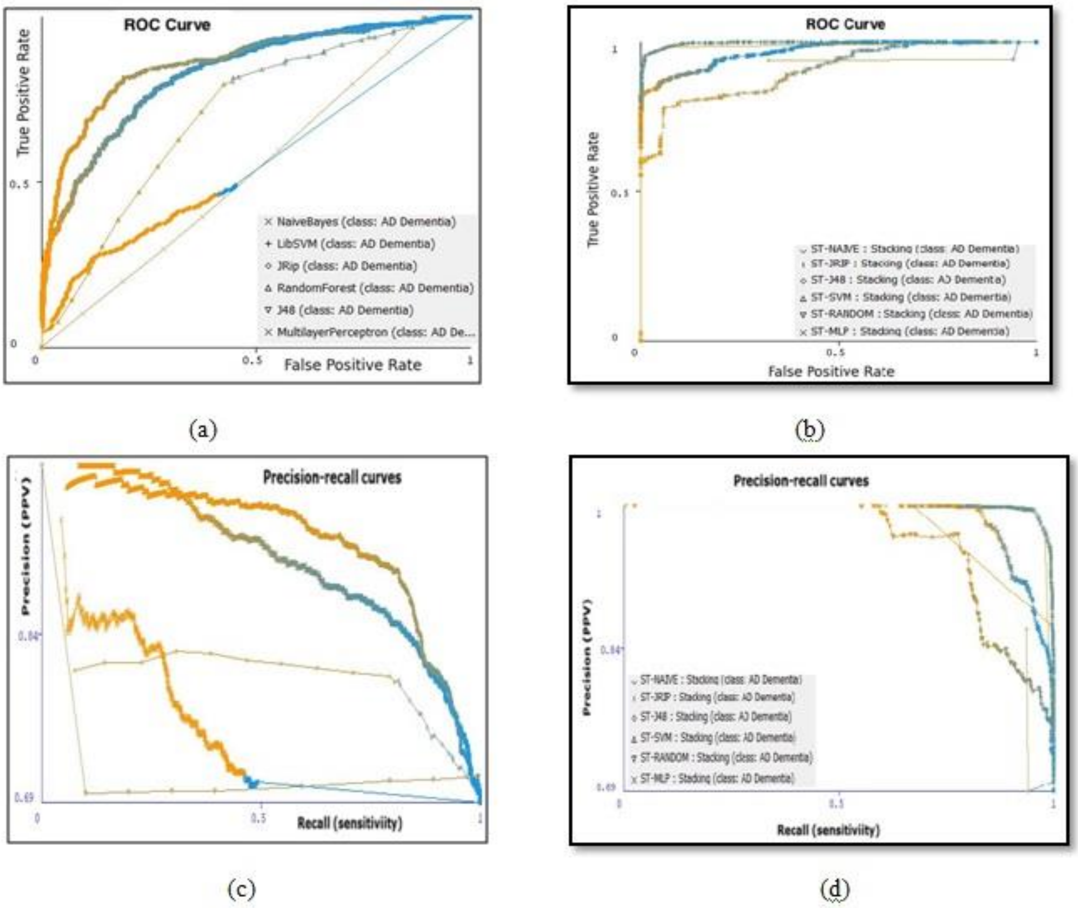
Traditional Classifier	Accuracy of traditional pre (CHFS)	Combined Classifiers with meta - classifier		Accuracy after(CHFS)
J48	71.92%	Random Forest and Jrip	J48	89.34%
SVM	68.83%		SVM	96.50%
Naive Bayes	65.17%		Naive Bayes	89.09%
JRIP	72.57%		JRIP	85.59%
Random Forest	78.27%		Random Forest	80.71%
Multiyear perceptron	68.83%		Multiyear perceptron	83%

Fig.14 Summarized results of pre-post stack classification based on (CHFS)



Therefore, The summarize of all the previous results in the ROC curve and recall –precision curve to evaluate all processes of our proposed stack hybrid classification model based on composite hybrid feature selection model (CHFS), as shown in fig 25.

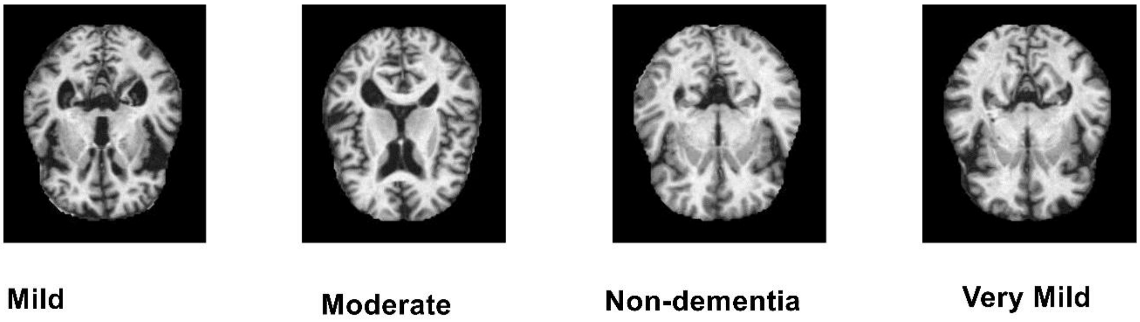
Fig.15 ROC curve and Precision - recall curve for all six classifiers on raw features (a) ROC -Pre- proposed hybrid classification model(b) ROC -Post- proposed hybrid classification model (c) Precision and recall curve pre- proposed hybrid classification model(d) Precision and recall curve post- proposed hybrid classification model



Source: (Weka open source software, 2018)

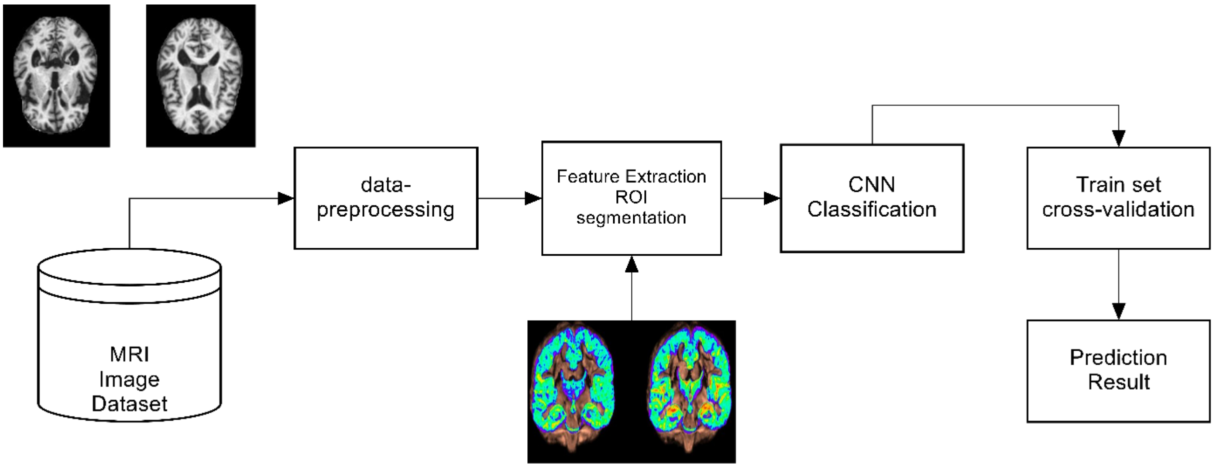
The author's aims to improve clinical data for early diagnose of AD and to prove that we make a comparison between clinical data diagnose and MRI diagnoses in Early-stage, We tested a set of Alzheimer's MRI images from the kaggle.com benchmark web of dataset science, and the dataset contains 5121 MRI image divided to 4 classes(mild, very mild,non, moderate) as shown in below figure 16 to compare with clinical data to perform the accuracy of early-disease diagnosis by the convolution neural network as shown in the below layout model in figure 17

Figure 16 the types of MRI classes for AD disease on the brain



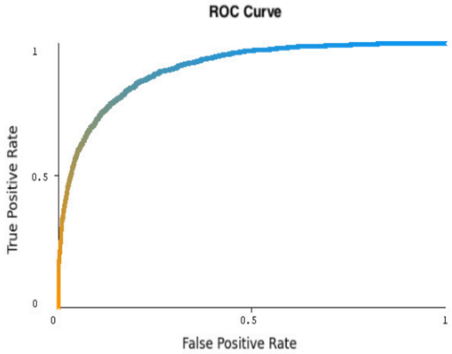
Mild **Moderate** **Non-dementia** **Very Mild**

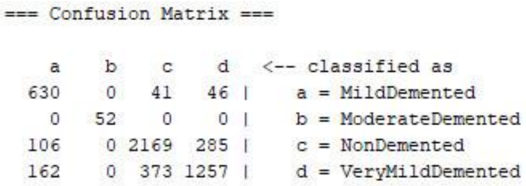
Figure 17 shows the prediction model for MRI images



In the below figure we propose the result of MRI classification on Kaggle 5121 images dataset by using the convolution neural network and the result evaluated by ROC curve – f - measure

Fig.15 ROC curve of MRI category and confusion matrix of CNN classification



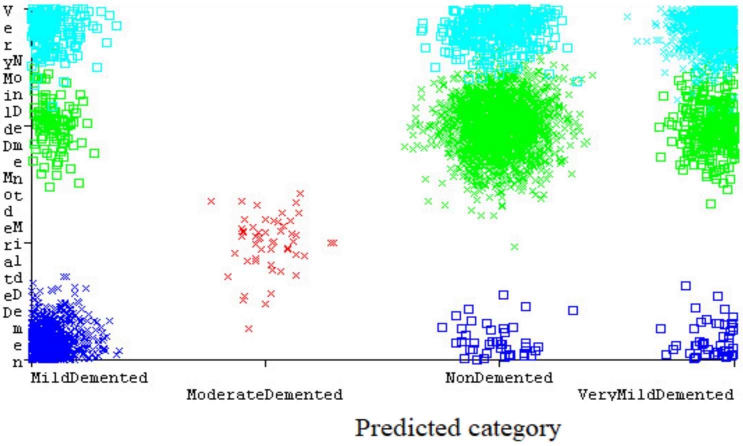


```
=== Confusion Matrix ===

   a    b    c    d   <-- classified as
 630    0   41   46 |    a = MildDemented
   0   52    0    0 |    b = ModerateDemented
 106    0 2169  285 |    c = NonDemented
 162    0  373 1257 |    d = VeryMildDemented
```

And in the below figure 16 shown the error curve of CNN classification.

Figure 16 shown the error curve of CNN classification process of MRI Images



The classification accuracy of the convolution neural network (CNN) was 80.21% and f-measure 80.1%, which considered least than the classification accuracy of our clinical data, which confirms the importance of our proposed feature extraction and hybrid classification of producing CHFS-SVM model of 96.5% in early diagnoses.

And the below table shows the comparison of our proposed model feature selection and hybrid classification with the last state of the art result.

Table 6: Comparison of our proposed model with the last state of the art result

	Sensitivity	specificity
(John and Langley 1995)	**89.2%**	**89.5%**
(Hagan, Demuth, and Beale1996)	**84%**	**85%**
(Yan et al. 2006)	**74%**	**79%**
(KO and Barkana 2014)	**67%**	**68%**
(Tejeswinee ,Shomana,2017)	**92%**	**85.7%**
ShaikBasheer,2019	**90.47%**	**86.66%**
Kaggle MRI dataset with **CNN classification**	**80.21%**	**67.5%**
Proposed **Model** **CHFS+SHC (SVM)**	**96.5%**	**91.8%**

7. Conclusion

In this work, the author aims to early diagnosis of AD by using a benchmark dataset on our proposed composed hybrid feature selection (CHFS) model. This combines the advantages of three filter feature selection approaches and optimizes the Genetic Algorithm (OGA) by improving the initial population generating and genetic operators.

Also, the results of the filter approach as some prior information using the J48 decision tree classifier as a fitness function instead of probability and random selection to speed up convergence and select the best features.

After that, using the selected feature in stack hybrid classification and combine three classifiers with improving the prediction and accuracy. The proposed model performs better than the traditional classification approaches for optimum feature selection and improvement of the classification process and effectively reduced the false-negative rate with high accuracy when using a support vector machine (SVM) as meta-classifier in a hybrid classification method with 96.50% compared to 68.83% of usage individually and the last state of the art result shown above in Table 6 with our experiment on Kaggle MRI dataset of CNN classification process with 80.21% accuracy result. The results of the proposed model show an accurate classify Alzheimer's clinical samples against at a low cost.

References

[1]Kalló, Gergő, Miklós Emri, et al. “Changes in the Chemical Barrier Composition of Tears in Alzheimer’s Disease Reveal Potential Tear Diagnostic Biomarkers.” PLoS One, vol. 11, no. 6, Public Library of Science, June 2016, p. e0158000.

[2] J. Escudero, et al. Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease. IEEE transactions on biomedical engineering 2013; 60(1): 164-168.

[3] Jyoti Soni et al. "Predictive Data Mining for medical diagnosis: An Overview of Heart Disease Prediction"International Journal of Computer Applications (0975 –8887), Volume 17, No. 8, March 2011."

[4] Williams, Jennifer A., Alyssa Weakley, Diane J. Cook, and Maureen Schmitter-Edgecombe. (2013) "Machine learning techniques for diagnostic differentiation of mild cognitive impairment and dementia." In Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence: 71-76...

[5] Chi, C. L., Oh, W., &Borson, S. (2015) "Feasibility Study of a Machine Learning Approach to Predict Dementia Progression," In Healthcare Informatics (ICHI), International Conference: 450-450

[6] Chyzhyk, D., &Savio, A. (2010) "Feature extraction from structural MRI images based on VBM: data from OASIS database," University of The Basque Country, Internal research publication.

[7] S. R Bhagya Shree and Dr. H. S. Sheshadri "An initial investigation in the diagnosis of Alzheimer's disease using various classification techniques" IEEE ICCIC, 2014.

[8] Tina R. Patil and Mrs. S. S. Sherekar "Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification" International Journal of Computer Science and Applications Vol. 6, No.2, Apr 2013

[9] Jihad Ali et al. "Random Forests and Decision Trees" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012

[10] Vemuri, P., Gunter, J. L., Senjem, M. L., Whitwell, J. L., Kantarci, K., Knopman, D. S., ... & Jack, C. R. (2008) "Alzheimer's disease diagnosis individual subjects using structural MR images: validation studies," Neuroimage, 39(3): 1186-1197.

[11] Datta, P., and Shankle, W. R., &Pazzani, M. (1996) "Applying machine learning to an Alzheimer's database," In Conference proceedings of the AAAI symposium: 25-27.

[12] Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scahill, R. I., Rohrer, J. D., ... &Frackowiak, R. S (2008)"Automatic classification of MR scans in Alzheimer's disease." Brain, 131(3): 681-689.

[13] So, A., Hooshyar, D., Park, K. W., & Lim, H. S. (2008) "Early Diagnosis of Dementia from Clinical Data by Machine Learning Techniques," Applied Sciences, 7(7): 651.

[14] Ramirez, J., Górriz, J. M., Salas-Gonzalez, D., Romero, A., López, M., Álvarez, I., & Gómez-Río, M. (2013) "Computer-aided diagnosis of Alzheimer's type dementia combining support vector machines and discriminant set of features." Information Sciences, 237: 59-72.

[15] Sheshadri, H. S., Shree, S. B., & Krishna, M. (2015) "Diagnosis of Alzheimer's Disease Employing Neuropsychological and Classification Techniques," In IT Convergence and Security (ICITCS), 2015 5th International Conference: 1-6.

[16]ALzehimer's clinical medical dataset available online at https://www.oasis-brains.org/#Data (accessed on 2 Jan 2019)

[17] Data mining: Practical machine learning tools and techniques by Ian H Witten and Eibe Frank, published by Elsevier, second edition 2008.

[18]Aouatif Amine Ali Elakad Mohammed Rziza Driss "GA-SVM and Mutual Information based Frequency Feature Selection for Face Recognition" GSCM-LRIT, Faculty of Sciences, Mohammed V University, B.P. 1014 Rabat, Morocco(2011).

[19] J.R. Quinlan, Induction of Decision Trees, Machine Learning 1: pp.81-106, Kluwer Academic Publishers, Boston, (1986).

[20] Leo Breiman. "Stacked Regressions" Machine Learning, 24, 49-64 (1996)

[21] Duarte Ferreira et al. "Applying data mining techniques to improve diagnosis in neonatal jaundice" BMC Medical Informatics and Decision Making, 0-5, 2012

[22] Jiawei Han, Micheline Kamber, Jian Pei "Data Mining: Concepts and Techniques" published by Elsevier, Third edition, 2012

[23] Anil Rajput, Ramesh Prasad Agarwal, Meghna Dubey S.P. Saxena, and Manmohan Raghuvanshi" J48 and JRIP Rules for EGovernance Data" International Journal of Computer Science and Security (IJCSS), Volume (5): Issue (2), Pp 201, 2011.

[24]Guyon I, Elisseeff A An introduction to variable and feature selection J Mach Learn Res 2003; 3(Mar):1157–82.

[25] Feature selection and classification systems for chronic disease prediction: A review Divya Jain, Vijendra SinghThe NorthCap University, Gurugram 12,2017, India, [online] http://creativecommons.org/licenses/by-nc-nd/4.0

[26] Huang, C., A hybrid stock selection model using genetic algorithms and support vector regression. Applied Soft Computing, 2012. 12(2): p.807-818

[27] K. Tejeswinee, S.G. Jacob. Binary classification of Cognitive Disorders: Investigation on the Effects of Protein Sequence Properties in Alzheimer's and Parkinson's disease. IAENG-IMECS 2017: 166-170
[28] Joliffe IT. Principal component analysis, series: Springer series in statistics. 2. Springer; NY: 2002
[29]O. Elsayed, K. Mahar, M. Kholief and H. A. Khater, "Automatic detection of the pulmonary nodules from CT images," 2015 SAI Intelligent Systems Conference (IntelliSys), IEEE, London, pp.742-746, 2015, DOI: 10.1109/IntelliSys.2015.7361223.
[30]Yasser Fouad, Hatem Khater, Mostafa Setta, Ashraf Alsaid, “proposed Approach for Automatic Underwater Object Classification,” International Journal of Research and Surveys, ICIC -EL, Volume 12, No. 12, pp. 1202-1212, 2018, DOI:10.24507/icicel.12.12.1205.
[31] The national center for health statistics report for 2019 (https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm).
[32] Weka open source software for data mining .availble online (https://www.cs.wakito.ac.nz/) accessed 26 Dec 2019.
[33] Chen, F. and F. Li, A combination of feature selection approach with SVM in credit scoring. Expert Systems with Applications, 2010. 37(7): p. 4902-4909.
[34] Chen, Y. -W., and Lin, C. -J., Combining SVMs with various feature selection Strategies, Available from http://www.csie.ntu.edu.tw/~cjlin/papers/features.pdf, 2005.
[35] Dash, M., and H. Liu, Feature Selection for Classification. Intelligent Data Analysis, 1997: p. 131C156.
[36] Devi, K.N., V.M. Bhaskaran and G.P. Kumar, Cuckoo ,Optimized SVM for Stock Market Prediction, IEEE Sponsored 2nd International Conference on Innovations in Information, Embedded, and Communication Systems (ICJJECS)2015, 2015.
[37] Dai, W., Y.E. Shao, and C. Lu, Incorporating feature selection method into support vector regression for stock index forecasting. Neural Computing and Applications, 2013. 23(6): p. 1551-1561.
[38] Yanan Mao, Zuoquan Zhang, Dingyuan Fan, “Hybrid feature selection based on improved genetic algorithm for stock prediction,” 2016 6th International Conference on Digital Home,2016.
[39] Dina A. Ragab, Maha Sharks and Omneya Attallah,” Breast Cancer Diagnosis Using an Efficient CAD System Based on Multiple Classifiers,” Diagnostics 2019, 9, 165; DOI:10.3390/diagnostics9040165.
[40] David P Salmon and Mark W. Bondi“Neuropsychological Assessment of Dementia" Access NIH public, PubMed central, US national library of medicine Institutes of Health, May 2010.
[41] Divya Jain, Vijendra Singh,” Feature selection and classification systems for chronic disease prediction: A review,” Egyptian Informatics Journal 19 (2018) 179–189,2018.
[42] R. Chaves, J. Ramírez, J.M., "Integrating discretization and association rule-based classification for Alzheimer’s disease diagnosis," science direct, Expert Systems with Applications, Volume 40, Issue 5, April 2013, Pages 1571-1578, https://doi.org/10.1016/j.eswa.2012.09.003.
[43] S. R. Bhagya Shree, Dr. H. S. Sheshadri, "An initial investigation in the diagnosis of Alzheimer’s disease using various classification techniques," 2014 IEEE International Conference on Computational Intelligence and Computing Research.