International Journal of Advanced Trends in Computer Science and Engineering

Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems

Sri Hari Nallamala[1], Dr. Pragnyaban Mishra[2], Dr. Suvarna Vani Koneru[3]

[1]Research Scholar, CSE Dept., K L E F (Deemed to be University), Vaddeswaram, Guntur, Andhra Pradesh, India, nallamala.srihari@gmail.com
[2]Associate Professor, CSE Dept., K L E F (Deemed to be University), Vaddeswaram, Guntur, Andhra Pradesh, India, pragnyaban@kluniversity.in
[3]Professor, CSE Dept., V.R. Siddhartha Engineering College, Vijayawada, Krishna Dist., Andhra Pradesh, India, suvarnavanik@gmail.com

ABSTRACT

The diagnosis of the Healthcare systems are playing prominent role with the advancement of latest technologies. Decision support systems may generate fruitful outcomes for better diagnosis for Breast cancer. The present context of the paper describes about BCD-NFIS that merely reduces usage of featured datasets using Fuzzy networks and produces enhanced accuracy of 98.24% much better results than the older approaches. The BCD-NFIS uses the methodology of Inference systems, Neural Fuzzy logic and BCD to overcome the problems. This would be very much helpful for rising physicians to simplify the diagnosis patterns through exploiting Information technology for Breast cancer.

Key words: Breast Cancer, Data Processing, Data Analysis, Fuzzy Logic, Neural Networks.

1. INTRODUCTION

1.1 The inference system for Neural Fuzzy (NFIS):

So far, the object's membership does not have specific criteria in this entire physical world. Ambiguity is a common issue regarding the object's membership. An object connected to many number of objects seems to be unpredictable easily to identify their membership relation, thus leading to ambiguity. Therefore, Professor L. Zadeh [1] noticed the ambiguity in the real numbers and introduced the concept of Fuzzy set which slowly gave path for the reduction in the ambiguity of real numbers in the year 1960. The fuzzy logic deals with the class of grades of object's membership. The ordinary way of dealing with issues that arise in precision paradigm rather than random variables is provided by this frame work. Later on, on the basis of fuzzy logic set theory Mamdani and Sugeno [2] designed Inference systems.

1.2 Fuzzy Sets:

Let us consider X as a space of points, consisting a generic element of X symbolized by x, thus X={x}. In x, the fuzzy set A is identified by a member ship function (M, F.) which is associated by each and every point in x separated by a real number within the time interval [0, 1]. The grade of membership of x in class A is represented by the value $f_A(x)$, 1 is for complete membership and 0 is for non-membership. The accomplishment of the membership of the features depends upon the grade value and is defined as the higher the grade of the features $f_A(x)$ resembles the nearest similarity.

The differential equations are involved in modeling the wide variety of systems. Modeling systems with most complex structures and unstable design may be a difficult task for the differential equations to execute the precision output. The use of Fuzzy inference systems (FIS) would be a probable solution for solving complete structures using various conditional statements like IF-THEN and IF-Else. The FIS applies all the statements to its logic transforming unstable to stable modeling. The FIS does not make use of Boolean logic to resolve complex structures. Though FIS is very consistent in solving complex problems, it falls behind with invariable inputs. On the contrary, new data sets and several parameters may be involved to model a structure where Neural Networks plays very helpful hand for all the parameters which keeps changing from time to time. Therefore, resolving the complex problems and changing parameters are aptly handled with the introduction and merging of both Neural Networks and FIS. The combination methodology reported favored results in exactly transforming the human ideologies in to specific rules to apprehend the acute structure and shaping the membership functions as to the expected output. In this context Takagi-Sugeno [2] studied the complete algorithms to shape the combinations of both Neural Networks and differential equations and thereafter the named it as ANFIS (Adaptive Neuro Fuzzy Inference Systems). R. Jang [4] bagged the pride in introducing the NFIS with FIS (Fuzzy Inference System) and Neural Networks in the year 1993.

Some of the previous chore on NFIS was accomplished by Ubelyi [3], who proposed combining Adaptive Neural Fuzzy inference system (NFIS) for breast cancer examination using

nine different characteristics. The results were very clear stating that NFIS was disciplined with a set of records. The system used Neural Network and extracted the capabilities for various calculations. Fuzzy modeling accounted much for this approach. Ubelys model on the other hand [NFIS] is a successful method resulted 88% precision in his approach. Using the Ubelys model, the BCD – NFIS approach has been introduced in the current context with fewer approaches of Data Mining techniques. The approach resulted best output with increase in the precision levels on all variants of data sets and features deployed for diagnosis of Breast Cancer.

2. NFIS STRUCTURE

The five layers constitute the layered structure of NFIS. The usage of the conditional statements in the NFIS is to overcome the drawback of complex computations. The NFIS takes all the inputs at a single stage and generates a single output. It uses the conditional statements like IF-THEN and IF-ELSE. For an instance, x and y are considered to be inputs to generate an output f. The below mentioned figure 1 gives the demonstration with rules.

*Rule*1: *IF x is A_1 and y is B_1, f=$p_1x+q_1y+r_1$*
*Rule*2: *IF x is A_2 and y is B_2, f=$p_2x+q_2y+r_2$*

In the figure 1, Rules 1 and 2 comprise of NFIS membership functions as {A1, B1} and {A2, B2} and parameters for the membership functions are {p1, q1, r1} and {p2, q2, r2}. The rules 1 and 2 are traced from the Takagi-Sugeno Fuzzy [2] Inference system (TSFIS). In this, the parameters are very adaptive in nature indicating that the nodes are also adaptive. This adaptability exists only in the foremost and fourth layer of the NFIS structure. On the contrary, third and fifth layers are very non-adaptive in nature, nodes are constant and fixed.

The below mentioned Figure 1(a) and 1(b) specifies the logical function and the functionality of each independent node in the specified layers [1].
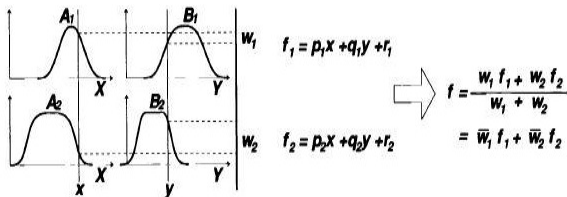

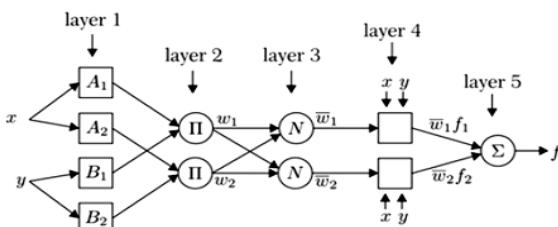**Figure 1. (a):** Fuzzy Reasoning

Figure 1. (b): Equivalent NFIS Structure

Layer 1: From the above Figure in layer 1 nodes are entitled by a square and they are denoted using letters. The membership$\mu_{A_i}(x)$ is derived from the input values of x and y from ANFIS.
$$O_i^1 = \mu_{A_i}(x) \tag{1}$$
It terms the membership function as O_i^1 for A_i and it also clearly identifies the degree for 'x' and resembles a closest similarity to the quantifier A_i.
$$\mu_{A_i}(x) = exp\left[-\left(\frac{x-c_i}{a_i}\right)^2\right] \tag{2}$$
Layer 2: For simplicity this layer 2 constitutes a circle and is rendered as a node.
The functionality of this node is to accept all the mere similarities obtained from the ancestral node 1. These similarities are categorized to obtain a specific weight.
$$\overline{w} = \mu_{A_i}(x) X \mu_{B_i}(x), i = 1,2 \tag{3}$$
The generated output constitutes the strength of each and every similarity and therefore the node holds "role of strength" for all similarities.
Layer 3: The Graphical representation of the third layer in the NFIS structure is a Circle. The notation of a node is 'N'. The layer computes the node's weight and termed as 'normalizing weight'.
$$w_i = \frac{w_i}{\sum_j w_j}, i = 1,2. \ j = 2 \tag{4}$$
Layer 4: The Graphical representation of the fourth layer in the NFIS structure is a square. Sugeno Inference Systems approach is considered in this layer to generate linear single output.
$$O_i^4 = \overline{w_i} f_i = \overline{w_i} = (p_i x + q_i y + r_i) \tag{5}$$
Where p_i, q_i are resultant parameters, r_i is the bias.
Layer 5: Justification process is amended for the applicability of all the rules from the ANFIS. The proposed approach results a single output from the proposed approach.
$$O_i^5 = final\ output = \sum_i \overline{w_i}\ f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{6}$$

3. BREAST CANCER DIAGNOSIS METHOD (BCD)

Prior and Post entropy techniques are aptly implemented in the BCD that enhances the quality factor of the attributes. This method also gave fruitful results in obtaining accurate text classification and categorization for ranking a simplest attribute.
$$H(x) = -\sum_x P(x) log_2 P(x) \tag{7}$$
The following equation specifies the entropy mechanism of a class used in the ANFIS. Assuming x is an attribute and C is class.
$$H(C/x) = -\sum_x P(x) \sum_C P(C|x) log_2 P(C|x) \tag{8}$$
The c in the below equation denotes the conditional entropy with x. The BCD method is observed as the difference of prior to the post entropy. The below equation defines the BCD approach.
$$H(c,x) = H(c) - H(C/x) \tag{9}$$
$$H(c,x) =$$
$$-\sum_c P(c) log_2 P(c) - \sum_x -(P(x) \sum_c P(C|x) log_2 P(C|x)) \tag{10}$$

4. THE BCD –NFIS APPROACH

The combination of both BCD method and NFIS method is a proposed approach for diagnosing Breast Cancer. The output of the BCD method is a set of features compounding more number of inputs for ANFIS. In order to train and test the proposed approach selected features are applied to ANFIS. IG, the Information Gain method is used for the process of BCD. The output of IG is further integrated with one or many inputs for NFIS to produce a single output called z. The combination mechanism of BCD and NFIS is carried out in experimental validations for precise understanding. The original features in data set are $X=\{x_1,x_2,..,x_n\}$ & $Y=\{y_1,y_2,..,y_k\}$ The features after applying the BCD, and the final output after applying on NFIS is denoted by 'Z'
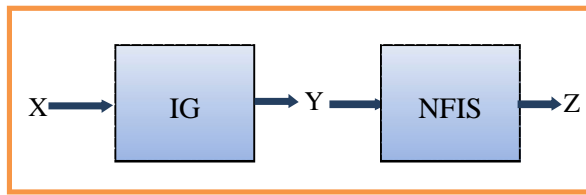
Figure 2: The general structure for the proposed approach

5. THE EXPERIMENTAL VALIDATIONS

William Wolberget.al created a Dataset, named "Wisconsis Breast cancer Dataset" (WBC) from the university of Wisconsin-Madison [6]. The complete set of database and its corresponding attributes are integrated from digital Fine needle aspirate (FNA) of breast mass. The data set also comprising of WBC with 699 records with every record consisting of a feature plus with the class attributes.

All the 341 records are implemented under training & simultaneously 342 records are used for testing. Out of which, 16 records contain missing values which are ignored and blind folded for no use. Under these circumstances the quality plays vital role. Hence, at such circumstances, to increase the quality standards for the attributes BCD method is aptly implemented. Further ranking is facilitated for attributes after applying the attribute evaluator Info Gain Attribute Val & the searching method Ranker –T-1 using WEKA on WBC dataset as shown in Table 1.

Table 1: BCD Ranking Using WEKA on WBC

Name of the Attribute	Rank
Cell Size Uniformity (UCSize)	0.636
Cell Shape Uniformity (UCShape)	0.633
Normal Nucleoli (NN)	0.555
Bare Nuclei (BN)	0.538
Single Epithelial Cell Size (SECS)	0.421
Clump Thickness (CT)	0.411
Marginal Adhesion (MA)	0.394
Bland Chromatin (BC)	0.316
Mitoses (MI)	0.278

The primary objective behind the proposed approach is to rank the designated features with certain numeric rankings. Total number of used features in the experiment account for good significance and productivity. The graphical representation revealed that first four features are exhibiting better outputs. Hence, the recommended features are gathered as input features for the NFIS structure. Cell size uniformity, Bare Nuclei, Cell shape Uniformity and Normal Nuclei are some of the features encountered in the experiment model for testing and training the experimental setup. Thus, figure.3 entails the reduction of usage of featured datasets from 0.7 to 0.25 on the y axis. Usage of more number of featured datasets increases complexity and less numbers of datasets usage increases robustness for Breast Cancer Diagnosis.
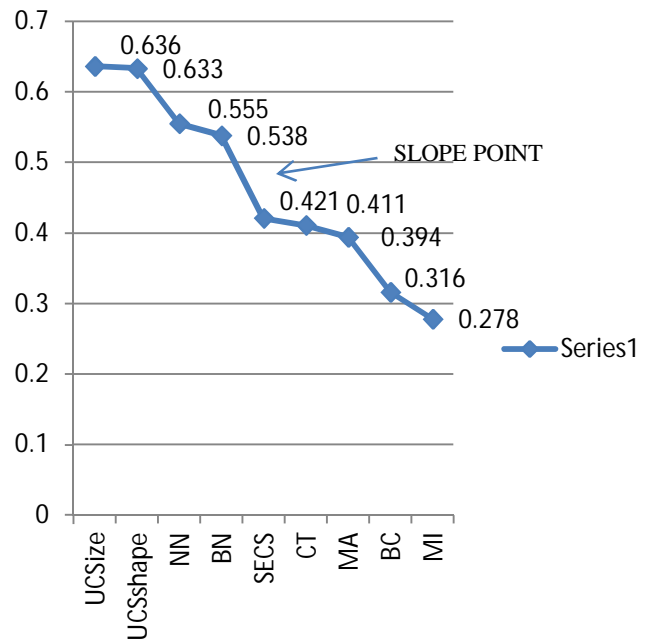
Figure 3: BCD Ranking on WBC

Initially, selected ranked features are accumulated as input process into ANFIS. Secondly, the features are projected to Fuzzy Inference System (FIS) by Sugeno and Mamdami [5]. On the contrary, Sugeno-Inference systems [2] are very effective and efficient that works better by using various adaptive and enhanced techniques with optimization methodologies.

The proposed approach called BCD-NFIS uses the technique to map all the featured features in to the member functions. The following entailed descriptions and Figure 4 clearly gives the awareness about the proposed approach and the method being involved for carrying out the experiment. The proposed approach clearly and precisely used Sugeno Fuzzy inference system for mapping all the featured features in to the member functions and thereafter member functions are termed as membership functions. Later, all the membership functions are then transformed in to several rules to produce a specific single output set. The produced output function is

again set to output member function. The figure 4 shows the membership function mapping input into membership values.

FIS not only has membership functions, but also consist of rules that enhance Machine intelligence by adding human reasoning abilities, The complex problems and the featured functions are being solved by using Boolean logic .The proposed approach comprise the rules that are taken from the real data. The purpose of these rules is to rank the features that are based completely by calculating the weight of these individual featured functions. The approach constitutes a total of 81 set of regulations that primarily deals with all the membership functions. The total 81 set regulations are derived with the calculations as total number of (81) rules=x^y where x represents the number of member functions existing and y denotes the total number of featured features i.e.3^4=81 rules).
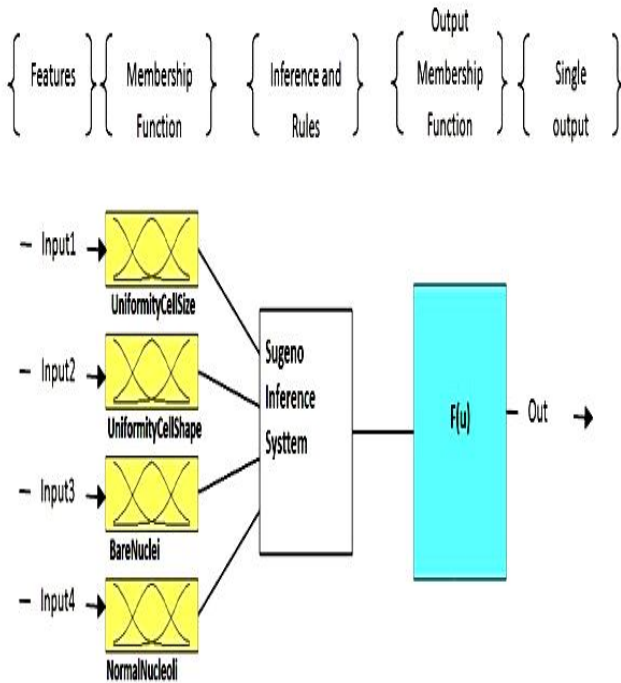

Figure 4: Sugeno Fuzzy Inference System generating single output with four inputs

For precision and clarity in the approach, two examples of among 81 rules are summed below:

IF AND (*Normal Nucleoli (NC) and Bare Nuclei (BN) is substantially poor, Cell Shape Uniformity (CSU) is Avg and Cell Size Uniformity is again poor,*) THEN (the predicted single output is as per the prediction).

IF AND (*Bare Nuclei (BN) and Cell Size Uniformity is substantially poor, Cell Shape Uniformity is gradually high and Normal Nucleoli is Average*) THEN (the predictable single output is not as per the actual prediction).

Cell size Uniformity play vital rule and application of rule is implemented widely which is represented visually as shown in Figure 4.

The final stage at the third step is that NFIS is loaded to the constructed Fuzzy Inference System with the new featured set to train and test the proposed approach as shown in Figure 5.
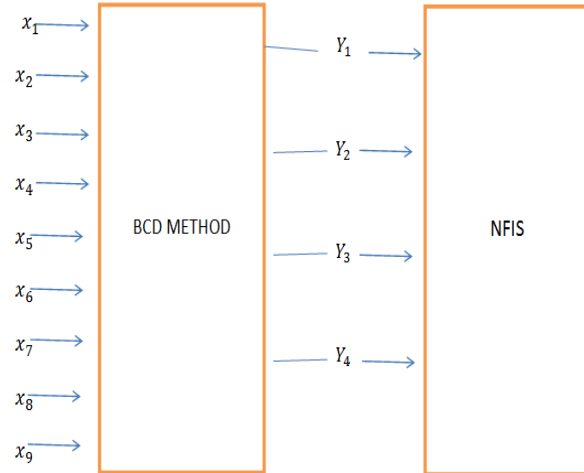

Figure 5: BCD-NFIS Structure

The MAT LAB is chosen to represent the complete structure of NFIS as detailed precisely in Figure 6.

The BCD Method was used for the result set on all variants of featured features. The BCD Method after implementing featured features towards a dataset of WBC, it has eminently generated 98.24% precision levels which is a remarkable and tremendous result set.

The Comparative Analysis on BCD-NFIS with previous methods as shown in Table 2.

Table 2: Comparative Analysis on BCD-NFIS with previous methods

The approach	Accuracy
AdaBoost	57.60%
NFIS	59.90%
SANFIS	96.07%
FUZZY	96.71%
FUZZY-GENETIC	97.07%
ILFN	97.23%
NNs	97.95%
ILFN&FUZZY	98.13%
BCD-NFIS (Current Method)	98.24%

The complete results of the proposed approach had a constructive comparison to the previous work with same dataset which is clearly represented in the Figure 7.
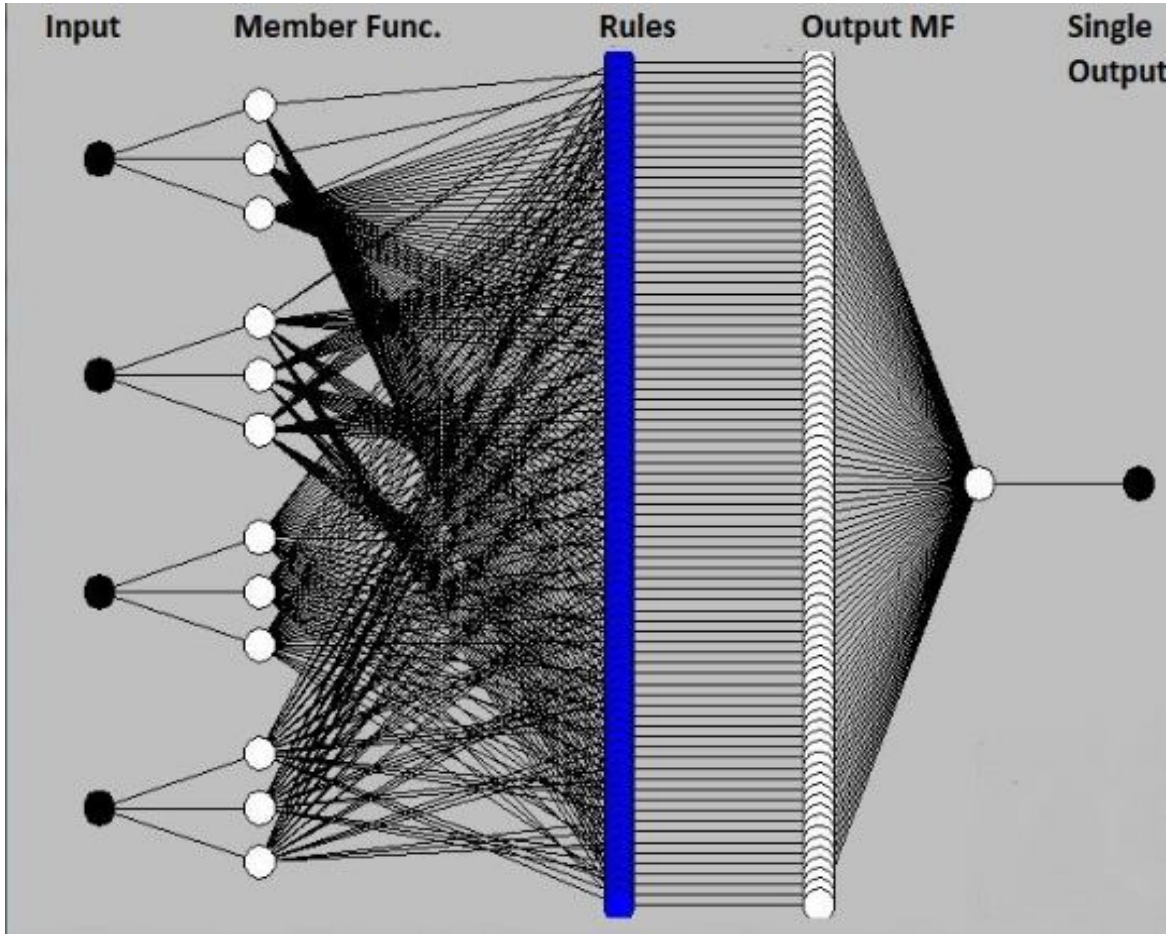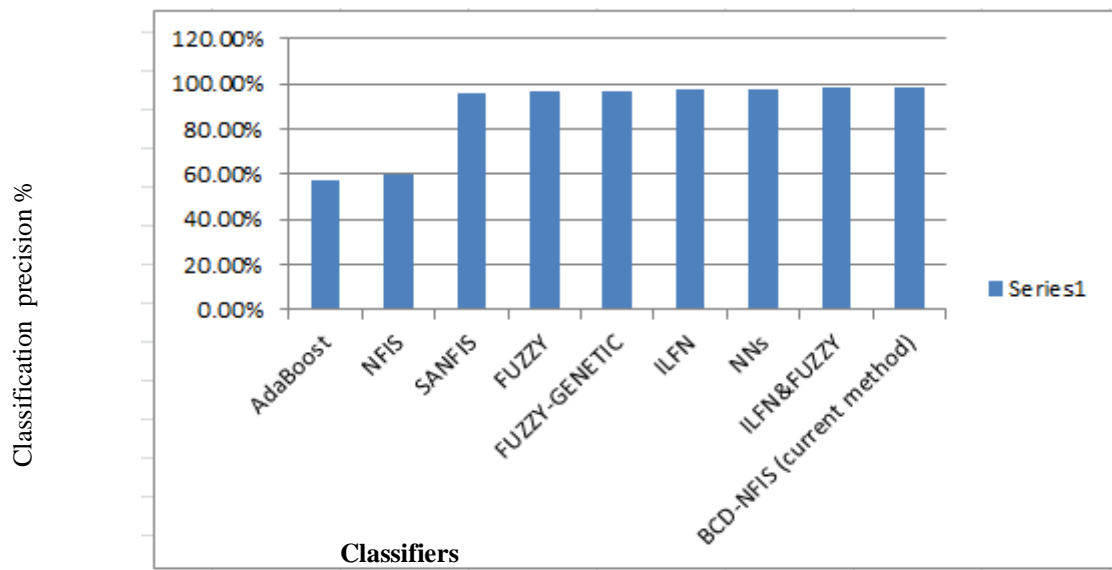
Figure 6: Generation of output using MATLAB

Figure 7: Precision levels with Graphical representation between BCD-NFIS and previous approaches.

6. CONCLUSION

The proposed concept discussed about the diagnosis of Breast cancer by using effective methodology called BCD-NFIS by reducing complexities existing in Sugeno and Mamdani's Fuzzy Inference systems. The usage of featured features is minimized to large extent in the BCD Method. The approach also used enormous data sets and applied all the data sets to NFIS. The implementation result sets in the proposed methodology revealed that the precision range is up to 98.24% compared to the previous methodologies. The competitive and challenging approach may be very useful to information technologies in future days ahead for simplifying the diagnosis process for fast growing problems in Breast cancer patients round the clock.

REFERENCES

1. L. Zadeh, **"Fuzzy Sets"**, Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers, 1996.
   https://doi.org/10.1142/9789814261302_0001
2. T. TAKAGI and M. SUGENO, **"Fuzzy identification of systems and its applications to modeling and control"**, IEEE transactions on systems, man, and cybernetics, vol. 15, no. 1, pp. 116–132, 1985.
   https://doi.org/10.1109/TSMC.1985.6313399
3. Ubeyli ED, E.D., **"Adaptive Neuro-Fuzzy Inference Systems for Automatic Detection of Breast Cancer"**, J. Med. Syst., 2009. 33(5): p. 353-358
   https://doi.org/10.1007/s10916-008-9197-x
4. R. Jang, **"ANFIS: Adaptive-Network-Based Fuzzy Inference System"**, IEEE Transactions on systems, man, and cybernetics, vol. 23, No. 3, May/June 1993.
   https://doi.org/10.1109/21.256541
5. Saeys, Y., I. Inza, and P. Larrañaga, **"A review of feature selection techniques in bioinformatics"**, Bioinformatics, 2007. 23(19): p. 2507-2517.
   https://doi.org/10.1093/bioinformatics/btm344
6. UCI Machine Learning Repository
   [Sighted 2010; Available from:
   http://archive.ics.uci.edu/ml/about.html]