Online ISSN : 0975-4172
Print ISSN : 0975-4350

Global Journal of Computer Science and Technology: D

Neural & AI

Highlights

Adaptive Genetic Algorithm

Secret Key Generation Approach

Reasoning for Hematology

Hematology Malignancies Classification

Discovering Thoughts, Inventing Future

VOLUME 15 ISSUE 1 VERSION 1.0

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence

Volume 15 Issue 1 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

*(A Delaware USA Incorporation with "Good Standing"; **Reg. Number: 0423089**)*

Sponsors: Open Association of Research Society
Open Scientific Standards

Publisher's Headquarters office

Global Journals Headquarters
301st Edgewater Place Suite, 100 Edgewater Dr.-Pl,
Wakefield MASSACHUSETTS, Pin: 01880,
United States of America
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Journals Incorporated
2nd, Lansdowne, Lansdowne Rd., Croydon-Surrey,
Pin: CR9 2ER, United Kingdom

Packaging & Continental Dispatching

Global Journals
E-3130 Sudama Nagar, Near Gopur Square,
Indore, M.P., Pin:452009, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investors@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Contents of the Issue

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence
Volume 15 Issue 1 Version 1.0 Year 2015
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Enhance the Performance of Chaotic Generator in the Filed of Cryptography: A Secret Key Generation Approach

By Sona Mishra, Richa Shrivastava & Abhinav Tiwari
University of R.G.P.V., India

Abstract- The main focus of this research paper is to propose and improvement of the data security using encryption and decryption method in ANN based chaotic generator of original value. The Binary value sequence of ASCII CODE is converted with two initial parameter, and converted value is again decrypted with same initial parameter. In which consists of Binary value of ASCII Code, chaotic neural network algorithm was used for encryption and decryption and it generates the chaotic sequence of random value for each A to Z letter. The generated random value is the encrypted binary ASCII values of A to Z sequence of original ASCII Code binary value, with same initial parameter. For simulation MATLAB software is used. This paper also includes improved experimental results and complete demonstration that ANN Based Chaotic Generator is successfully perform the cryptography.

Keywords: ann based chaotic generator, chaotic neural network, cryptography.

GJCST-D Classification : D.4.6

ENHANCETHEPERFORMANCEOFCHAOTICGENERATORINTHEFILEDOFCRYPTOGRAPHYASECRETKEYGENERATIONAPPROAH

Strictly as per the compliance and regulations of:

Enhance the Performance of Chaotic Generator in the Filed of Cryptography: A Secret Key Generation Approach

Sona Mishra [α], Richa Shrivastava [σ] & Abhinav Tiwari [ρ]

Abstract- The main focus of this research paper is to propose and improvement of the data security using encryption and decryption method in ANN based chaotic generator of original value. The Binary value sequence of ASCII CODE is converted with two initial parameter, and converted value is again decrypted with same initial parameter. In which consists of Binary value of ASCII Code, chaotic neural network algorithm was used for encryption and decryption and it generates the chaotic sequence of random value for each A to Z letter. The generated random value is the encrypted binary ASCII values of A to Z sequence of original ASCII Code binary value, with same initial parameter. For simulation MATLAB software is used. This paper also includes improved experimental results and complete demonstration that ANN Based Chaotic Generator is successfully perform the cryptography.

Keywords: ann based chaotic generator, chaotic neural network, cryptography.

I. Introduction

Cryptography is the exchange of information among the users without leakage of information to others. Many public key cryptography are available which are based on number theory but it has the drawback of requirement of large computational power, complexity and time consumption during generation of key [1].

Cryptosystems are commonly used for protecting the integrity, confidentiality, and authenticity of information resources. In addition to meeting standard specifications relating to encryption and decryption, such systems must meet increasingly stringent specifications concerning information security. This is mostly due to the steady demand to protect data and resources from disclosure, to guarantee the authenticity of data, and to protect systems from web based attacks. For these reasons, the development and evaluation of cryptographic algorithms is a challenging task [2].

This paper is study and performance of ANN Based Chaotic Generator in the filed of Cryptography. The rest of the paper is organized as follows: section 2 discusses background and related work in the field of chaotic neural network based cryptography, section 3 discusses implementation section 4 discusses experimental report and test result and finally section 5 discusses conclusion.

II. Background and Related Work

Ilker DALKIRAN, Kenan DANIS¸MAN introduced a research paper on Artificial neural network based chaotic generator for cryptology. In this paper, to overcome disadvantages of chaotic systems, the dynamics of Chua's circuit namely x, y and z were modeled using Arficial Neural Network (ANN). ANNs have some distinctive capabilities like learning from experiences, generalizing from a few data and nonlinear relationship between inputs and outputs. The proposed ANN was trained in diffrent structures using different learning algorithms. To train the ANN, 24 different sets including the initial conditions of Chua's circuit were used and each set consisted of about 1800 input-output data. The experimental results showed that a feed-forward Multi Layer Perceptron (MLP), trained with Bayesian Regulation back propagation algorithm, was found as the suitable network structure. As a case study, a message wasfirst encrypted and then decrypted by the chaotic dynamics obtained from the proposed ANN and a comparison was made between the proposed ANN and the numerical solution of Chua's circuit about encrypted and decrypted messages [3].

Jason L. Wright , Milos Manic Proposed a research paper on Neural Network Approach to Locating Cryptography in Object Code. In this paper, artificial neural networks are used to classify functional blocks from a disassembled program as being either cryptography related or not. The resulting system, referred to as NNLC (Neural Net for Locating Cryptography) is presented and results of applying this system to various libraries are described.[4].

Eva Volna, Martin Kotyrba, Vaclav Kocian,Michal Janosek developed a CRYPTOGRAPHY BASED ON NEURAL NETWORK. This paper deals with using neural network in cryptography, e.g. designing such neural network that would be practically used in the area of cryptography. This paper also includes an experimental demonstration [5].

T. SCHMIDT, H. RAHNAMA developed A REVIEW OF APPLICATIONS OF ARTIFICIAL NEURAL

Author α σ ρ : VITS Jabalpur, University of R.G.P.V. Bhopal M.P.
e-mails: sona.mishra2909@gmail.com, amit_2440@yahoo.co.in

NETWORKS IN CRYPTOSYSTEMS. This paper presents a review of the literature on the use of artificial neutral networks in cryptography. Different neural network based approaches have been categorized based on their applications to different components of cryptosystems such as secret key protocols, visual cryptography, design of random generators, digital watermarking, and steganalysis[2].

KARAM M. Z. OTHMAN , MOHAMMED H. AL JAMMAS introduced IMPLEMENTATION OF NEURAL - CRYPTOGRAPHIC SYSTEM USING FPGA. In this work, a Pseudo Random Number Generator (PRNG) based on artificial Neural Networks (ANN) has been designed. This PRNG has been used to design stream cipher system with high statistical randomness properties of its key sequence using ANN. Software simulation has been build using MATLAB to firstly, ensure passing four well-known statistical tests that guaranteed randomness characteristics [6].

An Empirical Investigation of Using ANN Based N-StateSequential Machine and Chaotic Neural Network in the Field of Cryptography. In this work using two artificial neural networks in the field of cryptography. First One is ANN based n-state sequential machine and Other One is chaotic neural network[10].

a) Problem Definition

In now a day's secret key generation in the field of cryptography continues to be an active research field, as shown by the large number of papers being published. In this previous published research work now we are study their performance and various different way to perform cryptography [9]. We considered chaos based cryptography [3] [10]. but I found that the adopted approaches of perform cryptography and generating secret key is very complex , time consuming and providing limited security during encryption in transmission end or decryption in receiving end. A Problem is arises Unauthorized user or hackers easily guess or predict secret key and view our confidential data and damage or modify it.

b) Proposed Work

In Cryptography, secret key generation scheme was proposed by ANN based Chaotic Generator. ANN based chaotic Generator system used chaotic neural network scheme for encryption and decryption [7]. In this paper ANN based chaotic generator is proposed for data encryption and decryption, it produces the outputs according to initial conditions and control parameter .We improve the level of performance of chaos based cryptography [10] using binary value of ASCII Code of A to Z letter instead of decimal value. A plain-text was encrypted and then obtained cipher text was decrypted by using the chaotic dynamics (control parameter and initial point), initial condition and control parameter act as a secret key in the field of cryptography. It is accepted that the initial conditions which were used in the training phase of the ANN model and the system parameters are known by both the transmitter and the receiver.

We adopted ANN based chaotic generator approach from et.al. [3] and increase the level of security from et. al. [10] and demonstrate by experimental result.

III. Implementation

a) Secret Key Cryptography Through chaotic neural network

A network is called chaotic neural network if its weights and biases are determined by chaotic sequence. In this section we use a algorithm for performing encryption and decryption using chaotic neural network.

ANN based chaotic generator Using CNN scheme for encryption and decryption.

Step 1. The chaotic Logistic map.

$$\mu x(i-1)(1-x(i-1))$$

Set the value of M.

Step 2. The secret key is the control parameter μ and the initial point x (0) of the Logistic map, which are all L-bit binary decimals. Determine parameter μ and initial point x (0).

Step 3. The initialization procedure:

Generate the chaotic sequence x(1), x(2), x(3)…. .. x(M) by the formula x(n + 1) = μx(n) (1 – x(n)) and create b(0), b(l), ..., b(8M – 1) from x(l), x(2),..., x(M) by the generating scheme that 0.b(8m – 8)b(8m – 7)….. b (8m – 2) b(8m – l) … is the binary representation of x(m) for m = 1, 2,. . . .,M.

Step 4. The encryption procedure:

Depending upon the chaotic sequence a weight matrix and a bias matrix is obtained and the net input is obtained. Then a hard limiter is applied as a transfer function in order to obtain the digital encrypted data. For decryption the same network is used and the same initial value is used to generate the chaotic sequence and for decrypting the data successfully.

For n=0 to M-1

$$g(n) = \sum_{i=0}^{7} d_i\ 2^i \sum_{i=0}^{7} d_i\ 2^i$$

For i=0 to 7

$$w_{ji} w_{ji} = \begin{cases} 1 & j=i, b(8n+i)=0 \\ -1 & j=i, b(8n+i)=1 \\ 0 & j \neq i \end{cases} \begin{cases} 1 & j=i, b(8n+i)=0 \\ -1 & j=i, b(8n+i)=1 \\ 0 & j \neq i \end{cases}$$

Global Journal of Computer Science and Technology (D) Volume XV Issue I Version I Year 2015

$_J \in \{ 0,1,2,,4,5,6,7 \} \in \{ 0,1,2,,4,5,6,7 \}$

$$\theta_i \theta_{i=} \begin{cases} -\frac{1}{2}\, b(8n+i) = 0 \\ \frac{1}{2}\, b(8n+i) = 1 \end{cases} \begin{cases} -\frac{1}{2}\, b(8n+i) = 0 \\ \frac{1}{2}\, b(8n+i) = 1 \end{cases}$$

End

For i=0 to 7

$$d_i' d_i' = f\ \left(\sum_{j=0}^{7} w_{ji}\, d_i + \theta_i\right)\left(\sum_{j=0}^{7} w_{ji}\, d_i + \theta_i\right)$$

Where f(x) is 1 if x>=0

End

Where f(x) is 1 if x>=0

End

$$g\,'(n) = \sum_{i=0}^{7} d_i'\, 2^i \sum_{i=0}^{7} d_i'\, 2^i$$

End

g = digital signal of length M and g (n) 0 ≤ M-1 , be the one- byte value of the signal g at position n.

Step 5. The decryption procedure

The decryption procedure is the same as the above one except that the input signal to the decryption Chaotic neural network should be g'(n) and its output signal should be g"(n).

b) ANN based chaotic Generator

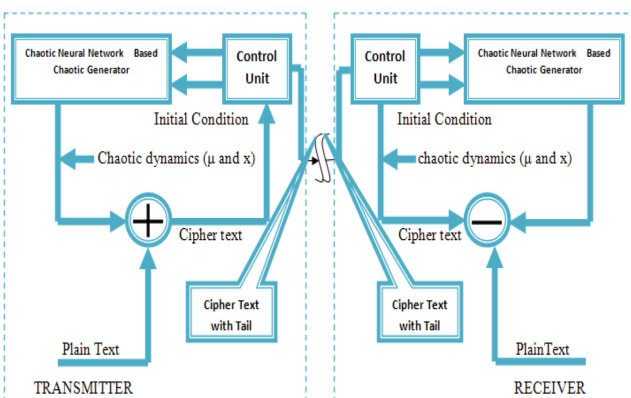


Figure : ANN based chaotic Generator

Working of Encryption and Decryption Using ANN Based Chaotic Generator

1. In the transmitter, software based control unit decides the values of initial conditions and the ANN based chaotic generator produces the outputs according to those initial conditions.
2. In our experiment ASCII Char A to Z Binary value was chosen as a plain-text.
3. The cipher-text is obtained by adding normalized data to the chaotic dynamic (μ and x).
4. A tail which consists of initial condition, and normalization parameters, is tied to the cipher-text by the control unit and then the cipher-text with tail is transmitted.
5. In the decryption phase, the tail is firstly discriminated from the cipher-text and the initial conditions are extracted from the tail by the control unit.
6. Then, the initial conditions are applied to the ANN based chaotic generator and the chaotic dynamic (μ and x) is produced by the generator, initial condition x and control parameter is act as secret key.
7. Finally the plain-text is obtained by subtracting the chaotic dynamic (μ and x) from the cipher-text. The size of cipher-text is equal to the size of plain-text. But tying the tail to the cipher-text enlarges the size of encrypted data. The size of tail is only 8 bytes

IV. Experiment and Test Result

a) Secret Key Cryptography by chaotic neural network

A chaotic network is a neural network whose weights depend on a chaotic sequence. The chaotic sequence highly depends upon the initial conditions and the parameters, x (0) and μ are set. It is very difficult to decrypt an encrypted data correctly by making an exhaustive search without knowing x (0) and μ ().

b) ANN based chaotic Generator

Example- Binary value of ASCII CODE 'A' Encryted with deiffrent Initial Conditions (Values of x (0) and μ ()).

			x(0)= 0.75 μ=3.9		**x(0)= 0.85 μ=3.5**		**x(0)= 0.90 μ=3.2**	
	DEC	**Bin**	**DEC**	**Bin**	**DEC**	**Bin**	**DEC**	**Bin**
A =	97	01100001	199	11000111	233	11011111204	204	11001100

Global Journal of Computer Science and Technology (D) Volume XV Issue I Version I Year 2015

Decryption with same Initial conditions as below --

x(0)= 0.75 μ=3.9

		DEC	Bin	DEC	Bin
A	=	199	11000111	97	01100001

x(0)= 0.85 μ=3.5

		DEC	Bin	DEC	Bin
A	=	233	11011111	97	01100001

x(0)= 0.90 μ=3.2

		DEC	Bin	DEC	Bin
A	=	204	11001100	97	01100001

It is clear from Example 01. that the binary value sequence of ASCII CODE A is encrypted and decrypted correctly by knowing the exact values of x (0) and μ otherwise we get the wrong value sequences and also cleared that the Binary values of ASCII Code is more strong as compared to decimal values of ASCII Code , no longer pridiction is possible of binanry values. Initial point x (0) and control parameter μ is act as a secret key.

V. Conclusion

It is clear that the binary value sequence of ASCII CODE is encrypted and decrypted correctly by knowing the exact values of x (0) and μ otherwise we get the wrong value sequences .And also clear that the binanry value is more strong enough as compre to decimal values. In this paper we successfully perform encryption and decryption with the help of Chaotic neural network and improve the level of security with the help of using binanry value of ASCII Code instead of decimal values . Network was trained with the help of back propagation algorithm in neural network. Above experiment clear that the Binary value of ASCII CODE is encrypted and its decrypted with same value of parameter , encrypted value is decrypt only correctly by knowing the exact values of x (0) and μ otherwise we get the wrong generated value sequences . ANN based Chaotic generator provide high range of security in the field of cryptography.

References Références Referencias

1. Shweta B. Suryawanshi and Devesh D. Nawgaje- a triple-key chaotic neural network for cryptography in image processing International Journal of Engineering Sciences & Emerging Technologies, April 2012. ISSN: 2231 – 6604 Volume 2, Issue 1, pp: 46-50 ©IJESET .
2. T. SCHMIDT, dept. of computer science, ryerson university, canada - a review of applications of artificial neural networks in cryptosystems.
3. Ilker DALKIRAN, Kenan DANIS¸MAN - Artificial neural network based chaotic generator for cryptology ,Turk J Elec Eng & Comp Sci, Vol.18, No.2, 2010, © T¨UB˙ITAK
4. Jason L. Wright , Milos Manic - Neural Network Approach to Locating Cryptography in Object Code. Emerging Technologies and Factory Automation INL Laboratory.
5. Eva Volna ,Martin Kotyrba ,Vaclav Kocian,Michal Janosek - CRYPTOGRAPHY BASED ON NEURAL NETWORK , Department of Informatics and Computers University of Ostrava Dvorakova 7, Ostrava, 702 00, Czech Republic.
6. KARAM M. Z. OTHMAN , MOHAMMED H. AL JAMMAS - IMPLEMENTATION OF NEURAL - CRYPTOGRAPHIC SYSTEM USING FPGA . journal of Engineering Science and Technology Vol. 6, No. 4 (2011) 411 – 428 © School of Engineering, Taylor's University
7. Harpreet Kaur , Tripatjot Singh Panag CRYPTOGRAPHY USING CHAOTIC NEURAL NETWORK International Journal of Information Technology and Knowledge Management July-December 2011, Volume 4, No. 2, pp. 417-422
8. E.C. Laskari , G.C. Meletiou, D.K. Tasoulis , M.N. Vrahatis , Studying the performance of artificial neural networks on problems related to cryptography , Nonlinear Analysis: Real World Applications 7 (2006) 937 – 942
9. Design and Realization of A New Chaotic Neural Encryption/Decryption Network" by Scott Su, Alvin Lin, and Jui-Cheng Yen.
10. Nitin Shukla, Abhinav Tiwari, —An Empirical Investigation of UsingANN Based N-State Sequential Machine and Chaotic Neural Network in the Field of Cryptography‖, Global Journal of Computer Science and Technology Neural & Artificial Intelligence, Vol. 12, Issue.10,No. 1,17-26, 2012.

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence
Volume 15 Issue 1 Version 1.0 Year 2015
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Fuzzy Inspired Case based Reasoning for Hematology Malignancies Classification

By Imianvan A.A. & Odigie B.B.
University of Benin, Nigeria

Abstract- Conventional approaches for collecting and reporting hematological data as well as diagnosing hematologic malignancies such as leukemia, anemia, e.t.c are based on subjective professional physician personal opinions or experiences which are influenced by human error, dependent on human-to-human judgments, time consuming processes and the blood results are non-reproducible. In the light of those human limitations identified, an automatic or semi-automatic classification and corrective method is required because it reduces the load on human observers and accuracy is not affected due to fatigue. Case-Based Reasoning (CBR) as a multi-disciplinary subject that focuses on the reuse of past experiences or cases to proffer solution to new cases was adopted and combined with the power of Fuzzy logic to design a software model that will effectively mine hematology data. This study aim at helping the medical practitioners to diagnose and provide corrective treatment to both normal patients and patients with hematology disorder at the early stage which can reduce the number of deaths. This aim is achievable by developing an intelligent expert system based on fuzzy logic and case-based reasoning for classification of hematology malignancy.

Keywords: *fuzzy logic, case base reasoning, hematology.*

GJCST-D Classification : *I.5.1 I.2.3*

FUZZYINSPIREDCASEBASEDREASONINGFORHEMATOLOGYMALIGNANCIESCLASSIFICATION

Strictly as per the compliance and regulations of:

Fuzzy Inspired Case based Reasoning for Hematology Malignancies Classification

Imianvan A.A. [α] & Odigie B.B. [σ]

Abstract- Conventional approaches for collecting and reporting hematological data as well as diagnosing hematologic malignancies such as leukemia, anemia, e.t.c are based on subjective professional physician personal opinions or experiences which are influenced by human error, dependent on human-to-human judgments, time consuming processes and the blood results are non-reproducible. In the light of those human limitations identified, an automatic or semi-automatic classification and corrective method is required because it reduces the load on human observers and accuracy is not affected due to fatigue. Case-Based Reasoning (CBR) as a multi-disciplinary subject that focuses on the reuse of past experiences or cases to proffer solution to new cases was adopted and combined with the power of Fuzzy logic to design a software model that will effectively mine hematology data. This study aim at helping the medical practitioners to diagnose and provide corrective treatment to both normal patients and patients with hematology disorder at the early stage which can reduce the number of deaths. This aim is achievable by developing an intelligent expert system based on fuzzy logic and case-based reasoning for classification of hematology malignancy.

Keywords: fuzzy logic, case base reasoning, hematology.

I. Introduction

The most commonly ordered or frequently requested test in medicine is hematology tests otherwise known as the Full Blood Picture (FBP). Larry et al (2012) indicated that the initial Hemogram results are Red Blood Cell (RBC) count, Hemogbobin (HB) concentration, Packed Cell Volume (PCV), Hematocrit, and total White Blood Cell count (WBC), to which platelet count, Lymprocytes and Neutrophit were added as the necessary vital dye stains and the resolution were substantially improved. Hematological or blood data are tightly regulated trails with high clinical relevance. They provide clinical indicators of health and diseases and they are affected by a number of factors even in apparently healthy patients (Kelada et al, 2012;). These factors include age sex, ethnic background, body build and social nutritional and environmental factors especially altitudes and genetic factors.

Several studies have shown that some of these hematological parameters exhibit considerably variations at different periods in life. Higher at birth than at any other period of life, the levels of these parameters decreases as human cells became hypochronic with the development of physiologic iron deficiency hematologic malignancies such as Anaemia. This explains that values outside the normal range are diagnostic for health disorders including cancer, immune diseases and other infections (Kelada et al, 2012). These parameters in turn have to be comprehended in context with vital signs and symptoms.

Conventional approaches for collecting and reporting hematological data as well as diagnosing and classifying hematologic malignancies using techniques such as gene expression signatures are based on subjective professional physicians opinions or experiences which are influenced by human error, dependent on human-to-human judgments, time consuming process and the blood results are non reproducible (Armstrong et al ,2002; Leven et al 2003; Pascal et al, 2008; Minakshi and Sourabh, 2013). So, an automatic or semi-automatic classification method is required because it reduces the load on human observer and accuracy is not affected due to fatigue. Based on this evidence, the application of soft computing techniques such as fuzzy logy and case base reasoning have continued to expand the horizons of intelligent system development in the medical domain.

Hullermeyer (2007) opted that Case based reasoning (CBR) methodology presents a foundation for new technologies for building intelligent computer aided diagnoses systems. These Technologies directly address the problems found in the traditional Artificial Intelligence (AI) techniques, e.g. the problems of knowledge acquisition, remembering, robust and maintenance. CBR solves a new problem by retrieving and adapting solutions or parts of solutions of a previously solved problem. The idea of CBR has strong appeal because it is recognized that much of human expertise in experimented database and CBR is considered to capture this idea. The main concept that characterized CBR is that expertise reports in a particular field is collected as a repository of cases, and each experience and solution or outcome is confined and archived in the case base for reuse and future reference (Hoda, 2008).

In another improvement, Livi et al., (2015) maintained that Zadeh (1988) has pioneered the concept of information granularity (IG) which has been

Author α : Department of Computer Science, Faculty of Physical Sciences, University of Benin, Benin City.
e-mail: ditsciences@yahoo.com
Author σ : Management Information System (MIS) Department, Auchi Polytechnic, Auchi. e-mail: tonyvanni@yahoo.com

conceived with the aim of developing the so-called human-centered computation. Pedrycz (2013) upheld that human perceive and process complex information by organizing existing knowledge along with available experimental evidence and structuring them in a form of some meaningful, semantically sound entities, which are central to all ensuring supporting decision-making activities. Zadeh has expanded IG to include fuzzy logic. Fuzzy Logic (FL) helps computers to paint vivid pictures of the uncertain world by representing and manipulating data that are not precise but rather fuzzy. Fuzzy logic provides an inference methodology that helps appropriate human reasoning capabilities to be applied to knowledge-based system (Obi and imianvan, 2014).

In this research using a CBR system will assist the physician to quickly diagnose the problem, provide pre-hospitalization aid to the patient and hence reduce the risk of death due to blood parameter in-balance, while fuzzy logic provides the conceptual framework for systematically analysis this un-précised and complex data of hematology parameter. It can recognize the types of blood parameter deficiency based on the evidence indentified from the blood data attributed, symptoms and their severity. The result and decision will be a suggestion that the human expert may accept or consider for further analysis, or completely reject it. We intend to implement this expert system using Matrix Laboratory (MATLAB) Software.

II. Review of Related Works

Artificial intelligent prediction is based on human-like learning ability in pattern recognition and generalization better known as machine learning and many machine learning algorithms have been designed and used to analyze medical datasets and proved to be indispensable tools for intelligent data analysis (Rosma, 2009; Garibaldi and Ifeachor, 2000; Butcher, 2004; Mendonca, 2004; Liu et al., 2005). Obi and Imianvan (2014) presented an interactive neural fuzzy expert system for leukemia diagnosis. They combine the strength of soft computing techniques of Neural network and fuzzy system to model a medical procedures for indentifying one of the four types of blood malignancies-leukemia based on available signs and symptoms. Bendi et al., (2011) performed a critical study of algorithms for Automatic classification used in medical fields and patients with liver diseases were analyzed. KNN, Back propagation and Support Vector Machine (SVM) are giving better results with all these features set combinations. Shahina et al., (2009) designed a computer-aided decision support system for analyzing and diagnosing stress-related disorders based upon finger temperature signals. The method of case-based reasoning is employed to make recommendations for stress diagnosis by retrieving and comparing with previous similar cases in terms of features extracted. Moreover, fuzzy techniques were incorporated to better accommodate uncertainty in clinicians reasoning as well as imprecision in case indexes. Hoda(2008) proposed a fuzzy case based reasoning for poison classification. The solution presented utilizes computer science field of artificial intelligence and it was realized through the Fuzzy Logic techniques and case based reasoning. The system developed was a complete stand-alone entity that can identify the types of poisoning and its' percentage. It still utilizes the human expert's intelligence. The identification it supplies is a suggestion that the human expert may accept or consider for further analysis, or completely reject. Guessoum et al. (2012) presented the combination of CBR and RBR in order to gather their powers within the same system. RBR has proved its performance in modeling of reasoning which can be explain by humans, that is why we adopted it to conceive the reuse phase of CBR process by an expert system. The inference engine associated to knowledge base and forwarding chaining ensure adaptation task by drawing inferences starting from the diagnosis found in the retrieval phase and basing on some attributes having highest weights whose values will compose the set of facts of our expert system. They also propose some heuristics traced for estimating the similarity on missing data and symbolic descriptors in the retrieval phase. Results show also that these heuristics functions are benefit for the system because they optimize the result of the retrieval phase. Nilsson et al. (2003) addressed the domain of psychophysiological dysfunctions, a form of stress. The system is classifying physiological measurements from sensors. The system is divided into smaller distinct parts. Measurements, like signals from an ECG, are filtered and improved. A case library of models of distortions etc. is applied to the filters. Features are extracted from the filtered signals (measurements). An additional set of features are extracted from the first set, for trend analysis etc. The features from the first and second set, and patient specific data, are used as a case. The cases are classified with a k-nearest neighbour match. Schmidt and Gierl (2002) proposed a prognostic model to forecast waves of influenza epidemics, based on earlier observations done in previous years. TeCoMED combines CBR with Temporal Abstraction to handle the problem of the cyclic but irregular behaviour of epidemics.

III. Methodology

The traditional process for the medical diagnosis of hematology malignancies or diseases such as Anemia, Leukemia, e.t.c starts when an individual consults a physician (doctor) and presents a set of complaints (symptoms). The physician then requests further information from the patient or from others close to him who knows about the patient's medical history in severe cases. Medical laboratory results such as the full blood count which comprises of the White Blood Cell, Red blood Cell, Packet Cell Count (PCV), Hemoglobin count, Hematocrit count e.t.c is requested and compared with values shown in table 1. Other data collected include patient's previous state of health, living condition and other medical conditions. A physical examination of the patient condition is conducted and in most cases, a medical observation along with other medical test(s) is carried out on the patient prior to medical treatment. From the symptoms presented by the patient, the physician narrows down the possibilities of the illness that corresponds to the apparent symptoms and make a list of the conditions that could account for what is wrong with the patient. These are usually ranked in the order (Low, Moderate and high). The physician then conducts a physical examination of the patient, studies his or her medical records and ask further questions, as he goes in an effort to rule out as many of the potential conditions as possible. When the list has been narrowed down to a single condition, it is called differential diagnosis and this provides the basis for a hypothesis of what is ailing the patient. Until the physician is certain of the condition present; further medical test are performed or schedule such as medical imaging, scan, X-rays in part to conform or disprove the diagnosis or to update the patient medical history. Other Physicians, specialist and expert in the field may be consulted (sought) for further advices. Despite all these complexities, most patient consultations are relatively brief because many diseases are obvious or the physician's experience may enable him to recognize the condition quickly. Upon the Completion of the diagnosis by the physician, a treatment plan is proposed, which includes therapy and follow-up (further meeting and test to monitor the ailment and progress of the treatment if needed). Review of diagnosis may be conducted again if there is failure of the patient to respond to treatment that would normally work. The procedure of diagnosing a patient suffering from hematology malignancies is synonymous to the general approach to medical diagnosis. The physician may carry out a precise diagnosis, which requires a complete physical evaluation to determine whether the patient have the diagnosed disease.

Since this traditional method involve the physician to continuously repeat this processes for several patients which will over time form a case base, adopting our intelligent fuzzy inspired case based reasoning (CBR) model depicted in figure 1 will help in providing solutions to new problem based on past experiment can accurately and promptly assist medical professional in drawing accurate conclusion on type of hematology disease.

IV. Hematological Data Selection

Hamilton and Bickle (2013) and Kim (2011) indicated that most frequently requested hematology test requested in medicine is the full blood picture (FBP). This contains a wealth of information about the components of blood. The typical constituent parts of the FBP are Hemoglobin concentration (Hb), Mean Cell Volume (MCV), Mean Corpuscular HEMOGLOBIN (MCH), Packed Cell Volume (PCV) Red Blood Cell (RBC) distribution width, White Blood Cell (WBC) Count incorporating a differential white cell count, Platelet Count (PC), and Reticulocyte Count (RC) ad their acceptable or normal values are shown in table 1.

Any hematology data outside these ranges are either caused by blood infections or other sources of hematology malignancies (Kim, 2011).

Global Journal of Computer Science and Technology (D) Volume XV Issue I Version I Year 2015

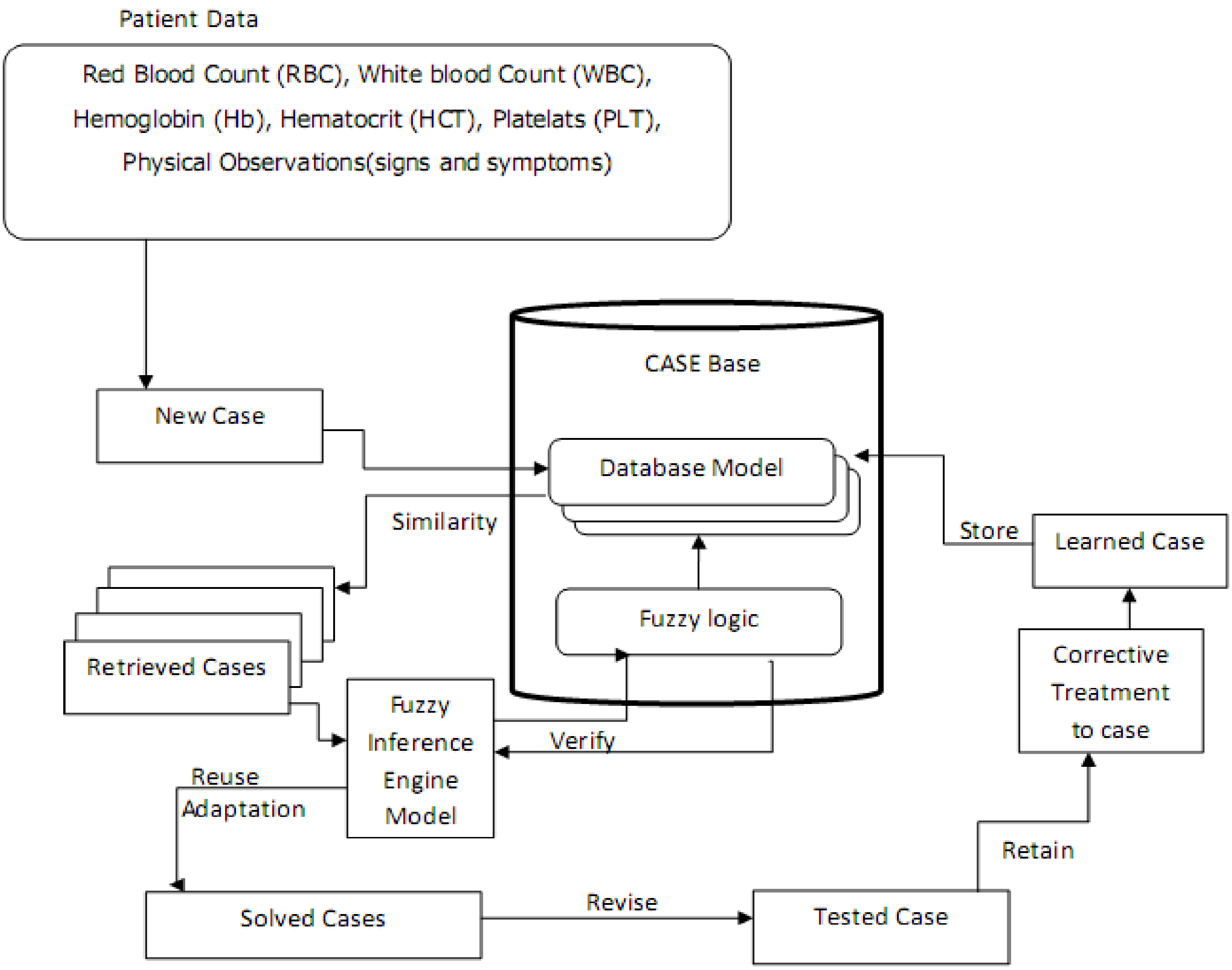


Figure 1: Fuzzy Inspired Case-base Expert System for Hematology malignancy classification and corrective treatment

Table 1: Standard Hematology Parameter values

Parameter	Age Group & Sex	Standard Range
Red Blood Count (RBC)	Newborn Children Adult-Male Adult-Female	5.00-6.50 (10^6/cmm) 3.70-5.50 (10^6/cmm) 4.40-5.80 (10^6/cmm) 3.80-5.20 (10^6/cmm)
White blood Count (WBC)	Newborn Children Adult-Male/Female	9.0-30.0 (10^3/uL) 6.0-17.0 (10^3/uL) 4.0-11.0 (10^3/uL)
Hemoglobin (HB)	Newborn Children Adult-Male Adult- Female	14.0-25.0 (d./Ul) 11.0-14.0 (d./Ul) 13.0-17.0 (d./Ul) 11.5-15.0 (d./Ul)
Hematocrit (HCT)	Newborn Children Adult-Male Adult-Female	44.0-64.0 (%) 34.0-42.0 (%) 37.0-51.0 (%) 35.0-46.0 (%)
Platelet Count (PLT)	Newborn /Children Adult-Male/Female	150-399 (10^3/cmm)
Reticulocyte count (RC)	Newborn Children /Adult-Male or Female	2.0-6.0 (%) 0.5-2.0 (%)

Mean Cell Volume (MCV)	Newborn /Children Adult-Male/Female	HCT/Total RBC (fl)
Mean Corpuscular HEMOGLOBIN (MCH)	Newborn /Children Adult-Male/Female	HB/Total RBC (pg)
Mean Corpusular hemoglobin concentration (MCHC)	Newborn /Children Adult-Male/Female	HB/HCT *100 in %

Source: Sykora (2008)

V. Result and Discussion

To design our fuzzy inspired case base expert system for hematology classification and corrective treatment (see figure 1), we designed a software system which consists of a set of symptoms (physical observations) and hematology parameters of patient which form a new case needed for the diagnosis.

The case base model consists of the database, which consist of cases of past experiences of hematology parameters (Full Blood Count), available signs and symptoms, identified disease and possible corrective treatment (solutions). Signs and symptoms are pre-medical data collected to assist the physicians carry an in-depth medical test on the patient. But the true picture of the patient heath condition is determined by collecting hematology parameter values. The values of the parameters are often vague (fuzzy) and imprecise hence the adoption of fuzzy logic in the model as means of analyzing these data. The fuzzy set of hematology parameters (attributes) is represented by 'Ha' which is defined as

$$Ha=\{ Ha_1, Ha_2, Ha_3,\ldots, Ha_n\}$$

where Ha_n represent the n^{th} parameter or attribute of Ha and n is the total number of parameter in Ha(here n=6). For each of the parameter, the set of constraint have been defined which makes it ambiguous to scale properly. You will recall that for each parameter a set of acceptable range for each parameter was given in table 1. Therefore, we have modeled a set of linguistic values that ensures the proper evaluation of the constraints using the linkert scale. The scale denoted as 'HL' is given thus:

HL= {Low, Normal, High}

$$\begin{cases} \text{Low if } Ha_i < \text{parameter range} \\ \text{Normal if } Ha_i \text{ fall within parameter range} \\ \text{High if } Ha_i > \text{parameter range} \end{cases}$$

Using the above context, assuming Ha1 for new born child is RBC with range from 5.00-6.50 (10^6/cmm), Ha_{1Left} and Ha_{1Right} equals 5.00 10^6/cmm and 6.50 10^6/cmm respectively. A trapezium fuzzy membership function can be use to classify the hematology data into low, normal or high. If the MF for the Ha is at critical lab value (CLV) range (i.e MF>0 as Ha increasing from 0 to Ha_{1Left}), we say the parameter is abnormal (low). If Ha is within normal range (MF=1), we say the parameter is in normal condition. When Ha drift away from Ha_{1Right} to the right (∞), MF decreases from 1 to zero), we also experience an abnormal case (high). Both low and high are abnormal cases which indicate a particular hematology disease or malignancy (see table 2 for groupings). Table 2 present the symbolic representation of the various parameter in a search space of S (-1,0,1) which represent Low, Normal and High respectively and it can be represented using a triangular fuzzy Membership function (MF) on the specified range for each parameter (see equation 1). A total membership function value (equation 2) is computed for each case (patient).

Table 2 : Fuzzy rules for classifying hematological cases

Rule No	RBC	WBC	MCV	PLT	HB	HCT	Hematology Classification
RO0	0	1	0	0	0	0	Leukocytosis
RO1	0	-1	0	0	0	0	Leukopenia, Sepsis, Marrow hypoplasia
RO2	0	0	0	1	0	0	Thrombocytosis
RO3	0	0	0	-1	0	0	Thrombocytopenia
RO4	1	0	0	0	1	0	Polycythemia
RO5	1	0	0	0	0	1	Polycythemia
RO6	-1	0	0	0	0	-1	Thalassemia
RO7	0	0	-1	0	0	0	Microcytic Anemia
RO8	0	0	1	0	0	0	Macrocytic Anemia
RO9	0	0	0	0	0	-1	Cardiac Failure
RO10	0	0	0	0	0	1	Spontaneous Bleeding
RO11	0	0	0	0	-1	0	Heart Failure
RO12	0	0	0	0	0	0	Normal/Healthy Condition

Key: -1(Low), 0(Normal), 1(High)

$$\mu_{\text{trap}}(x) = \begin{cases} 0 & x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \alpha \le x \le \beta \\ -\frac{x-\gamma}{\gamma-\beta} & \beta \le x \le \gamma \\ 0 & x > \gamma \end{cases} \quad \text{..................eq. 1}$$

$$\text{TMFV} = \sum_{i=1}^{n=6} \mu_{\text{tran}}(x_i) \quad \text{..................eq. 2}$$

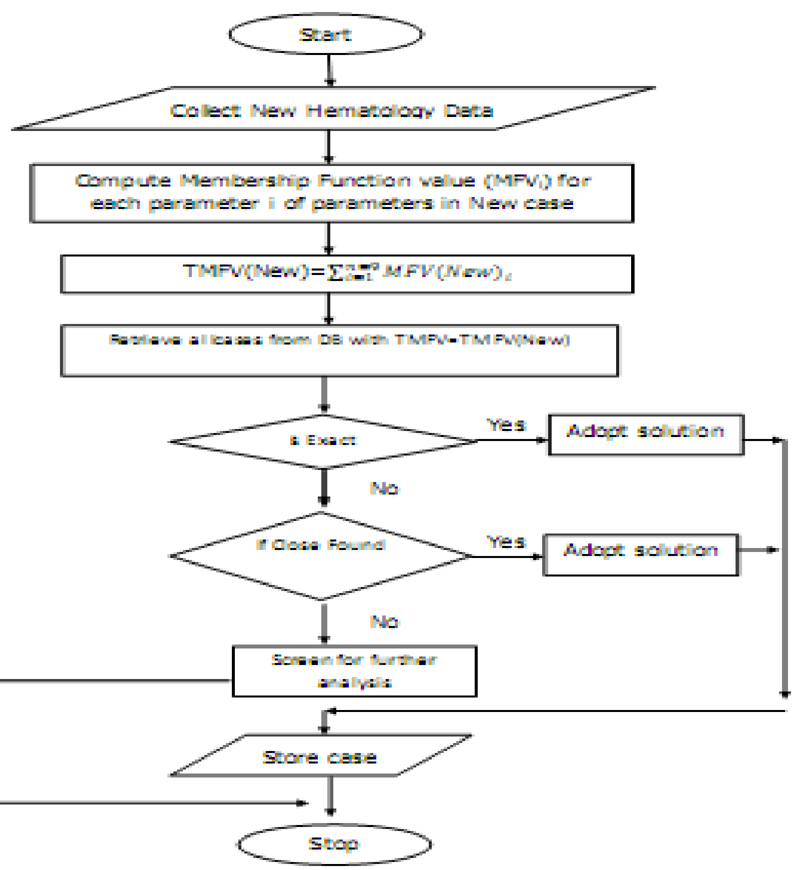
where x=ith Hematology parameter value, α=zero to Ha_{1Left}, β= Ha_{1Left} to Ha_{1Right}, γ= Ha_{1Right} to ∞

The CBR retrieval process takes place by the application of TMFV fitness function and k-nearest-neighbor matching. The goal of retrieval in the CBR system is to retrieve not only exact matches (equivalent cases) but partial matches (similar cases) as well. During the similarity assessment, an explicit context is used; therefore, the retrieval algorithm is based on incremental context transformations. Figure 2 shows a flowchart for the case retrieval and fuzzy inference engine for hematological classification. The advantage of using fuzzy logic is that it allows one to represent the concepts that could be considered to be in more than one category (or from another point of view, it allows representation of overlapping categories).

The fuzzy inference system consists of four modules as shown in figure 3. The Fuzzification module transform the hematology parameter data of each cases (inputs features), which are crisp numbers, into fuzzy sets. This is done by applying a fuzzification function. Here we propose to use the trapezoidal or triangular MF (see figure 4). The Knowledge base store the IF-THEN rules provided by the experts. Table 3.2 shows the logic representation which will be combined with fuzzy union (OR) and Interception (AND) operators. Each rule number will be translated into fuzzy rule. Here, the reuse phase is implemented. The Inference engine simulates the human reasoning process by making fuzzy inference on the inputs and IF-THEN rules so as to identify solution that best describe a rule. Here, the solution is a particular hematology disease. Here, the revise phase of the case base approach is performed. The Defuzzification module transforms the fuzzy set obtained by the inference engine into a crisp value. This crisp value is stored in the case base data bank which is retained into the system case base thereby completing the cycle of the proposed fuzzy inspired case base expert system.

Further, we create the necessary pre and post processing. As inputs are received by the system, the rule based is evaluated (See table 2). The antecedent, which is the (IF X AND Y) construct test the input and produces a conclusion-solution. The consequent (THEN Z) are satisfied which identify or classify the hematology malignancy. The conclusion is combined to form logical sums. Defuzzification coverts the rules base fuzzy output into non-fuzzy numerical values. It reflects the interpretation of the logic of the different linguistic variable.

Figure 2 : A Fuzzy logic inference flowchart for hematology classification



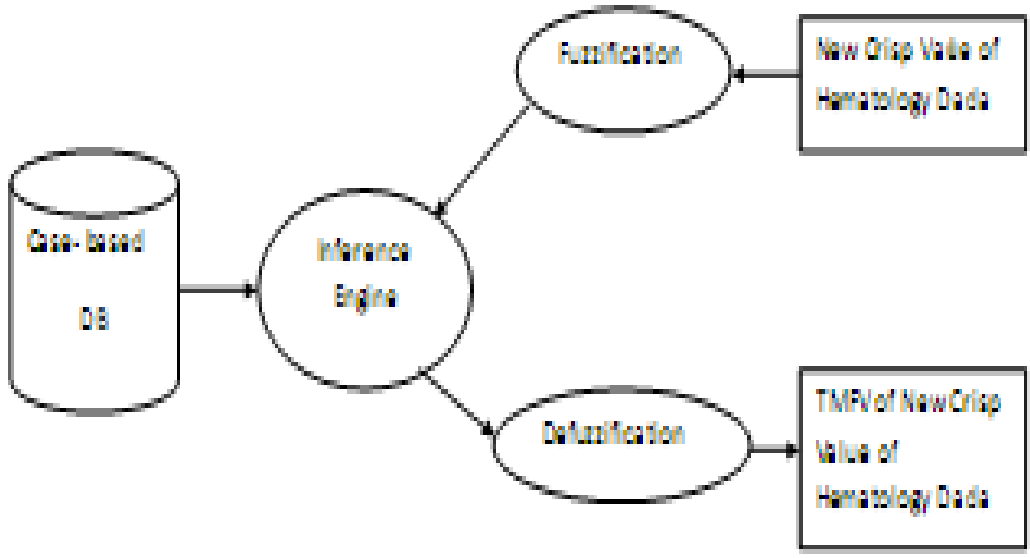


Figure 3 : Structure of fuzzy inference system

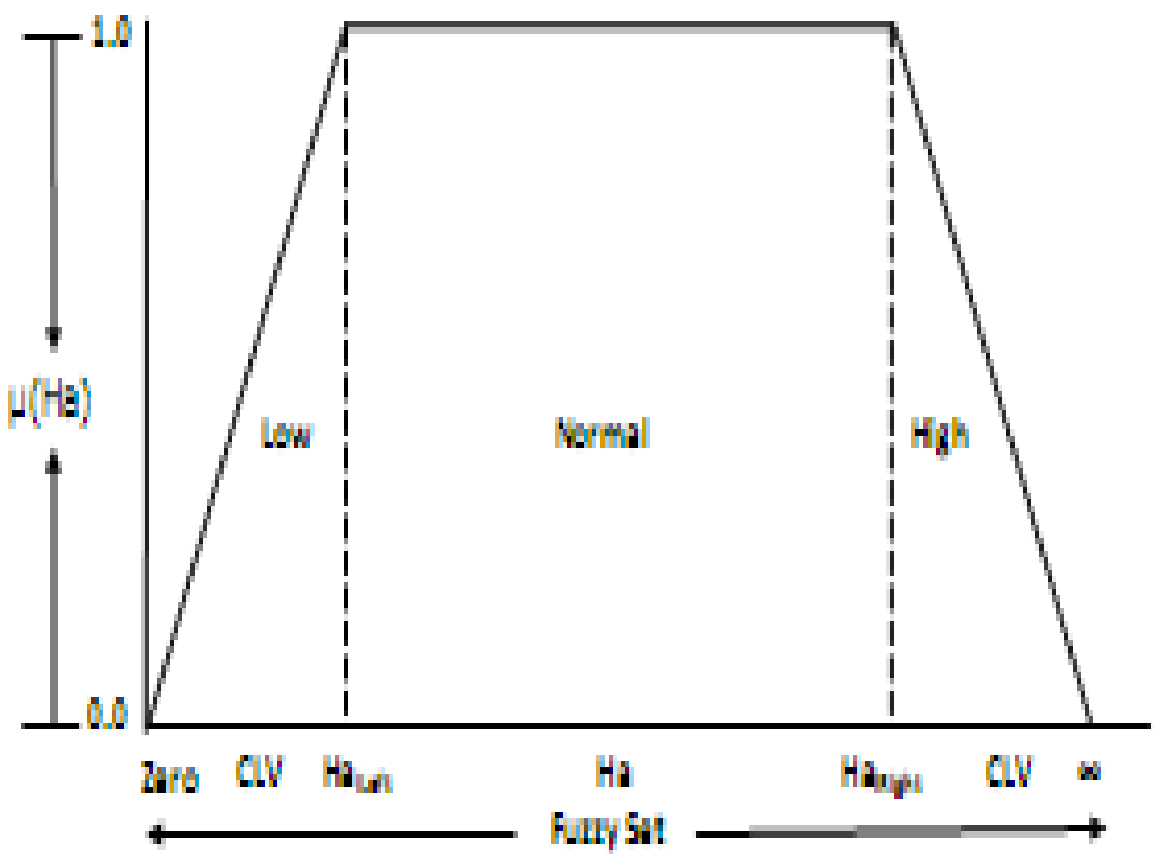


Figure 4 : Membership function for hematology parameter (Ha) classification

VI. Conclusion

In this study, the amazing gains of combining case based reasoning and fuzzy logic was used to design a software model that would assist health care provider to accurately classify hematology data base on both medical laboratory test and physical observation of patients. This system which uses a set of fuzzified data set incorporated case base provide more precise solution based on expert report of pervious experiences or experiment. The system designed is an interactive expert system that tells the physicians' patient's current condition as regards to evidence provided from the hematology data analysed. Hematology malignancies such as Anaemia, Leukemia. Thrombocytosis, Thrombocytopenia, Polycythemia, Polycythemia and Thalassemia can be identified and properly classified. It should however be noted that the system was able to give prescription of drugs or advice to patients as a corrective treatment. A system of this nature should be introduced in health care delivery centers and hospitals to help ease the work of physicians.

VII. Future Work

Base on the remarkable strength of fusing Fuzzy logic and CBR technique, we shall develop a robust intelligent expert system that will intuitively provide corrective treatment of various hematology malignancies based on the causes, signs and symptoms and the hematology parameter collected.

References Références Referencias

1. Armstrong S.A. Staunton J.E., Silverman L.B., Pieters R., deBoer M.L, Minden M.D. Sallan S.F. Lander E.S. Golub T.R, Korsmeyer S.J. (2002): MLL transactions specify a distinct gene expression profile that distinguishes a unique leukemia; *National Genet*, 30:41-7
2. Bendi V.R. Surendra M. P. B, Venkateswarlu N. B(2011): A Critical Study of Selected Classification Algorithms for Liver Disease Diagnosis, *International Journal of Database Management Systems (IJDMS)*, Vol.3, No.2
3. Butcher, J. (2004) Computers in clinical assessment: historical, developments, present status and future challenges, *Journal of Clinical Psychology*, 60, 331-345.
4. Garibaldi, J. M. & Ifeachor, E. C. (2000): The development of a Fuzzy Expert System for the Analysis of Umbilical Cord Blood, *Fuzzy Systems in Medicine*, 652-668.
5. Guessoum S., Tayeb M.L., Hayet D. and Tarek M.K. (2012): Combining Case and Rule Based Reasoning for the Diagnosis and Therapy of Chronic Obstructive Pulmonary Disease, *International Journal of Hybrid Information Technology* Vol. 5, No. 3
6. Hamilton P.K. and Bickle I.C. (2013): *Data interpretation for medical students*, 2nd edition, Belfast, PasTest.
7. Hoda S. (2008): An application of fuzzy case based reasoning for poison classification, master thesis, faculty of computer science and information technology, University of Malaya
8. Hullermeyer E. (2007). Case-Based Approximate Reasoning. Springer.
9. Kelada et al, (2012): *Genetic Analysis of Hematology parameters in incipient lines of the collaborative cross.* Available from: http://www.g3journal.org/lookup/suppl/doi:1534/g3.111.001/76
10. Kim M. (2011): *Lab value, Interpreting Chemistry and Hematology for Adult Patients*, AMN Healthcase Education Services, retrieved 28/08/2015, www.rn.com
11. Larry H.B. Gil D., James R. and Ronald R.C. (2012): *Converting Hematology Based Data into an inferential interpretation*, Hematology-Science and Practice, Dr. Charles (ed.), In-Tech. Available from: http://www.intechopen.com/books/hematology-science-and-practice/converting-hematology-based-data-to-useful-information
12. Leven A.P., Morgan G.J. and David F.F. (2003): The use of genetic microarray analysis to classify and predict prognosis in hematological malignancies, *clinical hematology*, 25: 209-22
13. Liu, H., Kshirsagar, A. & Niederberger, C. (2005): The Application of Machine Learning Techniques to the Prediction of Erectile Dysfunction, *Fourth International Conference on Machine Learning and Applications.*
14. Livi L., Rizzia A., and Sadeghianba A.(2014): *Granular modeling and computing approaches for intelligent analysis of non-geometric data,* Applied Soft Computing 27 567–574, available from www.elsevier.com/locate/asoc
15. Mendonca, E. A. (2004) Clinical Decision Support Systems: Perspective in Dentistry, *Journal of Dental Education*. Retrieved from: http://www.jdentaled.org/content/68/6/589.full
16. Minakshi S. and Sourabh M., (2013): Brain Tumor Segmentation Using Genetic Algorithm and Artificial Neural Network Fuzzy Inference System (ANFIS)*, Advances in Intelligent Systems and Computing* Volume 177, 2013, pp 329-339
17. Nilsson, M.; Funk, P.; and Sollenborn, M. (2003): Complex measurement classification in medical applications using a case-based approach. In Workshop on CBR in the Health Sciences, 63–72. ICCBR'03.
18. Obi J.C., and Imianvan A.A. (2011) Interactive Neuro-Fuzzy Expert System for Diagnosis of

Leukemia, *Global Journal of Computer Science and Technology* Volume 11 Issue 12 Version 1.0

19. Pascal L.E, True L.D., Campbell D.S., Deutsch E.W., Risk M., Coleman I.M., Eicher I.J. Nelson P.S., and Liu A.Y. (2008): Correction of mRNA and protein level:Cell type-specific gene expression of cluster designation antigens in the prostate, *BMC Genomics*, 9:246
20. Pedrycz W. (2013) *Granular Computing: Analysis and Design of Intelligent Systems* (Industrial Electronics Series), Taylor & Francis Group, 2013, ISBN978143988681
21. Rosma M.D. (2009): *A fuzzy regression model for the prediction of oral cancer susceptibility*, faculty of computer science & information technology, University of Malaya, Kuala Lumpur.
22. Schmidt, R., and Gierl, L. (2002). Prognostic Model for Early Warning of Threatening Influenza Waves. In: Minor M, Staab S (Hrsg.): Proceedings of 1st German Workshop on Experience Management, Lecture Notes in Informatics, Gesellschaft für Informatik, Köllen-Verlag, Bonn, ISBN 3-88579-340-7 (2002) 39-46
23. Shahina B., Mobyen U.A., Peter F., Ning X., Bo Von S. (2009): A case-based decision support system for individual stress diagnosis using fuzzy similarity matching, *Computational Intelligence* (ci), vol 25, nr 3, p180-195(16), Blackwell Publishing
24. Sykora L. (2008): *CLS500 Routine Hematology Tests*; University of Nebraska. Retrieved 12/08/2014 from www.nebraska.org
25. Zadeh, L.A. (1988), Fuzzy Logic, *IEEE Computer*, 1988, pp. 83-89.

This page is intentionally left blank

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence
Volume 15 Issue 1 Version 1.0 Year 2015
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Adaptive Genetic Algorithm Based Artificial Neural Network for Software Defect Prediction

By Racharla Suresh Kumar & Bachala Satyanarayana
Sri Krishnadevaraya University, India

Abstract- To meet the requirement of an efficient software defect prediction,in this paper an evolutionary computing based neural network learning scheme has been developed that alleviates the existing Artificial Neural Network (ANN) limitations such as local minima and convergence issues. To achieve optimal software defect prediction, in this paper, Adaptive-Genetic Algorithm (A-GA) based ANN learning and weightestimation scheme has been developed. Unlike conventional GA, in this paper we have used adaptive crossover and mutation probability parameter that alleviates the issue of disruption towards optimal solution. We have used object oriented software metrics, CK metrics for fault prediction and the proposed Evolutionary Computing Based Hybrid Neural Network (HENN)algorithm has been examined for performance in terms of accuracy, precision, recall, F-measure, completeness etc, where it has performed better as compared to major existing schemes. The proposed scheme exhibited 97.99% prediction accuracy while ensuring optimal precision, F-measure and recall.

Keywords: software defect prediction, machine learning, genetic algorithm, artificial neural network, object oriented software metrics.

GJCST-D Classification : *C.1.3 F.1.1*

Adaptive Genetic Algorithm Based Artificial Neural Network for Software Defect Prediction

Racharla Suresh Kumar [α] & Bachala Satyanarayana [σ]

Abstract- To meet the requirement of an efficient software defect prediction,in this paper an evolutionary computing based neural network learning scheme has been developed that alleviates the existing Artificial Neural Network (ANN) limitations such as local minima and convergence issues. To achieve optimal software defect prediction, in this paper, Adaptive-Genetic Algorithm (A-GA) based ANN learning and weightestimation scheme has been developed. Unlike conventional GA, in this paper we have used adaptive crossover and mutation probability parameter that alleviates the issue of disruption towards optimal solution. We have used object oriented software metrics, CK metrics for fault prediction and the proposed Evolutionary Computing Based Hybrid Neural Network (HENN)algorithm has been examined for performance in terms of accuracy, precision, recall, F-measure, completeness etc, where it has performed better as compared to major existing schemes. The proposed scheme exhibited 97.99% prediction accuracy while ensuring optimal precision, F-measure and recall.

Keywords: software defect prediction, machine learning, genetic algorithm, artificial neural network, object oriented software metrics.

I. Introduction

As per high pace rise in software applications and major dependency on it, the fault prediction has become one of the inevitable parts of software development life cycle (SDLC) that can play significant role in reducing the probability of software failure.

Software defect prediction (SDP) can be performed while planning to identify fault-prone modules in software product that as a result can provide the insight to the need for increased quality of monitoring during software development. In addition, it can also facilitate necessary approaches to incorporate certain proper fault verification schemes leading to enhanced software quality [1, 2, 3, 4] and reliability. SDP can be functional based on certain software metrics [3, 4, 5], such as source code changes, previous defects, etc. In fact software metrics are the quantitative data that are employed for characterizing the properties of source code and can be significant for predicting software quality. The efforts made through many generations have facilitated a number of schemes to mitigate defects, but the continuation of researches still indicates towards search for certain optimal SDP solution to ensure optimal performance, reliability, cost optimization and minimal maintenance. A number of efforts have been made for SDP using machine learning and neural network [6, 7, 8, 9, 10], clustering techniques, statistical method, mining and random forest [44, 45, 50] etc. In recent years, majority of software are being developed based on Object-Oriented (OO) paradigm. Thus, the quality of the software can be optimally assessed by employing software metrics, such as Abreu MOOD metric suite [11], QMOOD metrics suite [12], Bieman and Kang [13], Briand et al. [14], Etzkorn et al. [15], Halstead [16], Henderson-sellers [17], Li and Henry [18], McCabe [19], Tegarden et al. [20], Lorenz and Kidd [21] and CK metric [22] suite. These software metrics plays significant role in assessing the quality of software such as precision, accuracy, fault-resilience and sensitivity etc. The significance of these object oriented software metrics lies in their capability to predict the software quality in terms of adaptability, functionality, usability, portability, supportability, reliability and cost effectiveness. Predominantly two data driven algorithms, support vector machine (SVM) and artificial neural network (ANN) algorithms have been employed for fault detection. ANN approach functions on the basis of the human brain behaviorand possesses neurons and directed edges with certain weights existing between input and output layers. ANN employs output as the input so as to learn complex non-linear input-output relationship and can be stated to be a complex nonlinear mapping model between input and output layer. The processes in ANN comprise data sets to enhance the weight parameters, risk minimization scheme for stopping training as soon as the learning error enters in expected margin level. In fact, ANN has been employed in numerous utilities, but still it possesses certain limitations in terms of slow learning ability, local minima etc and hence require further optimization to achieve certain optimal SDP efficiency and performance. Thus, there is the requirement of further optimization of ANN approaches to accomplish a potential SDP solution. Some researches [23, 24] advocate the implementation of evolutionary computing techniques for SDP optimization. This paper proposes a novel evolutionary computing based enhanced ANN algorithmnamed Hybrid Evolutionary Computation

Author α: Research Scholar, Department of Computer Science, Sri Krishnadevaraya University, Andhra Pradesh, India.
e-mail: suresh_sku@yahoo.com
Author σ: Professor, Department of Computer Science, Sri Krishnadevaraya University, Andhra Pradesh, India.
e-mail: bachalasatya@yahoo.com

based Neural Network (HENN) for defect prediction and classification. HENN system employs Adaptive Genetic Algorithm (A-GA) for optimal weight estimation so as to enhance weight update and learning efficiency of the ANN.In this paper, the object oriented software metrics, CK metrics [22] have been employed as a fault classification data and the respective performance has been analyzed using confusion matrix.

The remaining sections discusses, related work in Section II, problem definition is briefed in Section III, which has been followed by proposed research discussion in Section IV. Section V presents the results and analysis and conclusion has been discussed in Section VI.

II. Related Work

The emergence of software applications and associated need of quality and reliability has motivated software practitioners as well as academia to develop certain novel scheme for defect prediction.With an objective to examine the relation between software metrics and associated faults some initiatives were made in [25, 26, 27, 28, 29, 30] where machine learning mechanism were used for fault detection. With an enthuse to compare the performance of varied other schemes such as decision trees, naïve Bayes, and 1-rule [31] performed fault detection using NASA MDP project. Chug et al [32] performed data mining based fault estimation using conventional J48, Random Forest, and Naive Bayesian Classifier (NBC) schemes but still couldn't employ the benefits of advanced classification schemes. With an objective to enhance conventional schemes Pushphavathi et al [33] introduced hybrid scheme of random forest (RF) and Fuzzy C Means (FCM) clustering. Then while, these systems were found limited for unbalanced data sets, which motivated author [34] to propose an approach called AdaBoost.NC that explored varied kinds of class imbalance learning schemes comprising resampling techniques, threshold moving, and ensemble algorithms. With an objective to explore SVM optimization in [35, 36] a dynamic SVM model was proposed for fault detection in source code using with error data and faulty code execution. Researcher in [37] developed an ANN based SDP system. This is the matter of fact that SVM refers the functional paradigm of single layer perceptron's NN which on addition with kernels behaves like multilayered perceptron's [38]. Till available systems based neural network with conventional learning and weight estimation suffers from local optima and convergence issue, which has not been discussed dominantly. On contrary, these days the software are developed and examined for faults using object oriented software metrics which even being significant has not been explored in depth to ensure optimal solution for reliability oriented defect prediction. This paper intends to provide an optimal solution for software defect prediction using evolutionary computing based neural network for efficient fault classification.

III. Problem Definition

In software development life cycle the reliability assurance is of great significance and to achieve it, the defect prediction is an inevitable need. The defect prediction can be performed using software metrics data, in which either it is predicted whether the code is defective or not or the magnitude of the probable defect and its severity is examined. In this research work, the predominant questions are whether evolutionary computing schemes, specifically GA can optimize neural network based artificial intelligence (AI) to achieve optimal software defect prediction. An another question that this research paper considers is that whether the conventional Genetic Algorithm can be further enhanced to deal with a scenario where multiple chromosomes are having similar fitness, and how this enhancement would perform classification or fault prediction?. In order to explore the answers of this significant question, in this paper it has been intended to optimize ANN learning and respective optimal weight estimation using GA, which has further being optimized to behave as an Adaptive GA (A-GA) scheme that ensures adaptive GA parameters (Crossover and mutation) estimation. Here, considering requirements of object oriented software metrics, CK metrics [22] have been considered that characterizes overall features of software in terms of varied component features. In this paper, the key software metrics considered are WMC, NOC,DIT,CBO,RFC,LCOM, which can be considered for defect prediction in certain class or data model. Based on the proposed model, the defect can be predicted which can be useful for ensuring quality and reliability of the software product. Given a training data, certain learning model can be developed that can classify the data for its faulty or non-faulty status. The artificial intelligence technique neural network has been used extensively so far for classification utilities, but being conventional these approaches do suffer from local minima and weight update issues. Thus, to enhance the systems, certain global optimization schemes like evolutionary computing can be considered. Since Particle Swarm Optimization suffers due to optimal minima and convergence issues, here we proposed an adaptive GA (A-GA) for ANN weight estimation where the weights are estimated dynamically in each iteration. Here, mean square error has been considered as the fitness value for A-GA. Further, the GA parameters such as crossover probability and mutation probability can be adaptively updated to make the overall system more robust and efficient. The optimization of ANN with A-GA can make it more effective and can be a potential candidate for fault detection in SDLC applications. The performance evaluation for these two approaches can be done in terms of accuracy, precision, recall, specificity etc.

IV. Proposed System

This section discusses the proposed evolutionary computing based hybrid neural network (HENN) for software defect prediction.

HENN: Evolutionary Computing Based Neural Network for Software Defect Prediction

Neural networks (NN) have seen an explosion of interest over the years, and are being successfully applied across a range of problem domains. Indeed, anywhere dealing with the problem of classification and prediction, neural networks are being used. For software defect prediction, ANN can be employed with learning approaches such as Gradient Descent (GD), Gauss Newton, and Levenberg Marquardt (LM) etc. Unlike conventional approach, in this paper, we have proposed an evolutionary computing technique called Adaptive Genetic Algorithm for ANN learning optimization and weight estimation, which has been further employed for fault prediction. Here, we intend to find relation between object oriented software metrics and fault prone classes and six CK metrics; WMC, NOC, DIT, RFC, CBO, LCOM have been taken as independent variable while fault data has been considered as dependent data. To design ANN, six inputs have been considered which do receive CK metrics individually as input having multiple classes, as per benchmark data (here PROMISE data). In this paper we have considered 8 hidden layers. Since, in the proposed SDP model, only FAULTY and NON-FAULTY are the results expected for prediction, therefore only one output node. The overall design of the proposed ANN model can be presented as follows:

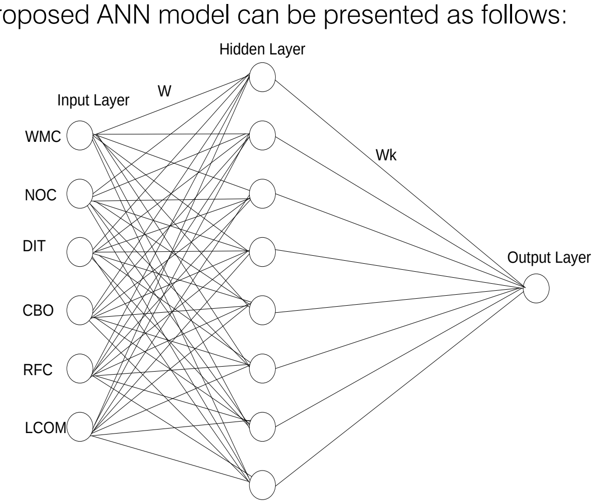


Figure 1 : ANN model for Defect prediction

The above mentioned figure illustrates the architecture of ANN containing three layers i.e., input layer, hidden layer and output layer. In the considered ANN model, the linear activation function has been used for input layer i.e., the output of the output layer is treated as input of the input layer($O_o = I_i$). Further, the sigmoid function has been employed for hidden layerO_h. Hence, the result of the hidden nodes O_h with the fed input of I_h is estimated mathematically as $O_h = \frac{1}{1+e^{-I_h}}$and final outcome of output nodes O_o is presented mathematically by$O_o = \frac{1}{1+e^{-O_i}}$.

In general, ANN is represented by a function $Y' = f(W, X)$ where Y' represents the output vector and$W$ and X are the weight vector and the input vector respectively. In process, the weight factor $W$is updated in each iterations to reduce Mean Square Error (MSE), which is estimated as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i' - y_i\right)^2 \tag{1}$$

Where y represents the actual output while the expected output is given byy_i'.

In order to process the datasets using ANN, at first the normalization of data is required. A discussion of the proposed data normalization technique in this paper is given as follows:

a) Data normalization

In the proposed model, initially the normalization has been performed before data processing that strengthens the system for better readability and defect prediction. Here the data normalization has been done over the range of [0, 1] for adjusting the defined range of input feature value and avoid the saturation of neurons. A number of schemes such as Min-Max normalization, Z-Score normalization and decimal scaling can be employed for the purpose of data normalization. In this paper, Min-Max normalization approach has been used that performs a linear transformation on the original data and maps each of the actual data x_i of attribute X to normalized value x'_i that exists in the range of [0, 1]. The Min-Max based normalized data has been obtained by the following expression:

$$Normalized(x_i) = x_i'' = \frac{x_i - min(X)}{max(X) - min(X)} \tag{2}$$

Where $max(X)$ and $min(X)$ represent the maximum and minimum value of the attribute X respectively. Performing data normalization the ANN model has been employed for fault classification and SDP functions.

In ANN based artificial intelligence systems, the efficient weight estimation is of great significance and till existing approaches have explored techniques such as Gauss Newton, Gradient descent, Levenberg Marquardt etc. Unfortunately these approaches couldn't be enhanced by scientific society to make weight estimation effective by means of certain global optimization techniques such as Genetic Algorithm. Efficient weight estimation during ANN learning can make classification optimal. This requirement motivated us to employ genetic algorithm for dynamic weight estimation during ANN learning. A brief discussion of the proposed Adaptive Genetic Algorithm (A-GA) is given in the following section.

b) Adaptive Genetic Algorithm(A-GA)

Genetic Algorithm (GA) is an adaptive search method for finding optimal or near optimal solutions, premised on the evolutionary ideas of natural selection. The fundamental concept of GA is emphasized on simulating processes in the natural system required for evolution, distinctively those that consider the Charles Darwin principles representing the terms of the survival of the fittest. Considering procedural flow, GA at first generates the initial population arbitrarily, where population refers a set of solutions. These solutions are nothing else but a chromosome that possesses a form of binary strings where all the comprising parameters are supposed to be encoded. Generating the population, GA estimates the fitness function of individual chromosome. Here the fitness function states toward a user-defined function that returns the evaluation results of each chromosome, thus a higher fitness value means its chromosome is a dominant gene. As per retrieved fitness values, offspring are generated using genetic operators—crossover and mutation. Applying these genetic operators the generations of the population are repeated iteratively until the stopping criteria are satisfied and an optimal solution is achieved. As illustrated in Figure-1, in this paper, the proposed HENN model comprises$i-h-o$ network configuration with i input layer, h hidden layer and O output layer or neurons. In the proposed ANN model, all the six considered CK metrics or feature vector are fed as input to the individual input node, where each feature vector metrics accompanies the number of classes available in datasets. Considering Figure-1 and relevant network configuration, there is N weight required to be estimated. Mathematically, the number of weight vectors is:

$$N = (i+O)*h \tag{3}$$

Here, the individual weight, which is considered as gene in the chromosomes of the A-GA, is a real number. Considering the gene length or the number of digits be$l$. Then the length of the chromosome L_{Chrom} can be estimated by the following expression:

$$L_{Chrom} = N*l = (i+O)*h*l \tag{4}$$

These all chromosomes are considered as the population of the genetic algorithm. In the proposed model to estimate the fitness value of the individual chromosome, the weights are required to be extracted from the individual chromosome. In our proposed model, the weights (W_k) are estimated by the following expression:

$$W_k = \begin{cases} & if\ 0 \le x_{kl+1} < 5 \\ -\dfrac{x_{kl+2}*10^{l-2}+x_{kl+3}*10^{l-3}+\dots+x_{(k+1)l}}{10^{l-2}} & \\ & if\ 5 <= x_{kl+l} <= 9 \\ +\dfrac{x_{kl+2}*10^{l-2}+x_{kl+3}*10^{l-3}+\dots+x_{(k+1)l}}{10^{l-2}} & \end{cases} \tag{5}$$

In order to process the Adaptive Genetic Algorithm (A-GA), the fitness values for each chromosome are required to be estimated. The fitness generation algorithm for the proposed A-GA system is given in Figure-2. The fitness values of each chromosome is estimated on the basis of the derived fitness function and the considered algorithm for deriving fitness function are $\bar{I}_i = (I_{1i}, I_{2i}, I_{3i}, \cdots, I_{li})$ and $\bar{T}_i = (T_{1i}, T_{2i}, T_{3i}, \cdots, T_{ni})$. For each chromosome $C_i, i = 1,2,3,\dots,p$, belonging to the current population P_i of size$P$.

Algorithm for Fitness Estimation
Input:$\bar{I}_i = (I_{1i}, I_{2i}, I_{3i}, \cdots, I_{li})$ **Output:**$\bar{T}_i = (T_{1i}, T_{2i}, T_{3i}, \cdots, T_{ni})$ Where $\bar{I}_i, \bar{T}_i$ represent the input and output pairs of the $i-h-o$ configuration of neural network. **Phase-1** : Weights $\bar{W}_i$ from C_i can be estimated by $W_k = \begin{cases} & if\ 0 \le x_{kd+1} < 5 \\ -\dfrac{x_{kd+2}*10^{d-2}+x_{kd+3}*10^{d-3}+\cdots+x_{(k+1)d}}{10^{d-2}} & \\ & if\ 5 <= x_{kd+d} <= 9 \\ +\dfrac{x_{kd+2}*10^{d-2}+x_{kd+3}*10^{d-3}+\cdots+x_{(k+1)d}}{10^{d-2}} & \end{cases}$ **Phase-2:** Considering $\bar{W}_i$ as a constant weight, train the network for N input instances and estimate output O_i **Phase-3:** Estimate error E_j for each input instance j $E_j = (T_{ji} - O_{ji})$ **Phase-4: Estimate** Root mean square error (RMSE) of chromosomeC_i $E_i = \sqrt{\dfrac{\sum_{j=1}^{j=N} E_j}{N}}$ Where N is the total number of training data set **Phase-5:** Estimate the fitnessvalue for chromosomeC_i $F_i = \dfrac{1}{E_i} = \dfrac{1}{\sqrt{\dfrac{\sum_{j=1}^{j=N} E_j}{N}}}$

Figure 2 : Algorithm for Fitness generation using A-GA

This is the matter of fact that the evolutionary computing scheme named Genetic Algorithm has established itself as a potential optimization technique for various application scenarios, still this approach possess scopes for further optimization that specifically depends on the working environment. In this paper, there might be the possibility that after every generation to achieve optimal fitness, certain new population would be generated and thus the processing data might be increased after each iterations, thus resulting into certain restraints such as premature convergence caused due to local optima and low convergence speed, which is common in other evolutionary techniques such as Particle Swarm Optimization. In order to alleviate these issues, the parameters like cross over probability (P_c) and mutation probability (P_m) can be made dynamic and weight adaptive. In addition, such novelty can deal with a common scenario, where there is the possibility of multiple chromosomes having similar fitness value, causing degraded classification accuracy. Taking into consideration of these all factors and motivations, in this paper a weight adaptive genetic algorithm (A-GA) has been developed where the genetic parameters (Crossover and mutation) are updated dynamically. In the proposed approach the parameters P_c and P_m have been dynamically updated by means of the following mathematical model:

$$(P_c)_{k+1} = (P_c)_k - \frac{C_1 * n}{5} \qquad (6)$$

$$(P_m)_{k+1} = (P_m)_k - \frac{C_2 * n}{5}$$

Where $(P_c)_{k+1}$ and $(P_m)_{k+1}$ represent the updated probability of cross over and mutation, $(P_c)_k$ and $(P_m)_k$ are the current probability of cross over and mutation, C_1 and C_2 can be the positive constant and n is the number of chromosome having same fitness value. Thus, implementing these discussed approaches, if the final output estimated is greater than 0.5, then the class is labeled as FAULTY otherwise NON-FAULTY. Figure-3 represents the overall process of software defect prediction using Adaptive Genetic Algorithm (A-GA).

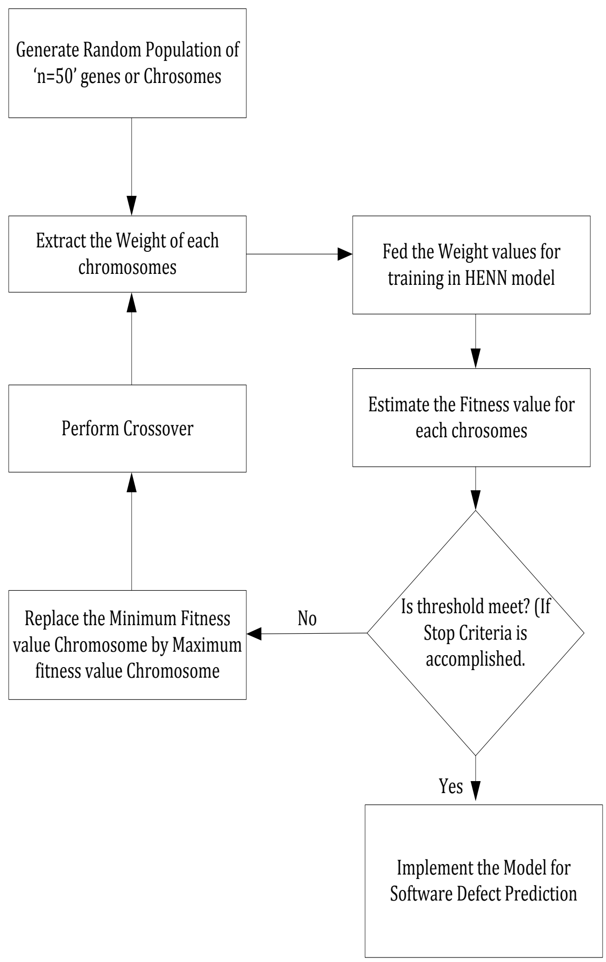


Figure 3 : Proposed HENN Scheme for Software Defect Prediction

The discussion of Figure-3 for HENN simulation is given as follows:

i. *HENN-SDP Simulation*

As illustrated in Figure-1, in the proposed HENN model, three layers of neural network has been considered comprising six input nodes, eight hidden nodes and one output node. The overall process of HENN based fault estimation is discussed as follows:

Chromosome Initialization: Initial chromosome selection with population size is 50 has been considered using random selection process.

Weight Estimation: Obtained the weight vector W_kfor each chromosome as the input to hiddenlayer and hidden layer to output layer and thus the weight of input to hidden node and hidden node to output are estimated using equation 5.

Fitness Estimation: On the basis of weights retrieved, the fitness value is estimated for each chromosome, where the proposed HENN intends to minimize the mean square error as defined in Figure 2.

Ranking of Chromosomes: Perform the ranking of each chromosomes based on respective fitness value and substitute the chromosomes with minimum fitness value by the chromosomes with highest fitness value chromosome.

Crossover: Perform two point crossover processdynamically vary the GA parameters $P_c$and P_m till reaching optimal criteria using equation(6).

In the simulation model, the initial P_c and P_m are 0.6 and 0.1 respectively and n signifies the number of chromosome having similar fitness value.

Stopping Criteria: The developed system terminates once the 95% chromosomes in the gene pool accomplishes its unique fitness value and beyond this the fitness level of chromosomes gets saturated.

Classify Faults: If the final weight is greater than 0.5, then the class is labeled as FAULTY otherwise NON-FAULTY. Confusion Matrix Generate the confusion matrix for each classes of OO-SM and classify fault/non-fault distribution for performance evaluation.

Thus, employing the proposed HENN model, the fault classification and prediction has been done. The simulation, results and discussion is provided in the following section.

V. Result and Analysis

This section discusses the research variables, simulation setups, results obtained and respective performance analysis.

a) Data collection

In this paper, the CK metric suites have been employed which have been defined for varied objectives such as software fault detection/prediction, effort evaluation, re-usability and maintenance. Considering the robustness of CK metric suite [27], it has been used as object oriented software metrics which has been processed using Chidamber and Kemerer Java Metrics tool (CKJM) tool that extracts software metrics by executing byte code of compiled Java cases and assigns a definite weight of the comprising classes having feature vectors. In this paper, PROMISE fault benchmark data [39] and NASA MDP datasets [40] and PROMISE repository to evaluate the performance of the proposed fault prediction scheme. We intended to establish the relationship between Object-Oriented software metrics (OO-SM) and the fault proneness at the class level. In order to perform defect prediction using regression analysis paradigm, we have considered fault as a dependent variable while the CK metric as the independent variable. The predominant OO-SM metrics are given in Table-1.

Table 1 : Object Oriented Software Metrics (CK Metrics [22])

WMC	Overall complexities of the methods in comprising classes
NOC	Number of sub-classes subordinate to a class in the class hierarchy
DIT	Maximum height of the class hierarchy
CBO	Number of other classes to which it is allied with
RFC	A set of approaches that can be executed in response to a message received by an object of that class
LCOM	Dissimilarity measurement of varied methods in a class using instanced attributes/variables
NOM	Number of methods (in a class)
NOA	Number of attribute (in a class)
NOAI	Number of attributes inherited by subclasses.
NOMI	Number of methods inherited by subclasses.
Fan-in	Total number of local flows in certain process and data structures from where it retrieves information
Fan-out	Total number of local flows in certain process and data structures from where it retrieves information
NOPM	Total number of private methods in a class
NOPA	Total number of private attribute in a class
NOPM	Total number of public methods in a class
NOPA	Total number of public attribute in a class
NLOC	Size of program by counting the number of lines in the source code.

In our work, we have developed a function to explore the relation between Object-Oriented software metrics (OO-SM) (WMC, NOC, DIT, RFC, CBO and LCOM) and faults existing in class under consideration. The minimization of faults can be of great significance towards optimization of software equality, and to ensure optimal defect prediction, the fault has been derived as the function of software metrics as illustrated as follows:

$$Faults = f(WMC, NOC, DIT, CBO, RFC, LCOM)$$

We used four public domain defect datasets from the PROMISE repository [9][39]. The considered data sets are *JEdit, IVY, Ant* and *Camel* which contain static code measures along with varied modules sizes, defective modules and defect rates. In our simulation model, the respective extracted weights and features of the data classes are taken as input. The datasets with respective classes or modules are given in Table-2.

Table 2 : PROMISE Data and modules

PROMISE	JEdit	IVY	Ant	Camel
Number of modules	492	352	744	965

In this paper, HENN algorithm has been developed for simulation using MATLAB 2012b software tool having artificial intelligence and ANN toolboxes. The proposed models examined defect datasets and the FAULTY and NON FAULTY data have been classified. Here on the basis of FAULT distribution by proposed model, a confusion matrix has been generated that encompasses two rows and columns comprising true negatives, true positive, false negative and false positive variables. The respective values of True negatives (TN) refer the modules which are NON FAULTY or fault-free on the other hand, true positives (TP) represents for those modules which are classified as FAULTY. False negatives (FN) are those modules which are FAULTY and are classified incorrectly as NON FAULTY. Similarly, false positives (FP) modules are those modules which are faultless but are classified incorrectly as FAULTY. A matrix presentation of confusion matrix is given in Table 3.

Table 3 : Confusion Matrix

	Predicted Defective	Predicted Defect Free
FAULTY	True Positive	False Negative
NON-FAULTY	False Positive	True Negative

Generally, the meanings of the values of the binary variables are not needed to be defined, however, in our work, especially for performance assessment the variables have been labeled as positive and negative. The positive levels refer towards the results as FAULTY in that specific simulation scenario. In this paper, we have measured the performance of the proposed HENN SDP in terms of correctness, precision, F-measures, accuracy, recall, specification and cost factor analysis. A brief mathematical definition of these variables is given as follows:

Table 4 : Performance Parameters

Construct	Mathematical Expression	Description
Recall	TP/(TP+FN)	Proportion of defective units correctly classified
Precision	TP/(TP+FP)	Proportion of Units correctly predicted as defective
Specification	TN/(TN+FP)	Proportion of correctly classified non defective units
F-measure	$2.\frac{Recall.Precision}{Recall + Precision}$	Defined as harmonic mean of precision and recall
Accuracy	$(TN+TP)/(TN+FN+FP+TP)$	Proportion of correctly defined units

Table 5 : Performance analysis for proposed HENN SDP model and other existing models

SDP Techniques	Accuracy (%)	Precision (%)	F-Measure (%)
LLE-SVM[41]	81.1	82.5	80.4
SVM [41]	69.4	68.1	69.7
SVM [42]	55.3	88.0	83.2
Natural Gas [48]	94.25	-	-
Symbolic Regression [48]	89.50	-	-
RBP-NN [48]	80.0	-	-
MLP [42]	86.6	86.6	87.4
Naive Based [42]	85.6	83.1	83.9
CPSO[43]	69.2	67.6	-
T-SVM [44]	75.8	84.1	80.91
GANN[43]	73.4	81.6	-
AdaBoost [43]	79.1	82.3	-
Random Forest [50]	91.4	-	-
k-NN [47]	91.8	-	-
C4.5 [47]	88.39	-	-
J 48 [47]	90.90		
Levenberg-Marquardt-NN [47]	88.0	-	-
NNEP-Evolutionary [43]	88.8	81.2	-
PSO [46]	78.78	-	-
PSO-NN [48]	97.75	-	-
HENN SDP*	**97.9***	**1**	**98.9**

**- The best performance of HENN*

Thus, the results obtained exhibit that the optimization made by means of Adaptive Genetic Algorithm has enhanced ANN learning for fault detection. The ultimate results obtained for HENN represents the most effective and optimal results as compared to other existing approaches, especially neural network based SDP models. The performance analysis for the proposed systems is given in Table-6.

Table 6 : Performance analysis for proposed HENN SDP Model

Technique	Data	Modules	Accuracy	Precision	F-Measure	Recall	Specification
HENN	JEdit	492	**0.9799**	1	**0.9897**	1	0.9756
HENN	IVY	352	0.8835	0.9936	0.9380	0.8883	0.3333
HENN	Ant	744	0.8145	0.9343	0.8867	0.8438	0.6346
HENN	Camel	965	0.8114	1	0.8952	0.8102	1

Global Journal of Computer Science and Technology (D) Volume XV Issue I Version I Year 2015

VI. Conclusion

Software defect prediction has become an inevitable need for organizations to ensure quality and reliability of software products. The early defect prediction can facilitate managers to rectify and enrich reliability of product. Approaches such as machine learning and neural network have become eminent solution for training and classification of data and can be significant for defect prediction. However, these approaches need optimization in terms of weight update, parametric enhancement while performing defect prediction. The local minima and convergence issue of ANN can be significantly dealt with employing evolutionary computing schemes and the implementation of genetic algorithm can be the dominant candidate. In this paper, Adaptive Genetic Algorithm (A-GA) has been used for ANN optimization, where A-GA functions for optimal weight estimation. The proposed HENN model has been tested with PROMISE data sets, where the average accuracy for HENN was retrieved as 87.23 % while the best classification performance was observed with JEdit datasets where HENN exhibited 97.99% accuracy while ensuring 100% precision. Performance in terms of F-measure using HENN was obtained as 98.97%. The results also depicted that A-GA based ANN can outperform Particle Swarm Optimization based defect prediction schemes, regression techniques, AdaBoost, and other conventional weight estimation based ANN models. In future, the efficiency of the proposed HENN model can be examined in comparison with optimized machine learning such as SVM with varied kernel functions.

References Références Referencias

1. Zuse H., "A Framework of Software Measurement, Walter de Grutger Publish" 1998.
2. University of Texas, Software Quality Institute Report, May 2002.
3. Rosenberg, L., S. B., Sheppard, "Metrics in Software Process Assessment, Quality Assurance and Risk Assessment", 2nd International Symposium on Software Metrics, London, October, 1994.
4. Boehm, B. W., Software Engineering Economics, Prentice-Hall, 1981.
5. L.C. Briand, W.L. Melo, J. Wu st, "Assessing the Applicability of Fault-Proneness Models Across Object-Oriented Software Projects," IEEE Trans. Software Eng., vol. 28, no. 7, pp. 706-720, July 2002.
6. Kutlubay O., A. Bener, "A Machine Learning Based Model for Software Defect Prediction," working paer, Boaziçi University, Computer Engineering Department 2005.
7. Bo. Yang, Xiang Li, "A study on software reliability prediction based on support vector machines", The Annual IEEE International Conference on Industrial Engineering and Engineering Management, pp. 1176-1180, 2-4 Dec. 2007.
8. Sandhu, Parvinder Singh, Sunil Kumar, Hardeep Singh, "Intelligence System for Software Maintenance Severity Prediction", Journal of Computer Science, Vol. 3 (5), pp. 281-288, 2007.
9. Gondra, "Applying machine learning to software fault-proneness prediction," Journal of Systems and Software, vol. 81, no. 2, pp. 186-195, Feb. 2008.
10. Q. P Hu, Y. S. Dai, M. Xie, S. H. Ng.,"Early software reliability prediction with extended ANN model," Proceedings of the 30th Annual International Computer Software and Applications Conference (COMPSAC'06), Vol. 2, pp. 234-239, September 2006.
11. F. B. E. Abreu, R. Carapuca, "Object-Oriented software engineering: Measuring and controlling the development process," in Proceedings of the 4th International Conference on Software Quality, vol. 186, 1994.
12. J. Bansiya, C. G. Davis, "A hierarchical model for Object-Oriented design quality assessment," ACM Transactions on Programming Languages and Systems., vol. 128, pp. 4–17, August 2002.
13. B. K. Kang and J. M. Bieman, "Cohesion and reuse in an Object-Oriented system," in Proceedings of the ACM SIGSOFT Symposium on software reusability, pp. 259–262, Seattle, March 1995.
14. L. C. Briand, J. Wust, J. W. Daly, D. V. Porter, "Exploring the relationships between design measures and software quality in Object-Oriented systems," The Journal of Systems and Software, vol. 51, pp. 245–273, May 2000.
15. L. Etzkorn, J. Bansiya, and C. Davis, "Design and code complexity metrics for Object-Oriented classes," Object-Oriented Programming, vol. 12, no. 10, pp. 35–40, 1999.
16. [16] M. Halstead, Elements of Software Sciencel. New York, USA: Elsevier Science, 1977.
17. B. Henderson-Sellers, Software Metrics. UK: Prentice-Hall, 1996.
18. W. Li and S. Henry, "Maintenance metrics for the Object-Oriented paradigm," in Proceedings of First International Software Metrics Symposium, pp. 52–60, 1993.
19. T. J. McCabe, "A complexity measure," IEEE Transactions on Software Engineering, vol. 2, pp. 308–320, December 1976.
20. D. P. Tegarden, S. D. Sheetz, D. E. Monarchi, "A software complexity model of Object-Oriented systems," Decision Support Systems, vol. 13, no. 3, pp. 241–262, 1995.
21. M. Lorenz and J. Kidd, Object-Oriented Software Metrics. NJ, Englewood: Prentice-Hall, 1994.
22. S. R. Chidamber and C. F. Kemerer, "A metrics suite for Object-Oriented design," IEEE

Transactions on Software Engineering on June 1994, vol. 20, pp. 476–493.
23. M. Harman, “Why the Virtual Nature of Software makes it Ideal for Search Based Optimization”, Fundamental Approaches to Software Engineering, 2010.
24. C. Grosan, and A. Abraham, “Hybrid Evolutionary Algorithms: Methodologies, Architectures, and Reviews”, Studies in Computational Intelligence, vol. 75, pp. 1-17, 2011.
25. Saida Benlarbi, Khaled El Emam, Nishith Geol (1999), “Issues in Validating Object-Oriented Metrics for Early Risk Prediction”, by Cistel Technology 210 Colonnade Road Suite 204 Nepean, Ontario Canada K2E 7L5.
26. Lanubile F., Lonigro A., Visaggio G. “Comparing Models for Identifying Fault-Prone Software Components”, Proceedings of Seventh International Conference on Software Engineering and Knowledge Engineering, pp. 12-19, June 1995.
27. Fenton, N. E. and Neil, M., “A Critique of Software Defect Prediction Models”, Bellini, I. Bruno, P. Nesi, D. Rogai, University of Florence, IEEE Trans. Softw. Engineering, vol. 25, Issue no. 5, pp. 675-689, 1999.
28. Giovanni Denaro, ”Estimating Software Fault-Proneness for Tuning Testing Activities” Proceedings of the 22nd International Conference on Software Engineering, Limerick, Ireland, June 2000.
29. Manasi Deodhar, “Prediction Model and the Size Factor for Fault-proneness of Object Oriented Systems”, MS Thesis, Michigan Tech. University, Dec. 2002.
30. Bellini, P., “Comparing Fault-Proneness Estimation Models”, 10th IEEE International Conference on Engineering of Complex Computer Systems (ICECCS'05), pp. 205-214, 2005
31. Khoshgoftaar, T.M., K. Gao and R. M. Szabo, “An Application of Zero-Inflated Poisson Regression for Software Fault Prediction. Software Reliability Engineering”, ISSRE 2001. Proceedings of 12th International Symposium, pp: 66 -73, 27-30 Nov. 2001.
32. Chug, A., Dhall, S., "Software defect prediction using supervised learning algorithm and unsupervised learning algorithm," Confluence 2013: The Next Generation Information Technology Summit, pp.173-179, 26-27 Sept. 2013.
33. Pushphavathi, T.P.; Suma, V.; Ramaswamy, V., "A novel method for software defect prediction: Hybrid of FCM and random forest," Electronics and Communication Systems (ICECS), 2014 International Conference, vol., no., pp.1,5, 13-14 Feb. 2014.
34. Wang, S.; Yao, X., "Using Class Imbalance Learning for Software Defect Prediction," Reliability, IEEE Transactions, vol.62, no.2, pp.434-443, June 2013.
35. Brun, Y. and D. E. Michael, “Finding Latent Code Errors via Machine Learning over Program Executions”, Proceedings of the 26th International Conference on Software Engineering, May, 2004.
36. F. Xing, P. Guo, M. R. Lyu, “A novel method for early software quality prediction based on support vector machine,” Software Reliability Engineering, International Symposium, pp. 213–222, 2005.
37. Cai K Y, 0n the Neura1 Network Approach in Software Reliability Modeling, Journal of Systems and Software, pp 47-62, 2001.
38. S.A. Rojas and D. Fernandez-Reyes, “Adapting multiple kernel parameters for support vector machines using genetic algorithms,” The 2005 IEEE Congress on Evolutionary Computation, vol. 1, pp. 626-631, September, 2005.
39. http://mdp.ivv.nasa.gov/.
40. http://promisedata.googlecode.com/svn/trunk/defect/
41. Chun Shan, Boyang Chen, Changzhen Hu, Jingfeng Xue1, Ning Li, “SOFTWARE DEFECT PREDICTION MODEL BASED ON LLE AND SVM” Communications Security Conference; pp 1-5, 22-24 May 2014.
42. Ye Xia, Guoying Yan, Xingwei Jiang, Yanyan Yang, "A new metrics selection method for software defect prediction," Progress in Informatics and Computing (PIC), International Conference, pp.433-436, 16-18 May 2014.
43. Malhotra, R., Pritam, N., Singh, Y., "On the applicability of evolutionary computation for software defect prediction," Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference, pp.2249-2257, 24-27 Sept. 2014.
44. Chug, A., Dhall, S., "Software defect prediction using supervised learning algorithm and unsupervised learning algorithm, "Confluence 2013: The Next Generation Information Technology Summit, pp.173-179, 26-27 Sept. 2013.
45. Armah, G.K., Guangchun Luo, Ke Qin, "Multilevel data preprocessing for software defect prediction," Information Management, Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference, vol.2, pp.170-174, 23-24 Nov. 2013.
46. Verma, R., Gupta, A., "Software defect prediction using two level data pre-processing," Recent Advances in Computing and Software Systems (RACSS), International Conference, pp.311-317, 25-27 April 2012.
47. Malkit Singh, Dalwinder Singh Salaria, “Software Defect Prediction Tool based on Neural Network”, International Journal of Computer Applications, Volume 70– No.22, pp-0975 – 8887, May 2013.
48. Anurag Shrivastava, Vishal Shrivastava, “A Hybrid Model of Soft Computing Technique for Software

Fault Prediction", International Journal of Current Engineering and Tech. Vol. 4, No. 4, Aug 2014.
49. Riju Kaushal,Sunil Khullar, "PSO based neural network approaches for prediction of level of severity of faults in NASAs public domain defect dataset", International Journal of Information Technology and Knowledge Management, Volume 5, No. 2, pp. 453-457, July-December 2012.
50. Mohamad Mahdi Askari, Vahid Khatibi Bardsiri, "Software Defect Prediction using a High Performance Neural Network", International Journal of Software Engineering and Its Applications, Vol. 8, No. 12, pp. 177-188, 2014.

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence
Volume 15 Issue 1 Version 1.0 Year 2015
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

A New Approach to Adaptive Neuro-Fuzzy Modeling using Kernel based Clustering

By Sharifa Rajab & Vinod Sharma
University of Jammu Campus, India

Abstract- Data clustering is a well known technique for fuzzy model identification or fuzzy modelling for apprehending the system behavior in the form of fuzzy if-then rules based on experimental data. Fuzzy c-Means (FCM) clustering and subtractive clustering (SC) are efficient techniques for fuzzy rule extraction in fuzzy modeling of Adaptive Neuro-fuzzy Inference System (ANFIS). In this paper we have employed a novel technique to build the rule base of ANFIS based on the kernel based variants of these two clustering techniques which have shown better clustering accuracy. In kernel based clustering approach, the kernel functions are used to calculate the distance measure between the data points during clustering which enables to map the data to a higher dimensional space. This generalization makes data set more distinctly separable which results in more accurate cluster centers and therefore a more precise rule base for the ANFIS can be constructed which increases the prediction performance of the system. The performance analysis of ANFIS models built using kernel based FCM and kernel based SC has been done on three business prediction problems viz. sales forecasting, stock price prediction and qualitative bankruptcy prediction. A performance comparison with the ANFIS models based on conventional SC and FCM clustering for each of these forecasting problems has been provided and discussed.

Keywords: *fuzzy modelling, kernel function, neuro-fuzzy model, fuzzy inference system, business prediction.*

GJCST-D Classification : *I.5.1*

ANEWAPPROACHTOADAPTIVENEUROFUZZYMODELINGUSINGKERNELBASEDCLUSTERING

Strictly as per the compliance and regulations of:

A New Approach to Adaptive Neuro-Fuzzy Modeling using Kernel based Clustering

Sharifa Rajab [α] & Vinod Sharma [σ]

Abstract- Data clustering is a well known technique for fuzzy model identification or fuzzy modelling for apprehending the system behavior in the form of fuzzy if-then rules based on experimental data. Fuzzy c-Means (FCM) clustering and subtractive clustering (SC) are efficient techniques for fuzzy rule extraction in fuzzy modeling of Adaptive Neuro-fuzzy Inference System (ANFIS). In this paper we have employed a novel technique to build the rule base of ANFIS based on the kernel based variants of these two clustering techniques which have shown better clustering accuracy. In kernel based clustering approach, the kernel functions are used to calculate the distance measure between the data points during clustering which enables to map the data to a higher dimensional space. This generalization makes data set more distinctly separable which results in more accurate cluster centers and therefore a more precise rule base for the ANFIS can be constructed which increases the prediction performance of the system. The performance analysis of ANFIS models built using kernel based FCM and kernel based SC has been done on three business prediction problems viz. sales forecasting, stock price prediction and qualitative bankruptcy prediction. A performance comparison with the ANFIS models based on conventional SC and FCM clustering for each of these forecasting problems has been provided and discussed.

Keywords: fuzzy modelling, kernel function, neuro-fuzzy model, fuzzy inference system, business prediction.

I. Introduction

The concept of fuzzy logic was introduced by Lofti Zadeh (1969) based on fuzzy set theory in early 60's as an innovative approach to characterize the non-probabilistic uncertainty. since then this field has evolved into a productive realm encompassing various domains viz. fuzzy reasoning, fuzzy topology, fuzzy modelling and fuzzy inference systems. A Fuzzy inference system (FIS) is referred by a number of names like fuzzy model, fuzzy expert system, fuzzy associative memory and so on. FIS is composed of three conceptual parts: a fuzzy rule base containing fuzzy rules, database defining the membership functions used in the fuzzy rules and a reasoning procedure for performing inference upon the rules and provided facts to obtain the output. The fuzzy if-then rules are used to represent the input-output relationships of the modeled system and are helpful to present the qualitative aspect of human reasoning without using any accurate mathematical model for the system. The fuzzy rule base for FIS can be constructed directly by domain experts a method that is usually error prone or by using fuzzy modelling approach. Fuzzy modelling also called fuzzy identification is an important technique to capture the behavior of a system to be modeled in the form of fuzzy if- then rules using its quantifiable characteristics. It has been addressed in a number of studies (Mamdani, 1976; Tong et al., 1980; Larsen, 1980) but was first discussed systematically by Takagi et al. (1985) as an effective technique for the estimation of dynamic fuzzy systems for problems of non linear uncertain nature. The important issue of determining the number of fuzzy rules in the rule base and the values of parameters of membership function in fuzzy rules are dealt with using fuzzy modelling. Fuzzy modelling is nowadays successfully applied in control, prediction and other applications for the identification of fuzzy models using observed input output datasets. The fuzzy models possess the capability to provide insights into the relationships between various variables in the model which is not possible with several black box techniques such as neural networks. The fuzzy models also allow integrating the information obtained from the observed numerical input output data with the prior expert knowledge. A standalone FIS however does not have the ability to learn and can be extended by using optimization and adaptive methods for performance improvements. ANFIS proposed by Jang (1995) is a widely employed fuzzy model based on the concept of integrating fuzzy inference systems and neural networks that uses learning to fine tune its fuzzy rule base for optimizing the system inference process. It combines the human like reasoning method of fuzzy systems based on fuzzy rules with the learning capability and connectionist structure of neural networks.

For fuzzy modelling different data partitioning techniques like clustering such as FCM clustering are used to obtain the partitions of the dataset to capture the internal trends of the input output data samples. These partitions or clusters are then used to construct the fuzzy if-then rules for the fuzzy model being build and then a method is used to fine tune the initial rule base to obtain the final rule base. The fuzzy modelling of ANFIS based on measured input-output data is generally performed using one of the three methods

Author α : Research scholar, Dept. of computer science and IT, University of Jammu, Jammu-180006, India.
e-mail: sharifa18mca@gmail.com
Author σ: Dept. of computer science and IT, University of Jammu, Jammu-180006, India. e-mail: vinodsharma@jammuuniversity.in

namely Grid Partitioning, Subtractive clustering and FCM clustering. The grid partitioning although an efficient method to partition the input space, has some disadvantages like curse of dimensionality of input, computation cost, exponential expansion of rule base etc. Data clustering is a handy alternative technique for fuzzy modelling where the clusters obtained from a clustering algorithm are used as a basis for fuzzy rule generation. Fuzzy identification using clustering consists of finding the clusters in the data space and using the obtained cluster centers to calculate the premise and consequent parts of the fuzzy rules. Therefore the accuracy of clustering determines the quality of the rule base and hence the performance of the resulting fuzzy model. Both the approaches viz. SC proposed by Chiu (1994) and FCM clustering technique proposed by Dunn (1974) and later improved by Bezdek et al. (1984) are efficient techniques for clustering data sets and hence effective modelling techniques. Recently, the kernel methods have achieved popularity for various classification and regression based problems. The accuracy of SC and FCM clustering is improved by incorporating kernel functions in the calculation of the distance measures between the data points during clustering process which results in more precise cluster centers (Kim et al., 2004; Qiang et al., 2004). Higher clustering accuracy is achieved as the kernel induced distance measures increase the data separability by using the higher dimensional space which reveals more precise data partitions.

In this paper we have a proposed novel modelling techniques for the fuzzy rule base construction of ANFIS based on kernel based SC (KSC) and kernel based FCM (KFCM) techniques. To build the prediction model the data set is first partitioned into clusters using kernel based clustering, the resulting cluster centers are then employed to build the initial fuzzy rule base for ANFIS and then the resulting rule base has been optimized using a hybrid learning algorithm consisting of standard Backpropagation and least square estimation. The effectiveness of the KSC and KFCM based ANFIS models has been tested on three business prediction problems namely qualitative bankruptcy prediction, sales forecasting and stock price prediction. A comparison with the ANFIS models based on original SC and FCM for these business prediction problems has also been presented.

This paper organization is as following. Section 2 deals with the review of the prior relevant research. In section 3 the research methodology is presented that gives the details of FCM and KFCM algorithms, provides an overview of ANFIS and fuzzy rule generation methods based on clustering and also presents and discusses the simulation results providing a comparison of performance with the ANFIS based on conventional FCM clustering and SC. Section 4 provides the concluding remarks on this study and various enhancements to this work.

II. Previous Work

Recently a number of studies (Yao et al., 2000; Dejan et al., 2011; Hossein et al., 2010; Kalhor et al., 2009; Suk et al., 2003) have addressed the problem of fuzzy model identification based on the data clustering algorithms. In several studies the kernel methods have also been employed along with various conventional clustering techniques for this purpose. Yang et al. (2008) proposed a novel method for fuzzy modelling of a Takagi-Sugeno system based on dual kernel-based method. The authors used a conventional FCM algorithm for partitioning data into various clusters. Then a kernel function independent of the parameter selection problem was used to locate the support vectors within each of the clusters. The experimental results from the study showed that the method lead to a fuzzy model with concise structure having good generalization capability. Also the performance of the system was not affected by the initial cluster number needed in FCM. Lukasik et al. (2008) presented a kernel-density gradient estimation technique for fuzzy rule extraction. The authors used clustering based on kernel density estimator. The cluster centers obtained on clustering were then used to construct the rule base. The assumption underlying the technique was that local maximum of a kernel estimator of a probability density function for m-dimensional data can be used as a basis for each cluster. But instead of the density function the authors used the gradient of the density function in cluster center identification. The neuro-fuzzy system based on this approach was experimented for non-linear function approximation and controller synthesis and showed good performance. Suga et al. (2006) used an iterative feature vector selection (FVS) based on kernel method to calculate membership function parameter values and the number of fuzzy rules for a Takagi-Sugeno fuzzy model. The kernel based FVS algorithm was used to obtain a basis of data space called feature vector into the feature space. This feature vector was then used as the center of a membership function in the antecedent part of the fuzzy rule. After finding the premise parts of the fuzzy rules, the coefficients of the consequent of the fuzzy rules were obtained using least square methods. The proposed system was applied to the modelling of a two input non-linear function which showed the effectiveness of the proposed fuzzy system for non-linear system modelling. Won et al. (2006) introduced a novel method based on kernel machine for fuzzy system modelling. The kernel machine was based on a Support Vector Machine, Feature Vector Selection and Relevance Vector Machine. The significance of the work was the method of reduction in the number of fuzzy rules by adjusting the parameter values or the

Global Journal of Computer Science and Technology (D) Volume XV Issue I Version I Year 2015

transformation matrix of the kernel function by employing a gradient descent based technique. The effectiveness of the system was shown by application to some benchmark non-linear problems.

Almost all of the above research studies proved that the kernel methods can be used to enhance the performance of the fuzzy rule based models but the use of KSC and KFCM for fuzzy modelling of popular ANFIS model was not explored which is undertaken in this paper.

III. Methodology

This study deals with the performance analysis of ANFIS built using kernel based clustering techniques viz. KSC and KFCM for business prediction problems. The prediction model for a problem is built in three stages: 1) dataset is partitioned into various clusters using one of these kernel based clustering techniques, 2) the cluster centers obtained from clustering are used to build the fuzzy rule base of ANFIS, and 3) the resulting ANFIS model is trained using the hybrid learning algorithm consisting of gradient descent method and least square method. The various techniques used have been discussed in the following sections.

a) *Techniques employed*

i. *Kernel based subtractive clustering*

The original SC is based on calculating the potential function called mountain value at each data point. It is an improved version of the mountain method and uses each input data point in the dataset as a potential cluster center rather than using grid based formulation in mountain clustering method thus leading to lower computational complexity for higher dimensional data sets. KSC was proposed by Kwang et al. (2004) as an improvement over conventional SC algorithm where kernel functions are employed in potential value calculation. In original subtractive clustering, for a dataset X={x_1, x_2, …,x_n}, the potential value at each data point x_i is given by:

$$\sum_{j=1}^{n} e^{-\alpha ||x_i - x_j||^2}, \quad \alpha = \frac{4}{r_a^2} \tag{1}$$

where r_a is a positive constant called cluster radius defining the range of influence of a cluster center along each data dimension and affects the number of clusters generated. The data point with the highest potential value P_1 is selected as the cluster center c_1. In order to find the subsequent cluster centers using the same procedure the potential value for each data point x_i is modified as:

$$P_i = P_i - P_1 * e^{-\beta ||x_i - c_1||^2}, \ \beta = \frac{4}{r_b^2}, \quad r_b = \eta r_a \tag{2}$$

Here r_b is a positive constant and η is the squash factor used to squash the potential values for the distant points to be considered as part of a cluster. It is evident that the reductions in potential values of data points near the newly found cluster center is more than the distant points and hence have a least chance of being selected as cluster centers.

Using kernel approach the kernel functions are employed in calculating the distance measure given by: $||x_i - x_j||^2$ and $||x_i - c_1||^2$ in Eqs. (1) and (2) so that the data points are mapped to a higher dimensional space which makes the dataset more distinctly separable resulting in more informative potential values. Therefore, the centers produced are more accurate and when used in fuzzy modelling can result in more useful fuzzy rule base for a fuzzy mode like ANFIS.

The basic notion in kernel methods is a non-linear mapping Ø to a higher dimensional space from the input space i.e. for a dataset X={x_1, x_2, …, x_n}:

$$\varnothing: x \rightarrow \varnothing(x) \tag{3}$$

Using this non-linear mapping the dot product '$x_i.x_j$' used as a similarity measure in various learning algorithms can be mapped to a more general measure: $\varnothing(x_i).\varnothing(x_j)$. This dot product in higher dimensional space is calculated using a kernel function $K(x_i,x_j)$ i.e.:

$$\varnothing(x_i).\varnothing(x_j) = K(x_i,x_j) \tag{4}$$

The distance measure $||x_i - x_j||^2$ in input space in terms of function Ø therefore is given by:

$$||x_i - x_j||^2 = ||\varnothing(x_i) - \varnothing(x_j)||^2 \tag{5}$$

where:

$$\begin{aligned} ||\varnothing(x_i) - \varnothing(x_j)||^2 &= (\varnothing(x_i) - \varnothing(x_j)) \cdot (\varnothing(x_i) - \varnothing(x_j)) \\ &= \varnothing(x_i).\varnothing(x_i) - 2\varnothing(x_i)\,\varnothing(x_j) + \varnothing(x_j).\varnothing(x_j) \\ &= K(x_i,x_i) - 2K(x_i,x_j) + K(x_j,x_j) \end{aligned} \tag{6}$$

Thus for KSC eq. (1) can be altered to incorporate kernel function by using eqs. (5) and (6):

$$\sum_{j=1}^{n} e^{-\alpha\left(K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j)\right)} \tag{7}$$

K(x, y) can be any kernel function like gaussian kernel, polynomial kernel, fisher kernel etc.

After a data point is selected as a cluster center in KSC, the potential function of other data points to find the subsequent centers is calculated as:

$$P_i = P_i - P^* * e^{-\left(\beta\ K(x_i, x_i) - 2K(x_i, x^*) + K(x^*, x^*)\right)} \tag{8}$$

where β is the positive constant in eq. (2) and x^* is the newly obtained cluster center with potential value P^*. After revising the potential of other data points, the data point with the highest potential is chosen as the second cluster center and the potential values of other data points are changed as in eq. (8). In general when nth cluster center $x_n{}^*$ is selected, the potential of other data points is revised as:

$$P_i = P_i - P_n^* * e^{-\beta\left(K(x_i, x_i) - 2K(x_i, x_n^*) + K(x_n^*, x_n^*)\right)} \tag{9}$$

After a data point with highest potential value is selected the following criteria is used to select the data point as the cluster center and to determine whether to repeat or terminate the clustering process:

If $P_n^* > \bar{\epsilon} P_1^*$ accept x_n^* as the next cluster center and repeat

else if $P_n^* < \underline{\epsilon} P_1^*$ reject x_n^* and stop the clustering process

else

let d_m = least distance between x_n^* and all earlier found cluster centers

if $\frac{d_m}{r_a} + \frac{P_n^*}{P_1^*} >= 1$ accept x_n^* as cluster center and repeat clustering

else reject x_n^* and set potential $P_n^* = 0$. Select the data point with the next highest potential as the cluster center and repeat the process.

end if

end if

where $\bar{\epsilon}$ is the Accept Ratio i.e. a threshold potential value below which the data point is rejected as the cluster center and $\underline{\epsilon}$ is the Reject ratio which specifies a threshold potential above which the data point is definitely accepted.

ii. *Kernel based FCM clustering*

KFCM was proposed by Qiang (2004) as an enhancement of the standard FCM clustering algorithm based on the use of kernel functions. For a dataset $X=\{x_1, x_2, \dots, x_n\}$, the conventional FCM algorithm calculates the fuzzy subsets of X by minimizing an objective function given by:

$$\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}||x_j - v_i||^2 \quad (10)$$

where *n* is the number of data points, *c* is the number of cluster centers, μ_{ij} is the membership of x_j in ith class, v_i is the ith cluster center and *m* is the quantity to control the fuzziness of clustering. In KFCM the distance measure is generalized by employing a non linear mapping Ø from input space to a higher dimensional space i.e.:

$$\emptyset: x \rightarrow \emptyset(x)$$

Therefore, as in KSC using (4) and (5) by the kernel approach the objective function in KFCM is given by:

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}||\emptyset(x_j) - \emptyset(v_i)||^2 \quad (11)$$

From eq. (6):

$$||\emptyset(x_j) - \emptyset(v_i)||^2 = K(x_j, x_j) - 2K(x_j, v_i) + K(v_i, v_i) \quad (12)$$

K(x, y) can be any kernel function for example gaussian kernel, polynomial kernel, fisher kernel etc. Using equation (12) eq. (11) becomes:

$$\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}(K(x_j,x_j) - 2K(x_j,v_i) + K(v_i,v_i)) \quad (13)$$

Gaussian function is a common kernel function given by:

$$K(x, y) = e^{(-||x-y||^2/\sigma^2)} \quad (14)$$

where K(x, x) = 1 and σ is an adjustable parameter. Using gaussian kernel function the eq. (13) becomes:

$$\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ij}^{m}(1 - K(x_j, v_i)) \quad (15)$$

Where

$$\mu_{ij} = \frac{(1/(1-K(x_j,v_i)))^{1/(m-1)}}{\sum_{k=1}^{c}(1/(1-K(x_j,v_k)))^{1/(m-1)}} \quad (16)$$

$$v_i = \frac{\sum_{j=1}^{n}\mu_{ij}^{m}K(x_j,v_i)\,x_j}{\sum_{j=1}^{n}\mu_{ij}^{m}K(x_j,v_i)} \quad (17)$$

Other kernel functions can also be used so that above equations can be modified accordingly.

Algorithm for KFCM

Step 1: set k=0, m > 1 and $\epsilon > 0$ for some positive constant.

Step 2: initialize the memberships μ_{ij}^{0}.

Step 3: I) Update all v_i^k using eq. (17).

II) Update all μ_{ij}^{k} using eq. (16).

If $\max\left(|\mu_{ij}^{k} - \mu_{ij}^{k-1}|\right) <= \epsilon$ Stop

else k=k + 1

go to step 3.

end if

iii. *Fuzzy rule base construction using clustering*

The fuzzy inference system has the capability of a non-linear system being modeled in terms of fuzzy if then rules. The fuzzy model identification therefore involves the determination of parameters for the premise membership functions and parameters in the consequences. Applying the clustering algorithm on the experimental dataset for the system to be modeled each of the resulting cluster centers essentially is an exemplary data point representing the system's characteristic behavior. Therefore, using clustering for fuzzy modelling each of the cluster centers is considered as a basis for a fuzzy rule for the initial rule base of the fuzzy inference system being modeled. Hence the number of cluster centers generated determines the number of the fuzzy rules for the modeled system.

The fuzzy model identification using data clustering techniques has been addressed in a number of studies (Han et al. 2010; Szymon 2008; Yao et al. 2000;). Babuska et al. (1994) provides an effective method for fuzzy rule generation from the FCM generated fuzzy clusters where premise membership functions are obtained using projection of fuzzy clusters which can be orthogonal or eigenvector projection. The consequent parameters using this method can be obtained using least square estimation. According to Degado et al. (1997) the antecedent and consequent parameters can be directly obtained from cluster centers instead of projections on domains of outputs and inputs.

A number of studies on fuzzy modelling using subtractive clustering have used the method presented in paper (Chiu 1994). With this method if k cluster centers $\{c_1,\ldots, c_k\}$ are generated in m-dimensional space, each of the vector c_i can be decomposed into two vectors X_i and Y_i where X_i represents the first n elements of c_i corresponding to the input variables and Y_i contains m-n output variables. For an input vector x the degree of fulfillment of rule i is given by:

$$\mu_i(x) = e^{-\alpha||x-X_i||^2} \qquad (18)$$

Where α is the positive constant used in eq. (1). The output vector y can be computed as:

$$y = \frac{\sum_{i=1}^{k} Y_i \mu_i}{\sum_{i=1}^{k} \mu_i} \qquad (19)$$

A typical fuzzy rule has the following form:

If x_1 is A_1 and … and x_n is A_n then y_1 is B_1 and … and y_n is B_n

Where x_i is the ith input variable, y_i is the ith output variable and A_i is the ith antecedent membership function and B_i is singleton. Each of the ith rule is determined by the cluster center c_i and each rule has multiple input variables and hence membership functions. If A_j is the jth membership function of a rule i, it is given by:

$$A_j(x) = e^{-\alpha(x-X_{ij})^2}$$

And the consequent B_i is given by:

$$B_j = Y_{ij} \qquad (21)$$

Where X_{ij} is the jth element of X vector and Y_{ij} is the jth element of Y vector of center c_i. This method of rule generation achieves significant accuracy if the Takagi-sugeno type fuzzy rules are used in which the consequent parameters are the linear combination of input variables (Chiu 1994).

iv. *ANFIS architecture*

ANFIS is an adaptive system that has the learning capability to optimize the performance based on finding the best parameters for the fuzzy rules within its rule base. Fig. 1 shows the architecture of ANFIS with two inputs x_1 and x_2 and a rule base consisting of consisting of two Sugeno type fuzzy rules:

If x_1 is A_1 and x_2 is B_1 then $f_1 = p_1x_1 + q_1x_2 + r_1$

If x_1 is A_2 and x_2 is B_2 then $f_2 = p_2x_1 + q_2x_2 + r_2$

The details of the functioning of each layer of the ANFIS are as follows:

Layer 1: This is the input layer and consists of nodes with adaptive node functions. Each node has an output equal to:

$$O_{1,i} = \mu A_i(x) \text{ for } i = 1,2 \qquad (22)$$

Here output of each node is the value of the membership function A of that node and $O_{k,l}$ is the node in the i-th position of the k-th layer.

Various types of membership function can be used, like gauss function, the bell-shaped function etc.

Layer 2: In this layer each node computes the product of incoming signals with output given by:

$$O_{2,i} = w_i = \mu A_i \mu B_i(y)\ ,\ i = 1,2 \qquad (23)$$

Layer 3: In this layer each j-th node computes the ratio of the firing strength of the j-th rule and the sum of all the firing strengths, with output:

$$O_{3,j} = w_j = \frac{w_j}{w1+w2}\ ,\ j = 1,2 \qquad (24)$$

Layer 4 : In this layer function for i-th node is:

$$O_{4,1} = \overline{w}_i f_i = \overline{w}_i(p_i x + q_i x + r_i) \qquad (25)$$

Layer 5 : This layer has a single node that computes the overall output as the sum of all incoming signals:

$$O_{5,1} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \qquad (26)$$

Where $O_{5,1}$ is the obtained output available to user.

For the optimization of the fuzzy rule base of ANFIS either standard back propagation or the hybrid learning algorithm can be used. The hybrid learning is mostly used and is an effective technique which uses gradient descent method to update the premise parameters of the fuzzy rules and LSE is used to identify the optimal consequent parameters.

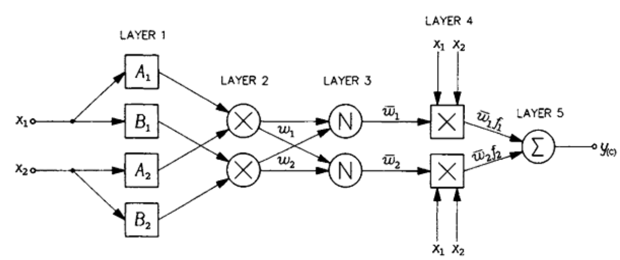


Fig. 1: ANFIS architecture

b) *Datasets used*

In order to test the effectiveness of the KSC and KFCM based ANFIS models we used three datasets one for each of the three business prediction problems viz. qualitative bankruptcy prediction, sales forecasting and stock price prediction. These problems were selected as these are the popular research problems in business field nowadays and ANFIS model has been extensively applied to these problems in numerous studies successfully. For the purpose of qualitative bankruptcy prediction the dataset has been collected from one of the largest banks in Korea consisting of 260 services and manufacturing companies for period 2001-2002. This dataset has also been used by Jong et al. (2003) to discover the bankruptcy decision rules based on experts, decisions using genetic algorithm. Half of the companies in this dataset are bankrupt and other half non-bankrupt according to the classification done by the experts having an experience of nine years in this area. This dataset is based on six qualitative risk factors as listed in fig. 1 (a). Each of the factors is assigned an appropriate level viz. positive (P) or negative (N) or average (A). The output is the class of the company i.e. bankrupt (B) and non-bankrupt (NB) as shown in Fig. 1 (b).

Risk Factor	Level
Industrial risk	(P,A,N)
Financial flexibility	(P,A,N)
Management risk	(P,A,N)
Credibility	(P,A,N)
Competitiveness	(P,A,N)
Operating risk	(P,A,N)

Table 1 : Six risk factors for qualitative bankruptcy

The dataset used in the stock price prediction is the daily BSE stock data obtained from Yahoo finance for a period eight years from 1/2/2007 to 30/12/2014 consisting of 1966 records. The dataset is composed of five fundamental stock quantities (open price, maximum price, minimum price, stock trading volume and close price). We have used 70% consisting of 1179 records of this dataset for training, 20% consisting of 394 records as checking data and rest 20% consisting of 394 records as testing data for all the ANFIS models.

For sales forecasting problem we have used the sales data of chocolate items of a major distributor in Jammu city (India) collected for a period of five months from 1/12/2014 to 30/12/2015 consisting of 150 records. The dataset has four attributes viz. present day sale amount, maximum daily temperature, minimum temperature and next day sale. The temperature attributes have been included as the sale of chocolate items is affected by the temperature during a period.

c) *Experimental results*

In this section simulation results of the application of the KSC and KFCM based ANFIS models for qualitative bankruptcy prediction, sales forecasting and stock price prediction have been presented. In all the experiments the performance of the KFCM and KSC based ANFIS models has been compared with the conventional FCM and SC based ANFIS models respectively. For all the ANFIS models online learning has been used with an initial step size 0.01 and gaussian membership function given by:

$$e^{-\frac{1}{2}\left(\frac{x-c}{\sigma}\right)^2} \tag{27}$$

has been used for input variables in the first layer of ANFIS . The parameter c in eq. (27) is the center of membership function, σ determines the width of the membership function and x is the input variable. The values used for various parameters explained in section 3.1.1 for both the SC and KSC algorithms were: accept ratio = .5, reject ratio = .15 and squash factor = 1.25 for all the simulation examples. For all the experiments we have used gaussian kernel function defined in eq. (13) as the kernel for the implementation of both KSC and KFCM technique. All the experiments were performed in MATLAB R2013a environment.

i. *Qualitative bankruptcy prediction*

The first business prediction problem considered is the qualitative bankruptcy prediction. Out of the total 260 records of the dataset used for this problem, 75% has been used for training, 15% as checking data and 20% as testing data for both KSC and KFCM clustering based ANFIS models.

The cluster radius r_a defined in eq. (1) for subtractive clustering takes values between 0 and 1 and strongly affects the number of clusters generated. A large value for r_a results in lesser number of clusters and therefore lesser fuzzy rules in rule base and vice versa. The number of fuzzy rules in the system in turn affects the forecasting performance of the system. Therefore finding the optimum value for r_a is important for a problem under consideration.

A value of .5 for parameter r_a for both KSC and SC produced the ANFIS systems with lesser training, checking and testing errors and therefore was optimum for this dataset. KSC resulted in 24 clusters so that resulting ANFIS system had 24 fuzzy rules with 5 gauss membership functions associated with each input variable. The SC resulted in 49 clusters and the ANFIS system based on it had 49 rules. Therefore, KSC resulted in a less complex ANFIS with lesser number of parameters to be optimized than ANFIS based on SC. Fig. 2 shows the training and checking RMSE for both the KSC and SC based ANFIS models. After the systems were trained for 150 epochs for this dataset the root mean square error (RMSE) for both the systems did not change significantly. Fig. 2 demonstrates that the KSC based ANFIS performs better than the SC based ANFIS in terms of the training and checking RMSE errors for this dataset.

Global Journal of Computer Science and Technology (D) Volume XV Issue I Version I Year 2015

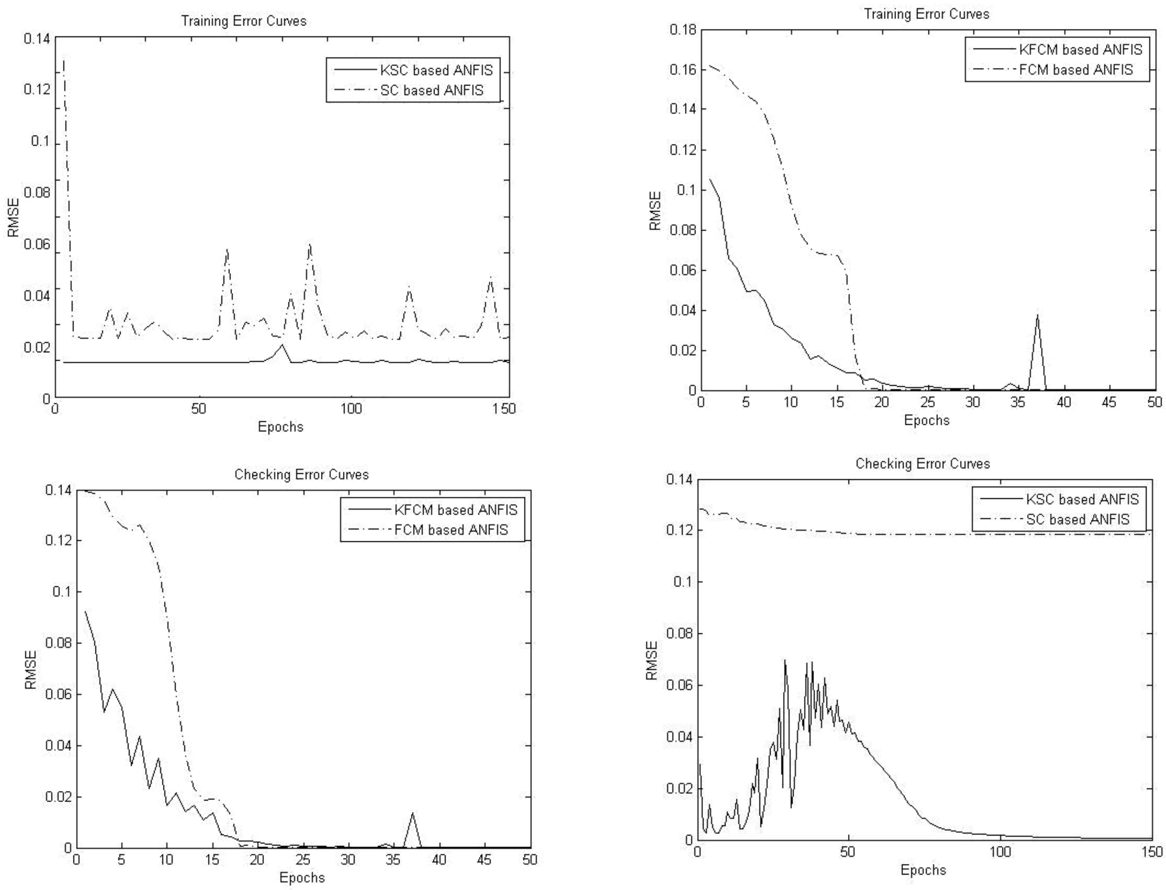


Fig. 2 : RMSE error curves for Qualitative Bankruptcy prediction (a) training error curves for KSC and SC based ANFIS. (b) Checking error curves for KSC and SC based ANFIS. (c) Training error curve for KFCM and FCM based ANFIS. (d) Checking error curve for KFCM and FCM based ANFIS

The number of clusters used as the input parameter to the FCM and KFCM algorithms results in an equal number of fuzzy rules for the ANFIS system to be implemented which in turn considerably affects the performance of the system. Both the FCM and KFCM based ANFIS systems containing three fuzzy rules with three membership functions of gaussian type for each input attribute showed the lowest RMSE. On training for 50 epochs the RMSE for both the systems remained constant. Fig. 2 shows the training and checking RMSE curves for both the systems where it is evident that KFCM based ANFIS is better than FCM based one for this problem.

A value of 150 was used for the adjustable parameter σ of gaussian kernel defined in (13) for both KSC and KFCM so that the resulting ANFIS models gave the lowest training, checking and testing errors.

ii. *Sales Forecasting*

The next simulation example is the sales forecasting based on the sales data collected by authors. 75% of this dataset containing total 150 records has been used for the training the models, 15% for checking and 25% for testing.

With parameter $r_a = .5$, the KSC resulted in 4 clusters for this dataset. The resulting ANFIS system contained four fuzzy rules with four membership functions of gaussian type associated with each input. The SC with $r_a = .5$ resulted in 2 fuzzy rules for this dataset so that the resulting SC based ANFIS contained two fuzzy rules in rule base. In case of KSC values near 10 for the adjustable parameter σ of gaussian kernel defined in (13) resulted in lowest errors. After training both the systems for 200 epochs we had training RMSE $= .0877$ and checking RMSE $= .01992$ for KSC based system and for SC based system we had training RMSE $= .117444$ and checking RMSE $= .186907$ as shown in table 1.

The KFCM and FCM based ANFIS with 2 fuzzy rules resulted in highest performance. A value of 150 for the parameter σ of the gaussian kernel function resulted in lowest error for the KFCM based ANFIS. After training for 200 epochs for KFCM based system we had training RMSE $= .117273$ and checking RMSE $= .187058$ and for SC based ANFIS training RMSE $= .10862$ and checking RMSE $= .18603$. The performance comparison of these systems on testing data based on RMSE and Mean average percentage error (MAPE) has been presented in Table 2.

iii. *Stock price prediction*

We have used the KFCM and KSC based ANFIS models for predicting the close price of the day based on the daily open price, trading volume, maximum and minimum stock price.

With $r_a = .5$ the KSC based ANFIS system has two fuzzy rules with two membership functions associated with each input and the conventional SC based system has three fuzzy rules with three membership functions associated with each input variable. Both the systems were trained for 200 epochs after which the RMSE remained constant. The training and checking RMSE curves for both the ANFIS systems are provided in fig. 3. After testing the KSC based ANFIS we had RMSE = .0034 and APE = 1.041%. For SC based ANFIS we had RMSE = .0037 and APE = 1.3961%.

The KFCM and FCM algorithm with 10, 5 and 3 cluster centers were used resulting in ANFIS systems with 10, 5 and 3 fuzzy rules respectively. But the FCM and KFCM based ANFIS containing 3 fuzzy rules were found to give the lowest errors. Fig. 3 shows the training and checking RMSE curves for these systems each containing 3 fuzzy rules. On testing the KFCM based ANFIS system resulted in RMSE = .0034 and APE = .7060% and the FCM based system showed RMSE = .0035 and .7471%.

For both KFCM and KSC algorithms a value of 150 for parameter σ for gaussian kernel function resulted in ANFIS systems with lowest errors.

IV. Conclusion

The conventional subtractive and FCM clustering techniques have been successfully used for fuzzy modelling. But the kernel based variations of these algorithms have shown better clustering accuracy by using higher dimensional space which results in better data space partitioning. Therefore when used in fuzzy modelling these techniques can result in a more useful fuzzy rule base for a fuzzy logic basis system. In this paper we have used the kernel based variants of these algorithms for extracting rules for initial fuzzy rule base for popular neuro-fuzzy model ANFIS. We used the kernel clustering based ANFIS models for three well known business prediction problems. For all the experiments the kernel based methods resulted in optimum number of fuzzy rules in the rule base of ANFIS, giving a lesser complex system. Moreover, the performance of these systems was mostly better in terms of training, checking and testing errors than the ANFIS models based on conventional subtractive and FCM clustering methods for these forecasting problems.

In this study we have used the gaussian kernel function for all the experiments. The major issue in using the kernel methods is to select the kernel function to be used, this work can be extended by using the multiple kernel based clustering approach which can overcome the problem of selecting the best kernel function for a particular data set. A performance comparison between the KFCM and KSC based ANFIS may be explored. Furthermore we have only considered ANFIS model which is currently most popular neuro-fuzzy system but other fuzzy models like type-2 fuzzy models can also be considered. Grid partitioning is also used for fuzzy modelling in ANFIS and gives satisfactory prediction accuracy but in this study we have not compared the performance of the kernel clustering based ANFIS models with such systems.

References Références Referencias

1. Zadeh, L.A. (1965), Fuzzy sets, Information and Control, vol. 8, pp. 338-353.
2. Takagi, T., Sugeno, M. (1985), Fuzzy identification of systems and its applications to modelling and control, IEEE Trans. Syst., Man, Cybern., vol. 15 pp. 116-132.
3. Mamdani, E. H. (1976), Application of fuzzy algorithms for control of simple dynamic plant, IEEE Proceedings of Institution Electrical Engineers, vol. 121, pp. 1585-1588.
4. Tong, R. M., Beck M. B., Latten, A. (1980), Fuzzy control of the activated sludge wastewater treatment process, Automatica, vol. 16, pp. 659-701.
5. Larsen, P. M. (1980), Industrial application of fuzzy logic control. Int. J. Man-Machine Studies, vol. 12, pp. 3-10.
6. Jyh-Shing Roger Jang (1993), ANFIS : Adaptive-Network-Based Fuzzy Inference System, IEEE Transactions On System, Man, And Cybernetics, vol. 23, pp. 665-685.
7. Dae-Won Kim, Ki Young Lee, Doheon Lee, Kwang H. Lee (2004), A kernel-based subtractive clustering method Pattern Recognition Letters, Elsevier, doi:10.1016/j. patrec. 2004.10.001.
8. Stephen chiu 1994), Fuzzy model identification based on cluster estimation, Journal of Intelligent and Fuzzy Systems, IOS press, vol. 2, pp. 267-278.
9. James, C., Bezdek, Robert Ehrlich, William Full (1984), FCM: THE FUZZY c-MEANS CLUSTERING ALGORITHM, Computers & Geosciences, vol. 10 pp. 191-203.
10. Dao- Qiang, Zhang Song, Can Chen (2004), A novel kernelized fuzzy C- means algorithm with application in medical image segmentation, Artificial intelligence in medicine, doi: http:// dx. doi.org /10.1016/j.artmed.2004.01.01

11. R., Babuska, H. B.,Verbruggen (1994), A new identification method for linguistic fuzzy models, at Open Meeting of the FALCON Working Group, Aachen,.
12. Degado, M., Gomez, A. F., Martin, F. (1997), Some Methods to Model Fuzzy Systems for Inference Purposes. International Journal of Approximate Reasoning, vol. 16, pp. 377-391.
13. Myoung Jong Kim and Ingoo Han (2003), The Discovery of experts' decision rules from Qualitative bankruptcy data using genetic algorithms, Expert Systems with Applications, Elsevier, vol. 25, pp. 637–646.
14. Pu Han, Jian-Zhong Shi, Dong-Feng Wang and Song-Ming Jiao (2010), FCM clustering algorithm for T-S Fuzzy model identification, International Conference on Machine Learning and Cybernetics, doi:10.1109/ICMLC.2010.5580478.
15. Szymon Łukasik, Piotr, A., Małgorzata Charytanowicz and Piotr Kulczycki (2008), Fuzzy Models Synthesis with Kernel-Density-Based Clusterin Algorithm, Fifth International Conference on Fuzzy Systems and Knowledge Discovery, DOI:10.1109/FSKD.2008.139.
16. Yao, J., Dash, M., Tan, S. T., Liu, H. (2000), Entropy-based fuzzy clustering and fuzzy modeling,Fuzzy Sets and Systems, Elsevier, vol. 113, pp. 381-388.
17. Dejan Dovžan and Igor Škrjanc (2011), Recursive fuzzy c-means clustering for recursive fuzzy identification of time-varying processes, ISA Transactions, Elsevier, vol. 50, pp.159–169.
18. Hossein Soleimani B., Caro Lucas and Babak N. Araabi (2010), Recursive Gath–Geva clustering as a basis for evolving neuro-fuzzy modeling, Evolving Systems, Springer-Verlag, vol. 1, pp. 59-71.
19. Kalhor, A. and Lucas C. (2009), Online Identification of a neuro-fuzzy model through indirect fuzzy clustering of data space, IEEE International Conference on Fuzzy Systems, DOI:10.1109/FUZZY.2009.5277139
20. Kim Sung-Suk, Kwak Keun-Chang, Ryu Jeong-Woong and Chun Myung-Geun (2003), A Neuro-Fuzzy Modelling using the Hierarchical Clustering and Gaussian Mixture Model, Journal of Korean Institute of Intelligent Systems, vol. 13, pp. 512-519.
21. Jongcheol Kim, Suga, Y. and Sangchul Won (2006), New Approaches to Fuzzy Inference System Using Kernel Machines, Japan Society For Fuzzy Theory and Intelligent Informatics, doi:10.14864/softscis.2006.0.1934.0
22. Szymon Łukasik, Piotr A. Kowalski, Małgorzata Charytanowicz and Piotr Kulczycki (2008), Fuzzy Model Synthesis with Kernel-Density-Based Clustering Algorithm, Fifth International Conference on Fuzzy Systems and Knowledge Discovery, doi: 10.1109/FSKD.2008.139.
23. Jongcheol Kim, Suga, Y. and Sangchul Won (2006), A new approach to fuzzy modelling of Nonlinear dynamic systems with noise, relevance Vector learning mechanism, IEEE Transactions on Fuzzy System, doi: 10. 1109 / TFUZZ.2005.864083.
24. Celikyilmaz, A., Turksen, I.B. (2007), Enhanced Fuzzy System Models With Improved Fuzzy Clustering Algorithm: IEEE Transactions on FuzzySystems, doi:10.1109/TFUZZ.2007.905919.
25. Wei Li and Yupu Yang(2008), A new approach to TS fuzzy modelling using dual kernel-based learning machines, Neurocomputing, 71 (2008), PP. 3660–3665.
26. Dunn, J.C. (1974), Some recent investigations of a new fuzzy partition algorithm and its application to pattern classification problems. J. of Cybernetics, vol. 4, pp. 1–15.

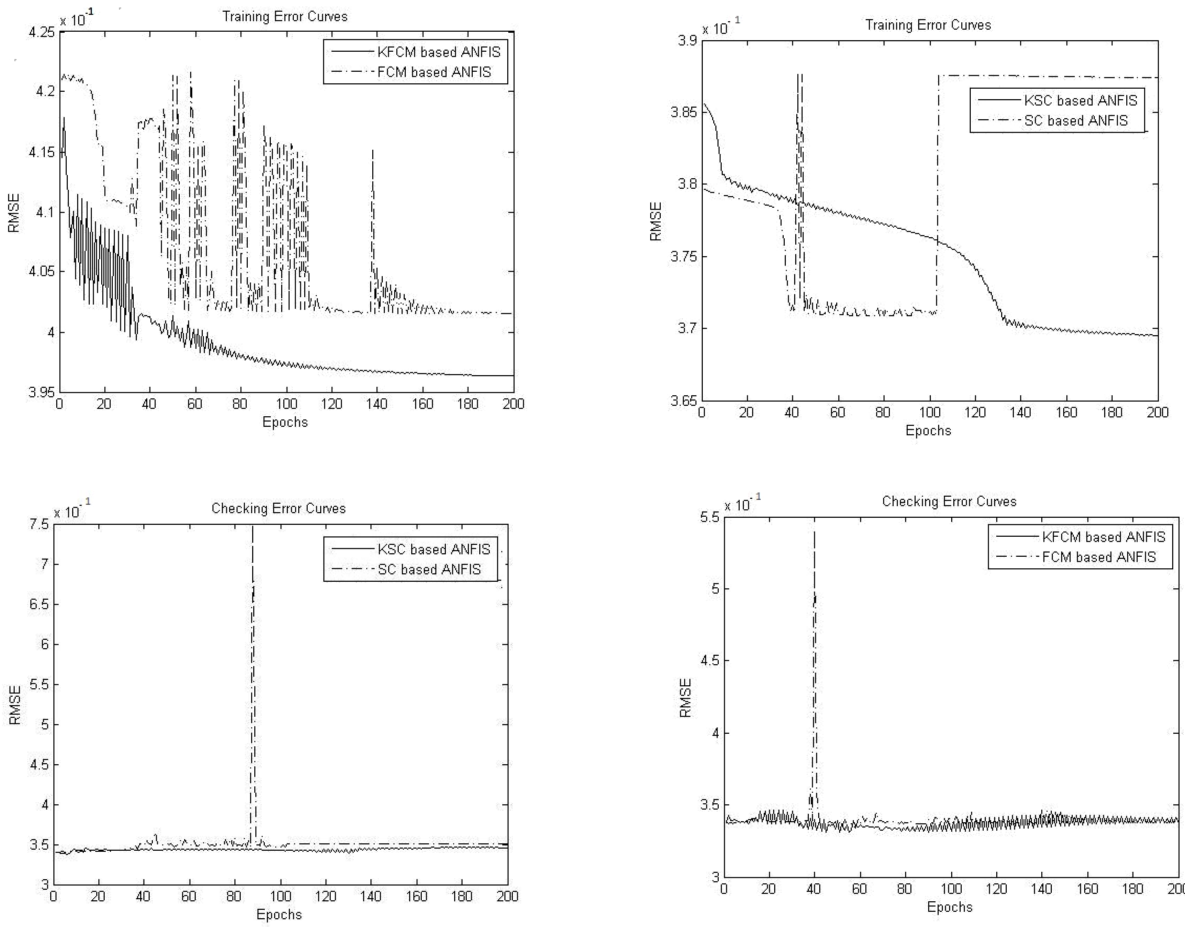


Fig. 3 : RMSE error curves for Stock price prediction (a) training error curves for KSC and SC based ANFIS. (b) Checking error curves for KSC and SC based ANFIS. (c) Training error curve for KFCM and FCM based ANFIS. (d) Checking error curve for KFCM and FCM based ANFIS

Global Journals Inc. (US) Guidelines Handbook 2015

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

Global Journals Incorporate (USA) is accredited by Open Association of Research Society (OARS), U.S.A and in turn, awards “FARSC” title to individuals. The 'FARSC' title is accorded to a selected professional after the approval of the Editor-in-Chief/Editorial Board Members/Dean.

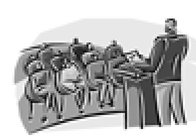
- The “FARSC” is a dignified title which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

FARSC accrediting is an honor. It authenticates your research activities. After recognition as FARSC, you can add 'FARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, and Visiting Card etc.

The following benefits can be availed by you only for next three years from the date of certification:

FARSC designated members are entitled to avail a 40% discount while publishing their research papers (of a single author) with Global Journals Incorporation (USA), if the same is accepted by Editorial Board/Peer Reviewers. If you are a main author or co-author in case of multiple authors, you will be entitled to avail discount of 10%.

Once FARSC title is accorded, the Fellow is authorized to organize a symposium/seminar/conference on behalf of Global Journal Incorporation (USA).The Fellow can also participate in conference/seminar/symposium organized by another institution as representative of Global Journal. In both the cases, it is mandatory for him to discuss with us and obtain our consent.

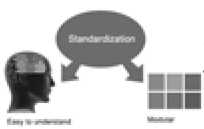


You may join as member of the Editorial Board of Global Journals Incorporation (USA) after successful completion of three years as Fellow and as Peer Reviewer. In addition, it is also desirable that you should organize seminar/symposium/conference at least once.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The FARSC can go through standards of OARS. You can also play vital role if you have any suggestions so that proper amendment can take place to improve the same for the benefit of entire research community.

As FARSC, you will be given a renowned, secure and free professional email address with 100 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.



The FARSC will be eligible for a free application of standardization of their researches. Standardization of research will be subject to acceptability within stipulated norms as the next step after publishing in a journal. We shall depute a team of specialized research professionals who will render their services for elevating your researches to next higher level, which is worldwide open standardization.

The FARSC member can apply for grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A. Once you are designated as FARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria. After certification of all your credentials by OARS, they will be published on your Fellow Profile link on website https://associationofresearch.org which will be helpful to upgrade the dignity.

The FARSC members can avail the benefits of free research podcasting in Global Research Radio with their research documents. After publishing the work, (including published elsewhere worldwide with proper authorization) you can upload your research paper with your recorded voice or you can utilize chargeable services of our professional RJs to record your paper in their voice on request.

The FARSC member also entitled to get the benefits of free research podcasting of their research documents through video clips. We can also streamline your conference videos and display your slides/ online slides and online research video clips at reasonable charges, on request.

The FARSC is eligible to earn from sales proceeds of his/her researches/reference/review Books or literature, while publishing with Global Journals. The FARSC can decide whether he/she would like to publish his/her research in a closed manner. In this case, whenever readers purchase that individual research paper for reading, maximum 60% of its profit earned as royalty by Global Journals, will be credited to his/her bank account. The entire entitled amount will be credited to his/her bank account exceeding limit of minimum fixed balance. There is no minimum time limit for collection. The FARSC member can decide its price and we can help in making the right decision.

The FARSC member is eligible to join as a paid peer reviewer at Global Journals Incorporation (USA) and can get remuneration of 15% of author fees, taken from the author of a respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

The ' MARSC ' title is accorded to a selected professional after the approval of the Editor-in-Chief / Editorial Board Members/Dean.
The “MARSC” is a dignified ornament which is accorded to a person’s name viz. Dr. John E. Hall, Ph.D., MARSC or William Walldroff, M.S., MARSC.

MARSC accrediting is an honor. It authenticates your research activities. After becoming MARSC, you can add 'MARSC' title with your name as you use this recognition as additional suffix to your status. This will definitely enhance and add more value and repute to your name. You may use it on your professional Counseling Materials such as CV, Resume, Visiting Card and Name Plate etc.

The following benefitscan be availed by you only for next three years from the date of certification.

MARSC designated members are entitled to avail a 25% discount while publishing their research papers (of a single author) in Global Journals Inc., if the same is accepted by our Editorial Board and Peer Reviewers. If you are a main author or co-author of a group of authors, you will get discount of 10%.

As MARSC, you will be given a renowned, secure and free professional email address with 30 GB of space e.g. johnhall@globaljournals.org. This will include Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.

The MARSC member can apply for approval, grading and certification of standards of their educational and Institutional Degrees to Open Association of Research, Society U.S.A.

Once you are designated as MARSC, you may send us a scanned copy of all of your credentials. OARS will verify, grade and certify them. This will be based on your academic records, quality of research papers published by you, and some more criteria.

It is mandatory to read all terms and conditions carefully.

Auxiliary Memberships

Institutional Fellow of Open Association of Research Society (USA)-OARS (USA)

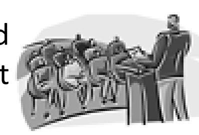
Global Journals Incorporation (USA) is accredited by Open Association of Research Society, U.S.A (OARS) and in turn, affiliates research institutions as "Institutional Fellow of Open Association of Research Society" (IFOARS).

The "FARSC" is a dignified title which is accorded to a person's name viz. Dr. John E. Hall, Ph.D., FARSC or William Walldroff, M.S., FARSC.

The IFOARS institution is entitled to form a Board comprised of one Chairperson and three to five board members preferably from different streams. The Board will be recognized as "Institutional Board of Open Association of Research Society"-(IBOARS).

The Institute will be entitled to following benefits:

The IBOARS can initially review research papers of their institute and recommend them to publish with respective journal of Global Journals. It can also review the papers of other institutions after obtaining our consent. The second review will be done by peer reviewer of Global Journals Incorporation (USA) The Board is at liberty to appoint a peer reviewer with the approval of chairperson after consulting us.
The author fees of such paper may be waived off up to 40%.



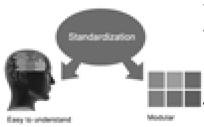
The Global Journals Incorporation (USA) at its discretion can also refer double blind peer reviewed paper at their end to the board for the verification and to get recommendation for final stage of acceptance of publication.

The IBOARS can organize symposium/seminar/conference in their country on behalf of Global Journals Incorporation (USA)-OARS (USA). The terms and conditions can be discussed separately.

The Board can also play vital role by exploring and giving valuable suggestions regarding the Standards of "Open Association of Research Society, U.S.A (OARS)" so that proper amendment can take place for the benefit of entire research community. We shall provide details of particular standard only on receipt of request from the Board.

The board members can also join us as Individual Fellow with 40% discount on total fees applicable to Individual Fellow. They will be entitled to avail all the benefits as declared. Please visit Individual Fellow-sub menu of GlobalJournals.org to have more relevant details.

We shall provide you intimation regarding launching of e-version of journal of your stream time to time. This may be utilized in your library for the enrichment of knowledge of your students as well as it can also be helpful for the concerned faculty members.



After nomination of your institution as "Institutional Fellow" and constantly functioning successfully for one year, we can consider giving recognition to your institute to function as Regional/Zonal office on our behalf.

The board can also take up the additional allied activities for betterment after our consultation.

The following entitlements are applicable to individual Fellows:

Open Association of Research Society, U.S.A (OARS) By-laws states that an individual Fellow may use the designations as applicable, or the corresponding initials. The Credentials of individual Fellow and Associate designations signify that the individual has gained knowledge of the fundamental concepts. One is magnanimous and proficient in an expertise course covering the professional code of conduct, and follows recognized standards of practice.

Open Association of Research Society (US)/ Global Journals Incorporation (USA), as described in Corporate Statements, are educational, research publishing and professional membership organizations. Achieving our individual Fellow or Associate status is based mainly on meeting stated educational research requirements.

Disbursement of 40% Royalty earned through Global Journals : Researcher = 50%, Peer Reviewer = 37.50%, Institution = 12.50% E.g. Out of 40%, the 20% benefit should be passed on to researcher, 15 % benefit towards remuneration should be given to a reviewer and remaining 5% is to be retained by the institution.

We shall provide print version of 12 issues of any three journals [as per your requirement] out of our 38 journals worth $ 2376 USD.

Other:

The individual Fellow and Associate designations accredited by Open Association of Research Society (US) credentials signify guarantees following achievements:

- The professional accredited with Fellow honor, is entitled to various benefits viz. name, fame, honor, regular flow of income, secured bright future, social status etc.

- In addition to above, if one is single author, then entitled to 40% discount on publishing research paper and can get 10%discount if one is co-author or main author among group of authors.
- The Fellow can organize symposium/seminar/conference on behalf of Global Journals Incorporation (USA) and he/she can also attend the same organized by other institutes on behalf of Global Journals.
- The Fellow can become member of Editorial Board Member after completing 3yrs.
- The Fellow can earn 60% of sales proceeds from the sale of reference/review books/literature/publishing of research paper.
- Fellow can also join as paid peer reviewer and earn 15% remuneration of author charges and can also get an opportunity to join as member of the Editorial Board of Global Journals Incorporation (USA)
- • This individual has learned the basic methods of applying those concepts and techniques to common challenging situations. This individual has further demonstrated an in–depth understanding of the application of suitable techniques to a particular area of research practice.

Note :

"

- In future, if the board feels the necessity to change any board member, the same can be done with the consent of the chairperson along with anyone board member without our approval.
- In case, the chairperson needs to be replaced then consent of 2/3rd board members are required and they are also required to jointly pass the resolution copy of which should be sent to us. In such case, it will be compulsory to obtain our approval before replacement.
- In case of "Difference of Opinion [if any]" among the Board members, our decision will be final and binding to everyone. "

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION **<u>(Must be strictly followed)</u>**

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, “Abstract” word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript’s Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications.

Research letters: The letters are small and concise comments on previously published matters.

5. STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

You must strictly follow above Author Guidelines before submitting your paper or else we will not at all be responsible for any corrections in future in any of the way.

Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

TECHNIQUES FOR WRITING A GOOD QUALITY RESEARCH PAPER:

1. Choosing the topic: In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. Pick a good study spot: To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. Think technically: Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page
- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.

- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.
- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.
- Do not give permission to anyone else to "PROOFREAD" your manuscript.
- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation) by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

Index

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350