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

Global Journal

of Computer Science and Technology: C

Software & Data Engineering

Risk Reduction using PCA

Job Stress in Software Companies

Highlights

Optimized Anomaly based Risk

PCA based Genetic Classifier

Discovering Thoughts, Inventing Future

VOLUME 14 ISSUE 7 VERSION 1.0

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Global Journal of Computer Science and Technology: C
Software & Data Engineering

Volume 14 Issue 7 (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)

Integrated Editorial Board
(Computer Science, Engineering, Medical, Management, Natural Science, Social Science)

Dr. Bart Lambrecht
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

Dr. Pina C. Sanelli
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

Dr. Roberto Sanchez
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

Dr. Wen-Yih Sun
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

Dr. Michael R. Rudnick
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

Dr. Bassey Benjamin Esu
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Aziz M. Barbar, Ph.D.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

Contents of the Issue

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 7 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

An under-Sampled Approach for Handling Skewed Data Distribution using Cluster Disjuncts

By Syed Ziaur Rahman, Dr. G Samuel Vara Prasad Raju & Dr. Ali MirzaMahmood

Andhra University, India

Abstract- In Data mining and Knowledge Discovery hidden and valuable knowledge from the data sources is discovered. The traditional algorithms used for knowledge discovery are bottle necked due to wide range of data sources availability. Class imbalance is a one of the problem arises due to data source which provide unequal class i.e. examples of one class in a training data set vastly outnumber examples of the other class(es). Researchers have rigorously studied several techniques to alleviate the problem of class imbalance, including resampling algorithms, and feature selection approaches to this problem. In this paper, we present a new hybrid frame work dubbed as Majority Under-sampling based on Cluster Disjunct (MAJOR_CD) for learning from skewed training data. This algorithm provides a simpler and faster alternative by using cluster disjunct concept. We conduct experiments using twelve UCI data sets from various application domains using five algorithms for comparison on six evaluation metrics. The empirical study suggests that MAJOR_CD have been believed to be effective in addressing the class imbalance problem.

Keywords : *classification, class imbalance, cluster disjunct, under sampling, MAJOR_CD.*

GJCST-C Classification : *H.2.8*

Strictly as per the compliance and regulations of:

An under-Sampled Approach for Handling Skewed Data Distribution using Cluster Disjuncts

Syed Ziaur Rahman [α], Dr. G Samuel Vara Prasad Raju [σ] & Dr. Ali Mirza Mahmood [ρ]

Abstract- In Data mining and Knowledge Discovery hidden and valuable knowledge from the data sources is discovered. The traditional algorithms used for knowledge discovery are bottle necked due to wide range of data sources availability. Class imbalance is a one of the problem arises due to data source which provide unequal class i.e. examples of one class in a training data set vastly outnumber examples of the other class(es). Researchers have rigorously studied several techniques to alleviate the problem of class imbalance, including resampling algorithms, and feature selection approaches to this problem. In this paper, we present a new hybrid frame work dubbed as Majority Under-sampling based on Cluster Disjunct (MAJOR_CD) for learning from skewed training data. This algorithm provides a simpler and faster alternative by using cluster disjunct concept. We conduct experiments using twelve UCI data sets from various application domains using five algorithms for comparison on six evaluation metrics. The empirical study suggests that MAJOR_CD have been believed to be effective in addressing the class imbalance problem.

Keywords: classification, class imbalance, cluster disjunct, under sampling, MAJOR_CD.

I. Introduction

A dataset is class imbalanced if the classification categories are not approximately equally represented. The level of imbalance (ratio of size of the majority class to minority class) can be as huge as 1:99 [1]. It is noteworthy that class imbalance is emerging as an important issue in designing classifiers [2], [3], [4]. Furthermore, the class with the lowest number of instances is usually the class of interest from the point of view of the learning task [5]. This problem is of great interest because it turns up in many real-world classification problems, such as remote-sensing [6], pollution detection [7], risk management [8], fraud detection [9], and especially medical diagnosis [10]–[13].

There exist techniques to develop better performing classifiers with imbalanced datasets, which are generally called Class Imbalance Learning (CIL) methods. These methods can be broadly divided into two categories, namely, external methods and internal methods. External methods involve preprocessing of training datasets in order to make them balanced, while internal methods deal with modifications of the learning algorithms in order to reduce their sensitiveness to class imbalance [14]. The main advantage of external methods as previously pointed out, is that they are independent of the underlying classifier.

Whenever a class in a classification task is under represented (i.e., has a lower prior probability) compared to other classes, we consider the data as imbalanced [15], [16]. The main problem in imbalanced data is that the majority classes that are represented by large numbers of patterns rule the classifier decision boundaries at the expense of the minority classes that are represented by small numbers of patterns. This leads to high and low accuracies in classifying the majority and minority classes, respectively, which do not necessarily reflect the true difficulty in classifying these classes. Most common solutions to this problem balance the number of patterns in the minority or majority classes.

Resampling techniques can be categorized into three groups. Under-sampling methods, which create a subset of the original data-set by eliminating instances (usually majority class instances); oversampling methods, which create a superset of the original data-set by replicating some instances or creating new instances from existing ones; and finally, hybrids methods that combine both sampling methods. Among these categories, there exist several different proposals; from this point, we only center our attention in those that have been used in under sampling. Either way, balancing the data has been found to alleviate the problem of imbalanced data and enhance accuracy [15], [16], [17]. Data balancing is performed by, e.g., oversampling patterns of minority classes either randomly or from areas close to the decision boundaries. Interestingly, random oversampling is found comparable to more sophisticated oversampling methods [17]. Alternatively, under-sampling is performed on majority classes either randomly or from areas far away from the decision boundaries. We note that random under-sampling may remove significant patterns and random oversampling may lead to over-

Author α: Research Scholar, Department of CS & SE, Andhra University, Vishakhapatnam, Andhra Pradesh India.
email: sdzrahman@gmail.com
Author σ: Professor in CS & SE department of Andhra University, Vishakhapatnam, Andhra Pradesh India.
Author ρ: Associate Professor in DMS SVH College of Engineering, Machlipatnam, Andhra Pradesh India.

fitting, so random sampling should be performed with care. We also note that, usually, selective under sampling of majority classes is more accurate than oversampling of minority class. In this paper, we are laying more stress to propose an external class imbalance learning method for solving the class imbalance problem by performing selective under sampling of majority class.

This paper is organized as follows. Section II presets the problem of cluster disjuncts. Section III briefly reviews the data balancing problems and its measures and in Section IV, we discuss the proposed method of MAJOR_CD (Majority Under-sampling based on Cluster Disjunct) for class imbalance learning. Section V presents the imbalanced datasets used to validate the proposed method, while In Section VI, we present the experimental setting and In Section VII discuss, in detail, the classification results obtained by the proposed method and compare them with the results obtained by different existing methods and finally, in Section VIII we conclude the paper.

II. Problem of Cluster Disjunct

In Class Imbalance learning, the numbers of instances in the majority class are outnumbered to the number of instances in the minority class. Furthermore, the minority concept may additionally contain a sub concept with limited instances, amounting to diverging degrees of classification difficulty [18-19]. This, in fact, is the result of another form of imbalance, a within-class imbalance, which concerns itself with the distribution of representative data for sub concepts within a class [20-22].

The existence of within-class imbalances is closely intertwined with the problem of small disjuncts, which has been shown to greatly depreciate classification performance [20-23]. Briefly, the problem of small disjuncts can be understood as follows: A classifier will attempt to learn a concept by creating multiple disjunct rules that describe the main concept [18-19], [23]. In the case of homogeneous concepts, the classifier will generally create large disjuncts, i.e., rules that cover a large portion (cluster) of examples pertaining to the main concept. However, in the case of heterogeneous concepts, small disjuncts, i.e., rules that cover a small cluster of examples pertaining to the main concept, arise as a direct result of underrepresented sub concepts [18-19], [23]. Moreover, since classifiers attempt to learn both majority and minority a concept, the problem of small disjuncts is not only restricted to the minority concept. On the contrary, small disjuncts of the majority class can arise from noisy misclassified minority class examples or underrepresented subconcepts. However, because of the vast representation of majority class data, this occurrence is infrequent. A more common scenario is that noise may influence disjuncts in the minority class. In this case, the validity of the clusters corresponding to the small disjuncts becomes an important issue, i.e., whether these examples represent an actual subconcept or are merely attributed to noise. To solve the above problem of cluster disjuncts we propose the method cluster disjunct minority oversampling technique for class imbalance learning.

III. Literature Review

In this section, we first review the major research about clustering in class imbalance learning and explain why we choose under-sampling as our technique in this paper.

The different imbalance data learning approaches are as follows:

Table 1 : Imbalanced Data learning Approaches

- ❖ SAMPLING METHODS
  - ✓ BASIC SAMPLING METHODS
    - ➢ Under-Sampling
    - ➢ Over-Sampling
  - ✓ ADVANCED SAMPLING METHODS
    - ➢ Tomek Link
    - ➢ The SMOTE approach
    - ➢ Borderline-SMOTE
    - ➢ One-Sided Selection OSS
    - ➢ Neighbourhood Cleaning Rule (NCL)
    - ➢ Bootstrap-based Over-sampling (BootOS)
- ❖ ENSEMBLE LEARNING METHODS
  - ✓ BAGGING
    - ➢ Asymmetric bagging, SMOTE Bagging
    - ➢ Over Bagging, Under Bagging
    - ➢ Roughly balanced bagging
    - ➢ Lazy Bagging
    - ➢ Random features selection
  - ✓ BOOSTING
    - ➢ Adaboost
    - ➢ SMOTEBoost
    - ➢ DataBoost-IM
  - ✓ RANDOM FORESTS
    - ➢ Balanced Random Forest BRF
    - ➢ Weighted Random Forest WRF
- ❖ COST-SENSITIVE LEARNING
  - ✓ Direct cost-sensitive learning methods
  - ✓ Methods for cost-sensitive meta-learning
  - ✓ Cost-sensitive meta-learning
  - ✓ Thresholding methods

- ✓ MetCost
- ✓ Cost-sensitive meta-learning sampling methods

❖ FEATURE SELECTION METHODS

- ✓ Warpper
- ✓ PREE (Prediction Risk based feature selection for Easy Ensemble)

❖ ALGORITHMS MODIFICATION

- ✓ Proposal for new splitting criteria DKM
- ✓ Adjusting the distribution reference in the tree
- ✓ Offset Entropy

Siti Khadijah Mohamad et al. [24] have conducted a review to look into how the data mining was tackled by previous scholars and the latest trends on data mining in educational research. Hongzhou Sha et al. [25] have proposed a method named EPLogCleaner that can filter out plenty of irrelevant items based on the common prefix of their URLs.

M.S.B. PhridviRaj et al. [26] have proposed an algorithm for finding frequent patterns from data streams by performs only one time scan of the database initially and uses the information to find frequent patterns using frequent pattern generation tree. Chumphol Bunkhumpornpat et al. [27] have a new over-sampling technique called DBSMOTE is proposed. DBSMOTE technique relies on a density-based notion of clusters and is designed to oversample an arbitrarily shaped cluster discovered by DBSCAN. DBSMOTE generates synthetic instances along a shortest path from each positive instance to a pseudo centroid of a minority-class cluster. Matías Di Martino et al. [28] have presented a new classifier developed specially for imbalanced problems, where maximum F-measure instead of maximum accuracy guide the classifier design.

V. Garcia et al. [29] have investigated the influence of both the imbalance ratio and the classifier on the performance of several resampling strategies to deal with imbalanced data sets. The study focuses on evaluating how learning is affected when different resampling algorithms transform the originally imbalanced data into artificially balanced class distributions. Table 2 presents recent algorithmic advances in class imbalance learning available in the literature. Obviously, there are many other algorithms which are not included in this table. A profound comparison of the above algorithms and many others can be gathered from the references list.

María Dolores Pérez-Godoy et al. [30] have proposed CO2RBFN, a evolutionary cooperative–competitive model for the design of radial-basis function networks which uses both radial-basis function and the evolutionary cooperative-competitive technique on imbalanced domains. CO2RBFN follows the evolutionary cooperative–competitive strategy, where each individual of the population represents an RBF (Gaussian function will be considered as RBF) and the entire population is responsible for the definite solution. This paradigm provides a framework where an individual of the population represents only a part of the solution, competing to survive (since it will be eliminated if its performance is poor) but at the same time cooperating in order to build the whole RBFN, which adequately represents the knowledge about the problem and achieves good generalization for new patterns.

Table 2 : Recent advances in Class Imbalance Learning

ALGORITHM	DESCRIPTION	REFERENECE
DCEID	Combining ensemble learning with cost-sensitive learning.	[27]
RUSBoost	A new hybrid sampling/boosting Algorithm.	[29]
CO2RBFN	A evolutionary cooperative–competitive model for the design of radial-basis function networks which uses both radial-basis function and the evolutionary cooperative-competitive technique.	[30]
Improved FRBCSs	Adapt the 2-tuples based genetic tuning approach to classification problems showing the good synergy between this method and some FRBCSs.	[33]
BSVMs	A model assessment of the interplay between various classification decisions using probability, corresponding decision costs, and quadratic program of optimal margin classifier.	[37]

Der-Chiang Li et al. [31] have suggested a strategy which over-samples the minority class and under-samples the majority one to balance the datasets. For the majority class, they build up the Gaussian type fuzzy membership function and a-cut to reduce the data size; for the minority class, they used the mega-trend diffusion membership function to generate virtual samples for the class. Furthermore, after balancing the data size of classes, they extended the data attribute dimension into a higher dimension space using classification related information to enhance the classification accuracy.

Enhong Che et al. [32] have described a unique approach to improve text categorization under class imbalance by exploiting the semantic context in text documents. Specifically, they generate new samples of rare classes (categories with relatively small amount of training data) by using global semantic information of

classes represented by probabilistic topic models. In this way, the numbers of samples in different categories can become more balanced and the performance of text categorization can be improved using this transformed data set. Indeed, this method is different from traditional re-sampling methods, which try to balance the number of documents in different classes by re-sampling the documents in rare classes. Such re-sampling methods can cause overfitting. Another benefit of this approach is the effective handling of noisy samples. Since all the new samples are generated by topic models, the impact of noisy samples is dramatically reduced.

Alberto Fernández et al. [33] have proposed an improved version of fuzzy rule based classification systems (FRBCSs) in the framework of imbalanced data-sets by means of a tuning step. Specifically, they adapt the 2-tuples based genetic tuning approach to classification problems showing the good synergy between this method and some FRBCSs. The proposed algorithm uses two learning methods in order to generate the RB for the FRBCS. The first one is the method proposed in [34], that they have named the Chi et al.'s rule generation. The second approach is defined by Ishibuchi and Yamamoto in [35] and it consists of a Fuzzy Hybrid Genetic Based Machine Learning (FH-GBML) algorithm.

J. Burez et al. [36] have investigated how they can better handle class imbalance in churn prediction. Using more appropriate evaluation metrics (AUC, lift), they investigated the increase in performance of sampling (both random and advanced under-sampling) and two specific modeling techniques (gradient boosting and weighted random forests) compared to some standard modeling techniques. They have advised weighted random forests, as a cost-sensitive learner, performs significantly better compared to random forests.

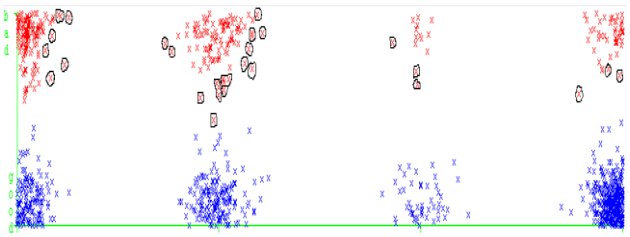
Che-Chang Hsu et al. [37] have proposed a method with a model assessment of the interplay between various classification decisions using probability, corresponding decis ion costs, and quadratic program of optimal margin classifier called: Bayesian Support Vector Machines (BSVMs) learning strategy. The purpose of their learning method is to lead an attractive pragmatic expansion scheme of the Bayesian approach to assess how well it is aligned with the class imbalance problem. In the framework, they did modify in the objects and conditions of primal problem to reproduce an appropriate learning rule for an observation sample. In [38] Alberto Fernández et al. have proposed to work with fuzzy rule based classification systems using a preprocessing step in order to deal with the class imbalance. Their aim is to analyze the behavior of fuzzy rule based classification systems in the framework of imbalanced data-sets by means of the application of an adaptive inference system with parametric conjunction operators. Jordan M. Malof et al. [39] have empirically investigates how class imbalance in the available set of training cases can impact the performance of the resulting classifier as well as properties of the selected set. In this K-Nearest Neighbor (k-NN) classifier is used which is a well-known classifier and has been used in numerous case-based classification studies of imbalance datasets.

The bottom line is that when studying problems with imbalanced data, using the classifiers produced by standard machine learning algorithms without adjusting the output threshold may well be a critical mistake. This skewness towards minority class (positive) generally causes the generation of a high number of false-negative predictions, which lower the model's performance on the positive class compared with the performance on the negative (majority) class.

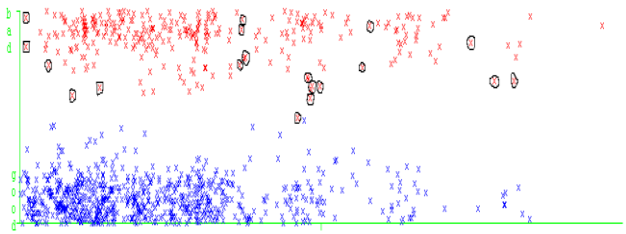
IV. Methodology

In this section, we follow a design decomposition approach to systematically analyze the different imbalanced domains. We first briefly introduce the framework design for our proposed algorithm.

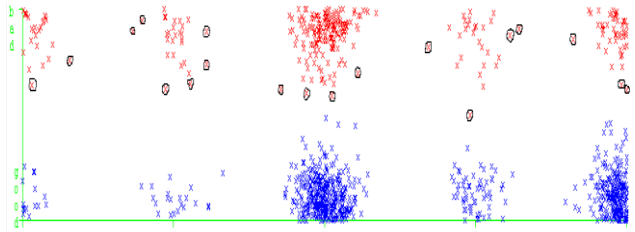
The working style of under-sampling tries to remove selective majority instances. Before performing selective under-sampling on the majority subset, the main cluster disjuncts has to be identified and the borderline and noise instances around the cluster disjuncts are to be removed. The number of instances eliminated will belong to the 'k' cluster disjuncts selected by visualization technique. The remaining cluster disjunct instances of the majority subset have to combined with minority set to form improved dataset. Here, the above said routine is employed o every cluster disjunct, which removes examples suffering from missing values at first and then removes borderline examples and examples of outlier category.



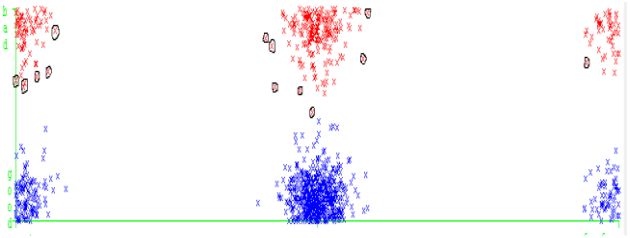
(a)



(b)



(c)

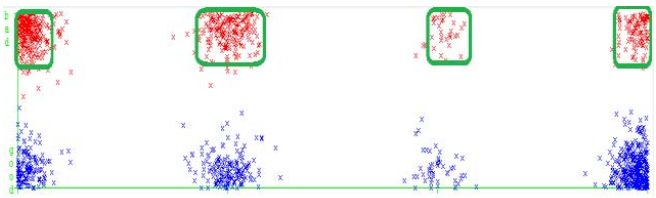


(d)

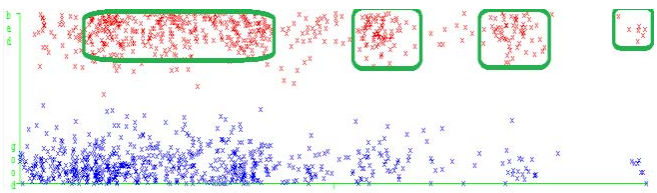
Figure 1 : Before (a). Checking Status (b). Duration (c). Credit History (d). Housing

The algorithm 1: MAJOR_CD can be explained as follows,

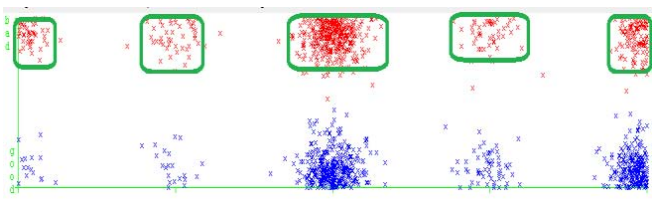
The inputs to the algorithm are majority subclass "p" and minority class "n" with the number of features j. The output of the algorithm will be the average measures such as AUC, Precision, F-measure, TP rate and TN rate produced by the MAJOR_CD methods. The algorithm begins with initialization of k=1 and j=1, where j is the number of cluster disjuncts identified by applying visualization technique on the subset "n" and k is the variable used for looping of j cluster disjuncts.



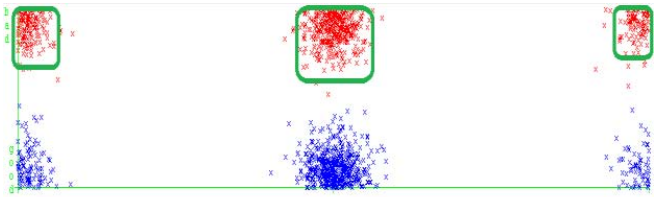
(a)



(b)



(c)



(d)

Figure 2 : After Applying MAJOR_CD: (a). Checking Status (b). Duration (c). Credit History (d). Housing

The 'j' value will change from one dataset to other, and depending upon the unique properties of the dataset the value of k can be equal to one also i.e no cluster disjunct attributes can be identified after applying visualization technique on the dataset.

In another case attributes related cluster disjunct oversampling can also be performed to improve the skewed dataset. In any case depending on the amount of minority examples generated, the final "strong set" can or cannot be balanced i;e number of majority instances and minority instances in the strong set will or will not be equal.

The presented MAJOR_CD algorithm is summarized as below.

Algorithm 1: MAJOR_CD

Input: A set of major subclass examples P, a set of minor subclass examples N, jPj < jNj, and Fj, the feature set, j > 0.

Output: Average Measure { AUC, Precision, F-Measure, TP Rate, TN Rate}

Phase I: Initial Phase:

1: begin

2: k ← 1,j←1.

3: Apply Visualization Technique on subset P,

4: Identify cluster disjunct Cj from P, j= number of cluster disjunct identified in visualization

Phase II: Under sampling Phase

5: Apply Oversampling on Cj cluster disjunct from P,

6: repeat

7: k=k+1

8: Remove 'Cj × s' noisy, borderline instances from the majority examples in each cluster disjunct Cj.

9: Until k = j

Phase III: Validating Phase

10: Train and Learn A Base Classifier (C4.5) using Improved P and N

11: end

The different components of our new proposed framework are elaborated in the next subsections

a) Preparation of the Majority and Minority subsets

The datasets is partitioned into majority and minority subsets. As we are concentrating over sampling, we will take minority data subset for further visualization analysis to identify cluster disjuncts.

b) Improve cluster disjunct by removing noisy and borderline instances

Minority subset can be further analyzed to find the noisy or borderline instances so that we can eliminate those. For finding the weak instances one of the ways is that find most influencing attributes or features and then remove ranges of the noisy or weak attributes relating to that feature.

How to choose the noisy instances relating to that cluster disjunct from the dataset set? We can find a range where the number of samples are less can give you a simple hint that those instances coming in that range or very rare or noise. We will intelligently detect and remove those instances which are in narrow ranges of that particular cluster disjunct. This process can be applied on all the cluster disjuncts identified for each dataset.

c) Forming the strong dataset

The minority subset and majority subset is combined to form a strong and balance dataset, which is used for learning of a base algorithm. In this case we have used C4.5 or Naïve Bayes as the base algorithm.

V. Evaluation Metrics

To assess the classification results we count the number of true positive (TP), true negative (TN), false positive (FP) (actually negative, but classified as positive) and false negative (FN) (actually positive, but classified as negative) examples. It is now well known that error rate is not an appropriate evaluation criterion when there is class imbalance or unequal costs. In this paper, we use AUC, Precision, F-measure, TP Rate and TN Rate as performance evaluation measures.

Let us define a few well known and widely used measures:

The Area under Curve (AUC) measure is computed by equation (1),

$$AUC = \frac{1 + TP_{RATE} - FP_{RATE}}{2} \quad (1)$$

The Precision measure is computed by equation(2),

$$\Pr ecision = \frac{TP}{(TP) + (FP)} \quad (2)$$

The F-measure Value is computed by equation(3),

$$F - measure = \frac{2 \times \Pr ecision \times \mathrm{Re}\, call}{\Pr ecision + \mathrm{Re}\, call} \quad (3)$$

The True Positive Rate measure is computed by equation (4),

$$TruePositiveRate = \frac{TP}{(TP) + (FN)} \quad (4)$$

The True Negative Rate measure is computed by equation (5),

$$TrueNegativeRate = \frac{TN}{(TN) + (FP)} \quad (5)$$

VI. Experimental Framework

In this study MAJOR_CD are applied to twelve binary data sets from the UCI repository [40] with different imbalance ratio (IR). Table 3 summarizes the data selected in this study and shows, for each data set, the number of examples (#Ex.), number of attributes (#Atts.), class name of each class (minority and majority) and IR. In order to estimate different measure (AUC, precision, Fmeasure, TP rate and TN rate) we use a tenfold cross validation approach, that is ten partitions for training and test sets, 90% for training and 10% for testing, where the ten test partitions form the whole set. For each data set we consider the average results of the ten partitions.

Table 3 : Summary of benchmark imbalanced datasets

S.no	Datasets	# Ex.	# Atts.	Class (_,+)	IR
1.	Breast	268	9	(recurrence; no-recurrence)	2.37
2.	Breast_w	699	9	(benign; malignant)	1.90
3.	Colic	368	22	(yes; no)	1.71
4.	Credit-g	1000	21	(good; bad)	2.33
5.	Diabetes	768	8	(tested-potv; tested-negtv)	1.87
6.	Hepatitis	155	19	(die; live)	3.85
7.	Ionosphere	351	34	(b;g)	1.79
8.	Kr-vs-kp	3196	37	(won; nowin)	1.09
9.	Labor	56	16	(bad ; good)	1.85
10.	Mushroom	8124	23	(e ; p)	1.08
11.	Sick	3772	29	(negative ; sick)	15.32
12.	Sonar	208	60	(rock ; mine)	1.15

To validate the proposed MAJOR_CD algorithm, we compared it with the traditional Support Vector Machines (SVM), C4.5, Functional Trees (FT), SMOTE (Synthetic Minority Oversampling TEchnique) and CART algorithm.

VII. Results

For all experiments, we use existing prototype's present in Weka [41]. We compare the following domain adaptation methods:

We compared proposed method MAJOR_CD with the SVM, C4.5 [42], FT, SMOTE [43] and CART state-of -the-art learning algorithms. In all the

experiments we estimate AUC, Precision, F-measure, TP rate and TN rate using 10-fold cross-validation. We experimented with 12 standard datasets for UCI repository; these datasets are standard benchmarks used in the context of high-dimensional imbalance learning. Experiments on these datasets have 2 goals. First, we study the class imbalance properties of the datasets using proposed MAJOR_CD learning algorithm. Second, we compare the classification performance of our proposed MAJOR_CD algorithm with the traditional and class imbalance learning methods based on all datasets.

Following, we analyze the performance of the method considering the entire original algorithms, without pre-processing, data sets for SVM, C4.5, FT and CART. we also analyze a pre-processing method SMOTE for performance evaluation of MAJOR_CD. The complete table of results for all the algorithms used in this study is shown in Table 4 to 9, where the reader can observe the full test results, of performance of each approach with their associated standard deviation. We must emphasize the good results achieved by MAJOR_CD, as it obtains the highest value among all algorithms.

Table 4, 5, 6, 7, 8 and 9 reports the results of AUC, Precision, F-measure, TP Rate, TN Rate and accuracy respectively for fifteen UCI datasets. Tables 4-9 provide both the numerical average performance (Mean) and the standard deviation (SD) results. If the proposed technique is better than the compared technique then '●' symbol appears in the column. If the proposed technique is not better than the compared technique then 'o' symbol appears in the column. The mean performances were significantly different according to the T-test at the 95% confidence level. The results in the tables show that MAJOR_CD has given a good improvement on all the measures of class imbalance learning. This level of analysis is enough for overall projection of advantages and disadvantages of MAJOR_CD. A two-tailed corrected resampled paired t test is used in this paper to determine whether the results of the cross-validation show that there is a difference between the two algorithms is significant or not. Difference in accuracy is considered significant when the p-value is less.

Table 4 : Summary of tenfold cross validation performance for Accuracy on all the datasets

Datasets	SVM	C4.5	FT	SMOTE	CART	MAJOR_CD
Breast	**67.21±7.28●**	74.28±6.05○	**68.58±7.52●**	**69.83±7.77●**	**70.22±5.19●**	72.42±6.32
Breast_w	96.75±2.00○	95.01±2.73○	95.45±2.52○	96.16±2.06○	94.74±2.60	94.61±2.39
Colic	**79.78±6.57●**	85.16±5.91	**79.11± 6.51●**	88.53±4.10○	85.37±5.41	85.00±5.97
Credit-g	**68.91±4.46●**	71.25±3.17○	71.88±3.68○	76.50±3.38○	73.43±4.00○	70.39±4.19
Diabetes	76.55±4.67○	74.49±5.27○	**70.62± 4.67●**	76.08±4.04○	74.56±5.01○	73.45±5.07
Hepatitis	81.90±8.38○	79.22±9.57○	81.40±8.55○	78.35±9.09○	77.10±7.12○	75.29(8.95)
Ionosphere	90.26±4.97○	89.74±4.38○	**87.10±5.12●**	90.28±4.73○	88.87±4.84	88.70(5.31)
Kv-rs-kp	99.02±0.54	99.44±0.37	**90.61±1.65●**	99.66±0.27	99.35±0.43	99.41(0.49)
Labor	92.40±11.07○	**78.60±16.58●**	84.30±16.24○	80.27±11.94	80.03±16.67	80.60(17.16)
Mushroom	100.0±0.00	100.0±0.00	100.0±0.000	100.0±0.00	99.95±0.09	100.00(0.00)
Sick	99.26±0.04○	**98.72±0.55●**	**96.10±0.92●**	**97.61±0.68●**	98.85±0.54	98.68(0.55)
Sonar	75.46±9.92○	73.61±9.34○	86.17±8.45○	82.42±7.25○	**70.72±9.43●**	71.70(9.00)

Table 5 : Summary of tenfold cross validation performance for AUC on all the datasets

Datasets	SVM	C4.5	FT	SMOTE	CART	MAJOR_CD
Breast	**0.586±0.102●**	**0.606±0.087●**	**0.604±0.082●**	0.717±0.084○	**0.587±0.110●**	0.611±0.095
Breast_w	0.977±0.017○	0.957±0.034○	**0.949±0.030●**	0.967±0.025○	**0.950±0.032●**	0.954±0.030
Colic	**0.802±0.073●**	**0.843±0.070●**	**0.777±0.072●**	0.908±0.040○	**0.847±0.070●**	0.850±0.065
Credit-g	**0.650±0.075●**	**0.647±0.062●**	**0.655±0.044●**	0.778±0.041○	0.716±0.055○	0.656±0.065
Diabetes	0.793±0.072○	0.751±0.070	**0.668±0.051●**	0.791±0.041○	0.743±0.071	0.743±0.067
Hepatitis	0.757±0.195○	0.668±0.184○	0.678±0.139○	0.798±0.112○	**0.563±0.126●**	0.631(0.182)
Ionosphere	0.900±0.060○	0.891±0.060○	**0.831±0.067●**	0.904±0.053○	0.896±0.059○	0.885(0.070)
Kr-vs-kp	**0.996±0.005●**	0.998±0.003	**0.906±0.017●**	0.999±0.001	**0.997±0.004●**	0.998(0.002)
Labor	**0.971±0.075●**	0.726±0.224●	**0.844±0.162●**	**0.833±0.127●**	**0.750±0.248●**	0.802(0.200)

Mushroom	1.000±0.00	1.000±0.00	1.000±0.00	1.000±0.00	0.999±0.001	1.000±0.00
Sick	0.990±0.014○	0.952±0.040○	**0.795±0.053●**	0.962±0.025○	0.954±0.043○	0.948(0.042)
Sonar	0.771±0.103○	0.753±0.113○	0.859±0.086○	0.814±0.090○	**0.721±0.106●**	0.725(0.100)

Table 6 : Summary of tenfold cross validation performance for Precision on all the datasets

Datasets	SVM	C4.5	FT	SMOTE	CART	MAJOR_CD
Breast	0.745±0.051○	0.753±0.042○	0.762±0.051○	**0.710±0.075●**	**0.728±0.038●**	0.732±0.043
Breast_w	0.988±0.019○	0.965±0.026○	0.964±0.026○	0.974±0.025○	0.968±0.026○	0.961±0.027
Colic	0.845±0.060○	0.851±0.055○	**0.839±0.062●**	0.853±0.057○	0.853±0.053○	0.843±0.061
Credit-g	0.776±0.033○	0.767±0.025○	0.791±0.027○	0.768±0.034○	0.779±0.030○	0.758±0.030
Diabetes	0.793±0.037○	0.797±0.045●	**0.764±0.036●**	**0.781±0.064●**	0.782±0.042	0.782±0.048
Hepatitis	0.604±0.271○	0.510±0.371○	0.546±0.333○	0.709±0.165○	**0.232±0.334●**	0.429(0.325)
Ionosphere	0.906±0.080○	0.895±0.084	0.938±0.073○	0.934±0.049○	**0.868±0.096●**	0.894(0.080)
Kr-vs-kp	**0.991±0.008●**	0.994±0.006	**0.905±0.021●**	0.996±0.005○	**0.993±0.007●**	0.994(0.006)
Labor	0.915±0.197○	**0.696±0.359●**	0.802±0.250○	0.871±0.151○	**0.715±0.355●**	0.738(0.300)
Mushroom	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.002	1.000±0.000
Sick	0.997±0.003○	0.992±0.005	**0.975±0.007●**	**0.983±0.007●**	0.992±0.005	0.992(0.005)
Sonar	0.764±0.119○	0.728±0.121○	0.883±0.100○	0.863±0.068○	**0.709±0.118●**	0.715(0.108)

Table 7 : Summary of tenfold cross validation performance for F-measure on all the datasets

Datasets	SVM	C4.5	FT	SMOTE	CART	MAJOR_CD
Breast	**0.781±0.059●**	0.838±0.040○	**0.776±0.057●**	**0.730±0.076●**	0.813±0.038○	0.823±0.043
Breast_w	0.965±0.019○	0.962±0.021○	0.975±0.016○	0.960±0.022○	0.959±0.020	0.958±0.019
Colic	**0.833±0.055●**	0.888±0.044○	**0.838±0.054●**	**0.880±0.042●**	0.890±0.040○	0.883±0.046
Credit-g	0.802±0.027	0.805±0.022○	**0.779±0.034●**	**0.787±0.034●**	0.820±0.028○	0.794±0.032
Diabetes	**0.778±0.037●**	0.806±0.044○	0.827±0.038○	**0.741±0.046●**	0.812±0.040○	0.794±0.041
Hepatitis	0.469±0.265○	0.409±0.272○	0.557±0.207○	0.677±0.138○	**0.179±0.235●**	0.375(0.258)
Ionosphere	0.787±0.098○	0.850±0.066○	0.855±0.079○	0.905±0.048○	0.841±0.070●	0.843(0.078)
Kv-rs-kp	**0.911±0.016●**	0.995±0.004	**0.991±0.005●**	0.995±0.004	0.994±0.004	0.994(0.005)
Labor	0.794±0.211○	**0.636±0.312●**	0.879±0.195○	0.793±0.132○	**0.660±0.316●**	0.734(0.280)
Mushroom	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.001	1.000±0.000
Sick	**0.979±0.005●**	**0.993±0.003●**	0.996±0.003○	**0.987±0.004●**	0.994±0.003	0.993(0.003)
Sonar	0.844±0.099○	**0.716±0.105●**	0.753±0.102○	0.861±0.061○	**0.672±0.106●**	0.704(0.105)

Table 8 : Summary of tenfold cross validation performance for TP Rate (Recall) (Sensitivity) on all the datasets

Datasets	SVM	C4.5	FT	SMOTE	CART	MAJOR_CD
Breast	**0.806±0.091●**	0.947±0.060○	**0.815±0.095●**	**0.763±0.117●**	**0.926±0.081●**	0.941±0.061
Breast_w	0.967±0.025○	0.959±0.033○	0.962±0.029○	**0.947±0.035●**	**0.952±0.034●**	0.956±0.032
Colic	**0.832±0.075●**	**0.931±0.053●**	**0.835±0.077●**	**0.913±0.058●**	0.932±0.050	0.931±0.062
Credit-g	**0.815±0.041●**	0.847±0.036○	**0.783±0.052●**	**0.810±0.058●**	0.869±0.047○	0.835±0.055
Diabetes	**0.795±0.054●**	0.821±0.073○	0.868±0.065○	**0.712±0.089●**	0.848±0.066○	0.811±0.067
Hepatitis	0.448±0.273○	0.374±0.256○	0.573±0.248○	0.681±0.188○	**0.169±0.236●**	0.371(0.272)
Ionosphere	**0.689±0.131●**	0.821±0.107○	0.820±0.114○	0.881±0.071○	0.830±0.112○	0.807(0.115)
Kv-rs-kp	**0.916±0.021●**	0.995±0.005	**0.990±0.007●**	0.995±0.006	0.995±0.006	0.994(0.007)
Labor	0.845±0.243○	**0.640±0.349●**	0.885±0.234○	**0.765±0.194●**	**0.665±0.359●**	0.775(0.321)
Mushroom	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000
Sick	**0.984±0.006●**	0.995±0.004	0.995±0.004	**0.990±0.005●**	0.996±0.003○	0.994(0.004)
Sonar	0.820±0.131○	0.721±0.140○	0.757±0.136○	0.865±0.090○	**0.652±0.137●**	0.708(0.147)

Table 9 : Summary of tenfold cross validation performance for TN Rate (Specificity) on all the datasets

Datasets	SVM	C4.5	FT	SMOTE	CART	MAJOR_CD
Breast	0.260±0.141	0.335±0.166○	**0.151±0.164●**	0.622±0.137○	**0.173±0.164●**	0.259±0.134
Breast_w	0.932±0.052○	0.977±0.037○	0.931±0.060○	0.975±0.024○	0.940±0.051○	0.928±0.053
Colic	**0.717±0.119●**	0.734±0.118○	0.731±0.121○	0.862±0.063○	**0.720±0.114●**	0.727±0.125
Credit-g	**0.398±0.085●**	0.469±0.098○	**0.371±0.105●**	0.713±0.056○	0.421±0.102○	0.419±0.092
Diabetes	0.603±0.111○	**0.574±0.095●**	**0.567±0.105●**	0.807±0.077○	**0.554±0.113●**	0.601±0.117
Hepatitis	0.900±0.097○	**0.882±0.092●**	0.942±0.093○	**0.837±0.109●**	0.928±0.094○	0.867(0.100)
Ionosphere	0.940±0.055○	0.949±0.046○	**0.933±0.063●**	**0.928±0.057●**	**0.921±0.066●**	0.936(0.054)
Kv-rs-kp	**0.993±0.007●**	**0.990±0.009●**	**0.987±0.010●**	0.998±0.003○	**0.992±0.008●**	0.994(0.007)
Labor	0.865±0.197○	0.945±0.131○	0.843±0.214○	0.847±0.187○	0.877±0.192○	0.827(0.192)
Mushroom	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.999±0.002	1.000±0.000
Sick	0.875±0.071	0.974±0.026○	**0.846±0.080●**	**0.872±0.053●**	**0.876±0.078●**	0.874(0.074)
Sonar	0.749±0.134○	0.752±0.148○	0.762±0.145○	0.752±0.113○	0.756±0.121○	0.724(0.122)

than 0.05 (confidence level is greater than 95%). In discussion of results, if one algorithm is stated to be better or worse than another then it is significantly better or worse at the 0.05 level.

We can make a global analysis of results combining the results offered by Tables from 4–9:

- Our proposal, MAJOR_CD are the best performing one when the data sets are no preprocessed. It outperforms the pre-processing SMOTE methods and this hypothesis is confirmed by including standard deviation variations. We have considered a complete competitive set of methods and an improvement of results is expected in the benchmark algorithms i;e SVM, C4.5, FT and CART. However, they are not able to outperform MAJOR_CD. In this sense, the competitive edge of MAJOR_CD can be seen.
- Considering that MAJOR_CD behaves similarly or not effective than SMOTE shows the unique properties of the datasets where there is scope of improvement in minority subset and not in majority subset. Our MAJOR_CD can only consider improvements in majority subset which is not effective for some unique property datasets.

The contributions of this work are twofold:

A general strategy to handle class imbalance problem: This is scalable, flexible, and modular, allowing the many existing supervised methods to be as a base algorithm. The method achieves competitive or better results compared to state-of-the-art baselines.

We emphasize that our approach is learner-independent: visualization can be used in conjunction with many of the existing algorithms in the literature. Furthermore, the fact that we select samples in the model space, as opposed to the feature space, is novel and sets it apart from many previous approaches to transfer learning (for both classification and ranking). This allows us to capture the ''functional change'' assumption and incorporate labeled information in the transfer learning process.

Finally, we can say that MAJOR_CD are one of the best alternatives to handle class imbalance problems effectively. This experimental study supports the conclusion that a cluster disjunct approach for cluster detections and elimination can improve the class imbalance learning behavior when dealing with imbalanced data-sets, as it has helped the MAJOR_CD method to be the best performing algorithms when compared with four classical and well-known algorithms: SVM, C4.5, FT and CART and a well-established pre-processing technique SMOTE.

VIII. Conclusion

Class imbalance problem have given a scope for a new paradigm of algorithms in data mining. The traditional and benchmark algorithms are worthwhile for discovering hidden knowledge from the data sources, meanwhile class imbalance learning methods can improve the results which are very much critical in real world applications. In this paper we present the class imbalance problem paradigm, which exploits the cluster disjunct concept in the supervised learning research area, and implement it with C4.5 as its base learners. Experimental results show that MAJOR_CD have performed well in the case of multi class imbalance datasets. Furthermore, MAJOR_CD is much less volatile than C4.5.

In our future work, we will apply MAJOR_CD to more learning tasks, especially high dimensional feature learning tasks. Another variation of our approach in future work is to analyze the influence of different base classifier effect on the quality of synthetic minority instances generated.

References Références Referencias

1. J. Wu, S. C. Brubaker, M. D. Mullin, and J. M. Rehg, "Fast asymmetric learning for cascade face detection," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 3, pp. 369–382, Mar. 2008.

2. N. V. Chawla, N. Japkowicz, and A. Kotcz, Eds., Proc. ICML Workshop Learn. Imbalanced Data Sets, 2003.
3. N. Japkowicz, Ed., Proc. AAAI Workshop Learn. Imbalanced Data Sets, 2000.\
4. G. M.Weiss, "Mining with rarity: A unifying framework," ACM SIGKDD Explor. Newslett., vol. 6, no. 1, pp. 7–19, Jun. 2004.
5. N. V. Chawla, N. Japkowicz, and A. Kolcz, Eds., Special Issue Learning Imbalanced Datasets, SIGKDD Explor. Newsl., vol. 6, no. 1, 2004.
6. W.-Z. Lu and D.Wang, "Ground-level ozone prediction by support vector machine approach with a cost-sensitive classification scheme," Sci. Total. Enviro., vol. 395, no. 2-3, pp. 109–116, 2008.
7. Y.-M. Huang, C.-M. Hung, and H. C. Jiau, "Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem," Nonlinear Anal. R. World Appl., vol. 7, no. 4, pp. 720–747, 2006.
8. D. Cieslak, N. Chawla, and A. Striegel, "Combating imbalance in network intrusion datasets," in IEEE Int. Conf. Granular Comput., 2006, pp. 732–737.
9. M. A. Mazurowski, P. A. Habas, J. M. Zurada, J. Y. Lo, J. A. Baker, and G. D. Tourassi, "Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance," Neural Netw., vol. 21, no. 2–3, pp. 427–436, 2008.
10. A. Freitas, A. Costa-Pereira, and P. Brazdil, "Cost-sensitive decision trees applied to medical data," in Data Warehousing Knowl. Discov. (Lecture Notes Series in Computer Science), I. Song, J. Eder, and T. Nguyen, Eds.,
11. K.Kilic¸,O¨ zgeUncu and I. B. Tu¨rksen, "Comparison of different strategies of utilizing fuzzy clustering in structure identification," Inf. Sci., vol. 177, no. 23, pp. 5153–5162, 2007.
12. M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," Comput.Med. Imag. Grap., vol. 31, no. 6, pp. 362–373, 2007.
13. X. Peng and I. King, "Robust BMPM training based on second-order cone programming and its application in medical diagnosis," Neural Netw., vol. 21, no. 2–3, pp. 450–457, 2008.Berlin/Heidelberg, Germany: Springer, 2007, vol. 4654, pp. 303–312.
14. Rukshan Batuwita and Vasile Palade (2010) FSVM-CIL: Fuzzy Support Vector Machines for Class Imbalance Learning, IEEE TRANSACTIONS ON FUZZY SYSTEMS, VOL. 18, NO. 3, JUNE 2010, pp no:558-571.
15. N. Japkowicz and S. Stephen, "The Class Imbalance Problem: A Systematic Study," Intelligent Data Analysis, vol. 6, pp. 429-450, 2002.
16. M. Kubat and S. Matwin, "Addressing the Curse of Imbalanced Training Sets: One-Sided Selection," Proc. 14th Int'l Conf. Machine Learning, pp. 179-186, 1997.
17. G.E.A.P.A. Batista, R.C. Prati, and M.C. Monard, "A Study of the Behavior of Several Methods for Balancing Machine Learning Training Data," SIGKDD Explorations, vol. 6, pp. 20-29, 2004.1.
18. J.R. Quinlan, "Induction of Decision Trees," Machine Learning, vol. 1, no. 1, pp. 81-106, 1986.
19. T. Jo and N. Japkowicz, "Class Imbalances versus Small Disjuncts," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 40-49, 2004.
20. N. Japkowicz, "Class Imbalances: Are We Focusing on the Right Issue?" Proc. Int'l Conf. Machine Learning, Workshop Learning from Imbalanced Data Sets II, 2003.
21. R.C. Prati, G.E.A.P.A. Batista, and M.C. Monard, "Class Imbalances versus Class Overlapping: An Analysis of a Learning
22. System Behavior," Proc. Mexican Int'l Conf. Artificial Intelligence, pp. 312-321, 2004.
23. G.M. Weiss, "Mining with Rarity: A Unifying Framework," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 7-19, 2004.
24. Siti Khadijah Mohamada, Zaidatun Tasir. "Educational data mining: A review", Procedia - Social and Behavioral Sciences 97 (2013) 320 – 324.
25. Hongzhou Sha, Tingwen Liu, Peng Qin, Yong Sun, Qingyun Liu." EPLogCleaner: Improving Data Quality of Enterprise Proxy Logs for Efficient Web Usage Mining" Procedia Computer Science 17 (2013) 812 – 818.
26. M.S.B. PhridviRaj, C.V. GuruRao." Data mining – past, present and future – a typical survey on data Streams", Procedia Technology 12 (2014) 255 – 263.
27. Chumphol Bunkhumpornpat, Krung Sinapiromsaran, Chidchanok Lursinsap." DBSMOTE: Density-Based Synthetic Minority Over-sampling Technique" Appl Intell (2012) 36:664–684.
28. Matías Di Martino, Alicia Fernández, Pablo Iturralde, Federico Lecumberry." Novel classifier scheme for imbalanced problems", Pattern Recognition Letters 34 (2013) 1146–1151.
29. V. Garcia, J.S. Sanchez , R.A. Mollineda," On the effectiveness of preprocessing methods when dealing with different levels of class imbalance", Knowledge-Based Systems 25 (2012) 13–21.
30. María Dolores Pérez-Godoy, Alberto Fernández, Antonio Jesús Rivera, María José del Jesus," Analysis of an evolutionary RBFN design algorithm, CO2RBFN, for imbalanced data sets", Pattern Recognition Letters 31 (2010) 2375–2388.

31. Der-Chiang Li, Chiao-WenLiu, SusanC.Hu," A learning method for the class imbalance problem with medical data sets", Computers in Biology and Medicine 40 (2010) 509–518.
32. Enhong Che, Yanggang Lin, Hui Xiong, Qiming Luo, Haiping Ma," Exploiting probabilistic topic models to improve text categorization under class imbalance", Information Processing and Management 47 (2011) 202–214.
33. Alberto Fernández, María José del Jesus, Francisco Herrera," On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets", Information Sciences 180 (2010) 1268–1291.
34. Z. Chi, H. Yan, T. Pham, Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition, World Scientific, 1996.
35. H. Ishibuchi, T. Yamamoto, T. Nakashima, Hybridization of fuzzy GBML approaches for pattern classification problems, IEEE Transactions on System, Man and Cybernetics B 35 (2) (2005) 359–365.
36. J. Burez, D. Van den Poel," Handling class imbalance in customer churn prediction", Expert Systems with Applications 36 (2009) 4626–4636.
37. Che-Chang Hsu, Kuo-Shong Wang, Shih-Hsing Chang," Bayesian decision theory for support vector machines: Imbalance measurement and feature optimization", Expert Systems with Applications 38 (2011) 4698–4704.
38. Alberto Fernández, María José del Jesus, Francisco Herrera," On the influence of an adaptive inference system in fuzzy rule based classification systems for imbalanced data-sets", Expert Systems with Applications 36 (2009) 9805–9812.
39. Jordan M. Malof, Maciej A. Mazurowski, Georgia D. Tourassi," The effect of class imbalance on case selection for case-based classifiers: An empirical study in the context of medical decision support", Neural Networks 25 (2012) 141–145.
40. A. Asuncion D. Newman. (2007). UCI Repository of Machine Learning Database (School of Information and Computer Science), Irvine, CA: Univ. of California [Online]. Available: http://www.ics. uci. edu/~mlearn/MLRepository.html.
41. Witten, I.H. and Frank, E. (2005) Data Mining: Practical machine learning tools and techniques. 2nd edition Morgan Kaufmann, San Francisco.
42. J. R. Quinlan, C4.5: Programs for Machine Learning, 1st ed. San Mateo, CA: Morgan Kaufmann Publishers, 1993.
43. N. Chawla, K. Bowyer, and P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. Artif. Intell. Res., vol. 16, pp. 321–357, 2002.

This page is intentionally left blank

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 7 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Gray Scale and Color Medical Image Compression by Lifting Wavelet; Bandelet and Quincunx Wavelets Transforms: A Comparison Study

By Boukli Hacene Ismail & A. Bessaid

University of Tlemcen, Algeria

Abstract- The Quincunx wavelet , the lifting Scheme wavelet and the Second generation bandelet transform are a new method to offer an optimal representation for image geometric; we use this transform to study medical image compressed using the Quincunx transform coupled by SPIHT coder. We are interested in compressed medical image, In order to develop the compressed algorithm we compared our results with those obtained by this transforms application in medical image field. We concluded that the results obtained are very satisfactory for medical image domain. Our algorithm provides very important PSNR and MSSIM values for medical images compression.

Index Terms: medical image, quincunx wavelet, biorthogonal wavelet 9/7 , lifting scheme ,bandelet transform, optical flow, quadtree segmentation, compression, SPIHT coder .

GJCST-C Classification : I.4.0

Strictly as per the compliance and regulations of:

Gray Scale and Color Medical Image Compression by Lifting Wavelet; Bandelet and Quincunx Wavelets Transforms: A Comparison Study

Boukli Hacene Ismail [α] & A. Bessaid [σ]

Abstract- The Quincunx wavelet , the lifting Scheme wavelet and the Second generation bandelet transform are a new method to offer an optimal representation for image geometric; we use this transform to study medical image compressed using the Quincunx transform coupled by SPIHT coder. We are interested in compressed medical image, In order to develop the compressed algorithm we compared our results with those obtained by this transforms application in medical image field. We concluded that the results obtained are very satisfactory for medical image domain. Our algorithm provides very important PSNR and MSSIM values for medical images compression.

Index Terms: *medical image, quincunx wavelet, biorthogonal wavelet 9/7, lifting scheme, bandelet transform, optical flow, quadtree segmentation, compression, SPIHT coder .*

I. Introduction

Today the massive use of numerical methods in medical imaging generates increasingly important volumes of data .One of the most important problems in such applications is how to store and transmit images [1].It is well established that the accuracy and precision of diagnostic are initially related to the image quality.

Over the past ten years, the wavelets (DWT), have had a huge success in the field of image processing such as encoding, weaknesses have been noted in its use in the detection and representation of the objects' contours, and have been used to solve many problems such as image compression and restoration [2].Image representations in separable orthonormal bases such as Fourier, local Cosine or Wavelets can not take advantage of the geometrical regularity of image structures. Standard wavelet bases are optimal to represent functions with piecewise singularities; however, they fail to capture the geometric regularity along the singularities of edges or contours because of their isotropic support. To exploit the anisotropic regularity along edges, the basis must include elongated functions that are nearly parallel to the edges. Multi-scale geometric analysis (MGA) developed recently provides a group of new basis that has anisotropic supports such as Curvelets, contourlet [15-16].

To overcome this problem, In this paper, we introduce tree new type of transform, the first is devoted to representation of the Lifting scheme, and then we present the biorthgonal wavelet CDF 9/7,the second called bandelet transform by Pennec and Stéphane Mallat [17], this transform is more recently developed method of compression technique, which decompose the image along multiscale vectors that are elongated in the direction of a geometric flow, and the third transform multi resolution decompositions called quincunx wavelets which are better adapted to the image representation. This structure of decomposition allows the construction of a no separable transform. No separable wavelets, by contrast, offer more freedom and can be better tuned to the characteristics of images. Their less attractive side is that they require more computations. The quincunx wavelets are especially interesting because they are nearly isotropic [3].

II. Lifting Scheme Wavelet Transforms

In [4], Calderbank et al.introduced how to use the lifting scheme presented in [5], where sweldens showed that the convolution based biorthogonal WT can be implemented in a lifting-based scheme as shown in figure (1) for reducing the computational complexity. The lifting-based WT consists of splitting, lifting, and scaling modules and the WT is treated as prediction-error decomposition.

It provides a complete spatial interpretation of WT. In figure (1), let X denote the input signal, and XL_1 and $X H_1$ be the decomposed output signals.

Author α σ: Biomedical engineering research laboratory, Department of electrics and Electronics Technology Faculty. University of Tlemcen, Algeria. e-mail: ismaill80@yahoo.fr, a.bessaid@gmail.com

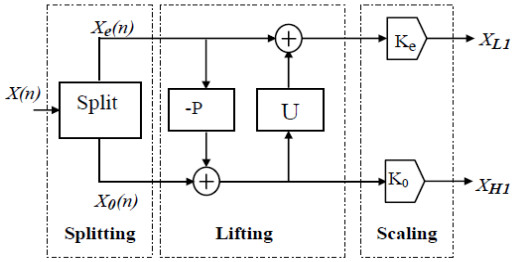


Figure 1 : The Lifting -based wavelet

This article deals with biorthogonal wavelet 9/7. These wavelets are part of the family of symmetric biorthogonal wavelet CDF. The low pass filters associated with wavelet 9/7 have p=9 coefficients in the analysis, p=7 coefficients to synthesize.

The wavelets 9/7 have a great number of null moments for a relatively short support. They are more symmetrical and very close to orthogonality. Antonini and Barlaud were the first [7] to show the superiority of the biorthogonal wavelet transform 9/7 for the decorrelation of natural images. It has been widely used in image coding [8],[9] ,[25] and is used by the JPEG-2000 codec [10].

III. The Bandelet Transform

Bandelet transform, introduced by Pennec and Mallat [18] built a base adapted to the geometric content of an image.

Bandelet transform is an analysis tool which aims at taking advantage of sharp image transitions in images. A geometric flow, which indicates directions in which the image gray levels have regular variations,is used to form bandelet bases in bandelet transform. The bandelet bases lead to optimal approximation rates for geometrically regular images and are proven to be efficient in still image compression [24], video compression, and noise-removal algorithms[6-10].

Apparently, bandelet transform is appropriate for the analysis of edges and texture of images.

In bandelet transform, a geometric flow of vectors is defined to represent the edges of image. These vectors give the local directions in which the image has regular variations. Orthogonal bandelet bases are constructed by dividing the image support in regions inside which the geometric flow is parallel. Let Ω_i denote the ith region, which composes the image support $S = U_i\Omega_i$. Within each Ω_i the flow is either parallel horizontally or vertically.

Figure 2 shows an example of a vertically parallel geometric flow in a region of the hat of Lena image.

Figure 2 : geometric flow in a region of the hat of Lena image

The image is partitioned small enough into square regions, each region Ω_i includes at most one contour. If a region does not include any contour, the image intensity is uniformly regular and the flow is not defined. In bandelet transform, these regions are approximated in the separable wavelet basis of $L^2(\Omega)$ in:

$$\begin{Bmatrix} \phi_{j,n_1}(x_1)\psi_{j,n_2}(x_2) \\ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2) \\ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2) \end{Bmatrix}_{(j,n_1,n_2)\in I_\Omega}$$

where I_Ω is an index set that depends upon the geometry of the boundary of Ω , and x_1, x_2 denote the location of pixel in the image, $\phi_{j,n_1}(x_1)\psi_{j,n_2}(x_2)$, $\psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2)$ and $\psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2)$ are the modified wavelets at the boundary. If a geometric flow is calculated in Ω , this wavelet basis is replaced by a bandelet orthonormal basis of $L^2(\Omega)$ in

$$\begin{Bmatrix} \phi_{j,n_1}(x_1)\psi_{j,n_2}(x_2-c(x_1)) \\ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2-c(x_1)) \\ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2-c(x_1)) \end{Bmatrix} \quad (4)$$

The horizontal wavelet $\psi^H_{j,n}$ have not vanishing moments along contour, to be replaced by new functions:

$$\psi_{j,n_1}(x_1)\psi_{j,n_2}(x_2-c(x_1)) \quad (5)$$

This is called bandeletization [13], The orthonormal basis of bandelet of field warping is defined by:

$$\begin{Bmatrix} \psi_{j,n_1}(x_1)\psi_{j,n_2}(x_2-c(x_1)) \\ \psi_{j,n_1}(x_1)\phi_{j,n_2}(x_2-c(x_1)) \\ \psi_{j,n_1}(x_1)\psi_{j,n_2}(x_2-c(x_1)) \end{Bmatrix} = \begin{Bmatrix} \psi^H_{l,n} \\ \psi^V_{j,n} \\ \psi^D_{j,n} \end{Bmatrix}, \; j,l \succ n_1,n_2 \quad (6)$$

IV. Quincunx Wavelets

The separable dyadic analysis require three families of wavelets, which is sometimes regarded as a disadvantage, in addition the factor of addition between

two successive scales is 4 which may seem high. It is possible to solve these two problems, but at the cost of the loss of filter separability and therefore a slightly higher computational complexity. An analysis has been particularly well studied to find a practical application, known as "quincunx" , [1].Quincunx decomposition results in fewer subbands than most other wavelet decompositions, a feature that may lead to reconstructed images with slightly lower visual quality.

The method is not used much in practice, but [14] presents results that suggest that quincunx decomposition performs extremely well and may be the best performer in many practical situations. Figure (3) illustrates this type of decomposition [3].

We notice that the dilation factor is not more than 2 between two successive resolutions, and that only one wavelet family is necessary [15,16]. It is noticed that the dilatation step is $\sqrt{2}$ on each direction and the geometry of the grid obtained justifies the name given to this multiresolution analysis.

First, we recall some basic results on quincunx sampling and perfect reconstruction filter banks, [17][18]. The quincunx sampling lattice is shown in figure(4). Let $x[\vec{n}]$ denote the discrete signal on the initial grid.

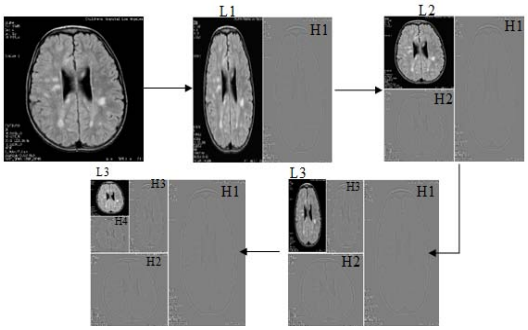
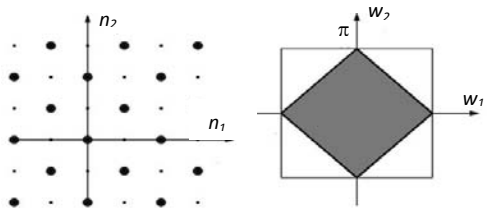


Figure 3 : Quincunx wavelet

Now, we write the quincunx sampled version of $x[\vec{n}]$ as:

$$[x]_{\downarrow M}[\vec{n}] = x[M\,\vec{n}] \quad where \quad M=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix} \tag{8}$$



(a) (a)

Figure 4 : (a) Quincunx lattice, (b) the corresponding Nyquist area in the frequency domain

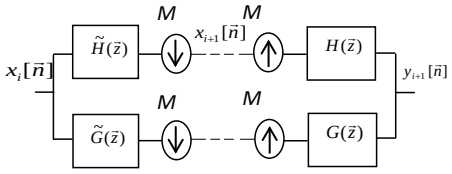


Figure 5 : Perfect reconstruction filter bank in a quincunx lattice

Since quincunx sampling reduces the number of image samples by a factor of two, the corresponding reconstruction filter bank has two channels (Fig.5). The low-pass filter $\tilde{H}$ reduces the resolution by a factor of $\sqrt{2}$; the wavelet coefficients correspond to the output of the high-pass filter $2\tilde{G}$ [15,16,17].
The conditions for a perfect reconstruction is:

$$\begin{cases}\tilde{H}(\bar{z})H(\bar{z})+\tilde{G}(\bar{z})G(\bar{z})=2\\ \tilde{H}(-\bar{z})H(\bar{z})+\tilde{G}(-\bar{z})G(\bar{z})=0\end{cases} \tag{14}$$

Where H and G (respectively $\tilde{H}$ and $\tilde{G}$) are the transfer functions of the synthesis (respectively analysis) filters. In the orthogonal case, the analysis and synthesis filters are identical up to a central symmetry; the wavelet filter G is simply a modulated version of the low-pass filter H .

To generate quincunx filters, we will use the standard approach which is to apply the diamond McClellan transform to map a 1D design onto the quincunx structure [19].
Thus, our quincunx refinement filter is given by

$$H_\alpha(e^{i\bar{\omega}})=\frac{\sqrt{2}\,(2+\cos\omega_1+\cos\omega_2)^{\frac{\alpha}{2}}}{\sqrt{(2+\cos\omega_1+\cos\omega_2)^{\alpha}+(2-\cos\omega_1-\cos\omega_2)^{\alpha}}} \tag{17}$$

V. SPIHT Coding Scheme

The SPIHT algorithm proposed by Said and Pearlman in 1996 [20,21], ameliorate progressive algorithm is compared to the EZW a lgorithm .The Set Partitioning in Hierarchical Trees (SPIHT) is one of the most advanced schemes available, even out performing the state-of-the-art JPEG 2000 in some situations, based on the creation of three list SCL, ICL and ISL with a calculated threshold T, each time you make a scan on both lists SCL and ISL andthat for the classified the significant coefficient in the list of significant coefficient.

The adapted for quincunx wavelet transform coupled by SPIHT is done in [26,27],

VI. Quality Evaluation Parameter

The Peak Signal to Noise Ratio (PSNR) is the most commonly used as a measure of quality of reconstruction in image compression. The PSNR were identified using the following formulae:

$$MSE^{2} = \frac{1}{M\,x\,N} \sum_{1=1}^{i=N} \sum_{j=1}^{j=M} \left(I(i,j) - \hat{I}(i,j)\right)^{2} \qquad (21)$$

Mean Square Error (MSE) which requires two MxN grayscale images I and where one of the images is considered as a compression of the other is defined as:

- *The PSNR is defined as:*

$$PSNR = 10 \log_{10} \left(\frac{(Dynamics \quad of \quad image)^{2}}{MSE} \right) \qquad (22)$$

Usually an image is encoded on 8 bits. It is represented by 256 gray levels, which vary between 0 and 255, the extent or dynamics of the image is 255.

- *The structural similarity index (SSIM):*

This parameter compares the similarity the brightness, contrast and structure between each pair of vectors, where the structural similarity index (SSIM) between two signals x and y is given by the following expression , [22].

$$SSIM\,(x,y) = l(x,y).c(x,y).s(x,y) \qquad (23)$$

The quality measurement can provide a spatial map of the local image quality, which provides more information on the image quality degradation, which is useful in medical imaging applications. For application, we require a single overall measurement of the whole image quality that is given by the following formula:

$$MSSIM\,(I,\hat{I}) = \frac{1}{M} \sum_{i=1}^{M} SSIM\,(I_i, \hat{I}_i) \qquad (24)$$

Where and are respectively the reference and degraded images, and are the contents of images at the i-th local window.
M: the total number of local windows in image. The MSSIM values exhibit greater consistency with the visual quality.

VII. Results and Discussion

We are interested in this work to the medical images compression, that we applied algorithm (QWT+SPIHT), (DWT9/7(lifting scheme) +SPIHT) ,(bandelet +SPIHT) , (DWT9/7 banc filter+SPIHT) . For this, we chose sets of medical images (MRI, CT ,ECHO and MAMOG) images gray level size 512x 512 encoded on 8 bits per pixel. These images are taken from the GE Medical System (database) [23].

The importance of our work lies in the possibility of reducing the rates for which the image quality remains acceptable. Estimates and judgments of the compressed image quality are given by the PSNR evaluation parameters and the MSSIM similarity Index.

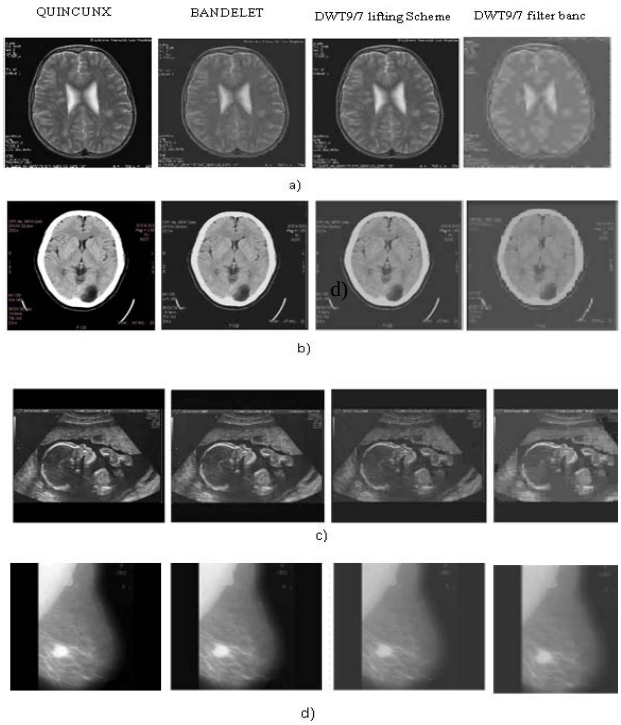
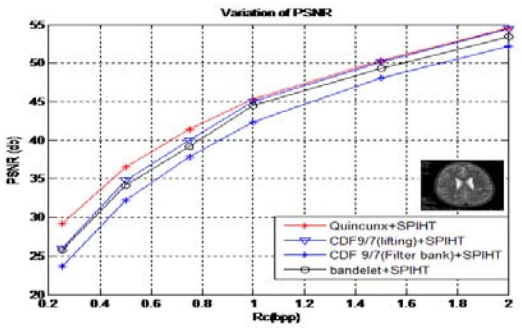
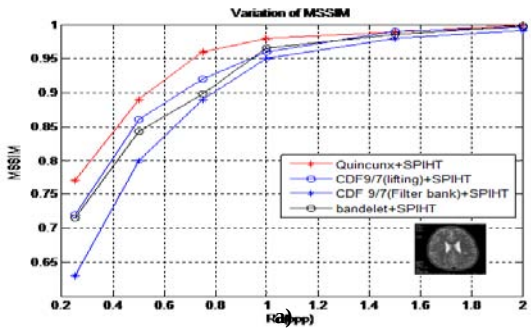
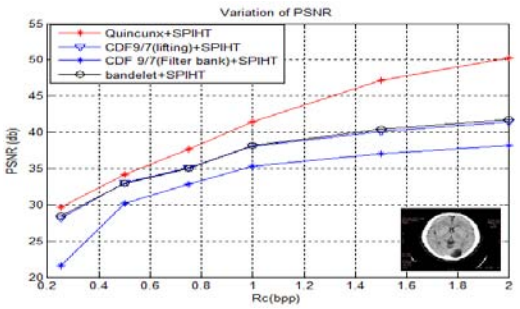
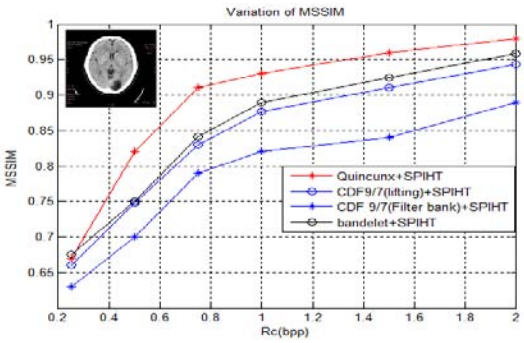


Figure 6 : MRI, CT ,ECHO and MAMOG compressed images by QWT , Bandelet ,DWT 9/7 (filter banc) and (DWT 9/7 (lifting scheme) coupled SPIHT coder for Rc=0.5Bpp (a) MRI image, (b) CT image, (c) ECHO image and (d) MAMOG image

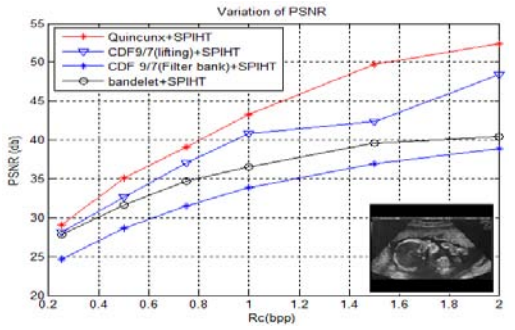


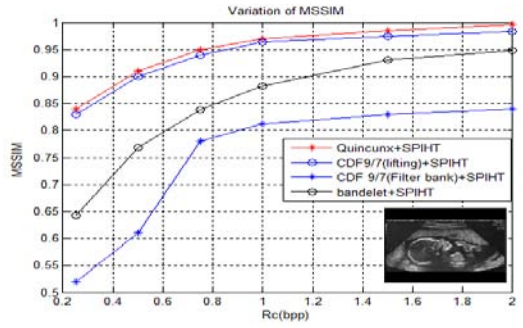




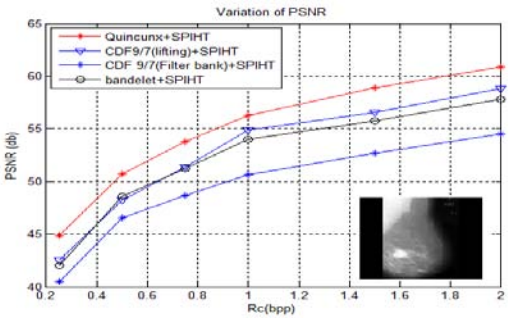


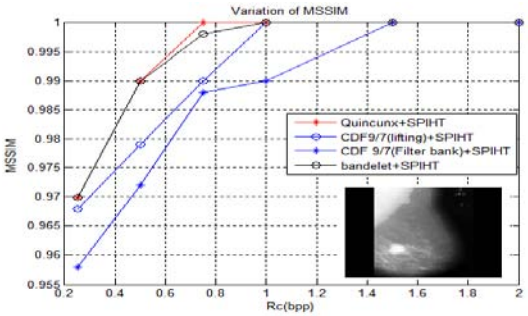
b)





c)





d)

Figure 7 : PSNR and MSSIM variation using different methods for : (a) MRI image, (b) CT image, (c) ECHO image and (d) MAMOG image

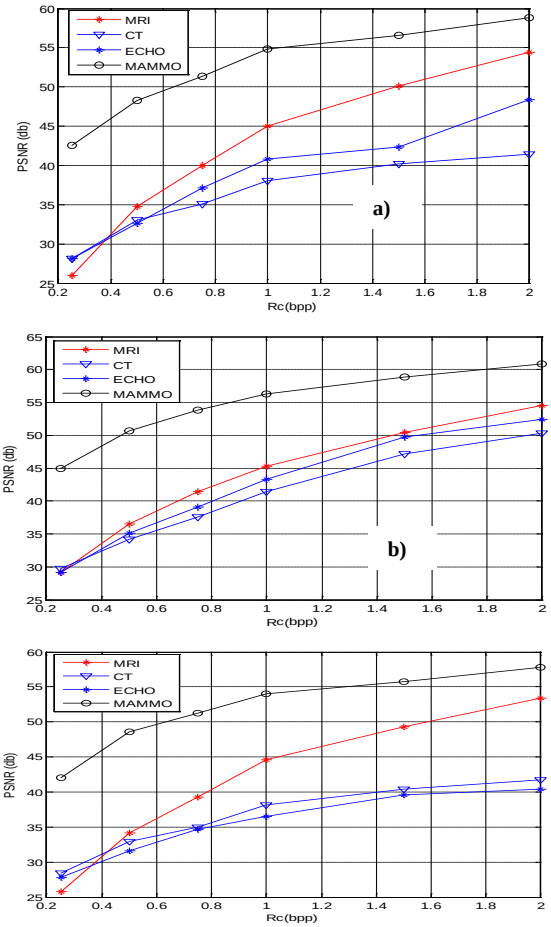


Figure 8 : Comparison results between MRI, CT ,ECHO and MAMMOGRAPHIC image compressed by / a) Quincunx transform ; b) biorthogonal wavelet 9/7 (lifting Scheme); and c)Bandelet transform coupled by SPIHT coder

Figure (6) shown below illustrates the compressed different modality of medical image quality using different transforms. According to the PSNR and MSSIM values, we note that from 0.5bpp, image reconstruction becomes almost perfect. We observe that compression degrades to a lessen extent the image

structure for a low compression bit-rate. However, for high compression bitrate, our algorithm better safe guards the various image structures. We note that our algorithm is adapted for the medical image compression.

We see from the figure 7, that the Quincunx transform coupled with SPIHT coder offer PSNR and MSSIM values better than to the other algorithms. In order to specify the type of medical image adapted to algorithm (Quincunx+SPIHT), we recapitulate the results for the four medical images (MRI, CT, ECHO and MAMOG) compressed by (Quincunx+SPIHT), (DWT9/7(lifting scheme) +SPIHT) ,(bandelet +SPIHT) algorithms in the following in figure 8.

Visually, from the two curves, it is clearly that the (QWT+SPIHT) algorithm allows us to have a good image reconstruction so a better image visual quality and this is proved by the large values of the parameters evaluation.

Now we have chosen a retinographic color image of size 512 x 512. In our application, we applied our algorithms for our color image for each layers after converted RGB space to YCrCb layers. We see from the figure 9, that the Quincunx transform coupled with SPIHT coder offer PSNR and MSSIM values better than to the other algorithms.

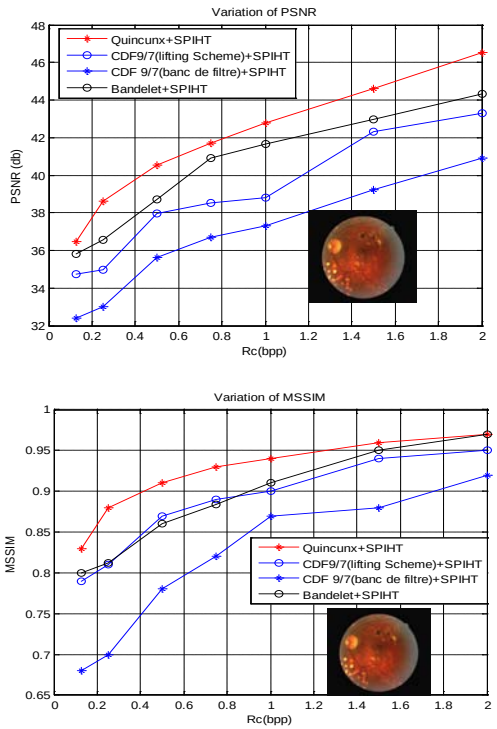


Figure 9 : PSNR and MSSIM variation using different methods for color image

VIII. Conclusion

The objective of this paper is undoubtedly the enhancement of medical images quality after the compression step. The latter is regarded as an essential tool to aid diagnosis (storage or transmission) in medical imaging. We compared the quincunx wavelet transform , bandelet transform and the biorthogonal wavelet 9/7 based on the filter banc and the lifting scheme coupled with the SPIHT coding for the gray scale and color medical image. After several applications for different modality medical images, we found that the algorithm for the quincunx wavelet transform gives better results than the other compression techniques. We have noticed that for 0.5 bpp bit-rate, the algorithm provides very important PSNR and MSSIM values from medical images. In perspective, we aspire to apply our algorithm to compress the medical video sequences.

References Références Referencias

1. V. Chappelier, Progressive coding of images by directed wavelet, Phd Thesis, Rennes1 University, 2005.
2. S. Mallat, Multifrequency channel decompositions of images and wavelet models, IEEE Transaction in Acoustic Speech and Signal Processing, Vol. 37, pp. 2091-2110, 1989.
3. M. Beladgham, A. Bessaid, A. Moulay Lakhdar and A. Taleb Ahmed, Improving quality of medical image compression using biorthogonal CDF wavelet based on lifting scheme and SPIHT coding, Serbian Journal of Electrical Engineering, Vol 8, Issue 2, ISSN 1451-4869 , pp 163-179, doi:10.2298/SJEE1102163B, 2011.
4. A.R Calderbank, I. Daubechies, W. Sweldens, and BL. Yeo, Wavelet transforms that map integers to integers, Appl Comput Harmon Anal, 5, pp. 332-369, 1998.
5. W. Sweldens, The lifting scheme: A custom-design construction of biorthogonal wavelets, Appl Comput Harmon Anal, pp. 186-200, 1996.
6. *Ismail Boukli Hacene Mohammed BELADGHAM, , Abdelhafid BESSAID,* " color Medical Image Compression Using Quincunx Wavelets and SPIHT Coding" Vol.8.No.2, pp.802-812.2013. International Journal of computer & Technology (IJCT)
7. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, Image coding using wavelet transform, IEEE Transactions on Image Processing, 1(2), 1992 pp. 205–220, 1992.
8. Said , W.A. Pearlman, A new fast and efficient image codec based on set partitioning in hierarchical trees, IEEE Trans. Circuits and Systems for Video Technology, Vol.6, pp. 243 – 250, 1996.
9. J.D. Villasenor, B. Belzer, and J. Liao, Wavelet filter evaluation for image compression, IEEE Transactions on Image Processing, 4(8), pp.1053–1060, 1995.

10. D.S. Taubman and M.W. Marcellin, JPEG-2000 : Image Compression Fundamentals, Standards and Practice, Kluwer Academic Publishers,2002.
11. Vladan Velisavljevi´c, Baltasar Beferull-Lozano, Martin Vetterli and Pier Luigi Dragotti. Directionlets: Anisotropic Multidirectional Representation with Separable Filtering. IEEE Trans. Image Processing,Dec, 2004.
12. E. Le Pennec et S. Mallat. Représentation d'images par bandelettes et application à la compression. Dans GRETSI, Toulouse, Septembre 2001.
13. E. Le Pennec, S. Mallat. Bandelettes et représentation géométrique des images. CMAP / école Polytechnique 91128 PALAISEAU – France.
14. Vetterli M. and Kovacevé J., "Wavelets and Subband Coding'', Upper Saddle River, NJ:Prentice-Hall, 1995.
15. Manuela F., Dimitri VD. and Michael U., "An Orthogonal Family of Quincunx Wavelets With Continuously Adjustable Order'', IEEE Transactions On Image Processing, Vol. 14, No. 4,APRIL 2005.
16. Dimitri VD, Thierry B. and Michael U. "On the Multidimensional Extension of the Quincunx Subsampling Matrix'', IEEE Signal Processing Letters, Vol. 12, No. 2, FEBRUARY 2005.
17. Chen Y. Michael AD. and Wu-Sheng L. "Design of Optimal Quincunx Filter Banks for Image Coding'', EURASIP Journal on Advances in Signal Processing, Vol. 2007.
18. Lee L., Oppenheir V.A., "Proprerties of approximate parks-McClellan filters'', IEEE, pp.2165-2168; 1997;
19. Yen-Yu C. and Shen-Chuan T., "Embedded medical image compression Using DCT based subband decomposition and modified SPIHT data organization'', Proceedings of the Fourth IEEE, (BIBE'04), 2004.
20. Miaou S.G., Chen S.T. and Chao S.N.,"Wavelet-based lossy-to-lossless medical image compression using dynamic
21. Said A. and Pearlman W. A., "A new fast and efficient image codec based on set partitioning in hierarchical trees'',IEEE Trans. Circuits and Systems for Video Technology, Vol. 6, p243 – 250,June 1996.
22. Wang Z., Bovik A. C., Sheikh H. R. and Simoncelli E.P, "Image quality assessment: From error visibility to structural similarity'', IEEE Transactions on Image Processing, Vol. 13, No. 4, APRIL 2004.
23. www.GE Medical System.com (database).
24. M. Beladgham, Habchi Yassine, A. Moulay Lakhdar and A. Taleb Ahmed, A Comparative Study between Bandelet and Wavelet Transform Coupled by EZW and SPIHT Coder for Image Compression, V5-N12-2 .DOI: 10.5815/ijigsp.2013.12.02 *I.J. Image, Graphics and Signal Processing,* 2013.
25. *Ismail Boukli Hacene ;Mohammed BELADGHAM, Abdelhafid BESSAID,* " lossy color Medical Image Compression Using CDF wavelet lifting scheme" . Vol.5.No.11, pp.53-60.2013. DOI:10-5815/ijigsp.2013.11.06. *I.J. Image, Graphics and Signal Processing,* 2013.

This page is intentionally left blank

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 7 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

SDLC and Development Methodologies

By Richard Scroggins

Capella University, United States

Introduction- The overall purpose of any design review is to make sure that the stakeholders understand the design, and that they can confirm that the project team understands their needs and that the project is on the right track. In specific terms that are relevant to SDLC, since SDLC, or Systems Development Life Cycle, is a cyclical methodology, phases repeat, so changes can be made to the design in the next cycle. This makes the process a little less rigid compared to the design process in a linear methodology like waterfall. That is one reason why it is important to know the methodology upfront, because the approach for each phase changes based on the methodology. The design phase is very important in any project and regardless of the methodology, so a lot of time should be dedicated to the design phase.

GJCST-C Classification : *D.2.9*

Strictly as per the compliance and regulations of:

SDLC and Development Methodologies

Richard Scroggins

I. Introduction

The overall purpose of any design review is to make sure that the stakeholders understand the design, and that they can confirm that the project team understands their needs and that the project is on the right track. In specific terms that are relevant to SDLC, since SDLC, or Systems Development Life Cycle, is a cyclical methodology, phases repeat, so changes can be made to the design in the next cycle. This makes the process a little less rigid compared to the design process in a linear methodology like waterfall. That is one reason why it is important to know the methodology upfront, because the approach for each phase changes based on the methodology. The design phase is very important in any project and regardless of the methodology, so a lot of time should be dedicated to the design phase. Once requirements have been gathered from the stakeholders, and the design process has started, you need to have a way to communicate back to the stakeholders that you understood their needs and that they have been incorporated into the project. They design review is an effective way to do this. By making sure that all team members and stakeholders meet and discuss the design and any needed changes, you are able to ensure that needs will be met, and that the project will be a success. The design phase is one of the five phases of the SDLC model; Analysis, Design, Implementation, Testing, and Evaluation. Once the requirements gathering is done, design is done to present to the stakeholders that their needs have been fully understood. Again, since the SDLC methodology is cyclical, changes can be made to the design in the next cycle if the feedback from stakeholders warrants it.

The Systems Development Life Cycle is more than just a theoretical concept; it is used every day in IT departments around the world. Sinason and Normand (2006) studied the Systems Development Life Cycle and the real world application for Omni Furniture Company as well as the benefits to student who study real world cases, " Organizations constantly adapt their information systems to reflect changes in the type of information needed because of changes in technology, the organization's business processes, the organization's structure, or the external environment. A process called the systems development life cycle (SDLC) has been developed to ensure that these changes are orderly and productive. Because of the importance of this process, all accounting information systems (AIS) textbooks present the SDLC as either a four- or five-stage cycle of activities. Thus, most students are introduced to the SDLC in the accounting systems course but few students have the chance to actually experience the process. The Omni Furniture Company Case helps students enrolled in an AIS course further their understanding of the SDLC by thinking through all stages of the process and designing a system that meets the users' information and internal control needs." (p. 01).

Author: *e-mail: mr_scroggins@yahoo.com*

II. Agile Methodology

What are the significant features of this approach?

The Agile approach is cyclical in nature and is based on iterative and incremental development, where requirements and solutions evolve through collaboration. Due to the cyclical nature, it allows for problems to be resolved as phases repeat. It allows for issues to be found and then addressed in the next cycle.

In what type of environment or situation will this approach be most appropriate?

Software development using standard methods and parameters.

What are the weaknesses of this approach, relative to other approaches?

With shorter phase time compared to Waterfall, some things can be missed early in the project.

III. Test Driven Development

What are the significant features of this approach?

Test Driven Development is another cyclical design methodology that is based on short cycles. Due to the cyclical nature, it allows for problems to be resolved as phases repeat. Studies have found this approach to be more productive, principally due to the hands on approach.

In what type of environment or situation will this approach be most appropriate?

This method is good when the direction is not clear and trial and error is required.

What are the weaknesses of this approach, relative to other approaches?

This method relies heavily of testing and short cycles. It could be described as brutish.

References Références Referencias

1. Introduction to Test Driven Design (n.d.). Retrieved April 12, 2013, from http://www.agiledata.org /essays/tdd.html
2. Project Management Methodologies (n.d.). Retrieved April 12, 2013, from http://infolific.com/technology/methodologies/
3. Gelowitz, C., Sloman, I., Benedicenti, L., Paranjape, R. (2003). Real-time extreme programming. XP'03: Proceedings of the 4th international conference on Extreme programming and agile processes in software engineering. Retrieved from The ACM Digital Library.
4. Sinason, D. H., & Normand, C. J. (2006). Omni furniture company: A systems development life cycle case. Journal of Information Systems, 20(2), 81-91. Retrieved from http://search.proquest.com.library.capella.edu/docview/235939890?accountid=27965

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 7 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Job Stress in Software Companies: A Case Study of HCL Bangalore, India

By Geeta Kumari, Dr. Gaurav Joshi & Dr. K. M. Pandey

Jharkhand Rai University, India

Abstract- In India software industry has become one of the fastest growing industries. The reason for choosing a particular software industry and its employees is that the level of stress these employees face is comparatively higher than employees in other private companies. Any kind of a job has targets and an employee becomes stressed when he or she is allotted with unachievable targets and are unable to manage the given situation. Thus the main aim of this article is to bring to lime light the level of stress with software employees in HCL and the total sample size for the study is 100 chosen by random sampling method in HCL. When the employees were asked how often they feel stressed while working while working at the company, 98 out 100 said they feel stressed about daily physically, mentally or emotionally. In a sample space of 100 employees, 16% of employees work for 4-6 hours a day, 32% works for 6-8 hours, 30% works for 8-10 hours while 22% works for 10-12 hours daily. When asked about the overtime they have to do, 28% said employees saying always were mainly from age group 20-29. 54% of employees said often while 18% of employees said they worked overtimes rarely. The other few aspects have also been studied about job satisfaction.

Keywords : stress, HCL, software industry, emotions, sample size, overtime.

GJCST-C Classification : K.4

Strictly as per the compliance and regulations of:

Job Stress in Software Companies: A Case Study of HCL Bangalore, India

Geeta Kumari [α], Dr. Gaurav Joshi [σ] & Dr. K. M. Pandey [ρ]

Abstract- In India software industry has become one of the fastest growing industries. The reason for choosing a particular software industry and its employees is that the level of stress these employees face is comparatively higher than employees in other private companies. Any kind of a job has targets and an employee becomes stressed when he or she is allotted with unachievable targets and are unable to manage the given situation. Thus the main aim of this article is to bring to lime light the level of stress with software employees in HCL and the total sample size for the study is 100 chosen by random sampling method in HCL. When the employees were asked how often they feel stressed while working while working at the company, 98 out 100 said they feel stressed about daily physically, mentally or emotionally. In a sample space of 100 employees, 16% of employees work for 4-6 hours a day, 32% works for 6-8 hours, 30% works for 8-10 hours while 22% works for 10-12 hours daily. When asked about the overtime they have to do, 28% said employees saying always were mainly from age group 20-29. 54% of employees said often while 18% of employees said they worked overtimes rarely. The other few aspects have also been studied about job satisfaction.

Keywords: stress, HCL, software industry, emotions, sample size, overtime.

I. Introduction

In the 1990s India emerged as a major player in the field of software engineering, information technology services and Web-based services. Presently the Indian information technology industry employs a little more than half million people and provides indirect employment to over a quarter of a million people (NASSCOM). In today's world, the degree of stress increased owing to urbanization, globalization that results into cut-throat competition. Stress is inescapable part of modern life, work place is becoming a volatile stress factory for most employees and it is rightly called as the Age of anxiety. Stress has becoming significantly with the result of dynamic social factors and changing needs of life styles. Stress is man's adaptive reaction to an outward situation which would lead to physical mental and behavioural changes. Brain cells create ideas, Stress may kill brain cells. The truth is that not all stresses are destructive in nature. Appropriate amount of stress can actually trigger your passion for work, tap your latent abilities and even ignite inspirations. Stress is the emotional and physical strain caused by our response to pressure from the outside world. Common stress reactions include tension, irritability, inability to concentrate, and a variety of physical symptoms that include headache and a fast heartbeat. Stress is a condition or feeling experienced when a person perceives that- demands exceed the personal and social resources the individual is able to mobilize. S= P>R i.e., stress occurs when the pressure is greater than the resources. Stress is our body's way of responding to any kind of demand. It can be caused by both good and bad experiences. When people feel stressed by something going on around them, their bodies react by releasing chemicals into the blood. These chemicals give people more energy and strength, which can be a good thing if their stress is caused by physical danger. But this can also be a bad thing, if their stress is in response to something emotional and there is no outlet for this extra energy and strength.

Stress is everywhere, but as a relatively new phenomenon. How can we define it and how can we explain its extraordinary cost to both business and government? The suffering induced by stress is no figment of the imagination but can we accurately examine the relationship between stress and ill-health? Whatever stress is, it has grown immensely in recent years, which brings us to question – what is happening in society that is causing stress? The report shows that stress has its greatest effects on those at the very top and those at the very bottom of the socio-economic ladder. The Indian Software industry has grown at a compounded annual growth rate (CAGR) of 28 % during the last 5 years. The key segments that have contributed significantly to the industry's exports include – software services - BPO sector is playing vital role in the growth of our country's economy. Due to liberalization of Indian economic policy, the growth of software industry is in commendable position. Due to cost advantage, availability of skilled manpower, quality services are the main reasons for the growth of IT industry in India. The perception of the effects of stress on an individual has changed. Stress is not always dysfunctional in nature, and, if positive, can prove one of the most important factors in improving productivity within an organization (Spielberger, 1980). If not positive, stress can create a

Author α: Research Scholar, Department of Management, Jharkhand Rai University, Ranchi, India. e-mail- geekumari@gmail.com
Author σ: Lal Bahadur Shastri Institute of Management, New Delhi, India. e-mail- gauravjoshi91983@gmail.com
Author ρ: Department of Mechanical Engineering National Institute of Technology, Silchar, Assam, India.
e-mail- kmpandey2001@yahoo.com

number of physical and psychological disorders among employees, and can be responsible for frustration, haste, and job dissatisfaction. As a result, the lack of work may cause complacency within the organization. Stress is, therefore, multidimensional, and its results depend on whether employees perceive it as a problem or a solution.

Different studies have classified occupational stress in terms of physical environment; role stressors, organizational structure, job characteristics, professional relationships, career development, and work-versus-family conflict (see Burke, 1993). Cooper and Marshall (1976) add to this list factors intrinsic to a job, the management's role, and professional achievements. Based on these complexities, stressors can be grouped into two main categories: (i) job-related stressors, and (ii) individual-related stressors. Stress is a slow and insidious malady which is an unavoidable one and a common problem in the workplace. The level of stress and its amount of consequences vary between organizations based on the nature and type of work practices. Organization must begin to manage people at work differently, treating them with respect and valuing their contribution. Recognition, participation and continuous training of employees are required to retain the skilled employees. Stress issue has become contemporary, being an occupational hazard in fast pacing IT profession, needs to be addressed without delay. Hence the importance of the study of stress at various levels, among IT employee is growing. At organizational level, well designed coping strategies have become the attention of companies like Tata Consultancy Services, Infosys, Wipro, Microsoft, and Cognizant etc., Stress can make an individual productive and constructive when it is identified and well managed. In times of great stress or adversity, it's always best to keep busy, to plough anger and energy into something positive. Positive attitude and meditation will be helpful for coping the stress. Having broader perspective of life will definitely change the perception of stress. Let us hope that we will be successful eliminating stress for our healthy lifestyle as well as organizational well-being. Stress is measured using a number of instruments. Our focus, however, is organizational role stress (ORS), which measures total role stress. We use Pareek's (1983) scale, which evaluates respondents' quantum of stress in terms of total ORS scores. It also measures the intensity of the following ten role stressors that contribute to the total ORS score:

1. Inter-role distance (IRD): Conflict between organizational and non-organizational roles.
2. Role stagnation (RS): The feeling of being "stuck" in the same role.
3. Role expectation conflict (REC): Conflicting expectations and demands between different role senders.
4. Role erosion (RE): The feeling that functions that should belong to the respondent's role are being transformed/performed or shared by others.
5. Role overload (RO): The feeling that more is expected from the role than the respondent can cope with.
6. Role isolation (RI): Lack of linkages between the respondent's role and that of other roles in the organization.
7. Personal inadequacy (PI): Inadequate knowledge, skills, or preparation for a respondent to be effective in a particular role.
8. Self-role distance (SRD): Conflict between the respondent's values/self-concepts and the requirements of his or her organizational role.
9. Role ambiguity (RA): Lack of clarity about others' expectations of the respondent's role, or lack of feedback on how others perceive the respondent's performance.
10. Resource inadequacy (RIn): No availability of resources needed for effective role performance.

II. Review of Literature

Darshan et al (2009)1 in their article, A study on professional stress, depression and alcohol use among Indian software professionals, observed that the software employees are professionally stressed and are at 10 times higher risk for developing depression and also significantly increase the incidence of psychiatric disorders. Preventive strategies like training in stress management, frequent screening to identify professional stress and depression at the initial stages and addressing these issues adequately might help the software professionals cope with their profession better without affecting their lifestyle and health. Saurabh Shrivastava and Prateek Bobhate (2010)2 in their study, Computer related health problems among software professionals in Mumbai: A cross-sectional study, investigated that Ocular discomfort, musculo-skeletal disorders and psycho-social problems form key category of health problems found among constant computer users. This study has also brought into focus factors contributing to the occurrence of these problems. Thus, the problem requires a multidisciplinary action and hence there is an immediate need for the concerned authorities to collaborate and enforce suitable preventive measures.

Jakkula Rao and Chandraiah (2011)3 in their article, Occupational stress, mental health and coping among information technology professionals, found that job satisfaction and mental health are correlated but not significant. However, job satisfaction was positively and significantly correlated with coping behaviour. The mental health is negatively and significantly correlated with occupational stress. It can be explained that as job satisfaction and mental health increases coping

behaviour increases. And as stress increases mental health decreases. Kesavachandran et al (2012)4 in their study, Working conditions and health among employees at information technology - enabled services: A review of current evidence, identified that muscular-skeletal disorders, ocular disorders and psycho-social problems were some of the key health problems observed among software professionals. There is a need for implementation of the programs that include the concepts of ergonomics, health education, training of personnel to prevent and overcome the morbidity, as well as psycho-social problems among workers in software industry.

Michael R. Frone (2008), the relationship of work stressors, those work over load and job insecurity, to employees alcohol use illicit drug use resulted, support the relation of work stressors to alcohol and illicit drug use before work, during the workday, and after work. Vijay V. Raghavan, (2010), the effect of flexible work schedule, employee support and training, and telecommuting as potential coping resources to relieve stress. Perceived workload, role ambiguity, work facilitation, and decision latitude are potential stressors of IT professionals. Removing role ambiguity and improving work facilitation reduce work-related stress and allowing employees to have flexible work schedules ease their perceptions of workload. Sahana Charan, (2007), High work pressure, long hours in front of the computer and a fast-paced lifestyle, if these factors team up to weaken your physical health, here is one more strong reason why they are simply unhealthy: mental health professionals are now convinced that an increasing number of persons working in the IT and IT-enabled services sector fall prey to depression, because of the high stress they undergo. Murali Raj, (2009), Depression is usually related to work and stress these people undergo because of the pressure to perform better, compete with other colleagues and meet tight deadlines. Most of their work is target-oriented and if targets are not met, it can lead to anxiety. Peers are not very supportive as they also competing in the same field. Moreover, insecurity about the job may lead to feelings of expression.

Elkin and Rosch (1990) have summarized a wide range of other strategies which are directed towards increasing worker autonomy, participation and control. These strategies include: redesigning tasks, redesigning the physical work environment, role definition and clarification, establishing more flexible work schedules, participative management, employee-centred career development programmes, providing feedback and social support for employees and more equitable reward system. These are approaches which could prevent stress at work rather than treat stress once it has developed. 75% to 90% of all visits to primary care physicians are for stress-related complaints.

- 40% of job turnover is due to stress; Up to 80% of on-the-job accidents are stress-related.
  -American Institute of Stress
- The annual cost to Canadian companies due to stress-related disorders is $12 billion Absenteeism due to stress has increased by over 300% since 1995
  -Statistics Canada
- Employees in extreme workplace stress conditions suffer from: more than triple the rate of cardiovascular problems; over five times the rate of colorectal cancer; up to three times the rate of back pain
  -Health Canada
- Problems at work are more strongly associated with health complaints than are any other life stressor; more so than even financial problems or family problems

St.Paul Fire and Marine Insurance Co.

Every year in Japan around 30,000 deaths occur because of Karoshi (over work). In a study conducted by Delhi based NGO - Saarthak in 30 Indian companies, it was found that 50% of the employees suffered from stress related problems. Further, in the studies conducted in the US and UK, it was found that more than 60% of employees complain to be stressed out in their jobs. Pestonjee and Singh (1983) study the psychodynamics of people working in the field of computers as software or hardware personnel. In this study job satisfaction and morale were taken, as dependent variables and alienation, participation, involvement and role stress were independent variables. It was hypothesized that personas scoring high on the role stress measure would be less satisfied and obtain lower scores on the morale measure in comparison to those who scored low on the role stress measure. Singh (1987) conducted another study related to computer professionals. While reviewing the literature, he noted that there are very few studies on computer professionals and foreign researchers using foreign samples conduct all of them. All such studies have reported that job dissatisfaction, high role stress and high rate of turnover are common phenomena related to computer professionals.

Mishra et al (1997) studied the nature and inter relationship between motivation and role stress on entrepreneurs in and around Delhi. The major findings of the study revealed that women entrepreneurs scored higher on the motivational variables namely safety, belongingness, self-esteem and self-actualisation as compared to role stagnation, role isolation and role ambiguity. Self-esteem was associated positively and significantly with role overload. Sharma et al (2001) found that gender related unequal division of domestic duties when coupled with a job, may not result in more severe psychological or subjective health impairments. It was found that job provides women with means of

feeling useful and important and provides an opportunity to interact with people and this could be the important source of satisfaction for women. The study also revealed that paid work enhances the status of the employee resulting in enhanced self-esteem. Matthews et al. (2006), in his study compared EI and the personality factors of the Five Factor Model (FFM) as predictors of task-induced stress responses. Results confirmed that low EI was related to worry states and avoidance coping, even with the FFM statistically controlled. However, EI was not specifically related to task-induced changes in stress state.

III. Objective of Study

1. To study on job stress among employees of software industries.
2. To examine the relaxation techniques practiced in the organization.
3. To study the relationship between self-esteem and stress.
4. To ascertain the impact of job stress on personal health of employees.
5. To give some suggestions for future studies.

IV. Conceptual basis of the Study

This qualitative study takes the lead from a recent survey (2010), published in the Journal of Occupational and Environmental Medicine, and noted that for those working 12 hours a day, there was a 37% increase in risk of illness and injury in comparison to those who work fewer hours. And another study done by North-western National Life, reports that one-fourth of employees view their jobs as the number one stressor in their lives. A St. Paul Fire and Marine Insurance Co. study concluded that problems at work are more strongly associated with health complaints than any other life stressor, even financial or family problems.

Sethy and Schulter (1996) outlined four major reasons why job stress and coping have become important issues:

1. *Concern for individual employee health and well-being:* (E.g. coronary heart disease, high blood pressure, job related accidents).
2. *The financial impact on organization:* (Including days lost due to stress related illness and injury).
3. *Organizational effectiveness:* for organizational health and wellbeing.
4. *Legal obligations:* on employers to provide safe and healthy working environment.

According to Global business and Economic Roundtable of Addiction and Mental Health, the top 9 stressors are:

a) Lack of Control

Less control employees have over their situations, the greater their stress. Solicit and consider employee suggestions, comments and input.

b) Lack of Communication

Try communicating early and often, making sure you listen as often as you deliver news or observations.

c) No Appreciation

When is the last time you praised an employee for a job well done? Say "Thank you" more often. Put it in writing for even greater impact. Corporate wellness is a good investment, with a strong return on investment.

d) No Feedback, good or bad

Don't wait until the annual review to let employees know how they're doing. They wonder every day. Career and Job ambiguity Uncertainty about opportunity within the company or job security can lead to a feeling of loss of control. Keep employees clear about performance goals, room for advancement and how your organization is doing.

e) Unclear Policies and no Sense of Direction

Clearly communicate policies and company goals, and alert top management if employees need further clarity.

f) Mistrust, Unfairness and Office Politics

It's important to treat everyone the same and perfectly appropriate to reprimand someone who is negative about other employees. Backbiting keeps everyone on edge.

g) Pervasive Uncertainty

This results from inadequately explained or unannounced changes. Meet with people individually to review changes. Follow those meetings with a written memo so everyone can review the facts after emotions have died down.

h) Random Interruptions

Telephone calls, e-mails, walk-ins and supervisor demands can keep employees from completing the work at hand. Consider time management training to help people prioritize and delegate.

i) The Treadmill syndrome

Having too much or too little to do results in self-defeating behaviour that can lead to high stress. Make sure work is evenly divided, and hire additional help where needed.

The research problem is formulated on the basis of vast study of related literature survey which provides theoretical background and conceptual frame work to this study which broaden knowledge base in this area of research. The research takes the lead from the following dimensions; those are the impact of stress on

body, mind, behaviour and emotions on the basis of review of literature.

Impact of stress in various dimensions

- *Impact on Body*
  1. Headaches, taut muscles, breathlessness
  2. Worrying, muddled thinking, night mares
  3. Accident prone, loss of appetite
- *Impact on Mind*
  1. Loss of confidence, more fussy
  2. Frequent infections, skin irritations
  3. Impaired judgment, indecisions
- *Impact on Behaviour*
  1. drinking and smoking more
  2. irritability, depression, apathy
  3. Fatigue, muscular twitches
- *Impact on Emotions*
  1. Negativity, hasty decisions
  2. Loss of sex drive, insomnia
  3. Alienation, apprehension

V. Hypothesis

The Hypotheses may be stated as:

1. The relationship between the profile of the individual and the level of job stress in the company.
2. The relationship between esteemed employees and stress.
3. The relationship between Gender and Personal health.
4. The relationship between marital status and personal health.
5. Organizational level of outcomes is dependent on occupational stress.

VI. Analysis and Discussion

The study is based on a survey conducted in HCL in Bangalore with sample size 100. Out of total sample, 69 were men and rest women. The respondents were software professionals at lower and middle levels with a mean of about 5 years in the company.

When the main cause of stress was asked from the employees the main reasons came out to be:-

1. Work environment
2. Supervision
3. Workloads
4. Social injustice
5. Organisational culture
6. Fear of loss of job
7. Operating style

Table1: shows sample description on the basis of employee's age group.

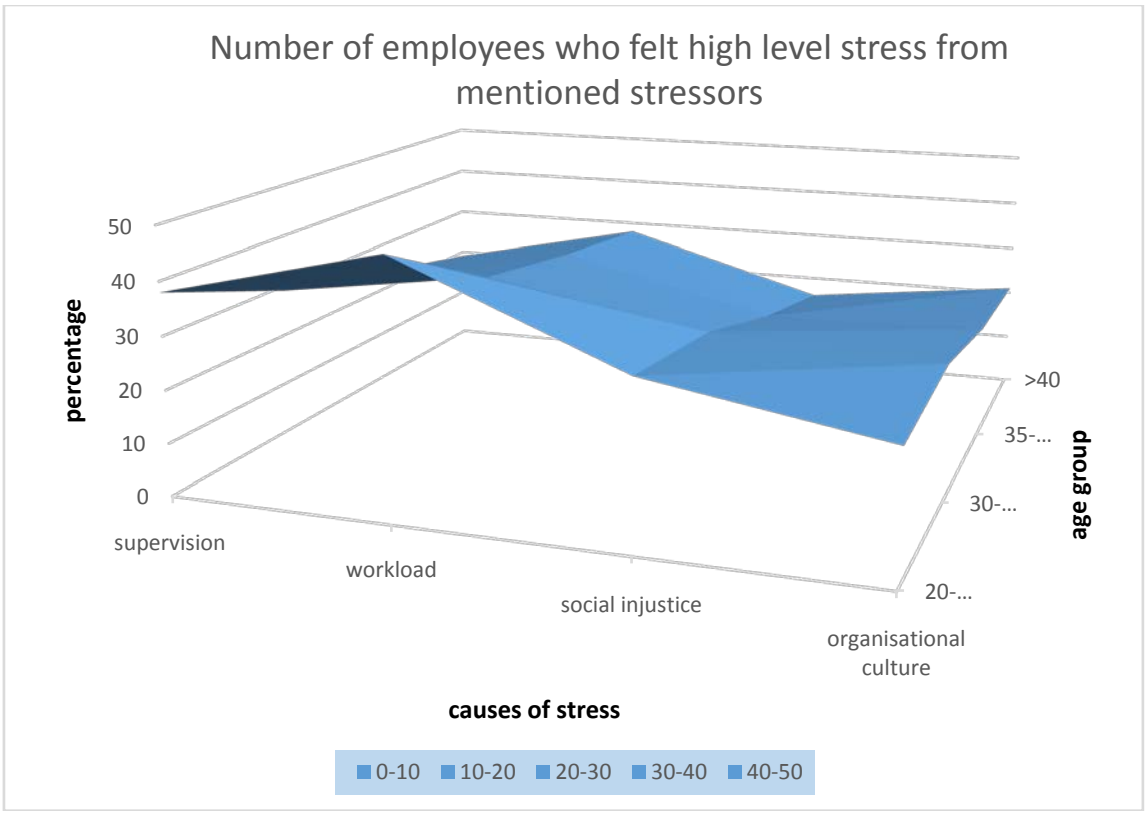
Table 1 : Sample description

Sample size=100

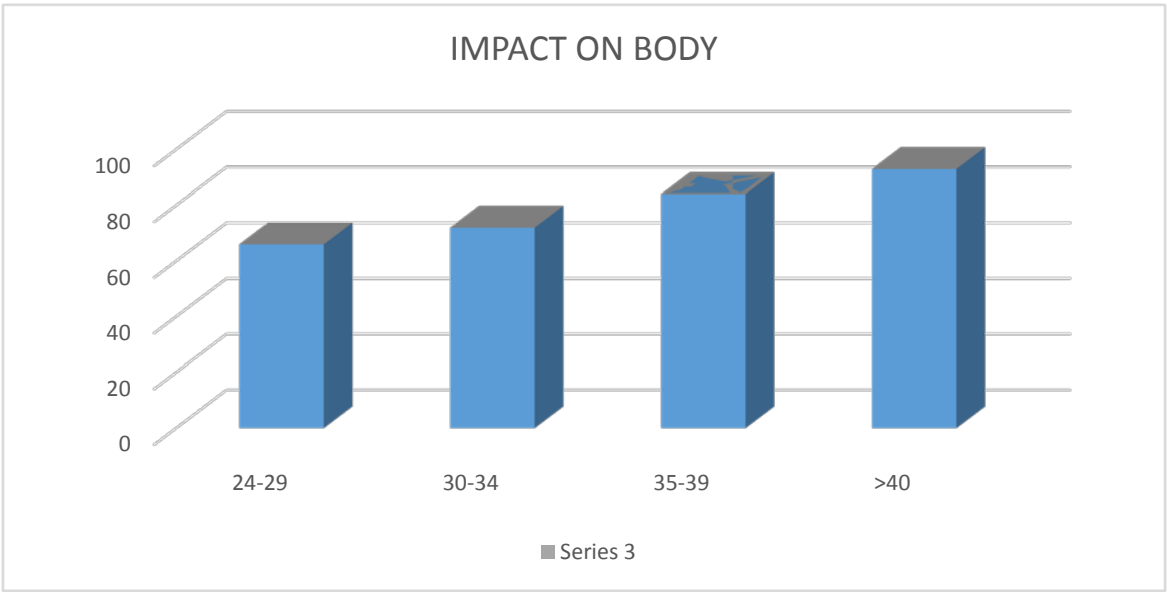
Analysis of the Table 1 statistics

The chart below shows the different causes of stress faced by different groups of people along with their percentage.

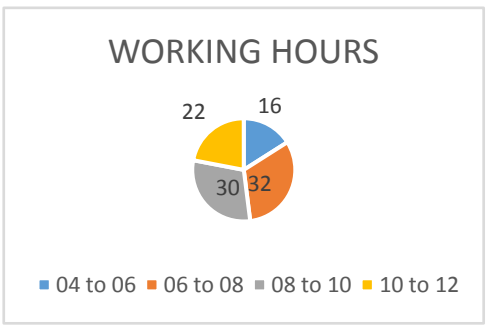
AGE GROUP	SAMPLE SIZE
20-29	18
30-34	24
35-39	36
>40	22



- Impact on Body

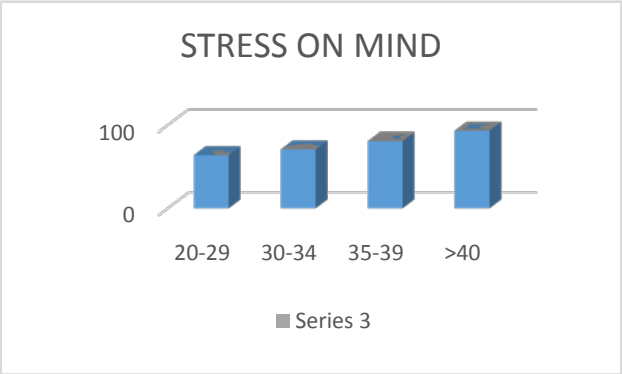


From the Table 1, 93% of employees from >40 age group felt high level of stress impact on body followed by 84% from 35 -39 age group, 72% from 30 - 34 age group and 66% from 20 -29 age group. Headaches, fatigue, Hypertension, Coronary artery diseases, Skin disease etc. are impacts of stress. The data obtained from the employees, draw the attention and alarming the individual as well as the company. Considering the physical stress when they were asked about the number of hours they were working daily the answer was-



In a sample space of 100 employees, 16% of employees work for 4-6 hours a day, 32% works for 6-8 hours, 30% works for 8-10 hours while 22% works for 10-12 hours daily. When asked about the overtime they have to do, 28% said employees saying always were mainly from age group 20-29. 54% of employees said often while 18% of employees said they worked overtimes rarely.

- Impact on Mind



VII. Conclusions

93% of employees from 35 -39 age group felt high level of stress impact on mind followed by 91% from the age group of 30 -34 and 80% from the rest age groups. Stress on mind causes Depressions, Anger, Irritability, Mood swings, Lack of self-confidence etc. which leads to vulnerable effect on individual. When asked that do they feel that they are constantly under pressure going from one deadline to another, 52% of the employees said always, 31% of the employees said often while 17% of the employees said sometimes. When asked that do they feel that they have lost or losing a sense of control in their life and that the balance they need is gone, 53% of the employees said always, 29% of the employees said often while 18% of the employees said sometimes. When asked about how they feel while working in the organization, the answers were unexpected. Only 26% of the people were feeling satisfied or great and rest were just working as if they were no other options. Remaining 74% were frustrated, depressed or unable to concentrate. The following pie chart shows their percentage.

References Références Referencias

1. Arnold, H.J. (1960). Moderator Variable: A classification of conceptual, analytic and psychometric issues; Organizational Behaviour and Human Performance, 29, pp 143 – 174.
2. Beehr, T.A. and NewMan, S.E (1978). Job Stress, employee health and organizational effectiveness: Facet analysis, Personnel Psychology, winter, pp 665 – 669.
3. Selye H (1956). The stress of life, Mc Graw Hill, New York.
4. Pestonjee D.M and Singh G.P (1987). Organizational Behaviour issues for managers and systems analysts. Working paper No. 660, Indian Institute of Management, Ahmedabad.
5. Palnitkar A and Helode R.D (1987). Stress and Coping, the Indian Experience, (II ED.). D.M.Pestonjee, Sage Publications, New Delhi, 2002, p 133.
6. Pareek A. (1999). Managing stress and coping. In D.M.Pestonjee, U.Pareek and A.Agarwal (Eds), Studies in Stress and its Management, New Delhi, Oxford and IBH.
7. Singh AP and Singh B (1997). Correlates of psychological strain and coping behaviour. In D.M Pestonjee and U.Pareek (Eds), Studies in Organizational role stress and coping. Jaipur / New Delhi; Rawat Publication.
8. Nadia Clarkin and Richard Wynne ,Work Research Centre, Dublin, Ireland, "Vocational rehabilitation and work resumption - A review of the literature".
9. Stephen Palmer and Kristina Gyllensten, "The role of gender in workplace stress: a critical literature review" Health Education Journal, Vol. 64, No. 3, 2005, pg.271-288.
10. Kumar Sunil and Rooprai K. Y. , Role of Emotional Intelligence in Managing Stress and Anxiety at workplace, Proceedings of ASBBS Annual Conference: Las Vegas February 2009 Volume 16 Number 1, Pg. 5-6.
11. Hansez, I., Bertrand, F. and Barbier, M., 'Evaluation des pratiques de diagnostic de stress au sein d'entreprises belges: Facteurs bloquants et facteurs stimulants' [Evaluation of stress diagnosis practices in Belgian companies: Blocking and stimulating factors], Le Travail Humain, Vol. 72, No. 2, 2009, pp. 127–153.
12. Palm M. Lehrel, Robert L. Woolfolk, Wepley E. sime, (2007), Stress Management, 3rd Edition, Guilford press. Journal of Information Technology and Economic Development 2(2), 30-48, October 2011 47.
13. Gregory Moorhead, Ricky W. Griffin, (2005), Organizational Behavior, 7th Edition Biztantra pub.
14. Nirmal singh, (2002), Human Relations and Organizational Behavior, Deep & Deep pub.
15. Ajay Shukla, (2006), Stress Management, 2nd edition Unicorn pub.
16. Suja R. Nair, (2005), Organizational Behavior, 1st Edition, Himalaya pub., India.

This page is intentionally left blank

Global Journal of Computer Science and Technology: C
Software & Data Engineering
Volume 14 Issue 7 Version 1.0 Year 2014
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Optimized Anomaly based Risk Reduction using PCA based Genetic Classifier

By C. Kavitha & Dr. K. Iyakutti

Pasumpon Muthuramalinga Thevar College, India

Abstract- Security risk analysis is the thrust area for the information based world. The researchers in this field deployed numerous techniques to overcome the information security oriented problem. In this paper the researcher tried for a approach of using anomaly detection for the risk reduction. The hub initiative for this work is that the anomalies are the deviation which could increase the percentage of risk. The anomaly detection is guided by the PCA and the genetic based multi class classifier is used. The classification is induced by the decision tree approach were the genetic algorithm is set out for the optimization in the process of finding the nodes of the tree. The proposed approach is evaluated with the bench mark on PCA based ANN classifier. The proposed approach outperforms the existing one. The results are demonstrated.

Keywords : *anomaly detection, PCA, genetic algorithm.*

GJCST-C Classification : *D.2.9*

Strictly as per the compliance and regulations of:

Optimized Anomaly based Risk Reduction using PCA based Genetic Classifier

C. Kavitha [α] & Dr. K. Iyakutti [σ]

Abstract- Security risk analysis is the thrust area for the information based world. The researchers in this field deployed numerous techniques to overcome the information security oriented problem. In this paper the researcher tried for a approach of using anomaly detection for the risk reduction. The hub initiative for this work is that the anomalies are the deviation which could increase the percentage of risk. The anomaly detection is guided by the PCA and the genetic based multi class classifier is used. The classification is induced by the decision tree approach were the genetic algorithm is set out for the optimization in the process of finding the nodes of the tree. The proposed approach is evaluated with the bench mark on PCA based ANN classifier. The proposed approach outperforms the existing one. The results are demonstrated.

Keywords: *anomaly detection, PCA, genetic algorithm.*

I. Introduction

The day to day operations of any enterprise is highly prone to hijack of information. It is very essential part of any enterprise is to put an eye on the security measures so as to minimize the risk of information hijack. The information leaking is not only affecting the day to day activities of the enterprise but also restricts the long term viability of the enterprise. Networks play an important role in the information interchange. Without the communication many of the operation in the gross root level to executive level will be affected. The communication is the back bone of the information and resource sharing. At the same time the network communications are highly riskier because of the intruders. In order to provide a secured information exchange among the enterprise, the first work to be done is to strengthen the network security. The core point of information security is risk management [1]. The people employed with this area use various control and counter measures to defend the security vulnerabilities, but such practices not fulfill the system to be protected against the terrorization due to the intrinsic weakness of the security systems.

Network security measure could be employed on two phases, either the proactive strategy or the reactive strategy. Always the proactive strategy plays a key role for the risk reduction in the information loss and thereby increasing the security level for the enterprise. This paper discuss about a proactive strategy which uses the anomaly detection for the risk reduction. Intrusion detection techniques can be categorized in misuse detection and anomaly detection. Misuse detection systems find intrusions by matching sample data to known intrusive pattern. Anomaly detection systems find intrusion by analyzing the deviation from normal activities profiles that are retrieved from historical data. Intrusion detection is a critical component of secure information systems [2].

The anomaly detection is based on a genetic classifier. This evolutionary strategy is used for the anomaly detection to detect the unusual patterns in the historical data and able to classify the test data. Evolutionary algorithms are a solid but computationally expensive heuristic. Nevertheless, progressively faster computational resources have allowed evolutionary algorithms to be increasingly used in a variety of applications over the years [3]. To facilitate the evolutionary algorithms to come out of the above said drawback Principal Component Analysis (PCA) is employed as a preprocessing measure in this paper.

The paper is organized in the following sections as section 2 talks about the background study of the concepts deployed in the paper, section 3 discuss on the existing methodology of the core problem. Problem formulation is thrashed out in the section 4. Section 5 deals with the proposed approach of this paper. The experimental details and the results obtained are depicted in the section 6. Discussion on the results is carried out in the section 7. Section 8 concludes the paper.

II. Background Study

This section briefly discuss on the concepts used in this paper. Information security, Anomaly detection, Evolutionary algorithms, Classification, Genetic algorithm, PCA is conferred in this section.

a) *Information security*

In today's fast-changing IT world, even the best available security is insufficient for the latest vulnerabilities in various products, and against malware/attacks created to target those vulnerabilities. While cyber-security cannot be 100 per cent fool-proof, we can still try to achieve the maximum security possible

Author α: Department of Computer Science, Pasumpon Muthuramalinga Thevar College, Madurai, Tamilnadu,India.
e-mail: kkavitha009@gmail.com
Author σ: Department of Physics & Nanotechnology, SRM University,Kattankulathur, Chennai, Tamilnadu, India .
e-mail: iyakutti@gmail.com

[4]. The purpose of the intrusion detection system is to help the computer system on how to deal with the attacks that IDS is collecting information from several different sources within the computer systems and networks and compares this information with pre existing patterns of discrimination as to whether there are attacks or weaknesses [5]. The goal of an intrusion detection system is to provide an indication of a potential or real attack. An attack or intrusion is a transient event, whereas vulnerability represents an exposure, which carries the potential for an attack or intrusion. The difference between an attack and vulnerability, then, is that an attack exists at a particular time, while vulnerability exists independently of the time of observation. Another way to think of this is that an attack is an attempt to exploit vulnerability [6]. An intrusion detection system examines system or network activity to find possible intrusions or attacks. Intrusion detection systems are either network-based or host-based. The information security is highly reliable on intrusion detection in the network since the network is the channel for the intrusion and provides the ways of mishandling the information. The intrusion could be detected in three ways

- Event or Signature-based Analysis
- Statistical Analysis
- Adaptive Systems

The event, or signature-based, systems function much like the anti-virus software with which most people are familiar. The vendor produces a list of patterns that it deems to be suspicious or indicative of an attack; the IDS merely scan the environment looking for a match to the known patterns. The IDS can then respond by taking a user-defined action, sending an alert, or performing additional logging. This is the most common kind of intrusion detection system.

A statistical analysis system builds statistical models of the environment, such as the average length of a telnet session, and then looks for deviations from "normal". After over 10 years of government research, some products are just beginning to incorporate this technology into marketable products.

The adaptive systems start with generalized rules for the environment, then learn, or adapt to, local conditions that would otherwise be unusual. After the initial learning period, the system understands how people interact with the environment, and then warns operators about unusual activities. There is a considerable amount of active research in this area.

Intrusion detection based on anomaly detection techniques has a significant role in protecting networks and systems against harmful activities [7].

b) Anomaly detection

Anomaly Detection is an important alternative detection methodology that has the advantage of defending against new threats not detectable by signature based systems. In general, anomaly detectors build a description of normal activity, by training a model of a system under typical operation, and compare the normal model at run time to detect deviations of interest. Anomaly Detectors may be used over any audit source to both train and test for deviations from the norm [8]. The goal of the anomaly detection is to find all objects that are different to other objects. Anomaly detection finds extensive use in a wide variety of applications such as fraud detection for credit cards, insurance or health care, intrusion detection for cyber-security, fault detection in safety critical systems, and military surveillance for enemy activities [9]. The challenges of the anomaly detection is listed in [9] as follows

- Defining a normal region which encompasses every possible normal behavior is very difficult. In addition, the boundary between normal and anomalous behavior is often not precise. Thus an anomalous observation which lies close to the boundary can actually be normal, and vice-versa.
- When anomalies are the result of malicious actions, the malicious adversaries often adapt themselves to make the anomalous observations appear like normal, thereby making the task of defining normal behavior more difficult.
- In many domains normal behavior keeps evolving and a current notion of normal behavior might not be sufficiently representative in the future.
- The exact notion of an anomaly is different for different application domains. Thus applying a technique developed in one domain to another is not straightforward.
- Availability of labeled data for training/validation of models used by anomaly detection techniques is usually a major issue Often the data contains noise which tends to be similar to the actual anomalies and hence is difficult to distinguish and remove.

c) Evolutionary algorithms

Evolutionary algorithms are stochastic search methods that mimic the metaphor of natural biological evolution [10]. Evolutionary algorithms operate on a population of potential solutions applying the principle of survival of the fittest to produce better and better approximations to a solution. At each generation, a new set of approximations is created by the process of selecting individuals according to their level of fitness in the problem domain and breeding them together using operators borrowed from natural genetics. This process leads to the evolution of populations of individuals that are better suited to their environment than the individuals that they were created from, just as in natural adaptation. Evolutionary computation (EC) techniques can be used in optimization, learning and design [11].

The most important components of any one of the variants of an evolutionary algorithm is given by [12] as

- Representation
- Evaluation function
- Population
- Parent selection mechanism
- Variation operators, recombination and mutation
- Survivor selection mechanism

d) Classification

Classification consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. The algorithm tries to discover relationships between the attributes that would make it possible to predict the outcome. Next the algorithm is given a data set not seen before, called prediction set, which contains the same set of attributes, except for the prediction attribute –not yet known. The algorithm analyses the input and produces a prediction. The prediction accuracy defines how "good" the algorithm is [13]. These data analysis help us to provide a better understanding of large data. Classification predicts categorical and prediction models predict continuous valued functions [14].

e) Genetic algorithms

A genetic algorithm is a class of adaptive stochastic optimization algorithms involving search and optimization [15]. Genetic algorithms were first used by Holland (1975). Genetic Algorithms (GAs) are adaptive heuristic search algorithm based on the evolutionary ideas of natural selection and genetics. As such they represent an intelligent exploitation of a random search used to solve optimization problems. Although randomized, GAs are by no means random, instead they exploit historical information to direct the search into the region of better performance within the search space. The basic techniques of the GAs are designed to simulate processes in natural systems necessary for evolution; especially those follow the principles first laid down by Charles Darwin of "survival of the fittest." Since in nature, competition among individuals for scanty resources results in the fittest individuals dominating over the weaker ones. [16]

f) Principal Component Analysis

Principal component analysis is appropriate when you have obtained measures on a number of observed variables and wish to develop a smaller number of artificial variables (called principal components) that will account for most of the variance in the observed variables. The principal components may then be used as predictor or criterion variables in subsequent analyses [17]. Principal component analysis (PCA) is a standard tool in modern data analysis - in diverse fields from neuroscience to computer graphics because it is a simple, non-parametric method for extracting relevant information from confusing data sets [18].

III. Existing Methodologies

Security risk assessment and mitigation are two vital processes that need to be executed to maintain a productive IT infrastructure [19]. Risk analysis is the basis of information protection, risk management, and risk in the process of information protection. Risk analysis includes process such as identification of activity, threat analysis, vulnerability analysis and guarantees [20]. The modern security analysis has employed the following techniques like Grey relational making approach [21], Fuzzy number arithmetic operational, Information entropy [22], Fuzzy weighted average approach [23] and Fuzzy measure and Evidence theory [24], Fuzzy AHP method [25-26].

IV. Problem Formulation

In the modern day of communication era Information security highly relies on the network security. In this paper the author tries to employ the anomaly detection mechanism as the base for the identification of the security risk factor. The core idea is the different behavioral pattern leads to the risk. So by the identification of the anomaly could be used to identify the risk.

The major issues in the anomaly detection are the classification mechanism employed. The input data to classifiers is an extremely large set of features, but not all of features are relevant to the classes to be classified. Hence, the learner must generalize from the given examples in order to produce a useful output in new cases. Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features [27].

The advantage of using this method is, it is simple to understand and to interpret. Trees can be visualized. It requires little data preparation. Performs well even if its assumptions are somewhat violated by the true model from which the data were generated. Nonlinear relationships between parameters do not affect tree performance. At the same time we have to concentrate on the issues of Decision tree. Decision-tree learners can create over-complex trees that do not generalize the data well. This is called over fitting. Mechanisms such as pruning, setting the minimum number of samples required at a leaf node or setting the maximum depth of the tree are necessary to avoid this problem.

Decision tree construction with GA Based classification takes longer time then the traditional one although it eliminates many of the unnecessary features; this is one area where more research can be done, to improve the response time [28].

This paper aims to develop a risk reduction mechanism through the employment of the anomaly deduction mechanism through the application of dimension reduction by the PCA and the multi class classifier uses the GA. The pruning is done by the GA based K means clustering.

V. Proposed Approach

The algorithm of the proposed approach is described as follows.

Input: NSL KDD dataset

Step 1: preprocessing is done by the normalization using Z score

/* a set of n scores each denoted by and whose mean is equal to and whose standard deviation is equal to is transformed in Z -scores as

$$Z_{n=} \frac{x_{n-\bar{x}}}{s}$$

*/

Step 2: Application of PCA for the dimensionality reduction

/* PCA algorithm

a. Mean center the data
b. Compute the covariance matrix of the dimensions
c. Find eigenvectors of covariance matrix
d. Sort eigenvectors in decreasing order of Eigen values
e. Project onto eigenvectors in order (The eigenvector with the highest Eigen value is the Principle component of the data)
f. Keep only the terms corresponding to the principle component.

*/

Step 3: Multi class classifier guided by genetic algorithm

/* ID3 algorithm to build the decision tree guided by Genetic algorithm

Step 3: (a)Tree construction

a. choose one attribute as the root with highest information gain and put all its values as branches
b. Apply GA for the choosing recursively internal nodes (attributes) with their proper values as branches.
c. Stop when
   - all the samples (records) are of the same class, then the node becomes the leaf labeled with that class
   - or there is no more samples left
   - or there is no more new attributes to be put as the nodes. In this case we apply MAJORITY VOTING to classify the node.

Step 3: (b) Tree pruning

- Identify and remove branches that reflect noise or outliers using GA based K means Clustering

Output: Multi Class classified dataset

VI. Experiments and Results

The experiment is carried out with the NSL KDD data set. Initially the normalization is done through the Z –score. This method preserve range (maximum and minimum) and introduce the dispersion of the series ie, standard deviation. With elementary algebraic manipulations, it can be shown that a set of Z -score has a mean equal of zero and a standard deviation of one. Therefore, Z -scores constitute a unit free measure which can be used to compare observations measured with different units [29]. PCA is employed for the linear projection of high dimensional data into a lower dimensional subspace. Genetic algorithm based multi class classifier using the decision tree induction is deployed for the classification. The tree pruning is done by the GA based K means clustering.

The results obtained are evaluated based on the following performance metrics

- Entropy
- F measure
- Feature reduction
- Accuracy
- Error Rate
- Time taken for the classification

The experiment is carried out and the results are evaluated against the classifier proposed in [30]. The results are shown as the graphical representation as follows

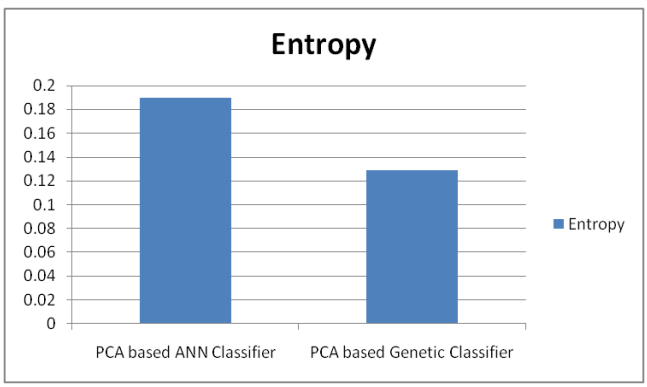


Figure 6.1 : Comparison based on entropy measure

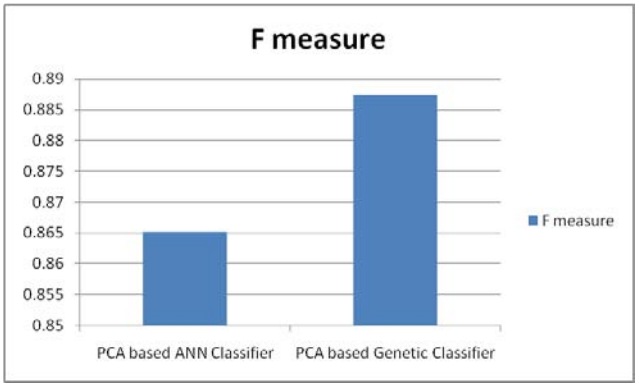


Figure 6.2 : Comparison based on F-measure

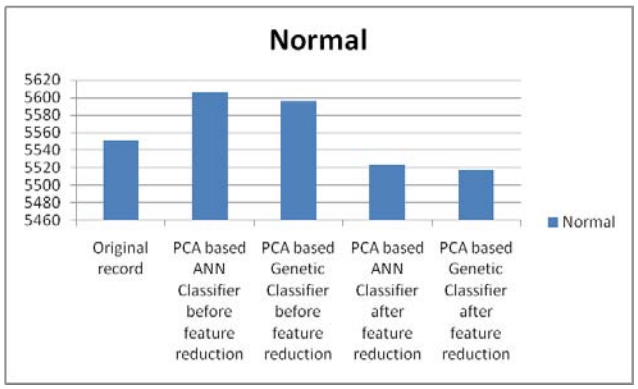


Figure 6.3 : Comparison based on feature reduction on Normal attack data

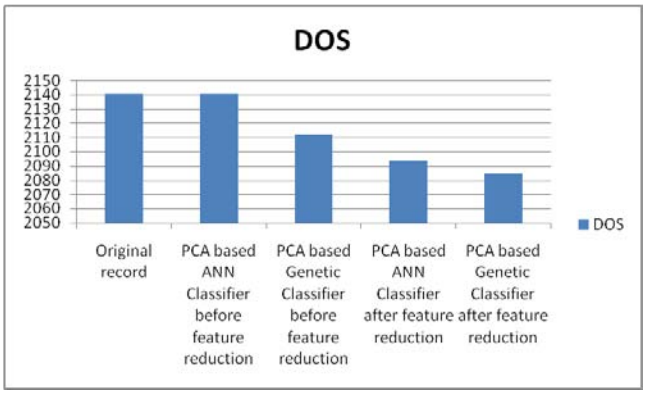


Figure 6.4 : Comparison based on feature reduction on DOS attack data

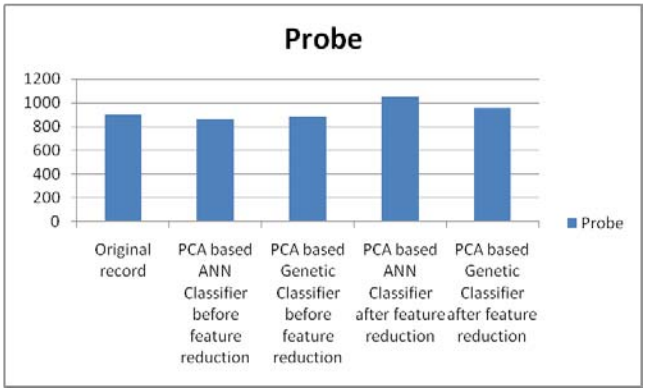


Figure 6.5 : Comparison based on feature reduction on Probe attack data

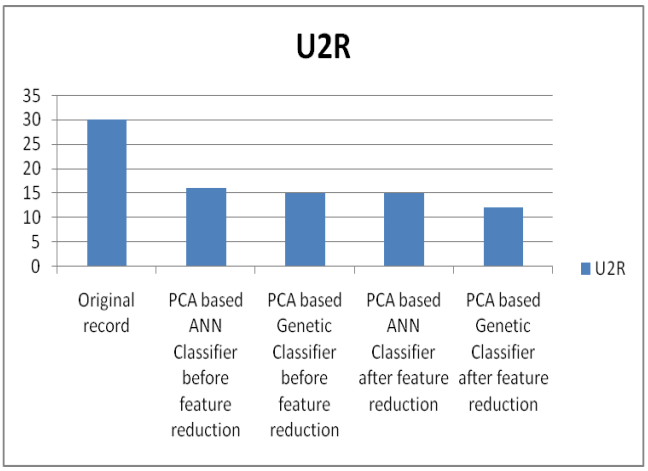


Figure 6.6 : Comparison based on feature reduction on U2R attack data

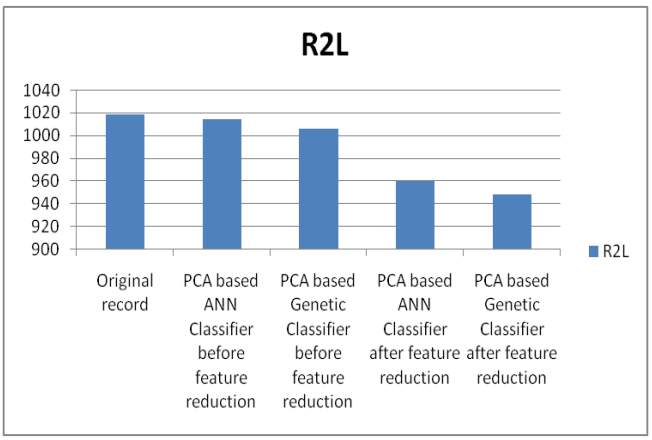


Figure 6.7 : Comparison based on feature reduction on R2L attack data

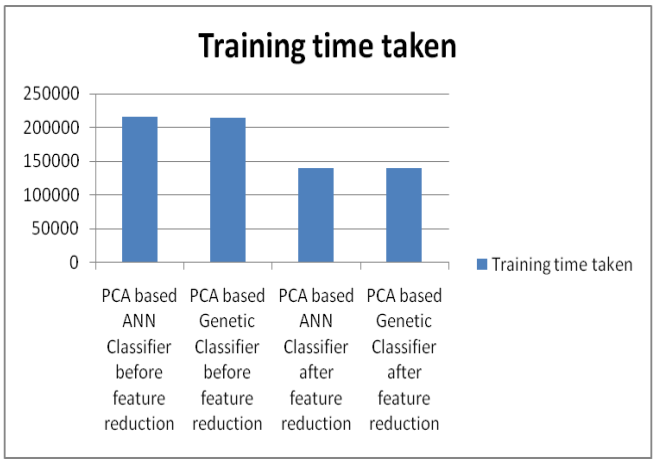


Figure 6.8 : Comparison based on Training time Taken

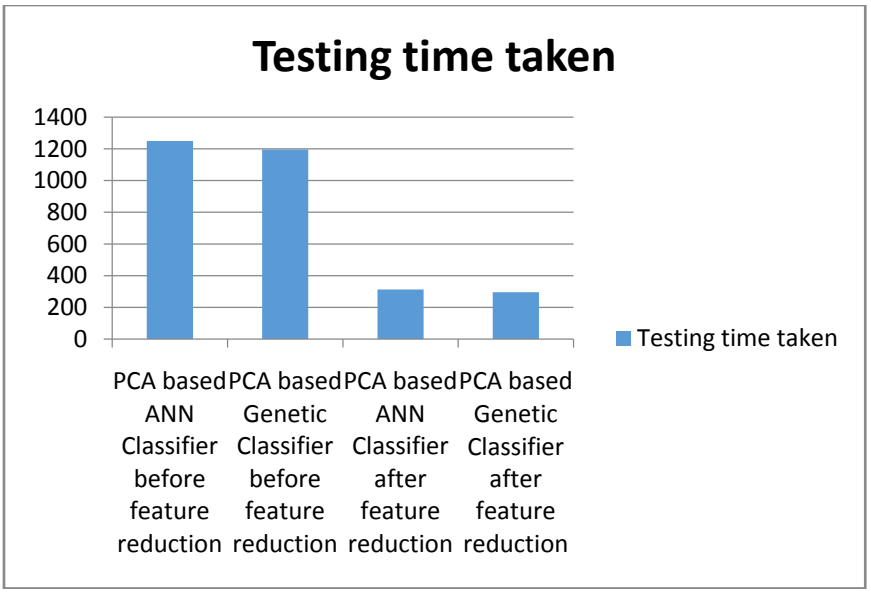


Figure 6.9 : Comparison based on testing time taken

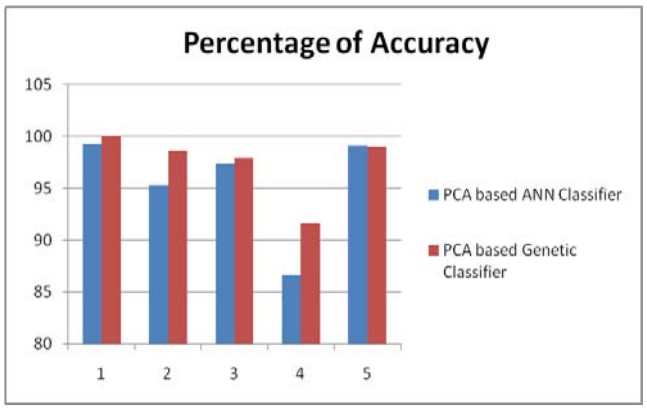


Figure 6.10 : Comparison based on prediction accuracy of the classifier

Figure 6.11 : Comparison based on error rate of the classifier

VII. Discussion

The proposed classifier based on the PCA for the dimensionality reduction and the genetic algorithm with K-means for the decision tree induction is outperforming the existing methodologies.

Table 7.1 : Percentage of improvement of the proposed approach based on Entropy

	PCA based ANN Classifier	PCA based Genetic Classifier	Percentage of Improvement
Entropy	0.1897	0.1293	31.83975

Table 7.2 : Percentage of improvement of the proposed approach based on F-Measure

	PCA based ANN Classifier	PCA based Genetic Classifier	Percentage of Improvement
F measure	0.8652	0.8874	2.50169

VIII. Conclusion

The risk reduction in the information security is being carried out by the anomaly detection. The classification task is the challenging work in this type of detection. In this paper an optimized approach for the classifier is proposed. PCA is employed for the dimensionality reduction. Genetic algorithm is employed for the construction of the tree nodes in the building process of the decision tree. The tree pruning is being employed by the clustering approach, which is being guided by the Genetic algorithm. The experimental

results are demonstrated and the proposed approach is proven to be the best suited from the classical ANN classifier.

References Références Referencias

1. Alireza Tamjidyamcholo ,Rawaa Dawoud Al-Dabbagh, Genetic Algorithm Approach for Risk Reduction of Information Security, International Journal of Cyber-Security and Digital Forensics (IJCSDF) 1(1): 59-66, 2012.
2. A. Chandrasekar, V. Vasudevan, P. Yogesh, Evolutionary Approach for Network Anomaly Detection Using Effective Classification, IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.1, January 2009.
3. Rodrigo C. Barros, M´arcio P. Basgalupp, Andr´e C. P. L. F. de Carvalho , Alex A. Freitas, A Survey of Evolutionary Algorithms for Decision Tree Induction, IEEE transactions on systems, man, and cybernetics - part c: applications & reviews, Volume:42 , Issue: 3, 2012.
4. http://www.opensourceforu.com/2011/01/importance-of-intrusion-prevention-systems
5. Asmaa Shaker Ashoor, Sharad Gore, Importance of Intrusion Detection System, International Journal of Scientific & Engineering Research, Volume 2, Issue 1, January-2011.
6. https://www.ipa.go.jp/security/fy11/report/contents/intrusion/ids-meeting/idsbg.pdf
7. Tamer F. Ghanema, Wail S. Elkilani, Hatem M. Abdul-kader, A hybrid approach for efficient anomaly detection using metaheuristic methods, Journal of Advanced Research, March 2014.
8. Salvatore J. Stolfo, Shlomo Hershkop, Linh H. Bui, Ryan Ferster, and Ke Wang, Anomaly Detection in Computer Security and an Application to File System Accesses, SMIS 2005, LNAI 3488, pp. 14–28, 2005.
9. Varun Chandola, Arindam Banerjee, and Vipin Kumar, "Anomaly Detection : A Survey", ACM Computing Surveys, Vol. 41(3), Article 15, July 2009.
10. http://www.geatbx.com/docu/algindex01.html#P153_5403
11. http://ewh.ieee.org/cmte/cis/mtsc/ieeecis/Xin_Yao.pdf
12. http://www.cs.vu.nl/~gusz/ecbook/Eiben-Smith Intro2EC-Ch2.pdf
13. http://courses.cs.washington.edu/courses/csep521/07wi/prj/leonardo_fabricio.pdf
14. http://www.tutorialspoint.com/data_mining/dm_classification_prediction.htm
15. http://mathworld.wolfram.com/GeneticAlgorithm.html
16. http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol1/hmw/article1.html#introduction
17. http://support.sas.com/publishing/pubcat/chaps/55129.pdf
18. Jonathon Shlens, A Tutorial on Principal Component Analysis, April 7, 2014; Version 3.02, http://arxiv.org/pdf/1404.1100v1.pdf
19. Nayot Poolsappasit, Rinku Dewri, and Indrajit Ray, Dynamic Security Risk Management Using Bayesian Attack Graphs, IEEE Transaction on Dependable and secure computing, Jan 2012.
20. Ming-Chang Lee, Information Security Risk Analysis Methods and Research Trends: AHP and Fuzzy Comprehensive Method, International Journal of Computer Science & Information Technology (IJCSIT) Vol 6, No1, February 2014.
21. Gao Y, Luo J. Z. (2009), Information security risk assessment based on grey relational decision making algorithm" , Journal of Southeast University , Vol. 39, No. 2, pp. 225-229.
22. Liu F, Dai K, Wang Z. Y. (2004), "Research on the technology of quantitative security evaluation based on fuzzy number arithmetic operation", Fuzzy Systems and Mathematics, Vol. 18, No. 4, pp. 51-54.
23. Chang, P. T., Hung K, C. (2005), "Applying the fuzzy weighted average approach to evaluation network security systems". Computers and Mathematic s with Application, Vol. 49, pp. 1797-1814.
24. Feng N, Li M. (2011), "An information systems security risk assessment model under uncertain environment". Applied Soft Computer, Vol. 11, No.7,pp. 4332-4340.
25. Syamsuddin, I. and Hwang, J., (2010), "The use AHP in security policy decision making: An open office calc application", Journal of Software, Vol. 5, No. 10, pp. 1162-1169.
26. Syamsuddin, I. (2012), "Evaluation of strategic information security with fuzzy AHP method".American Journal of Intelligence Systems, Vol. 2, No. 1, pp. 9-13.
27. http://scikit-learn.org/stable/modules/tree.html
28. Shanta Rangaswamy, Dr. Shobha G, , Sandeep R V, , Raj Kiran, Comparative Study of Decision Tree Classifier with and without GA based feature selection, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 1, January 2014.
29. Herve Abdi, Normalizing Data, In Neil Salkind (Ed.),Encyclopedia of Research Design.Thousand Oaks, CA: Sage. 2010.
30. Shilpa lakhina, , Sini Joseph, Bhupendra verma, Feature Reduction using Principal Component Analysis for Effective Anomaly–Based Intrusion Detection on NSL-KDD, International Journal of Engineering Science and Technology Vol. 2(6), 2010, 1790-1799.

Global Journals Inc. (US) Guidelines Handbook 2014

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 (Must be strictly followed)

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