http://www.icact.org/journal.asp

ICACT-TACT JOURNAL

Transactions on Advanced Communications Technology

Volume 7 Issue 3, May. 2018, ISSN: 2288-0003

Editor-in-Chief
Prof. Thomas Byeongnam YOON, PhD.

GIRI
Global IT Research Institute

Journal Editorial Board

■ Editor-in-Chief

Prof. Thomas Byeongnam YOON, PhD.
Founding Editor-in-Chief
ICACT Transactions on the Advanced Communications Technology (TACT)

■ Editors

Prof. Jun-Chul Chun, Kyonggi University, Korea
Dr. JongWon Kim, GIST (Gwangju Institute of Science & Technology), Korea
Dr. Xi Chen, State Grid Corparation of China, China
Prof. Arash Dana, Islamic Azad university , Central Tehran Branch, Iran
Dr. Pasquale Pace, University of Calabria - DEIS - Italy, Italy
Dr. Mitch Haspel, Stochastikos Solutions R&D, Israel
Prof. Shintaro Uno, Aichi University of Technology, Japan
Dr. Tony Tsang, Hong Kong Polytechnic UNiversity, Hong Kong
Prof. Kwang-Hoon Kim, Kyonggi University, Korea
Prof. Rosilah Hassan, Universiti Kebangsaan Malaysia(UKM), Malaysia
Dr. Sung Moon Shin, ETRI, Korea
Dr. Takahiro Matsumoto, Yamaguchi University, Japan
Dr. Christian Esteve Rothenberg, CPqD - R&D Center for. Telecommunications, Brazil
Prof. Lakshmi Prasad Saikia, Assam down town University, India
Prof. Moo Wan Kim, Tokyo University of Information Sciences, Japan
Prof. Yong-Hee Jeon, Catholic Univ. of Daegu, Korea
Dr. E.A.Mary Anita, Prathyusha Institute of Technology and Management, India
Dr. Chun-Hsin Wang, Chung Hua University, Taiwan
Prof. Wilaiporn Lee, King Mongkut's University of Technology North, Thailand
Dr. Zhi-Qiang Yao, XiangTan University, China
Prof. Bin Shen, Chongqing Univ. of Posts and Telecommunications (CQUPT), China
Prof. Vishal Bharti, Dronacharya College of Engineering, India
Dr. Marsono, Muhammad Nadzir , Universiti Teknologi Malaysia, Malaysia
Mr. Muhammad Yasir Malik, Samsung Electronics, Korea
Prof. Yeonseung Ryu, Myongji University, Korea
Dr. Kyuchang Kang, ETRI, Korea
Prof. Plamena Zlateva, BAS(Bulgarian Academy of Sciences), Bulgaria
Dr. Pasi Ojala, University of Oulu, Finland
Prof. CheonShik Kim, Sejong University, Korea
Dr. Anna bruno, University of Salento, Italy
Prof. Jesuk Ko, Gwangju University, Korea
Dr. Saba Mahmood, Air University Islamabad Pakistan, Pakistan
Prof. Zhiming Cai, Macao University of Science and Technology, Macau
Prof. Man Soo Han, Mokpo National Univ., Korea
Mr. Jose Gutierrez, Aalborg University, Denmark

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Dr. Youssef SAID, Tunisie Telecom, Tunisia
Dr. Noor Zaman, King Faisal University, Al Ahsa Hofuf, Saudi Arabia
Dr. Srinivas Mantha, SASTRA University, Thanjavur, India
Dr. Shahriar Mohammadi, KNTU University, Iran
Prof. Beonsku An, Hongik University, korea
Dr. Guanbo Zheng, University of Houston, USA
Prof. Sangho Choe, The Catholic University of Korea, korea
Dr. Gyanendra Prasad Joshi, Yeungnam University, korea
Dr. Tae-Gyu Lee, Korea Institue of Industrial Technology(KITECH), korea
Prof. Ilkyeun Ra, University of Colorado Denver, USA
Dr. Yong Sun, Beijing University of Posts and Telecommunications, China
Dr. Yulei Wu, Chinese Academy of Sciences, China
Mr. Anup Thapa, Chosun University, korea
Dr. Vo Nguyen Quoc Bao, Posts and Telecommunications Institute of Technology, Vietnam
Dr. Harish Kumar, Bhagwant institute of technology, India
Dr. Jin REN, North china university of technology, China
Dr. Joseph Kandath, Electronics & Commn Engg, India
Dr. Mohamed M. A. Moustafa, Arab Information Union (AIU), Egypt
Dr. Mostafa Zaman Chowdhury, Kookmin University, Korea
Prof. Francis C.M. Lau, Hong Kong Polytechnic University, Hong Kong
Prof. Ju Bin Song, Kyung Hee University, korea
Prof. KyungHi Chang, Inha University, Korea
Prof. Sherif Welsen Shaker, Kuang-Chi Institute of Advanced Technology, China
Prof. Seung-Hoon Hwang, Dongguk University, Korea
Prof. Dal-Hwan Yoon, Semyung University, korea
Prof. Chongyang ZHANG, Shanghai Jiao Tong University, China
Dr. H K Lau, The Open University of Hong Kong, Honh Kong
Prof. Ying-Ren Chien, Department of Electrical Engineering, National Ilan University, Taiwan
Prof. Mai Yi-Ting, Hsiuping University of Science and Technology, Taiwan
Dr. Sang-Hwan Ryu, Korea Railroad Research Institute, Korea
Dr. Yung-Chien Shih, MediaTek Inc., Taiwan
Dr. Kuan Hoong Poo, Multimedia University, Malaysia
Dr. Michael Leung, CEng MIET SMIEEE, Hong Kong
Dr. Abu sahman Bin mohd Supa'at, Universiti Teknologi Malaysia, Malaysia
Prof. Amit Kumar Garg, Deenbandhu Chhotu Ram University of Science & Technology, India
Dr. Jens Myrup Pedersen, Aalborg University, Denmark
Dr. Augustine Ikechi Ukaegbu, KAIST, Korea
Dr. Jamshid Sangirov, KAIST, Korea
Prof. Ahmed Dooguy KORA, Ecole Sup. Multinationale des Telecommunications, Senegal
Dr. Se-Jin Oh, Korea Astronomy & Space Science Institute, Korea
Dr. Rajendra Prasad Mahajan, RGPV Bhopal, India
Dr. Woo-Jin Byun, ETRI, Korea
Dr. Mohammed M. Kadhum, School of Computing, Goodwin Hall, Queen's University , Canada
Prof. Seong Gon Choi, Chungbuk National University, Korea
Prof. Yao-Chung Chang, National Taitung University, Taiwan
Dr. Abdallah Handoura, Engineering school of Gabes - Tunisia, Tunisia
Dr. Gopal Chandra Manna, BSNL, India

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Dr. Il Kwon Cho, National Information Society Agency, Korea
Prof. Jiann-Liang Chen, National Taiwan University of Science and Technology, Taiwan
Prof. Ruay-Shiung Chang, National Dong Hwa University, Taiwan
Dr. Vasaka Visoottiviseth, Mahidol University, Thailand
Prof. Dae-Ki Kang, Dongseo University, Korea
Dr. Yong-Sik Choi, Research Institute, IDLE co., ltd, Korea
Dr. Xuena Peng, Northeastern University, China
Dr. Ming-Shen Jian, National Formosa University, Taiwan
Dr. Soobin Lee, KAIST Institute for IT Convergence, Korea
Prof. Yongpan Liu, Tsinghua University, China
Prof. Chih-Lin HU, National Central University, Taiwan
Prof. Chen-Shie Ho, Oriental Institute of Technology, Taiwan
Dr. Hyoung-Jun Kim, ETRI, Korea
Prof. Bernard Cousin, IRISA/Universite de Rennes 1, France
Prof. Eun-young Lee, Dongduk Woman s University, Korea
Dr. Porkumaran K, NGP institute of technology India, India
Dr. Feng CHENG, Hasso Plattner Institute at University of Potsdam, Germany
Prof. El-Sayed M. El-Alfy, King Fahd University of Petroleum and Minerals, Saudi Arabia
Prof. Lin You, Hangzhou Dianzi Univ, China
Mr. Nicolai Kuntze, Fraunhofer Institute for Secure Information Technology, Germany
Dr. Min-Hong Yun, ETRI, Korea
Dr. Seong Joon Lee, Korea Electrotechnology Research Institute, korea
Dr. Kwihoon Kim, ETRI, Korea
Dr. Jin Woo HONG, Electronics and Telecommunications Research Inst., Korea
Dr. Heeseok Choi, KISTI(Korea Institute of Science and Technology Information), korea
Dr. Somkiat Kitjongthawonkul, Australian Catholic University, St Patrick's Campus, Australia
Dr. Dae Won Kim, ETRI, Korea
Dr. Ho-Jin CHOI, KAIST(Univ), Korea
Dr. Su-Cheng HAW, Multimedia University, Faculty of Information Technology, Malaysia
Dr. Myoung-Jin Kim, Soongsil University, Korea
Dr. Gyu Myoung Lee, Institut Mines-Telecom, Telecom SudParis, France
Dr. Dongkyun Kim, KISTI(Korea Institute of Science and Technology Information), Korea
Prof. Yoonhee Kim, Sookmyung Women s University, Korea
Prof. Li-Der Chou, National Central University, Taiwan
Prof. Young Woong Ko, Hallym University, Korea
Prof. Dimiter G. Velev, UNWE(University of National and World Economy), Bulgaria
Dr. Tadasuke Minagawa, Meiji University, Japan
Prof. Jun-Kyun Choi, KAIST (Univ.), Korea
Dr. Brownson ObaridoaObele, Hyundai Mobis Multimedia R&D Lab , Korea
Prof. Anisha Lal, VIT university, India
Dr. kyeong kang, University of technology sydney, faculty of engineering and IT , Australia
Prof. Chwen-Yea Lin, Tatung Institute of Commerce and Technology, Taiwan
Dr. Ting Peng, Chang'an University, China
Prof. ChaeSoo Kim, Donga University in Korea, Korea
Prof. kirankumar M. joshi, m.s.uni.of baroda, India
Dr. Chin-Feng Lin, National Taiwan Ocean University, Taiwan
Dr. Chang-shin Chung, TTA(Telecommunications Technology Association), Korea

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Dr. Che-Sheng Chiu, Chunghwa Telecom Laboratories, Taiwan
Dr. Chirawat Kotchasarn, RMUTT, Thailand
Dr. Fateme Khalili, K.N.Toosi. University of Technology, Iran
Dr. Izzeldin Ibrahim Mohamed Abdelaziz, Universiti Teknologi Malaysia , Malaysia
Dr. Kamrul Hasan Talukder, Khulna University, Bangladesh
Prof. HwaSung Kim, Kwangwoon University, Korea
Prof. Jongsub Moon, CIST, Korea University, Korea
Prof. Juinn-Horng Deng, Yuan Ze University, Taiwan
Dr. Yen-Wen Lin, National Taichung University, Taiwan
Prof. Junhui Zhao, Beijing Jiaotong University, China
Dr. JaeGwan Kim, SamsungThales co, Korea
Prof. Davar PISHVA, Ph.D., Asia Pacific University, Japan
Ms. Hela Mliki, National School of Engineers of Sfax, Tunisia
Prof. Amirmansour Nabavinejad, Ph.D., Sepahan Institute of Higher Education, Iran

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Editor Guide

■ Introduction for Editor or Reviewer

All the editor group members are to be assigned as a evaluator(editor or reviewer) to submitted journal papers at the discretion of the Editor-in-Chief. It will be informed by eMail with a Member Login ID and Password.

Once logined the Website via the Member Login menu in left as a evaluator, you can find out the paper assigned to you. You can evaluate it there. All the results of the evaluation are supposed to be shown in the Author Homepage in the real time manner. You can also enter the Author Homepage assigned to you by the Paper ID and the author's eMail address shown in your Evaluation Webpage. In the Author Homepage, you can communicate each other efficiently under the peer review policy. Please don't miss it!

All the editor group members are supposed to be candidates of a part of the editorial board, depending on their contribution which comes from history of ICACT TACT as an active evaluator. Because the main contribution comes from sincere paper reviewing role.

■ Role of the Editor

The editor's primary responsibilities are to conduct the peer review process, and check the final camera-ready manuscripts for any technical, grammatical or typographical errors.

As a member of the editorial board of the publication, the editor is responsible for ensuring that the publication maintains the highest quality while adhering to the publication policies and procedures of the ICACT TACT(Transactions on the Advanced Communications Technology).

For each paper that the editor-in-chief gets assigned, the Secretariat of ICACT Journal will send the editor an eMail requesting the review process of the paper.

The editor is responsible to make a decision on an "accept", "reject", or "revision" to the Editor-in-Chief via the Evaluation Webpage that can be shown in the Author Homepage also.

■ Deadlines for Regular Review

Editor-in-Chief will assign a evalaution group(a Editor and 2 reviewers) in a week upon receiving a completed Journal paper submission. Evaluators are given 2 weeks to review the paper. Editors are given a week to submit a recommendation to the Editor-in-Chief via the evaluation Webpage, once all or enough of the reviews have come in. In revision case, authors have a maximum of a month to submit their revised manuscripts. The deadlines for the regular review process are as follows:

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Evalution Procedure	Deadline
Selection of Evaluation Group	1 week
Review processing	2 weeks
Editor's recommendation	1 week
Final Decision Noticing	1 week

■ Making Decisions on Manuscript

Editor will make a decision on the disposition of the manuscript, based on remarks of the reviewers. The editor's recommendation must be well justified and explained in detail. In cases where the revision is requested, these should be clearly indicated and explained. The editor must then promptly convey this decision to the author. The author may contact the editor if instructions regarding amendments to the manuscript are unclear. All these actions could be done via the evaluation system in this Website. The guidelines of decisions for publication are as follows:

Decision	Description
Accept	An accept decision means that an editor is accepting the paper with no further modifications. The paper will not be seen again by the editor or by the reviewers.
Reject	The manuscript is not suitable for the ICACT TACT publication.
Revision	The paper is conditionally accepted with some requirements. A revision means that the paper should go back to the original reviewers for a second round of reviews. We strongly discourage editors from making a decision based on their own review of the manuscript if a revision had been previously required.

■ Role of the Reviewer

Reviewer Webpage:

Once logined the Member Login menu in left, you can find out papers assigned to you. You can also login the Author Homepage assigned to you with the paper ID and author's eMail address. In there you can communicate each other via a Communication Channel Box.

Quick Review Required:

You are given 2 weeks for the first round of review and 1 week for the second round of review. You must agree that time is so important for the rapidly changing IT technologies and applications trend. Please respect the deadline. Authors undoubtedly appreciate your quick review.

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Anonymity:

Do not identify yourself or your organization within the review text.

Review:

Reviewer will perform the paper review based on the main criteria provided below. Please provide detailed public comments for each criterion, also available to the author.

- How this manuscript advances this field of research and/or contributes something new to the literature?
- Relevance of this manuscript to the readers of TACT?
- Is the manuscript technically sound?
- Is the paper clearly written and well organized?
- Are all figures and tables appropriately provided and are their resolution good quality?
- Does the introduction state the objectives of the manuscript encouraging the reader to read on?
- Are the references relevant and complete?

Supply missing references:

Please supply any information that you think will be useful to the author in revision for enhancing quality of the paperor for convincing him/her of the mistakes.

Review Comments:

If you find any already known results related to the manuscript, please give references to earlier papers which contain these or similar results. If the reasoning is incorrect or ambiguous, please indicate specifically where and why. If you would like to suggest that the paper be rewritten, give specific suggestions regarding which parts of the paper should be deleted, added or modified, and please indicate how.

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Journal Procedure

Dear Author,

- **You can see all your paper information & progress.**

- **Step 1. Journal Full Paper Submission**

Using the Submit button, submit your journal paper through ICACT Website, then you will get new paper ID of your journal, and send your journal Paper ID to the Secretariat@icact.org for the review and editorial processing. Once you got your Journal paper ID, never submit again! Journal Paper/CRF Template

- **Step 2. Full Paper Review**

Using the evaluation system in the ICACT Website, the editor, reviewer and author can communicate each other for the good quality publication. It may take about 1 month.

- **Step 3. Acceptance Notification**

It officially informs acceptance, revision, or reject of submitted full paper after the full paper review process.

Status	Action
Acceptance	Go to next Step.
Revision	Re-submit Full Paper within 1 month after Revision Notification.
Reject	Drop everything.

- **Step 4. Payment Registration**

So far it's free of charge in case of the journal promotion paper from the registered ICACT conference paper! But you have to regist it, because you need your Journal Paper Registration ID for submission of the final CRF manuscripts in the next step's process. Once you get your Registration ID, send it to Secretariat@icact.org for further process.

- **Step 5. Camera Ready Form (CRF) Manuscripts Submission**

After you have received the confirmation notice from secretariat of ICACT, and then you are allowed to submit the final CRF manuscripts in PDF file form, the full paper and the Copyright Transfer Agreement. Journal Paper Template, Copyright Form Template, BioAbstract Template,

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Journal Submission Guide

All the Out-Standing ICACT conference papers have been invited to this "ICACT Transactions on the Advanced Communications Technology" Journal, and also welcome all the authors whose conference paper has been accepted by the ICACT Technical Program Committee, if you could extend new contents at least 30% more than pure content of your conference paper. Journal paper must be followed to ensure full compliance with the IEEE Journal Template Form attached on this page.

➢ How to submit your Journal paper and check the progress?

Step 1. Submit	Using the Submit button, submit your journal paper through ICACT Website, then you will get new paper ID of your journal, and send your journal Paper ID to the Secretariat@icact.org for the review and editorial processing. Once you got your Journal paper ID, never submit again! Using the Update button, you can change any information of journal paper related or upload new full journal paper.
Step 2. Confirm	Secretariat is supposed to confirm all the necessary conditions of your journal paper to make it ready to review. In case of promotion from the conference paper to Journal paper, send us all the .DOC(or Latex) files of your ICACT conference paper and journal paper to evaluate the difference of the pure contents in between at least 30% more to avoid the self replication violation under scrutiny. The pure content does not include any reference list, acknowledgement, Appendix and author biography information.
Step 3. Review	Upon completing the confirmation, it gets started the review process thru the Editor & Reviewer Guideline. Whenever you visit the Author Homepage, you can check the progress status of your paper there from start to end like this, " Confirm OK! -> Gets started the review process -> ...", in the Review Status column. Please don't miss it!

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Volume. 7 Issue. 3

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

A Supervised Machine Learning Approach using Different Feature Selection Techniques on Voice Datasets for Prediction of Parkinson's Disease

Satyabrata Aich[*], Hee-Cheol Kim[*], Kim younga[**], Kueh Lee Hui[***], Ahmed Abdulhakim Al-Absi[****] and Mangal Sain[*****]

[*]Department of Computer Engineering, Inje University, South Korea

[*]Department of Rehabilitation Science, Inje University, South Korea

[***]Department of Electrical Engineering, Dong-A University, South Korea

[****]Department of Computer Engineering, Kyungdong University- Global Campus, Gangwondo, South Korea

[*****]Department of Computer Engineering, Dongseo University, South Korea

satyabrataaich@gmail.com, heeki@inje.ac.kr, kya2664@hanmail.net, leehkueh@dau.ac.kr, absiahmed@kduniv.ac.kr, mangalsain1@gmail.com

Abstract—**Among the neurological diseases, parkinson's disease is the second most common disease, which affect the old age people over the age of 65 year. It is also mentioned that the number of people affected with Parkinson's disease will increase at a higher rate until 2050, and it will be a rising concern to many developed countries because the cost due to the healthcare service of these disease is really high. Parkinson's disease (PD) belongs to the group of neurological disorder, which directly affect the brain cells and the effect is shown in terms of movement, voice and other cognitive disabilities. Past few years researchers are working for detection and monitoring of the Parkinson's disease by using the speech analysis as well as the gait analysis data. Machine learning and artificial intelligence techniques are gaining popularity because these techniques are able to automate the pattern recognition process with high accuracy.**

However so far, no body has compared the performance metrics using different feature sets by applying nonlinear and linear classification approaches based on the voice data. So, in this paper we have proposed a new approach by comparing the performance metrics with different feature sets such as genetic algorithm-based feature sets as well as Principal Component Analysis based feature reduction technique for selecting the feature sets. We have used different classification approaches to compare the performance metrics. We have found an accuracy of 97.57% using SVM with RBF by using genetic algorithm-based feature sets. This analysis will help the clinicians to differentiate the PD group from healthy group based on the voice data.

Keywords— **Parkinson's disease, machine learning, feature selection, voice data, genetic algorithm**

I. INTRODUCTION

Past few years a lot of research has been going on the Parkinson's disease because the healthcare related cost due to this disease is keeping on increasing as the longevity of the population is increasing in the developed countries. Since this disease affect most of the old people, it is become necessary for the developed counties to detect the disease at the early stage. The early detection will help the developed country in economic perspective as well as social perspective because it can be assessed well. Parkinson's disease belongs to one of the category of neurodegenerative disease which directly as well as indirectly affects the brain cells that will affect the movement, speech and other cognitive parts [1, 2, and 3]. The Parkinson's disease is progressive in nature. As the disease progresses more than 90% of the patients has the speech disorder [4]. The symptom related to the vocal impairment of Parkinson's disease patients is called dysphonia**.** The clinicians measured some indicators related to dysphonia to assess the PD patients. The measures related to dysphonia

Manuscript received on January 2, 2018. This work was supported by Kyungdong University grant, 2018 and a follow-up the invited journal to the outstanding paper of the 20th International Conference on Advanced Communication Technology (ICACT 2018).

Satyabrata Aich is with the Department of Computer Engineering, Inje University, South Korea (e-mail: satyabrataaich@gmail.com)

Hee-Cheol Kim is with the Department of Computer Engineering, Inje University, South Korea (e-mail: heeki@inje.ac.kr)

Kim younga is with the Department of Rehabilitation Science, Inje University, South Korea (e-mail: kya2664@hanmail.net)

Kueh Lee Hui is with the Department of Electrical Engineering, Dong-A University, South Korea (e-mail: leehkueh@dau.ac.kr)

Ahmed Abdulhakim Al-Absi is with the Department of Computer Engineering, Kyungdong University- Global Campus, Gangwondo, South Korea (e-mail: absiahmed@kduniv.ac.kr)

Mangal Sain is with the Department of Computer Engineering, Dongseo University, South Korea. He is the corresponding author of this paper. (Corresponding author phone: +8251-320-2009; e-mail: mangalsain1@gmail.com).

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

could be treated as an important and most reliable tool to assess the voice related problem and monitor it at different stage [5, 6]. Usually the measures have lot of features which does not helpful for machine learning approaches, so feature selection method has been used for proper assessment. The feature selection method will help to evaluate the important contribution of the features in the assessment of the disease at different stage and also it helps to achieve good accuracy [7, 8].

The traditional diagnosis needs lot of observations related to the daily living activities, motor skills and other neurological parameters to assess the progression of PD, but this process is not suitable for the early detection of the PD. With respect to the past research it is found that artificial intelligence and machine learning techniques have good potential for the classification and it also found that the classification system helps to improve the accuracy and the reliability of the diagnosis and also minimize the errors as well as make the system more efficient [9]. improvement on the prediction of accuracy on the progression of PD is getting lot of attention these days [10, 11].

So in this paper an attempt has been made to check the improvement in the accuracy while classifying the PD group from the healthy control group by using different machine learning algorithm with different feature sets such as the genetic algorithm (GA) based feature set as well as the PCA based feature sets. Finally, a comparison has made in terms of performance metrics using different feature sets. The structure of the paper is organized as follows: Section 2 presents the past work related to classification model used for voice datasets. Section 3 describes about the methodologies used for this research work. Section 4 describes about the result of feature selection as well as the result of classification. Section 5 describes about the conclusion and future work.

II. RELATED WORK

Some of the Past research works are mentioned below to give overall ideas about the amount of work has been done in this field. Shahbaba and Neal used nonlinear based approach for classification of PD. They have used Dirichlet process mixtures and compared the results with other classification model such as decision trees, support vector machine and multinomial logit model and they found Dirichlet process-based method provides best classification approach of 87.7% compared to the other model [12]. Sakar and Kursun used feature selection method as well as machine learning based method for diagnosis of PD. They have used mutual information-based feature selection and support vector machine as the classification approach and they found their approach gives an accuracy of 92.75% [13].

Li et al used fuzzy based method to extend the classification related information and then they have used principal component analysis-based method for feature selections and the optimal features has been integrated with SVM based method provides a good accuracy of 93.47% [14]. Spadoto et al used evolutionary base techniques for feature selection and they have used Optimum-path Forest Classifier to detect the Parkinson's disease and they found this approach provides a best accuracy of 84.01% while detecting the PD [15].

Luukka have proposed a feature selection method based on the fuzzy entropy measure and used similarity classifiers to classify PD. The best classification accuracy obtained by that method was 85.03% [16]. AStröm and Koker proposed a method that is used parallel feed-forward neural network-based approach to predict the PD. They have found the model is robust and the best classification accuracy obtained from that approach is 91.20% [17]. Nilashi et al proposed a method for the prediction of PD progression using clustering and prediction methods. They have applied Adaptive Neuro-Fuzzy Inference system (ANFIS) and Support Vector Regression for prediction of PD progression. They found this proposed method helps to improve the accuracy of the progression of PD [18].

Abdulhay et al proposed a method to investigate gait and tremor by using machine learning techniques based on the gait data. They have extracted various gait features using the peak detection and pulse duration and they found accuracy of 92.7% for diagnosis of parkinson's disease [19]. Er et al proposed a method to distinguish dementia patients (AD) from the age-related cognitive decline (ARCD) by using machine learning techniques. They found that these techniques are able to distinguish ARCD and AD at a success rate of 100% based on neurocognitive tests [20]. Zeng et al proposed a method to classify the patients with PD. They have used deterministic learning method to distinguish PD from healthy control group. They found accuracy, sensitivity and specificity of 96.39%, 96.77% and 95.89% respectively [21]. Armañanzas et al proposed a method to find the non-motor related PD severity based on the machine learning approach. They have used Hoehn & Yahr index and clinical impression of severity index as the measure of the severity for the PD patients. They have used classification with feature selection using evolutionary algorithm and found accuracy ranging from 72-92% [22]. Kubota et al. discussed about the benefits of using wearable sensors for measuring the gait parameters and also mentioned about signal processing and machine learning approaches used for extracting meaningful information from data. This kind of information helps to understand the potentials of the technology on PD research and practices [23].

Abós et al proposed a method distinguishing the PD patients based on cognitive status using machine learning technique. They have used randomized logistic regression for feature selection and SVM for machine learning and found accuracy of 82.6%. They found connection-wise patterns of functional connectivity may be helpful for distinguishing the PD patients. [24]. Singh and Samavedham proposed a methodology using machine learning tools as statistical measure for determining most important features to distinguish different neurological disease(ND). They have found accuracy of 99.93% while distinguishing various ND such as HC, PD and SWEDD patients.

The above past works motivated us to try a different approach. In this paper we have tried different feature

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

selection approach such as principal component analysis (PCA) approach and genetic algorithm-based approach and then the performance measures are compared with different machine learning classifiers.

III. PROPOSED TECHNIQUE

The flow chart of the proposed methodologies is shown in the Fig 1. In this paper we have used the dataset created by Max little University Oxford, in collaboration with the National Centre for Voice and Speech, Denver, Colorado, who recorded the speech signals [26]. The original data collected from the dataset composed of voice measurements from 31 people out of which 23 were diagnosed with PD. We have used Principal Component Analysis (PCA) algorithm on the original feature sets.

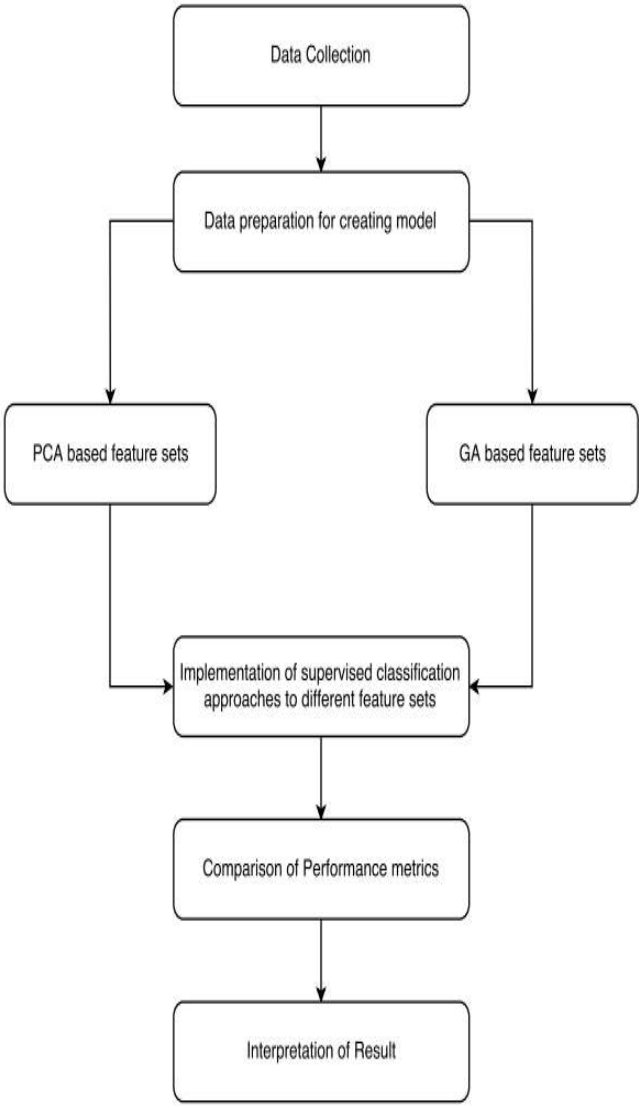

Fig. 1. Flowchart of the proposed method

The most widely applied technique for data reduction and feature selection is done using PCA. In PCA, the principal components or the latent variables are obtained from the variance of the data by maximizing it. The number of principal components is lesser than the regular variables. PCA reduces the dimensionality of the space so that the data can be visualized in the low dimensional space. The feature selection process is done by removing the redundant variables. [27] We have found 11 features after implementing the algorithm to the original feature sets. We have also used genetic algorithm (GA)-based feature sets for feature selection. Pledsoe first presented an adaptive optimization search methodology is called genetic algorithm and Holland mathematically presented the genetic algorithm-based approach by getting inspiration from Darwin's theory of evolution. A variable is mentioned as a gene. A chromosome is nothing but a sequence of gene. An initialization is done randomly by using population of chromosome. The quality of the chromosomes is evaluated according to a predefined fitness function. High performance chromosomes are used to produce the offspring. The genetic operators such as mutation and crossover are used to form the offspring. In this process the chromosomes are competing with each other and the fittest one survives at the end. The optimal solution comes after a series of iterative computations. [28, 29].

We have found 10 features using GA based feature sets. We have used different classification approach such as RPART, C4.5, PART, Bagging classification and Regression tree (Bagging CART), Random Forest, Boosted C5.0 and SVM.

Except SVM all other classifiers are belongs to nonlinear decision tree-based classifier. Decision tree-based classifiers are famous for giving good accuracy. SVM is most popular linear classification method. It uses different kernel function while doing classification. In this paper we have used radial basis function kernel function because it is recommended by many well-known data scientists. After implementing different classification approach, we have compared the performance measures such as accuracy, sensitivity, specificity, PPV and NPV.

Performance Metrics

The parameters used to compare the performance and validations of classifier are as follows: accuracy, sensitivity, specificity, positive predictive value (ppv), negative predictive value (npv). The sensitivity is defined as the ratio of true positives to the sum of true positives and false negatives. The specificity is defined as the ratio of true negatives to the sum of false positives and true negatives. In our research we have used the Positive predictive value and negative predictive value to check the present and absent of disease. So, the ppv is the probability that the disease is present given a positive test result and npv is the probability that the disease is absent given a negative test result [30]. Accuracy is defined as the ratio of number of correct predictions made to the total prediction made and the ratio is multiplied by 100 to make it in terms of percentage.

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

IV. RESULT AND DISCUSSIONS

We have used R programming language to write the code. We have reduced the original feature sets by feature selection techniques. We have chosen the PCA and GA as feature selection technique because these techniques are more widely used for feature selection without affecting the performance. As we have already discussed about the PCA method, after implementing it removes some redundant features and the reduced features are able to provide 99% variance, without affecting the performance. The GA method is different, which needs lot of iteration to reach the optimal result. With lot of iteration we have found optimized result with fewer features compared to the original features. We have trained each classifier based on the trained data and predict the power of classifier on the test data. The ratio of train data to test data is 70:30. In case of SVM we have used different Kernel function such as linear, polynomial, and radial basis function. So, each classifier able to show all the performance metrics based on the test data. We have plot the graphs shown below.

A. *Comparison of Accuracy*

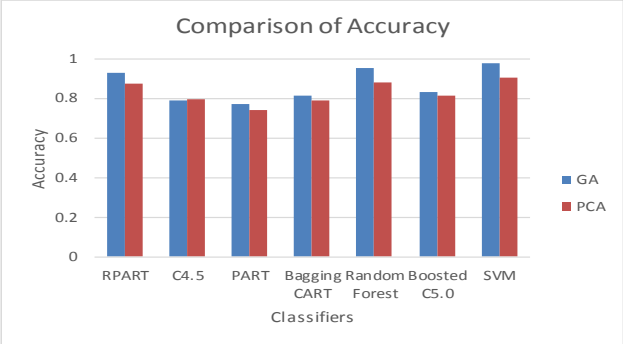

Fig. 2. Accuracy of different classifiers

Fig. 2 shows that SVM with RBF has highest accuracy of 97.57% with GA based feature sets followed by random forest and RPART classifiers with the same feature sets. Even though the other classifiers have less accuracy than SVM, but the difference in accuracy is not much.

B. *Comparison of Sensitivity*

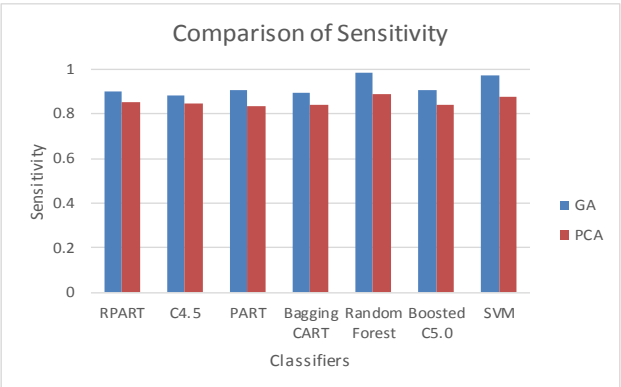

Fig. 3. Sensitivity of different classifiers

Fig .3 shows that random forest has highest sensitivity of 0.9985 with GA based feature sets followed by SVM with RBF classifiers with the same feature sets. The SVM classifier has a sensitivity of 0.9756.

C. *Comparison of Specificity*

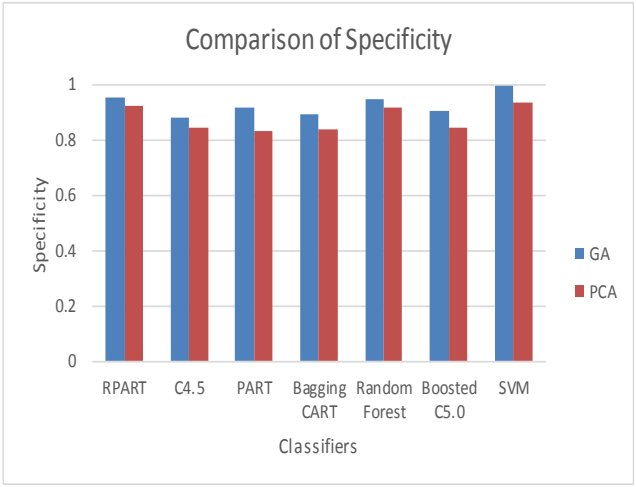

Fig. 4. Specificity of different classifiers

Fig. 4 shows that SVM with GA based feature set has highest specificity of 0.9987. RPART and random forest also follows SVM with good specificity with the same feature sets.

D. *Comparison of PPV*

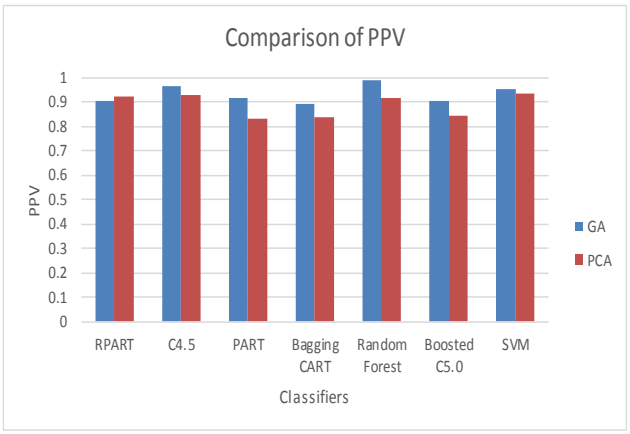

Fig. 5. PPV of different classifiers

Fig. 5 shows that the GA based feature set perform better compared to the PCA based feature set with random forest classifier. The maximum specificity achieved with this combination is 0.9934.

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

E. Comparison of NPV

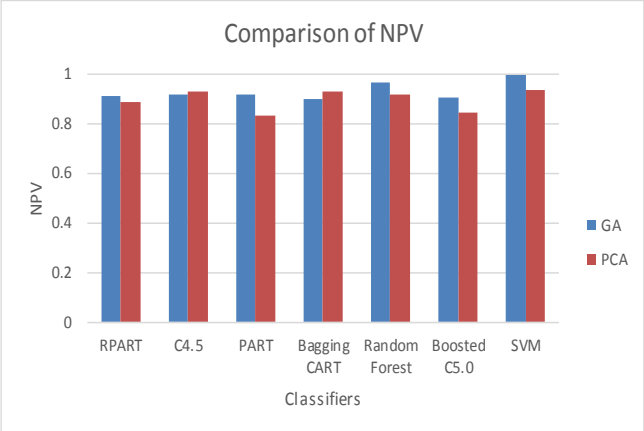

Fig. 6. NPV of different classifiers

Fig. 6 shows that SVM with GA based feature sets performs better compared to the PCA based feature sets. GA based feature sets shows maximum npv of 0.9995.

From the above plots we have seen that most of the classifiers are performing better while distinguishing between the Parkinson's patients to the control group. Basically, the feature selection processes help to save the time and space by removing the redundant sets, which has almost no impact on the performance. In our case we have seen the SVM and random forest are doing well in terms of the performance metric comparison with the GA based feature sets. We have seen that the GA based feature sets are performing well compared to PCA based feature sets. In theoretical case we may give a suggestion based on the accuracy while choosing the classifier, however in practical case lot of other performance measures is also come into picture. In our case we can suggest SVM is the best classifier and GA is best for feature selection, however it may provide different result with bigger dataset. This result actually gives an idea about the performance comparison and also gives an impression to analyze more deeply for implementing in practical life.

V. CONCLUSION AND FUTURE WORK

This paper outlined some of the new feature selection technique as well as some of the supervised machine learning approach for distinguishing Parkinson's patient from the control group. We have seen the capability of the PCA based feature selection technique and GA based feature selection technique. We have also seen the performance of different classification approach when combined with different feature sets. We have found that GA based feature selection technique performs better when combined with SVM classifier in terms of all performance measure compared to the PCA based feature set. Mostly while looking for performance measure we usually concentrate on accuracy, in that respect our SVM classifiers provides highest accuracy of 97.57% with GA based feature sets. This type of analysis will save the time and efficiency while doing pattern classification comprising two groups such as PD and control groups. This will help the clinician to distinguish PD and control groups. This research can be further improvised by adding more features and trying other classification techniques.

REFERENCES

[1] S.Przedborski, M.Vila, AND V.Jackson-Lewis, "Series Introduction: Neurodegeneration: What is it and where are we?", *Journal of Clinical Investigation*, 111(1), pp. 3-10,2003.

[2] Y.Xu, X.Wei, X.Liu, J.Liao, J.Lin, C.Zhu ...andM.Cheng, "Low cerebral glucose metabolism: a potential predictor for the severity of vascular Parkinsonism and Parkinson's disease", *Aging and disease*, 6(6), pp. 426-436, 2015.

[3] K.Tjaden, "Speech and swallowing in Parkinson's disease.", *Topics in geriatric rehabilitation*, 24(2), pp. 115-126, 2008.

[4] A. K. Ho, R. Iansek, C. Marigliani, J. L. Bradshaw, and S. Gates, "Speech impairment in a large sample of patients with Parkinson's disease", *Behavioural Neurology*, 11(3), pp. 131–137,1998.

[5] Little, M. A., McSharry, P. E., Hunter, E. J., Spielman, J., & Ramig, L. O. (2009). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4), 1015–1022.

[6] Rahn, D. A., Chou, M., Jiang, J. J., & Zhang, Y. (2007). Phonatory impairment in Parkinson's disease: evidence from nonlinear dynamic analysis and perturbation analysis. *Journal of Voice*, 21, 64–71.

[7] T. Hastie, R. Tibshirani, and J. H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction: *With 200 Full-Color Illustrations. New York: Springer-Verlag*, 2001.

[8] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Mach. Learn. Res.*, vol. 3, pp. 1157–1182, 2003.

[9] D.Gil, and D.J.Manuel, "Diagnosing Parkinson by using artificial neural networks and support vector machines", *Global Journal of Computer Science and Technology*, 9(4), pp.63-71, 2009.

[10] W.Froelich, K.Wrobel, and P. Porwik, "Diagnosis of Parkinson's disease using speech samples and threshold-based classification", *Journal of Medical Imaging and Health Informatics*, 5(6), pp.1358-1363, 2015.

[11] M.Hariharan, K.Polat, and R.Sindhu, A new hybrid intelligent system for accurate detection of Parkinson's disease. *Computer methods and programs in biomedicine*, 113(3), pp.904-913, 2014.

[12] Shahbaba, B., & Neal, R. (2009). Nonlinear models using Dirichlet process mixtures. *The Journal of Machine Learning Research*, 10, 1829–1850..

[13] Sakar, C. O., & Kursun, O. (2010). Telediagnosis of Parkinson's disease using measurements of dysphonia. *Journal of Medical Systems*, 34, 1–9

[14] Li, D. C., Liu, C. W., & Hu, S. C. (2011). A fuzzy-based data transformation for feature extraction to increase classification performance with small medical data sets. *Artificial Intelligence in Medicine*, 52, 45–52

[15] Spadoto, A. A., Guido, R. C., Carnevali, F. L., Pagnin, A. F., Falcao, A. X., & Papa, J. P. (2011). Improving Parkinson's disease identification through evolutionarybased feature selection. *In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the* IEEE (pp. 7857–7860).

[16] Luukka, P. (2011). Feature selection using fuzzy entropy measures with similarity classifier. *Expert Systems with Applications*, 38, 4600–4607

[17] AStröm, F., & Koker, R. (2011). A parallel neural network approach to prediction of Parkinson's Disease. *Expert Systems with Applications*, 38, 12470–12474.

[18] M. Nilashi, O. Ibrahim, A. Ahani, "Accuracy Improvement for Predicting Parkinson's Disease Progression," *Scientific Reports*,vol.6,34181,2016

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

[19] Abdulhay, E., Arunkumar, N., Narasimhan, K., Vellaiappan, E., & Venkatraman, V. (2018). Gait and tremor investigation using machine learning techniques for the diagnosis of Parkinson disease. *Future Generation Computer Systems.*

[20] Er, F., Iscen, P., Sahin, S., Çinar, N., Karsidag, S., & Goularas, D. (2017). Distinguishing age-related cognitive decline from dementias: A study based on machine learning algorithms. *Journal of Clinical Neuroscience*, 42, 186-192.

[21] Zeng, W., Liu, F., Wang, Q., Wang, Y., Ma, L., & Zhang, Y. (2016). Parkinson's disease classification using gait analysis via deterministic learning. *Neuroscience letters*, 633, 268-278.

[22] Armañanzas, R., Bielza, C., Chaudhuri, K. R., Martinez-Martin, P., & Larrañaga, P. (2013). Unveiling relevant non-motor Parkinson's disease severity symptoms using a machine learning approach. *Artificial intelligence in medicine*, 58(3), 195-202.

[23] Kubota, K. J., Chen, J. A., & Little, M. A. (2016). Machine learning for large scale wearable sensor data in Parkinson's disease: Concepts, promises, pitfalls, and futures. *Movement disorders*, 31(9), 1314-1326.

[24] Abós, Alexandra, Hugo C. Baggio, Bàrbara Segura, Anna I. García-Díaz, Yaroslau Compta, Maria José Martí, Francesc Valldeoriola, and Carme Junqué. "Discriminating cognitive status in Parkinson's disease through functional connectomics and machine learning." *Scientific reports 7* (2017): 45347.

[25] Singh, G., & Samavedham, L. (2015). Unsupervised learning-based feature extraction for differential diagnosis of neurodegenerative diseases: a case study on early-stage diagnosis of Parkinson disease. *Journal of neuroscience methods*, 256, 30-40.

[26] M. A.Little, P.E. McSharry, E. J.Hunter, J.Spielman, and L. O.Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease", *IEEE transactions on biomedical engineering*, 2009, 56(4), pp.1015-1022.

[27] Guo, Q., Wu, W., Massart, D. L., Boucon, C., & De Jong, S. (2002). Feature selection in principal component analysis of analytical data. *Chemometrics and Intelligent Laboratory Systems*, 61(1-2), 123-132.

[28] Bledsoe, W. W. (1961). The use of biological concepts in the analytical study of systems. *In the ORSA-TIMS National Meeting*.

[29] Holland, J. H. (1992). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. *MIT press.*

[30] H.B.Wong, G.H.Lim, Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV, *Proceedings of Singapore healthcare*, vol.20, no.4, pp.316-318, 2011.

Satyabrata Aich is working as a researcher in the field of computer engineering He has over four years of teaching, research and industry experience in India and abroad. He has published many research papers in journals and conferences in the realms of Supply Chain Management and data analytics. His research interests are natural language processing, Machine learning, supply chain management, data mining.

Hee-Cheol Kim received his BSc at Department of Mathematics, MSc at Department of Computer Science in SoGang University in Korea, and PhD at Numerical Analysis and Computing Science, Stockholm University in Sweden in 2001. His primary concerns are u-Healthcare, smart home technology, e-learning and Human Computer Interaction (HCI). He is associate professor at Department of Computer Engineering, Inje University in Korea. He has published many research papers in journals and conferences in the realms of HCI and CSCW.

Young a Kim is a licensed Physical therapist and has received her Masters degree in Counseling and Psychology in 2008. She is working as a lecturer for Clinical Kinesiology and Functional Anatomy at Inje University. She is also a PhD candidate in the Department of Rehabilitation Science at the same University. Her research focuses on assessment tools and rehabilitation for the elderly, especially for the detection and treatment of age related conditions. She is a co-author of the book 'Community Occupational Therapy' and a frequent contributor to the journal of 'Physical Therapy Science' and a member in the Korean Ageing Friendly Industry Association as well as the Society of Occupational Therapy for the Aged and Dementia.

Kueh Lee Hui is working as an assistant professor at the department of Electrical Engineering, Dong-A University since 2012. She completed her PhD Degrees from Department of Electrical Engineering, Dong-A University, Korea. In 2009 she completed her BS degree in Electronic and Communication, Department of Electronic Engineering, University Malaysia of Sarawak, Malaysia. She also done MS in 2007 from Malaysia. Her research interests are image processing, face recognition, digital image forensic, intelligent control and control application, power system.

Ahmed Abdulhakim Al-Absi is an assistant professor in Department of Computer Engineering (Smart Computing) at Kyungdong University in South Korea. He earned a Ph.D. in computer science from Dongseo University in 2015. He received M.Sc. degree in information technology at University Utara Malaysia in 2011, and B.Sc. degree in computer applications at Bangalore University in 2008. His research interests include Big Data processing, Hadoop, Cloud computing, IoT, Distributed systems, Parallel computing, Bioinformatics, Security, and VANETs.

Mangal Sain received the M.Sc. degree in computer application from India in 2003 and the Ph.D. degree in computer science in 2011. Since 2012, he has been an Assistant Professor with the Department of Computer Engineering, Dongseo University, South Korea. His research interest includes wireless sensor network, cloud computing, Internet of Things, embedded systems, and middleware. He has authored over 50 international publications including journals and international conferences. He is a member of TIIS and a TPC member of more than ten international conferences.

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Prediction of Quality for Different Type of Wine based on Different Feature Sets Using Supervised Machine Learning Techniques

Satyabrata Aich[*], Ahmed Abdulhakim Al-Absi[**], Kueh Lee Hui[***], and Mangal Sain[****]

[*]*Department of Computer Engineering, Inje University, South Korea*

[**]*Department of Computer Engineering, Kyungdong University- Global Campus, Gangwondo, South Korea*

[***]*Dept. of Electrical Engineering, Dong-A University, South Korea*

[****]*Department of Computer Engineering, Dongseo University, South Korea*

satyabrataaich@gmail.com, absiahmed@kduniv.ac.kr, leehkueh@dau.ac.kr, mangalsain1@gmail.com

***Abstract*—In recent years, most of the industries promoting their products based on the quality certification they received on the products. The traditional way of assessing the product quality is time consuming, however with the invent of machine learning techniques the processes has become more efficient and consumed less time than before. In this paper we have explored, some of the machine learning techniques to assess the quality of wine based on the attributes of wine that depends on quality. We have used white wine and red wine quality dataset for this research work. We have used different feature selection technique such as genetic algorithm (GA) based feature selection and simulated annealing (SA) based feature selection to check the prediction performance. We have used different performance measure such as accuracy, sensitivity, specificity, positive predictive value, negative predictive value for comparison using different feature sets and different supervised machine learning techniques. We have used nonlinear, linear and probabilistic classifiers. We have found that feature selection-based feature sets able to provide better prediction than considering all the features for performance prediction.**

We have found accuracy ranging from 95.23% to 98.81% with different feature sets. This analysis will help the industries to access the quality of the products at less time and more efficient way.

Keywords—machine learning; feature selection; classifiers; performance metrics;wine qulaity

I. INTRODUCTION

In recent years there is a modest increase in the wine consumption as it has been found that wine consumption has a positive correlation to the heart rate variability [1]. With the increase in the consumption wine industries are looking for alternatives to produce good quality wine at less cost. Different wines have different purposes. Although most of the chemicals are same for different type of wine based on the chemical tests, the quantity of each chemicals have different level of concentration for different type of wine. These days it is really important to classify different wine for quality assurance [2]. In the past due to lack of technological resources it become difficult for most of the industries to classify the wines based on the chemical analysis as it takes lot of time and also need more money. These days with the advent of the machine learning techniques it is possible to classify the wines as well as it is possible to figure out the importance of each chemical analysis parameters in the wine and which one to ignore for reduction of cost. The performance comparison with different feature sets will also help to classify it in a more distinctive way. In this paper an intelligent approach is proposed by considering genetic algorithm (GA) based feature selection as well as simulated annealing-based feature selection considering the nonlinear classifiers, linear classifiers and probabilistic classifiers to predict the quality in red wine as well as the white wine.

The structure of the paper is organized as follows: Section II presents the past work related to this field. Section 3 describes about the methodologies used for this research work. Section 4 describes about the result of feature selection as well

Manuscript received January 02, 2018. This work was supported by Kyungdong University grant, 2018 and a follow-up the invited journal to the outstanding paper of the 20th International Conference on Advanced Communication Technology (ICACT 2018).

Satyabrata Aich is with the Department of Computer Engineering, Inje University, South Korea (e-mail: satyabrataaich@gmail.com)

Ahmed Abdulhakim Al-Absi is with the Department of Computer Engineering, Kyungdong University- Global Campus, Gangwondo, South Korea (e-mail: absiahmed@kduniv.ac.kr)

Kueh Lee Hui is with the Department of Electrical Engineering, Dong-A University, South Korea (e-mail: leehkueh@dau.ac.kr)

Mangal Sain is with the Department of Computer Engineering, Dongseo University, South Korea. He is the corresponding author of this paper. (Corresponding author phone: +8251-320-2009; e-mail: mangalsain1@gmail.com).

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

as the result of classification. Section 5 describes about the conclusion and future work.

II. RELATED WORKS

In the past few attempts have been made to use different machine learning approaches and feature selection techniques to the wine dataset. Er and Atasoy proposed a method to classify the quality of wines using three different classifier such as support vector machines, Random forest and k-nearest neighborhood. They have used principal component analysis for feature selection and they found good result using Random forest algorithm [3]. Chen et al proposed an approach that will predict the grade of wine using the human savory reviews. They have used hierarchical clustering approach and association rule algorithm to process the reviews and predict the wine grade and they found an accuracy of 85.25% while predicting the grade [4]. Appalasamy et al proposed a method to predict wine quality based on physiochemical test data. They have pointed out that classification approach helps to improve the quality of wine during the production [5]. Beltrán et al proposed an approach to classify the wine based on aroma chromatograms and they have used PCA for dimensionality reduction and wavelet transform for feature extraction and classifiers such as neural network, linear discriminant analysis and support vector machine and found that support vector machine with wavelet transforms perform better than other classifiers [6]. Thakkar et al., used analytical hierarchy process (ahp) to rank the attributes and then used different machine learning classifiers such as support vector machine and random forest and they found accuracy of 70.33% using random forest and 66.54% using SVM [7]. Reddy and Govindarajulu used a user centric clustering approach to recommend the product. They have used red wine data set for the survey purpose. They have allocated relative voting to the attributes based on the literature review. Then they assigned weight to the attributes using Gaussian Distribution Process. They judged the quality based on the user preference group [8]. The above past work motivated us to try different feature selection algorithm as well as different classifiers to compare the performance metrics. This paper proposed GA based feature selection and SA based feature selection and used different classifiers such as PART, RPART, Bagging, C5.0, random forest, svm, lda, naïve bayes etc.

III. METHODOLOGIES

The flow chart of the proposed methodology is shown in the Fig. 1.

A. Data Collection

The wine data set is publicly available in the database of UCI. The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. This data set contains the physiochemical variables as well as sensory variables; altogether there are 12 attributes [9]. We have used genetic algorithm (GA)-based feature sets for feature selection. Pledsoe first presented an adaptive optimization search methodology is called genetic algorithm and Holland mathematically presented the genetic algorithm-based approach by getting inspiration from Darwin's theory of evolution. A variable is mentioned as a gene. A chromosome is nothing but a sequence of gene. An initialization is done randomly by using population of chromosome. The quality of the chromosomes is evaluated according to a predefined fitness function. High performance chromosomes are used to produce the offspring. The genetic operators such as mutation and crossover are used to form the offspring. In this process the chromosomes are competing with each other and the fittest one survives at the end. The optimal solution comes after a series of iterative computations. [10, 11].

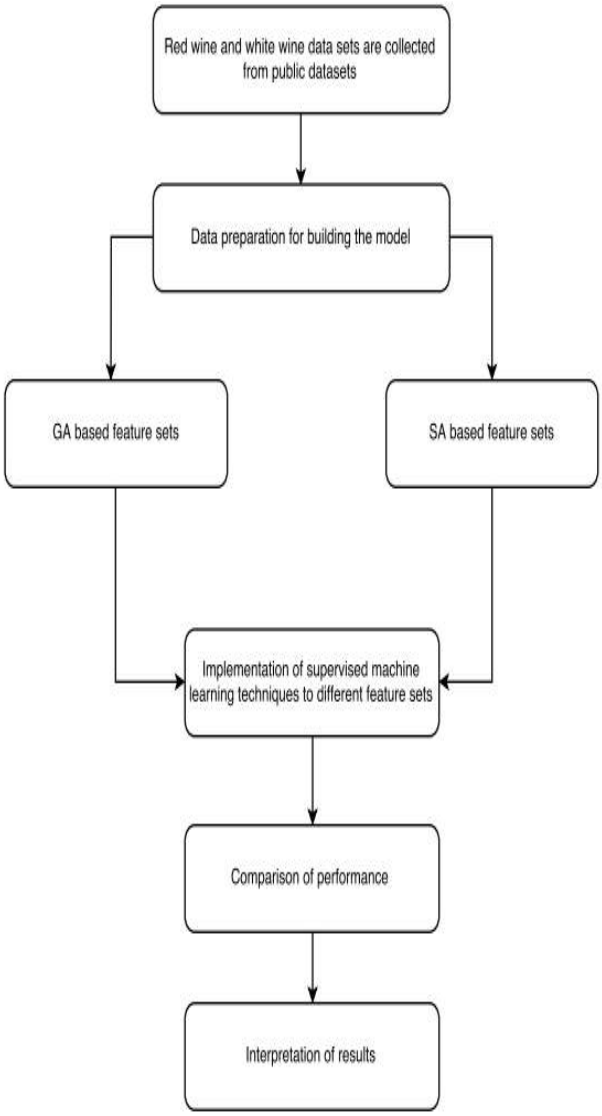

Fig. 1. Flow chart of proposed method

We have also used simulated annealing-based feature set for feature selection. The widely used combinatorial optimization method is called simulated annealing. It is one of the most popular search algorithms. This method used probabilistic technique to find the local optima that ultimately find a better solution [12]. This method is widely used for feature selection method. The simulated algorithm procedure is mentioned below. It runs based on the number of classes. If the number of classes is n then it runs n^{th} times. In each run j, the

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

subset of the feature for the j^{th} class is found. All the j^{th} class patterns are taken into one class and other pattern belong to the other class while evaluating the current string. This process helps to give the features which classify patterns as belonging to class j or not class j [13]. After selecting the features by using simulated annealing (SA) and genetic algorithm (GA), we have implemented the data sets into various classifiers and compare the performance parameters.

B. *Performance Measure Metrics*

The parameters used to compare the performance and validations of classifier are as follows: accuracy, sensitivity, specificity, positive predictive value (ppv), negative predictive value (npv). The sensitivity is defined as the ratio of true positives to the sum of true positives and false negatives. The specificity is defined as the ratio of true negatives to the sum of false positives and true negatives. In our research we have used the Positive predictive value and negative predictive value to check the present and absent of one type of wine. So, the ppv is the probability that the one type of wine is present given a positive test result and npv is the probability that the one type of wine is absent given a negative test result [14]. Accuracy is defined as the ratio of number of correct predictions made to the total prediction made and the ratio is multiplied by 100 to make it in terms of percentage.

IV. RESULTS AND DISCUSSION

We have divided the data into two groups such as train data and test data. We trained each classifier based on the trained data and predict the power of classifier on the test data. So, each classifier able to show all the performance metrics such as accuracy, sensitivity, specificity, PPV, and NPV based on the test data. We have applied all the classification techniques to the GA based reduced feature sets for two types of wine as well as SA based reduced feature sets for two types of wine to measures the performance parameter with respect to each classifier. We separated each performance measures with respect to GA and SA sets and plot the column plot for better visualization. The results of each performance measure with respect to two feature sets are shown in the Fig. 2, 3, 4, 5, and 6 respectively for red wine and 7, 8, 9, 10, 11 for white wine.

A. *Comparison of Accuracy for red wine*

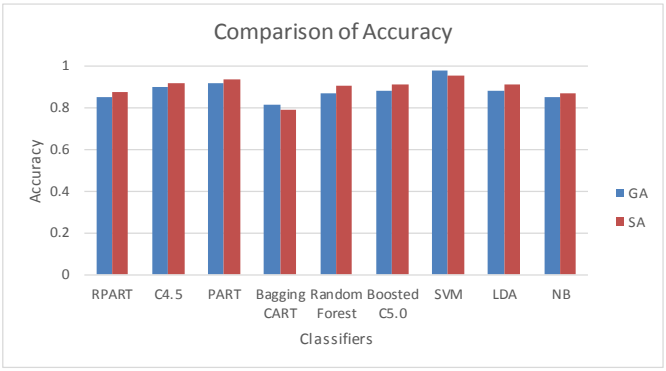

Fig. 2. Comparison of Accuracy on PCA and RFE sets

Fig. 2 show that SVM classifier shows maximum accuracy among all the classifiers. It is performed better with the SA based feature sets. The accuracy of SVM classifier with SA feature set found to be 95. 23%.

B. *Comparison of Sensitivity for red wine*

Fig. 3 shows the sensitivity plot of all the classifiers with two different feature sets. The plot shows SVM has the highest sensitivity compared to others and it was found to be 0.9717 with the SA based feature sets

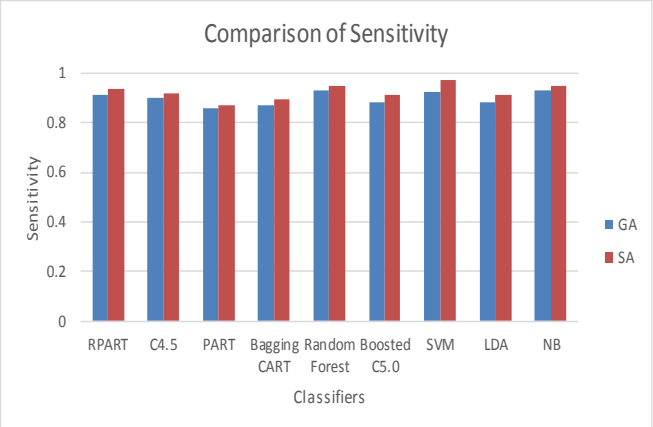

Fig. 3. Comparison of Sensitivity on PCA and RFE sets

C. *Comparison of Specificity for red wine*

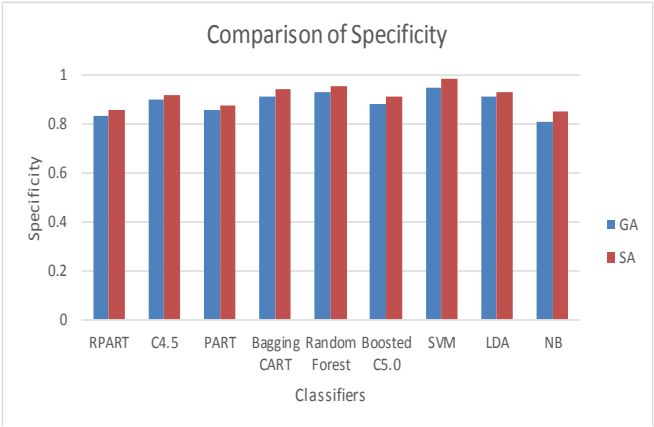

Fig. 4. Comparison of Specificity on PCA and RFE sets

Fig. 4 shows that SVM classifier shows maximum specificity among all the classifiers. It is performed better with the SA based feature sets. The specificity of SVM classifier with SA feature set found to be 0.9835

D. *Comparison of PPV for red wine*

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

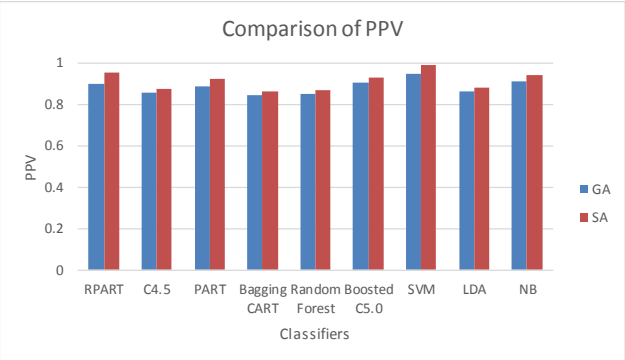

Fig. 5. Comparison of PPV on PCA and RFE sets

Fig. 5 shows the PPV plot of all the classifiers with two different feature sets. The plot shows SVM has the highest PPV compared to others and it was found to be 0.9912 with the SA based feature sets

E. Comparison of NPV for red wine

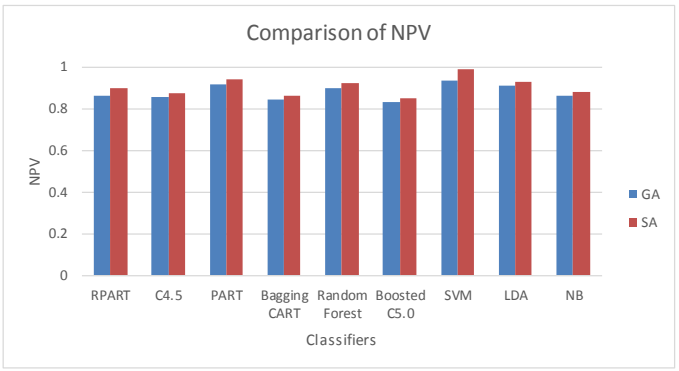

Fig. 6. Comparison of NPV on PCA and RFE sets

Fig. 6 shows that SVM classifier shows maximum NPV among all the classifiers. It is performed better with the SA based feature sets. The NPV of SVM classifier with SA feature set found to be 0.9907

F. Comparison of Accuracy for white wine

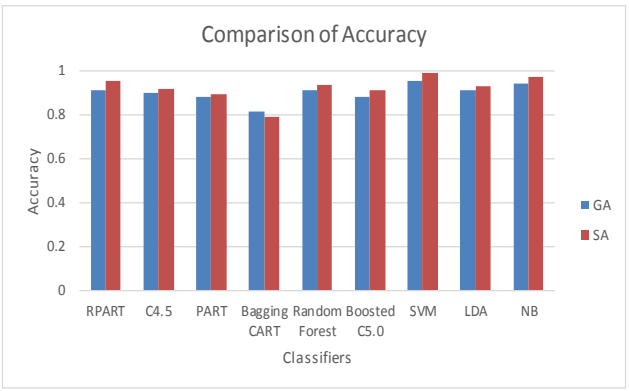

Fig. 7. Comparison of Accuracy on PCA and RFE sets

Fig. 7 shows that SVM classifier shows maximum accuracy among all the classifiers. It is performed better with the SA based feature sets. The accuracy of SVM classifier with SA feature set found to be 98. 81% for white wine data set.

G. Comparison of Sensitivity for white wine

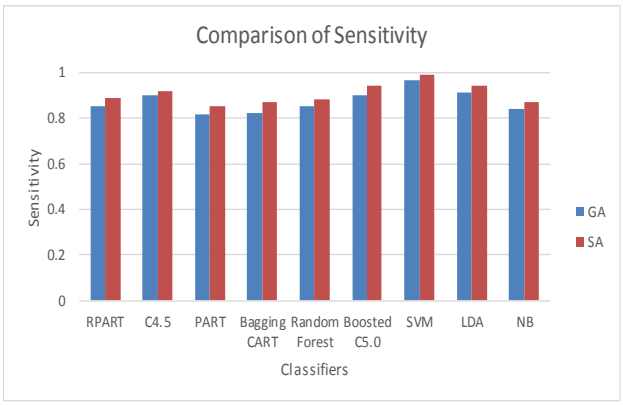

Fig. 8. Comparison of Sensitivity on PCA and RFE sets

Fig. 8 shows the sensitivity plot of all the classifiers with two different feature sets. The plot shows SVM has the highest sensitivity compared to others and it was found to be 0.9934 with the SA based feature sets for white wine data set.

H. Comparison of Specificity for white wine

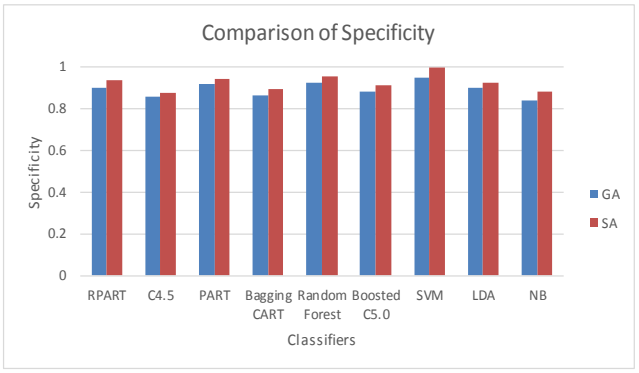

Fig. 9. Comparison of specificity on PCA and RFE sets

Fig. 9 shows that SVM classifier shows maximum specificity among all the classifiers. It is performed better with the SA based feature sets. The specificity of SVM classifier with SA feature set found to be 0.9956 for white wine data set.

I. Comparison of PPV for white wine

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

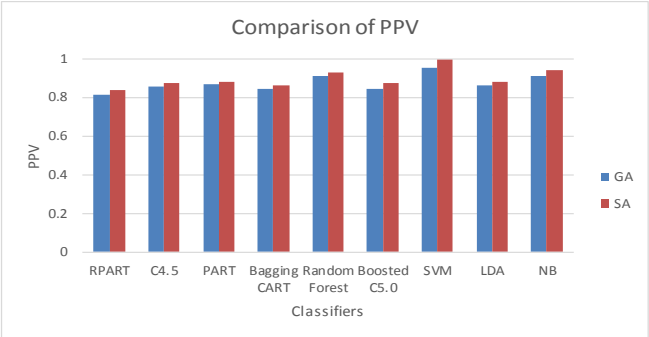

Fig. 10. Comparison of PPVon PCA and RFE sets

Fig. 10 shows the PPV plot of all the classifiers with two different feature sets. The plot shows SVM has the highest PPV compared to others and it was found to be 0.9987 with the SA based feature sets for white wine data set.

J. Comparison of NPV for white wine

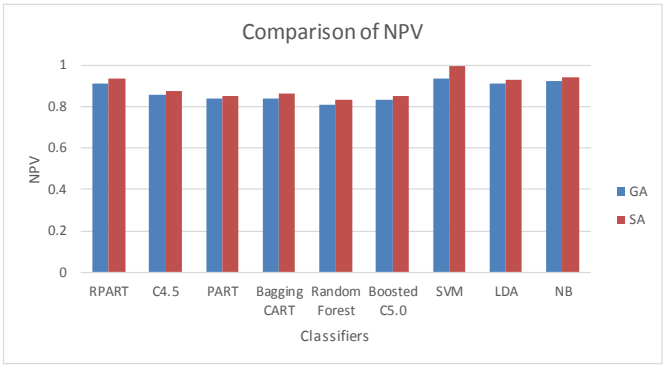

Fig. 11. Comparison of NPV onPCA and RFE sets

Fig. 11 shows that SVM classifier shows maximum NPV among all the classifiers. It is performed better with the SA based feature sets. The NPV of SVM classifier with SA feature set found to be 0.9992 for white wine data set.

The above plots show the performance metrics comparison of different type of wines based on the metrics parameters such as accuracy, sensitivity, specificity, ppv and npv on two different feature sets. The result shows that the SVM classifier performs better for both type of data sets. Specially it is performing better in SA based feature sets. Although it is easy to say based on our result that Simulated Annealing is better algorithm for feature selection compared to genetic algorithm-based feature selection method, however the result could be different for other datasets as well as it could be different for bigger datasets. Similarly based on our result we can say that SVM classifier is best, but in practical lot of other parameters also come into picture that could change the scenario completely. This analysis will give a clear idea about the important attributes for the prediction of quality as well as it saves lot of time and money for the industries.

V. CONCLUSION AND FUTURE WORK

This paper mentioned about potential of genetic algorithm as well as simulated annealing algorithm for feature selection as well as the potentials of the classifiers to predict accurately based on the new feature sets. The feature selection algorithm provided a clear idea about the importance of the attributes for prediction of quality, which was time consuming and expensive when done in the traditional way. We have also compared the performance metrics of linear, nonlinear, and probabilistic based classifiers and it was found that these classifiers performed well with the new feature sets. We have found that the SA based feature sets performed better than the GA based feature sets. We have also found that the SVM classifier performed better compared to all other classifiers for red wine and white wine data sets. We have found accuracy ranging from 95.23% to 98.81% with different feature sets. In future we can try other performance measures and other machine learning techniques for better comparison on results. This analysis will help the industries to predict the quality of the different type of wines based on certain attributes and also it will helpful for them to make good product in the future.

REFERENCES

[1] I.Janszky, M.Ericson, M.Blom, A. Georgiades, J.O.Magnusson, H.Alinagizadeh, and S.Ahnve, "Wine drinking is associated with increased heart rate variability in women with coronary heart disease," *Heart*, 91(3), pp.314-318,2005.

[2] V. Preedy, and M. L. R. Mendez, "Wine Applications with Electronic Noses," in Electronic Noses and Tongues in Food Science, Cambridge, MA, USA: *Academic Press*, 2016, pp. 137-151.

[3] Y.Er, and A.Atasoy, "The Classification of White Wine and Red Wine According to Their Physicochemical Qualities,"*International Journal of Intelligent Systems and Applications in Engineering*,4,pp.23-26,2016.

[4] B. Chen, C. Rhodes, A. Crawford, and L. Hambuchen, "Wineinformatics: applying data mining on wine sensory reviews processed by the computational wine wheel," *IEEE International Conference on Data Mining Workshop*, pp. 142-149, Dec. 2014.

[5] P.Appalasamy, A.Mustapha, N.D.Rizal, F.Johari, and A.F.Mansor, "Classification-based Data Mining Approach for Quality Control in Wine Production*," Journal of Applied Sciences*, 12(6), pp.598-601,2012

[6] N. H. Beltran, M. A. Duarte- MErmound, V. A. S. Vicencio, S. A. Salah, and M. A. Bustos, "Chilean wine classification using volatile organic compounds data obtained with a fast GC analyzer," Instrum. Measurement, *IEEE Trans*., 57: 2421-2436, 2008.

[7] K.Thakkar,J.Shah,R.Prabhakar,A.Narayan,A.Joshi, "AHP and MACHINE LEARNING TECHNIQUES for Wine Recommendations," *International Journal of Computer Science and Information Technologies*,7(5),pp. 2349-2352 ,2016

[8] Reddy, Y. S., & Govindarajulu, P. (2017). An Efficient User Centric Clustering Approach for Product Recommendation Based on Majority Voting: A Case Study on Wine Data Set. *IJCSNS*, 17(10), 103.

[9] M.Forina, R. Leardi, C. Armanino, and S. Lanteri, "PARVUS An Extendible Package for Data Exploration," *Classification and Correla*,1988.

[10] Bledsoe, W. W. (1961). The use of biological concepts in the analytical study of systems. In the ORSA-*TIMS National Meeting*

[11] Holland, J. H. (1992). Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. *MIT press.*

[12] Jeong, I. S., Kim, H. K., Kim, T. H., Lee, D. H., Kim, K. J., & Kang, S. H. (2018). A Feature Selection Approach Based on Simulated Annealing

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

for Detecting Various Denial of Service Attacks. *Software Networking*, 2018(1), 173-190.

[13] Devi, V. S. (2015, December). Class Specific Feature Selection Using Simulated Annealing. In International Conference on Mining Intelligence and Knowledge Exploration(pp. 12-21). *Springer, Cham.*

[14] H.B.Wong, and G.H. Lim, "Measures of diagnostic accuracy: sensitivity, specificity, PPV and NPV," *Proceedings of Singapore healthcare*, 20(4), pp.316-318,2011.

Satyabrata Aich is working as a researcher in the field of computer engineering He has over four years of teaching, research and industry experience in India and abroad. He has published many research papers in journals and conferences in the realms of Supply Chain Management and data analytics. His research interests are natural language processing, Machine learning, supply chain management, data mining.

Kueh Lee Hui is working as an assistant professor at the department of Electrical Engineering, Dong-A University since 2012. She completed her PhD Degrees from Department of Electrical Engineering, Dong-A University, Korea. In 2009 she completed her BS degree in Electronic and Communication, Department of Electronic Engineering, University Malaysia of Sarawak, Malaysia. She also done MS in 2007 from Malaysia. Her research interests are image processing, face recognition, digital image forensic, intelligent control and control application, power system.

Ahmed Abdulhakim Al-Absi is an assistant professor in Department of Computer Engineering (Smart Computing) at Kyungdong University in South Korea. He earned a Ph.D. in computer science from Dongseo University in 2015. He received M.Sc. degree in information technology at University Utara Malaysia in 2011, and B.Sc. degree in computer applications at Bangalore University in 2008. His research interests include Big Data processing, Hadoop, Cloud computing, IoT, Distributed systems, Parallel computing, Bioinformatics, Security, and VANETs.

Mangal Sain received the M.Sc. degree in computer application from India in 2003 and the Ph.D. degree in computer science in 2011. Since 2012, he has been an Assistant Professor with the Department of Computer Engineering, Dongseo University, South Korea. His research interest includes wireless sensor network, cloud computing, Internet of Things, embedded systems, and middleware. He has authored over 50 international publications including journals and international conferences. He is a member of TIIS and a TPC member of more than ten international conferences.

Copyright ⓒ 2018 GiRI (Global IT Research Institute)

Volume 7 Issue 3, May. 2018, ISSN: 2288-0003

ICACT-TACT JOURNAL

GIRI
Global IT Research Institute

1713 Obelisk, 216 Seohyunno, Bundang-gu, Sungnam Kyunggi-do, Republic of Korea 13591
Business Licence Number : 220-82-07506, Contact: tact@icact.org Tel: +82-70-4146-4991