

ISSN: 0973-6255 Vol. 12 No.2 July - December 2018

# Indian Journal of Engineering, Science, and Technology

A Refereed Research Journal



















Published by BANNARI AMMAN INSTITUTE OF TECHNOLOGY (Autonomous Institution Affiliated to Anna University of Technology, Coimbatore -Approved by AICTE - Accredited by NBA and NAAC with "A" Grade) Sathyamangalam - 638 401 Erode District Tamil Nadu India Ph: 04295-226340 - 44 Fax: 04295-226666

www.bitsathy.ac.in E-mail: ijest@bitsathy.ac.in



# Indian Journal of Engineering, Science, and Technology

IJEST is a refereed research journal published half-yearly by Bannari Amman Institute of Technology. Responsibility for the contents rests upon the authors and not upon the IJEST. For copying or reprint permission, write to Copyright Department, IJEST, Bannari Amman Institute of Technology, Sathyamanagalam, Erode District - 638 401, Tamil Nadu, India.

**Chief Patron** 

Editor

Associate Editors

Dr. M.P. Vijayakumar Trustee & Director Dr. C. Palanisamy Principal Dr. Lakshmi Narayana M Mohan Professor/ECE

Bannari Amman Institute of Technology, Sathyamangalam, Erode District - 638 401, Tamil Nadu, India

# **Editorial Board**

#### Dr. Srinivasan Alavandar

Department of Electronics and Computer Engineering Caledonian (University) College of Engineering PO Box: 2322, CPO Seeb-111, Sultanate of Oman

**Dr. H.S. Jamadagni** Centre for Electronics Design and Technology Indian Institute of Science Bangalore – 560 012

**Dr. V.K. Kothari** Department of Textile Technology Indian Institute of Technology–Delhi New Delhi – 110 016

**Dr. S. Mohan** National Institute of Technical Teachers Training and Research Taramani, Chennai - 600 113

**Dr. P. Nagabhushan** Department of Studies in Computer Science University of Mysore Mysore - 570 006

**Dr. Edmond C. Prakash** Department of Computing and Mathematics Manchester Metropolitan University Chester Street, Manchester M1 5GD, United Kingdom

**Dr. E.G. Rajan** Pentagram Research Centre Pvt. Ltd. Hyderabad – 500 028 Andhra Pradesh

**Dr. Seshadri S.Ramkumar** Nonwovens & Advanced Materials Laboratory The Institute of Environmental & Human Health Texas Tech University, Box 41163 Lubbock, Texas 79409-1163, USA

#### Dr. T.S. Ravi Sankar

Department of Electrical Engineering University of South Florida Sarasota, FL 34243, USA

#### Dr. T.S. Jagannathan Sankar

Department of Mechanical and Chemical Engineering North Carolina A&T State University NC 27411, USA

#### Dr. A.K. Sarje

Department of Electronics & Computer Engineering Indian Institute of Technology, Roorkee Roorkee - 247 667

#### Dr. R. Sreeramkumar

Department of Electrical Engineering National Institute of Technology - Calicut Calicut – 673 601

#### **Dr. Talabatulla Srinivas**

Department of Electrical & Communication Engineering Indian Institute of Science Bangalore - 560 012

#### Dr. Dinesh K. Sukumaran

Magnetic Resonance Centre Department of Chemistry State University of New York Buffalo, USA – 141 214

#### Dr. Prahlad Vadakkepat

Department of Electrical and Computer Engineering National University of Singapore 4 Engineering Drive 3, Singapore 117576

#### Dr. S. Srikanth

AU-KBC Research Centre Madras Institute of Technology Campus Anna University Chennai-600 044

# Indian Journal of Engineering, Science, and Technology

ISSN: 0973-6255

Volume 12 Number 2

July - December 2018

# **CONTENTS**

	*Excerpt from the Proceeding of National Conference						
S.No.	Title	Page.No.					
1	<b>Digitalization of Electoral Process</b> V. Alamelu, Priyadharshini, Suganthi and Suriya Varshini	01					
2	Secured Energy Efficient Wireless Data Transfer Using Li-Fi Technology C. Manipriya, A.S.Deepika, S. Indhumathi, S. DivyaPratheeba and S.Nithya	07					
3	<b>Denial of Service Using Software Puzzle</b> E.Anjana Devi, S.Raj Kumar, G.M.Madhu Mitha and S.Ponmutharthan	11					
4	Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques V. Kaviya Shree, R.Kalaivani, K.Gowri Nayaki and B.Gokulnath	16					
5	<b>Precise Inspection of Mechanical Systems Based Android APP</b> G.Angusamy, P.Prakash Kumar, K.Hirushi Kesan, R.Naveen Kumar, S.Rahul and V.Vignesh	20					
6	<b>Predictive Analytics for Diabetes Mellitus Using Classification Models</b> M. Munirathinam and K. Premalatha	22					
7	A Survey on Data Mining in Healthcare T. GitanjaliSimran, D.Sujeetha and M.Sushmitha	27					
8	<b>Survey on Application of Image Processing</b> J Gayathri and S Ramya	32					
9	Efficient Content Based Image Retrieval S.Ramya, R.Prathiba, S.Raja Ganesh and PCD.Kalaivaani	38					
10	Question Paper Generation System S.C. Lavanya, K.K. Aarthi, R. Devaprithika, S. Durgapiriyadharshini and K. Dinesh kumar	41					

# **Digitalization of Electoral Process**

#### V. Alamelu, Priyadharshini, Suganthi and Suriya Varshini

Department of Computer Science and Engineering Kumarasamy College of Engineering (Autonomous) Thalavapalayam, Karur - 639 113, Tamil Nadu E-mail: Alameluv.cse@mkce.ac.in, priyajack96@gmail.com, suganthikalaimani@gmail.com

### Abstract

A wide arrangement of structures require trustworthy individual affirmation wants to either certify or choose the discerning of a solitary invocation their organizations. The explanation behind alike plans is to protect that the furnish organizations are gotten too simply by a true blue customer, and no other person. Instances of alike approach consolidate firm ingress to arrangements, PC systems, PCs, telephones and ATMs. Without generous individual affirmation schemes, these structures are unprotected against the wiles of an pretender. Bioscience affirmation, or basically statistics, implies the modified affirmation of peculiar in light of their cytology and additionally observable qualities. By applying life measurement it is breeze to assert or develop a man's character in light of "her personality", as a substitute by "what she has" (prototype an ID card) or "what she remembers" (prototype a mystery word). This endeavor is proposing to reenact digitalization of voting system which is used to check the customer data in light of their fingerprints using novel stamp machine and keep up a key separation from fake voting.

Keywords: Recognition, Fake, Voting, Fingerprint, Identification, Individuals.

#### **1. INTRODUCTION**

The objective of voting is to empower voters to hone their qualification to express their choices as for specific issues, bits of sanctioning, national exercises, set up adjustments, audits and in addition to pick their organization and political specialists. Advancement is being used progressively as a device to help voters to cast their votes. To allow the action of this right, all voting structures the world over join the going with progresses.

- Voter ID and affirmation
- Voting and recording of votes cast
- Vote checking
- Publication of race comes to fruition

Biometric development into our optional system, every voter's exceptional check will be affirmed by the usage of a Biometric Verification Device (BVD). A voter will be allowed to vote just if his character is insisted by the (BVD) i.e. in case his/her fingerprints are seen by BVD as the person whose name is on the enroll.

A voter's fingerprints may not be seen by the Biometric Verification Device if his/her fingers are dingy or dirtied regardless. It is basic that all voters must ensure that their fingers are spotless before they encounter check. Recognizing individuals in perspective of their

unmistakable anatomical (novel stamp, stand up to, iris, hand geometry) and observable (indication, tone) temperament is termed bioscience. Since biometric modifier can't be mutual or lost, they normally address a man's identity. Biometrics is quickly transforming into a crucial piece of intense unmistakable confirmation courses of action. Affirmation of a man by their bag of bones, by then associating that bag of bones to a remotely settled "identity", shapes a skilled affirmation instrument. Among the assorted biometric recognizing verification techniques, one of kind check affirmation advancement has a better than average change of attributes including precision, thresh out, capacity and payment of perusers, improvement of development and solace of usage, composing it the transcendent life measurement development in business utilization.

#### 1.1 Fingerprint Classifying

The sheath innermost area of our palm, fingers, feet, and toes is "wrinkled" or secured including coextensive constructed cases. These edges are called rubbing edges and they give grinding composing it less requesting to us to deal with and grip inquiries and area left out slowpoke. The various qualifications in the procedure disintegration edges are outlined, cracked, and angled compose wrinkled sheath zones, along with dactylogram, specific. V. Alamelu, Priyadharshini, Suganthi and Suriya Varshini

The uniqueness of dactylogram is settled in. The concealed normal steadiness of special stamp characteristics is in like manner a settled sureness uncovered in various finger impressions ponders drove in different sensible fields over the earlier century Ideally, it would be a straightforward issue to decide if two fingerprints were from a similar finger the pictures would be indistinguishable or they would not. In any case, despite the fact that our fingerprints don't change after some time, the unique mark pictures can shift a ton, particularly for a few people. For instance, certain skin conditions and wear because of difficult work can influence unique mark pictures. This makes unique finger impression acknowledgment an exceptionally difficult issue that does not have a flawless arrangement. Thus, caught unique finger impression pictures are regularly not an ideal match to the put away picture from a similar finger.

Unique mark pictures from two distinct fingers of two distinct individuals can seem to be comparative, particularly when, due to worn fingerprints or brief wrinkles, there is next to no data left about the real unique mark. The bigger the populaces you are working with, the greater probability of comparative unique finger impression pictures.



Fig. 1.1 Fingerprint capture device

Fingerprint recognition works in two stages:

- First, customers are chosen with the system their fingerprints are gotten and secured in a database
- Next, when a man ought to be offered (e.g., to open a gateway or to login in to a PC), they simply yield their on the special finger impression peruser
- An Application that Enrolls Users
- Captures different fingerprints for no less than two fingers from a unique finger impression peruser
- Checks picture quality to ensure that a better than average quality scope is gotten
- Extracts the one of a kind finger impression particulars
- Saves the one of a kind finger impression pictures or conceivably particulars in a database

- A service(s)/application(s) that perceives/checks people
- Captures a one of a kind finger impression particulars
- Extracts the unique finger impression particulars
- Compares unique finger impression with selected fingerprints to distinguish a client from a rundown or confirm particular client

This SDK gives unique mark catch, extraction, and enlistment and recognizable proof/check capacities to enable you to build up these parts.

When fabricating a unique mark acknowledgment application, the information stream comprises of: Taking a dactylogram figure (check) in distinction to the special stamp peruser. The conclusion dactylogram figure input enclose no less than one of a kind stamp pictures, called a Fingerprint Image Views (FIVs). A regular FID for one of a kind check affirmation applications contains only a solitary FIV anyway we furthermore reinforce alternate points of view (e.g., if there are various thumb in distinction to only singular or different pictures from a singular thumb set away in a lone FID).

Extract the one of a kind stamp features. In the midst of extraction, Fingerprint Minutiae Data (FMD) is made, with every novel check set away in a Fingerprint Minutiae View (FMV) in the FMD. A FMV in a FMD takes near 1.5K (maybe less depending upon the interesting imprint). FMDs are used for recognizing customers in a social occasion and checking specific customers. The ANSI and ISO measures permit different points of view however this SDK makes simply single-see FIDs and FMDs. This data stream is showed up in Figure 1.2



Fig. 1.2 Data flow of fingerprint recognition

#### 2. LITERATURE REVIEW

# 2.1 A Development of IRIS-BLOB Map Approach for IRIS Recognition

A man in iris affirmation system is recognized in light of the unlikeness of the iris case of the eye. Generally iris is described as an indirect zone among understudy and sclera of personal eye which indicates comical that is excellent for each particular, important for executive the estimation and range of the understudy. This methodology handling a first rate camcorder to get huge assurance pictures of iris. A wonderful photo of the iris is gotten to earn each point by point feature while remaining non-nosy of the human overseer. Obstructions of iris analyze,

- Difficult to get for a couple of individuals
- To secure the iris configuration, requires a considerable measure of customer interest or to usage exorbitant and complicated information equipment
- The presentation of iris confirmation may be incapacitated by reflectors, shades and connection central focuses
- It isn't helpful for quantifiable utilization as confirmation of iris biometric on bad behaviour scene isn't gotten out

# 2.2 Speaker Recognition Identifying People by their Voices

Speaker affirmation is a dull term which insinuates any errand that isolates between people in perspective of their tone aspect. There are two limited endeavors that have been pondered generally. These are announcer recognizing confirmation and announcer check. The speaker conspicuous verification undertaking is to arrange an unlabeled voice movement as having a place with one of a game plan of N note announcers, while the announcer affirmation function is to pick regardless of in case an unlabeled tone flag has a place with a limited mention announcer. Announcer affirmation revolves around the verbal personality related the estimations of the verbal plot, nasal holes, mouth and the other talk dealing with part of the mortal frame that make talk. Announcer check does not revolve around the vibrant or the enunciation of talk itself. In tone affirmation "sound spectrogram" contraptions are used. Sound spectrogram makes a graph evident and display vibrant plentifulness on the steep center point and exhibits the time on the level center point. Weights of voice affirmation,

Voice acknowledgment requires a top notch sound flag with no clamor and undesirable flag impedance

# 2.3 Pattern Recognition Based DNA Sequence Compressor

The normal development of each living being is composed by billions of respective cells. The force focal point of all cells is the De-oxyribo Nucleic destructive (DNA) that involve an aggregate plan of rules anticipated that would facilitate the working of every single one of the cells. Each individual's DNA is stand-out beside indistinct twins. It would in this way have the capacity to be seen as a perfect technique for identity affirmation. This system for get still should be cultivated. So far the DNA examination has not been adequately customized to putrid the DNA examination as a biometric advancement, where examination of human DNA can be done inside 10 minutes. At the point when the development advances with the objective that DNA can be composed subsequently consistently it may exhibit more basic. Regardless, at demonstrate DNA System is significant set up in bad behavior acknowledgment in this manner will remain in the law execution domain. Burdens of DNA affirmation,

- It is extremely costly
- It's mind boggling and inclined to blunders
- It can damage individuals' protection

#### 2.4 Personal Identification Using Blood Vessels

A toned locality betwixt the understudy and sclera is familiar as iris while retina is arranged at back zone of eye. The authority of retinal affirmation is the vein outline. Retinal affirmation is the most strong and solid due to its fat and uncommon vascular illustration. It is moderately hard to produce a mortal retina as it is arranged at backward of eye and isn't introduced to foreign condition. Retina is separated testing a camera which is frequently worn by ophthalmologists for remedy of retinal sicknesses. Remembering the true objective to check retina, the individual requisite spot his eye adjacent to the point of convergence of camera. In the midst of separating action, he should persist still and besides ought to discard glasses to fend off hail impedance. Already the scanner is started, by then green shiny shift to outright ring and vascular case of retinal is gotten. As a result of a tangled acquiring process, retinal affirmation systems are not used by and large but instead in view of their unfaltering quality and relentless quality, retina biometrics structures are suited for most extraordinary security zones, for instance, Government and military [4]. A capable executive is needed and the individual being checked necessities to take after his or her course. In any case, retina checking schemes are said to be correct, and treasure trove its application in huge preservation involvement. Deterrents of Retina Recognition,

- Very meddlesome
- Very costly

# 2.5 Face Recognition

A leading affirmation facility usage a man's mask for unmistakable evidence reason. This is ended by means of therefore perceiving a man from a mechanized picture or a video diagram from a video source. The figuring can be geometric part established and presence-based. 3-Dimensional mask affirmation systems are for the utmost part used as they give correct results since they get the honest to goodness condition of faces. 3D exemplary of appearances are built by the system and the 3D faces are considered for affirmation. The appearance recycled to recognize an singular are spot of eyes, nose and mouth, cheekbones, expel betwixt eyes, eyebrow density and spot, mask design, jaw pattern. These appearances are then differentiated and the database pictures for conspicuous evidence. Shortcomings of Face Realization,

- Face can be deterred by hair, glasses, tops, et cetera tricky to change in lighting up, flag and explanation
- Face change with time
- A go up against ought to be adequately splendid by controlled light source in motorized face approval structure

# 2.6 Personal Verification and Identification using Hand Geometry

Hand geometry affirmation systems use different estimations taken in distinction to the human hand, counting its pattern, scope of perceive, range and thickness of the fingers, finger guidelines. About, condition of each one individual's hand one of a kind and that the condition of a man's hand does not adjustment later specific age. Picture obtainment for hand biometrics may be touch form and handled one, which demand a level stage to bring the hand and pegs to deal with the circumstance of the customer's hand, or stage complimentary, non-contact techniques. Its affirmation apparatus are monstrous yet recognizing evidence is done in a seconds. Weights of hand geometry,

- Wide magnitude of hand geometry devices required
- Single hand use just in one time
- Not extremely intriguing

# 2.7 A Survey on Writer Identification Schemes

It is the methodology used to see a man's stamp. It can be worked in two particular policy : immovable and movable or separated and on the web. Movable: In this approach, customers make their blemish on paper, which is log in over a visual scanner or a camera, and a short time later the biometric system is used to see the stamp by consider its pattern. This social event is generally called "disengaged mode". While the in interesting, customers form their check in a digitizing tablet. This approach is used as a ceaseless structure from now on noted as "on-line mode". The utilization of unmistakable confirmation of the reporter using his stamp is in deep regions, for instance, to deal with the ace issues in criminology by lawful ace essential administration systems, where a constrained once-over of recognized writers is given by the writer ID structure. Downside of Signature Recognition,

• Consistent method for check is needed for obtain and assertion of the separate

# 2.8 A Tool for Information Security

- Unlikeness is the way by which strong the biometric
- detaches independently from a different person
- Perpetuation measures how strong a biometric contradicts developing
- Performance precision, speed, and heartiness of innovation utilized

# 3. EXISTING SYSTEM

Before examining present and future bearings in voting, it merits looking over what voting techniques have beforehand been utilized. One of the most seasoned and most conventional types of voting is to make a choice utilizing a paper ticket. While paper votes may appear to be straightforward, there are various types of these polls and they have experienced enormous changes. Previous renditions of voting by paper poll have required the voter to write in the applicant's name and host enabled political

IJEST Vol.12 No.2 July - December 2018

gatherings to convey pre-printed tickets. For this last type of paper ticket, the main activities required with respect to the voter were to convey this pre-printed vote to the surveying spot and store it in a voting station.

Another framework is Electronic Voting Machines (EVMs) are by and large progressively utilized rather than voting stations to counteract decision extortion by means of stall catching, which is vigorously pervasive in specific parts of India. A permanent ink is connected for the most part on the left pointer of the voter as a marker that the voter has made his choice. Disservices of existing framework,

- It requires human venture, in tallying the votes that impacts the race to monotonous and slanted to human slip-up
- Deceitful race instrument
- Constant spending stores for the choices staff are given along these lines, the proposed electronic voting structure must be tended to with these issues

# 4. PROPOSED SYSTEM

The commitment of walk as a biometric for individuals ID in observation application has pulled in analysts from PC vision as Gait can be seen from a separation. Despite the fact that walk acknowledgment is as yet another biometric, it beats a large portion of constraint that other biometric experience the ill effects of, for example, confront, speaker, unique finger impression acknowledgment. Step can be characterized as the planned and cyclic blend of developments that outcome in human motion. Step is an unpredictable motion design which includes synchronized development of torso segmentation, joints. The walk include used to distinguish individual are walk mileage, rhythm, tallness, remove amongst head and pelvis, most extreme separation amongst pelvis and feet, the separation between feet. The decision of a specific mortal trademark, to be utilized as a biometric highlight relies upon the accompanying guideline,

Human interesting finger impression remains a greatly general identifier and the biometric system for choice among law prerequisite. These thoughts of human conspicuous verification have incite the progression of exceptional stamp scanners that serve to quickly recognize individuals and dole out access benefits. The basic reason for these devices is in like manner to investigate the one of a kind check data of an individual and balance it with a database of various fingerprints. About everybody on the planet is conceived with a unique finger impression that is one of a kind; a different and exhaustively recognizing trait that separates us from the other 6.5 billion individuals that possess this world. It is a direct result of this reality that the unique finger impression has demonstrated such a helpful piece of biometric security. The very reason that unique mark scanners are valuable can be found in this reality also. In any case, this is a long way from the main reason they are utilized. Points of interest of proposed framework,

- Accuracy It isn't possible for a vote to be adjusted
- Democracy It enables simply qualified voters to vote and, it ensures that qualified voters vote just once
- Privacy Neither authority nor some other individual can associate any survey to the voter
- Verifiability Independently affirmation of that the whole of what votes have been checked precisely
- Resistance No optional component (any server sharing in the choice) or social affair of components, running the race can work in an interest to familiarize votes or with shield voters from voting
- Availability The system works properly as long as the review stands and any voter can approach it from the soonest beginning stage to the complete of the study

#### 5. CONCLUSION AND FUTURE WORK

Biometrics is a to a great degree captivating and empowering field. It has been growing exponentially starting late. The structure that can be achieved to recognize and scale the natural facts from body of human is also known as life measurement system. Life measurement are mechanized techniques for confirming or perceiving the character of a person based on some physical highlights or appearances, similar to a unique finger impression or face example or a few qualities of conduct, such as penmanship or keystroke designs. The wide assortment of physically one of kind characteristics will soon enable us to live in an exceptionally secure secret word less world. In certifiable the ID of voted people is done physically, yet in digitalization of voting framework the distinguishing proof is finished utilizing aadhar points of interest so the phony voting can be lessened and security is expanded. As a future upgrade, more number of biometric modalities can be utilized to give more number of security administrations.

V. Alamelu, Priyadharshini, Suganthi and Suriya Varshini

# REFERENCES

- U. Jarujareet and N. Covavisaruch, "An Improvement of Iris-Blob map Approach for Iris Identification," Proceedings of ECTI-CON, E- ISBN 978-1- 4244-5607-9, Print ISBN 978-1-4244-5606-2, INSPEC Accession Number 114102752010, 2010, pp.846-850.
- [2] George R. Doddington, "Speaker Recognition-Identifying People by their Voices", Proceedings of The IEEE, Vol 73, No. 11, November 1935
- [3] Prof G. Ramaswamy, Vudasreenivasarao, Dr. P.Ramesh, D.Ravikiran, "A Novel Approach for Human Identification through Fingerprints", International Journal of Computer Applications (0975 – 8887) Vol.4, No.3, July 2010.
- [4] M. Usman Akram, Anam Tariqy and Shoab A.Khanz, "Retinal Recognition: Personal Identification using Blood Vessels", 6<sup>th</sup> International Conference on Internet Technology and Secured Transactions, 11-14 December 2011, Abu Dhabi, United Arab Emirates.
- [5] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, "FaceRecognition: A Literature Survey", ACM Computing Surveys, Vol. 35, No. 4, December 2003, pp. 399-458.

- [6] Nonglukcovavisaruch, Pipatprateepamornkul, Puripantruchikachorn, and Piyanaattaksaphan, "Personal Verification and Identification Using Hand Geometry", ECTITransactions on Computer and Information Technology, Vol.1, No.2, November 2005.
- [7] Manu Garg and Er.Harish Bansal, "Fingerprint Recognition System using Minutiae Estimation", IJAIEM, ISSN 2319-4847, Vol.2, 5 May 2013.
- [8] Asker M.Bazen, "Fingerprint Identification-Feature Extraction, Matching, and Database Search", Ph.D thesis from University of Twente, August 19, 2002.
- [9] Haiyun Xu, Raymond N.J.Veldhuis, Tom A.M.Kevenaar and Ton A.H.M.Akkermans, "A Fast Minutiae-Based Fingerprint Recognition Systems", in IEEE system Journal, Vol.3, No.4, December 2009.
- [10] D.Maltoni, D.Maio, A.K.Jain, S.Prabhakar, Handbook of Fingerprint Recognition (second edition), springer, London, 2009.
- [11] M.Sonka, V.Hlavac, R. Boyle, "Image Processing, Analysis, and Machine Vision", 2<sup>nd</sup> Edition, Pws. Pub.co, 1998.

# Secured Energy Efficient Wireless Data Transfer Using Li-Fi Technology

C. Manipriya, A.S. Deepika, S. Indhumathi, S. DivyaPratheeba and S.Nithya

Department of Electronics Communication Engineering, KPR Institute of Engineering and Technology, Coimbatore - 641 407, Tamil Nadu E-mail: manipriyachinnasamy@gmail.com

### Abstract

Wireless communications play a crucial role in today's society. Information through wireless network has the most desired value. A now a days many people are using internet through the wireless device such as Wi-Fi hotspot, modem, etc. to accomplish their tasks. So the number of users accessing the internet has been greatly increased. In the Wi-Fi case, each user or group of users directly competes for access to bandwidth. The net result is that the more connections there are, the slower the download speeds are for all. To remedy some such a type of limitation of Wi-Fi we introducing the new concept called as Li-Fi (Light Fidelity) technology. In the case of Li-Fi, with its greater number of available access points, each pool of light provides full channel data rates with fewer simultaneous users. The overall net benefit to each user is up to 1000 time's greater speeds. The transfer data rate is up to 10 gigabits per second, which is faster than the average broadband connection. In addition, and in contrast to radio waves, the light does not pass through the walls. Therefore, with minimal precautions to avoid leakage from windows, etc., security is fundamentally enhanced as compared with Wi-Fi. The technology works by creating a module which is transferring the data through light and communicating the one computer to another computer. LEDs are used to send digital information, which are invisible to the naked eye. It is similar to the Wi-Fi except that it uses Light waves instead of Radio waves.

Keywords: Visible light communication, Library module, Security, ASP.NET, C#, SQL server.

# **1. INTRODUCTION**

In this paper we show, it is possible to transfer the data from one computer to another computer. The LEDto-LED communication provides a unique opportunity to provide communication capabilities. This is developed using ASP.NET as front end, C# as coding language and SQL Server as backend. For Hardware interface RS232 Data logger has been used. This application is planned to develop in a digital library. This is because multiplying books leads and buying of more e books lead to more expensive. But sharing makes more economy. Readers need a laptop to view the books. The laptop will be connected with USB device for receiving the files from the server. These received files can also be converted into speech format which can be helpful for both the normal and blind people. Here c# is exclusively is used for the purpose of text to speech conversion process. The details information are furnished in the modules.

In early research the light will blink and flash during the time of data transfer that may irritate to the user during the data transfer. Also the upload speed is not getting match with the download speed, so that low data transfer rate will occur. There problems are overcame in this system like, light will not flash during the time of data transmission, continuous light emission will be transfer the data. Also upload speed is equal to the download speed, through this method data transfer rate will be higher.

The data transfer rate will be 9600 bits per second. It means this speed is more sufficient to the humans to read a file visibility.

#### 2. PROPOSED MODEL



IJEST Vol.12 No.2 July - December 2018

#### **3. DATA SERVER**

According to this application, a centralized library will be the data server. In this module a library application has been created according to the user need. This data server contains all the user details like user registration name, registration number and other user details. This data server will be a windows application which contains provision for hardware interface. This application supports COMM 32 port and 32 PIN Serial ports for efficient data transmission. Also supports USB 3.0 version also. So that this application will be a plug and play application and easy to handle. An interactive grid is available for user management along with a data grid. Admin will be available for all process like managing the files, managing the users, removal of unwanted users, add and remove files.

#### 4. ENVIRONMENT CREATION

As mentioned above this application is enhanced in a digital library. This is because sharing e book will be more economy instead of purchasing. Light is the basic mode of communication. Here the environment is challenge to setup. All the readers should have a laptop with a USB port and the provided application. The laptop will be connected with the receiver device. There will be 5 meters distance between the transmitter and the receiver. It looks the reader sit below a light, but the light transmits the data. The receiver receives the data and the data will be shown in the corresponding application in a human eye reading speed format. It looks like reading a book virtually. The transmitter will be connected from the server and the receiver will be connected to the system, the intermediate data transmitter will be the light medium. The light will be the sealing light or a table lamp, the data transmission will works on. The main condition is the environment is a bidirectional linear model. In this environment the light will be travel in a straight line.

#### 5. LIFI TRANSMITTER

The transmitter section is used for transmitting data from the transmitter PC. It consists of computer with the data server, Max 232 IC, LED switching circuit, Controller Board. The main component in a transmitter section is the visible light source. The LED is used as the light source to transmit the data transmission is performed using the serial communication technique of the computer. The serial port communication is performed to and fro using the RS 232 pin. For the easy and proper IJEST Vol.12 No.2 July - December 2018 analysis we transmitted a data from one PC to another. For the easy processing and manipulation windows application was used. LED light is the main component in the transmitter section. Here LED is used to transfer the data. In this module the data is changed to binary by using the application and then transmitted. The output of the computer is taken using the serial communication port. As the modern computers and laptops have a serial output port, here used a USB to Serial port converter. The output of the computer is made to a constant output by using a MAX 232 IC. While the LASER is in ON condition, it's considered to be as 1 at the receiving end else 0. The driver circuit of the LASER is designed by using a BC 547.

#### 6. LIFI RECEIVER

The output of the transmitter is the input of the receiver. The PC is taken using the serial communication port communication. As the modern computers and laptops have a serial output port, here used a USB to Serial port converter. The output of the computer is made to a constant output by using a MAX 232 IC. While the LED is in ON condition, it's considered to be as 1 at the receiving end else 0. The driver circuit of the LED is designed by using a BC 547. The silicon photodiode is used as the receiver. The MAX 232 IC is used to convert the TTL logic to RS 232 logic. So that the received data will be transferred into the system for data processing.

#### 7. DATA INTEGRATION

Data Integration is the process of data verification from the LI FI receiver. It is the process of bridging system between the LIFI receiver and the application interface of the system. Initially all the received data will be the binary values and digital values, the received values will be converted into machine language. This makes the file visible in the window application. Initially a binary value will be sent form the receiver to the system the COMM port receives the data and coverts into the machine language. The machine language will be shown in the system as the readable format. The data integration deals with bits and bytes of data. The data transmission rate will be higher than the Wi-Fi networks. Due to the conversion process, more security has been enhanced.

#### 8. SOFTWARE DEVELOPMENT 8.1 Front End (ASP.NET)

The .NET Framework (pronounced dot net) is a software framework developed by Microsoft that runs

primarily on Microsoft Windows. It includes a large library and provides language interoperability (each language can use code written in other languages) across several programming languages. Programs written for the .NET Framework execute in a software environment (as contrasted to hardware environment), known as the Common Language Runtime (CLR), an application virtual machine that provides services such as security, memory management, and exception handling. The class library and the CLR together constitute the .NET Framework. The .NET Framework's Base Class Library provides user interface, data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications. Programmers produce software by combining their own source code with the .NET Framework and other libraries. The .NET Framework is intended to be used by most new applications created for the Windows platform. Microsoft also produces an integrated development environment largely for .NET software called Visual Studio.

# 8.2 Coding Language (C#)

Microsoft Visual C# is Microsoft's implementation of the C# specification, included in the Microsoft Visual Studio suite of products. It is based on the ECMA/ISO specification of the C# language, which Microsoft also created. While multiple implementations of the specification exist, Visual C# is by far the one most commonly used. Visual C# is also heavily used by ASP.NET web sites and standalone applications based on the .NET Framework. C# is a new programming language, very similar to Java. An extensive class library is included, featuring all the functionality one might expect from a contemporize development platform - windows GUI development (Windows Forms), database access (ADO.NET), web development (ASP.NET), web services, XML etc.,

# 8.3 Back End (SQL Server)

Microsoft SQL Server is a relational database management system developed by Microsoft Inc. As a database, it is a software product whose primary function is to store and retrieve data as requested by other software applications, be it those on the same computer or those running on another computer across a network (including the Internet). There are at least a dozen different editions of Microsoft SQL Server aimed at different audiences and for different workloads (ranging from small applications that store and retrieve data on the same computer, to millions of users and computers that access huge amounts of data from the Internet at the same time). Its primary query languages are T-SQL and ANSI SQL.

# 9. RESULT

The LED-to-LED communication provides a unique opportunity to provide communication capabilities that is not noticed. It is practically possible to transmit higher quantity of data along with high quality using visible light as a medium. We are transferring data via one computer to another in a room using VLC system. On the basis of transmission of data in high speed, not penetrating outside of wall, low cost of LED, economical and easily accessible to all and etc., this lot of features of this technology can be used to replace the existing RF based Wi-Fi system by the visible light technology(LiFi).

# REFERENCES

- [1] S.Meshram and A.Wadhe, "Analysis For Secure Data Transfer Using Visible Light Communication Technique", IJERT, April 2016.
- [2] P. Pathak, X. Feng, P. Hu and P. Mohapatra, "Visible Light Communication, Networking and Sensing: A Survey, Potential and Challenges", IEEE Communications Surveys & Tutorials 2015.
- [3] B. Li. J. Wang, Senior Member, IEEE R. Zhang, Member, IEEE H. Shen Chunming Zhao, Member, IEEE L. Hanzo, Fellow, IEEE, "Multiuser MISO Transceiver Design for Indoor Downlink Visible Light Communication Under Per-LED Optical Power Constraints," IEEE Photonics Journal Vol. 7, No. 4, August 2015.
- [4] Y. Li. M. Safari, R. Henderson, H. Haas, "Optical OFDM with Single-Photon," IEEE PHOTONICS TECHNOLOGY LETTERS, Vol. 27, No. 9, 1st May 2015.
- [5] W. Yuanquan and C. Nan, "A High-Speed Bi-Directional Visible Light Communication System Based on RGB-ED", China Communications, 2014.
- [6] A. Kurup, V. Tiwari, Selvanathiya, "Implementation and Demonstration of Li-Fi Technology", IJRET, March 2014.
- [7] Tokuda *et al*, "Speech Synthesis Based on Hidden Markov Models", Proceedings of the IEEE Vol. 101, No. 5, May 2013, pp.1234-1252.
- [8] S. Schmid, G. Corbellini, S. Mangold, T. R. Gross and Disney Research "An LED-to-LED Visible Light Communication System with Software-Based

C. Manipriya, A.S.Deepika, S. Indhumathi, S. DivyaPratheeba and S.Nithya

Synchronization", 3rd IEEE Workshop on Optical Wireless Communications (OWC'12), 2012.

- [9] K. Wetzel, "Speech-recognizing Computers: A Written Communication tool for Students with Learning disabilities? Journal of Learning Disabilities, 1996.
- [10] Rita Singh, Bhiksha Raj and Richard M. Stern, 2002,"Automatic Generation of Subword Units for Speech Recognition Systems".

# **Denial of Service Using Software Puzzle**

E.Anjana Devi<sup>1</sup>, S.Raj Kumar<sup>2</sup>, G.M.Madhu Mitha<sup>3</sup> and S.Ponmutharthan<sup>4</sup>

Department of Computer Science and Engineering, Dr.Mahalingam College of Engineering and Technology, Pollachi - 642 003, Tamil Nadu

#### Abstract

Denial of service (DoS) and Distributed DoS (DDoS) are among the foremost threats to cyber security and the client's enigma, which requires a customer to implement computationally exclusiveprocedures before receiving services from a server, is a countermeasure for them. In the last few years, different defense mechanisms based on puzzles have been suggested against DoS (Denial of Service) attacks on networks. However, these mechanisms were not designed through formal methodologies and, therefore, some chief design problems, such as effectiveness and optimality, remained vague. This article uses game theory to propose a series of optimal riddle-based stratagems to handle progressively erudite alluvial attack scenarios. In this way, the concept of Nash equilibrium solution is used in a recommended way, where the defender takes his part in the result as a positive defense against the coherent attackers. This study concludes in a scheme to manage attacks disseminated by an anonymous number of bases.

#### 1. INTRODUCTION

A DoS attack is categorised by malevolent conduct that prevents genuine users of a network service from using that amenity. These exhausts some vital resources in the victim system so that authentic user requests are rejected for the same time. A resource can be the proficiency of a buffer, the CPU time to process requests, etc. A counter-measure of DoS and DDoS is to increase the ratio of generating the client's computing costs or decreasing the server's. The client's enigma is a sure way to address the expansion of customer spending, as it encourages customers to complete sizeable transactions before they are allowed to administer. A riddle plan for the client consists of three phases: creating riddles, understanding the riddles from the client, and verifying the riddle from the server.

Denial of the distributed service has two types of architectures. The architecture of the agent administrator and the architecture of the Internet relay chat. The agent's administrator architecture consists of clients, agents, and controllers. In the main passage, the attacker creates a computer network by learningfrail systems. The attacker then installs the attack tools on all the systems in the network. The machines that run these tools are called handlers. These tackles also install the tools on other frail systems or hosts. In the Internet Relay Chat architecture, a passage is created to connect clients to agents. An example of this type of attack is the Low Orbit Ion Cannon (LOIC).

#### 2. RELATED WORKS 2.1 DoS Attack Inflating Capacity

DoS / DDoS are the main extortions to computer security with software for rapid resolution of puzzles and / or integrated GPU hardware to pointedlyabate the efficacy of the customer's puzzle. DoS / DDoS attackers doesn't allow the users from improving their puzzlesolving skills. In the end, the customer's puzzle is known as the software enigma. The attacks are made to exhaust resources such as network bandwidth, memory and computing power, overwhelming amenities. Hackers spend negligible time to direct requests to the server, but the server needs to spend more time processing on the HTTPS link protocol.



Fig.1 GPU-Inflated DoS attack against data puzzle

E.Anjana Devi, S.Raj Kumar, G.M.Madhu Mitha and S.Ponmutharthan

#### 2.2 GPU Integration

The existing system assumes that the virulent client solves the puzzle using only the legacy CPU resources, but it is not every time. Today, a new dispensation unit is used that uses graphics that have many Streaming Multiprocessor (SM) that consists of many identical processing cores. This GPU can be incorporated for free or assimilated with the CPU to bloat the capacity for computing resources. DoS attackers typically target spots or services applied on prominent web servers, such as banks, credit card payment gateways, and even root name servers.



#### 2.3 Puzzle Auctions

In the attack model, opponents cannot transform the packets between the client and the server. In the DoS attack, the server is, at least, able to reject incoming packets and send response messages to the origination of these packets. You can perform IP satirizing and can overhear on all the packets sent by the server. It also coordinates perfectly with the activities to make the most of the resources. The safekeeping of the recommended puzzle auction mechanism can be analyzed and the hash function will be considered unplanned. It is provided with the TCP protocol stack of the Linux kernel and also deals with the exhaustion of the connection.

#### 2.4 Client Puzzle Approach

Hackers has begun to use more genuine or anticipated protocols and services as vehicles for packet transmissions. The consequential attacks are difficult to defend against the use of customary techniques since the malicious requests differ from the legitimate lines in the intention, but not in the content.

STEP 1: C S: service request is been sent

STEP 2: S: generating a puzzleSTEP 3: S C: send the description of the puzzleSTEP 4: C: solve the puzzleSTEP 5: C S: sending solution to the puzzleSTEP 6: S: the solution is verifiedIf the solution is correctSTEP 7: S: the service request is continued to process

#### 2.5 Time Lock Puzzle Construction

The time blocking scheme was proposed in "Temporary blocking jigsaw and timed release encryption". The notion was to encode a message so that it cannot be deciphered until a predefined time has arrived. Time blocking puzzles requires comparatively extra time to solve. The puzzle was designed not to be parallelizable to prevent the client from providing addedpossessions in a short period of time. A riddle will be provided to the client and the riddle solution will be the key to deciphering the message.

In this approach, for a random challenge a, of difficulty l, the client will be asked for a computer  $b = a^{2} t \pmod{1}$ 

Where n is the product of two large prime numbers p and q. The client does not know the value of n, and therefore the only way to calculate the value of b is to perform the quadrature of a. Because of the nonparallelizable properties of time lock puzzles, it has been suggested as a mechanism to prevent DoS attacks. However, this tactic has the problem of making authentic customers wait superfluously and the puzzle can be parallel and solved using the GPU.

#### 2.6 Server Verification

Verification of the answer as described above, cookies are used to support the verification of the unceasing state of the answers of the puzzle. The verification appliance is the same in all types of puzzles. Customers must submit the solution to the puzzle along with a previously issued cookie that the server has attached to the puzzle. To verify the fix, the server fetches the timestamp from the server, catalogues it to the server to obtain the corresponding nonce, checks the termination time, performs a hash of the customer response with the nonce and compares it with the cookie. These tasks are simple, allowing the server to verify and prevent two types of attacks: those aimed at depleting resources and overshadowing bandwidth, respectively. In an out-of-

IJEST Vol.12 No.2 July - December 2018

resource attack, the attacker sends false requests to the server, which causes the server to bestow all its resources to processing spiteful requests and rejecting requests from authentic clients.

# **3. EXISTING SYSTEM**

A DoS attack is characterized by malevolent behavior that prevents authentic users of a network service from using that service. There are two main classes of these attacks: flood attacks and logical attacks. A flood attack such as SYN flood, Smurf or TFN2K sends a huge number of requests for a service offered by the victim. These requests exhaust some of the victim's key resources so that the authentic requests of the users of the victim are precluded. A resource can be the capacity of a buffer, the CPU time to process requests, the available bandwidth of a communication channel, etc.

A logical attack such as Ping-of-Death or Teardrop falsify a terminal message recognised and processed by the victim's defenceless software and leads to the enervation of the victim's resources. Unlike flood attacks, the effects of a logical attack remain after the attack until suitable corrective action is taken. In common, responsive mechanisms suffer from the problem of scalability and the difficulty of identifying attack traffic. The disadvantages of the existing system are flood attacks such as the SYN flood. The reactive mechanism suffers from the problem of scalability and the main problem is in the identification of the traffic attack.

# 4. PROPOSED SYSTEM

Here the game theory is used to propose a series of defenses based on enigmas against flood attacks. Collaborations between an attacker launching a flood attack and a defender that neutralizes the attack using a riddles defense can be modelled as an infinitely repeated game of discounted pay-outs. The solution concepts of this type of games are implemented to find the solutions, i.e. the best strategy that a cogent defender can take against a cogent aggressor. This way, finest puzzle-based defense strategies are developed. More specifically, four defense mechanisms are proposed. PDM1 derives from the concept of open-cycle solution in which the defender chooses his actions nonetheless of what happened in the game's history. This defense is based on the hypothesis that the defender knows the dimensions of the attack combination. Finally, in the PDM4, the last defense mechanism is proposed in which it is assumed that the size of the attack association is unknown.

# 5. ALGORITHM 5.1 Nash Equilibrium Algorithm

The zero-sum games are those in which P1 + (P2)T = 0. The balance in these games can be calculated in polynomial time by direct programming and the ensuing balance strategies and the expected pay-outs for the two players are size numbers. Polynomial A balance strategy in a zero-sum game can be viewed simultaneously as a defensive strategy that maximizes a player's expected pay against the opponent's less desirable strategy, and also an attack strategy that forces the opponent to his payoff minimum expected.

While an efficient zero-sum game algorithm seems to handle a set of zero possible game sizes, we can use this proficient reckoning as an important component of a general-sum game algorithm. At the highest level, we can quickly discover that the game is essentially a zerosum game or that there are strategies that both players can play to get a better average payment than what they can guarantee in the worst case, stabilized by the threat of being forced to pay this lower payment.



Fig.3 Proposed system block diagram



Fig.4 Ip verification from the client

1	-		-//		-			
-								
er Hannel (65 erment (82								
unt Dis	Population			contribution				
ana Na 🛛 🕬	1909-0173		12	2	32		Ar may	1mg
	Puzzla				RRA		Contraction of the second seco	Million .
ni Datalo		_		Fig.5 I	Puzzle so	olving		
er Datals IPAddens 192 168 1.15	Protocol Top	Message Reguestionica	Personnel Tane 6:37 PM	Fig.5 F	Puzzle so	Nonge Name Nonge Name Nonge	Instruction 8:12:15:14:33:35:1:23:48; 8:32:35:14:33:55:1:23:48;	
er Datals IPAddess 192 168 1 15	Protocol Yop	Message ReguesService	Received Taxe CS7 PM	Fig.5 F	Puzzle so	Nage Name Nage Name Na ang	Instantino 8.1215/14.032557.23.48 8.321574.032553.23.48	
er Detab PRoblem 1922 163 1 15	Protocol Top	Message RegistService	Received Taxe 6:07 PM	Fig.5 I	Puzzle so	Nage Name Nage Name Nage 30	Politika 8.521514.13351.2.348 8.521514.13351.2.348 8.321514.13351.2.248 3	Process Cinco
er Dotali (Philabere 192168115 geter Detalb	Postecod Yep	Message Requestions	Received Time 6:07 PM	Fig.5 I	Puzzle so	Nonge Name Nonge Name Nonge Name Nonge Name	Manement 1	IngenetesCourt
er Detals Phadeso 1921 Hall 19 gate Detals	Pedagod Top	Meesage PlagaetGervice	Received Taxe CEP PM	Fig.5 I	Puzzle so	No ang No	Notification 8.12,15,14,33,35,1,2,3,4,8, 8.12,15,14,33,35,1,2,3,4,8, 8.12,15,14,33,35,1,2,2,4,8, Manamari 3	Topented Court

#### 6. CONCLUSION

The project uses game theory to intend a series of defenses based on enigmas against flood attacks. Communications between an attacker hurling a flood attack and a defender that neutralizes the attack using a riddles defense can be modelled as an interminably repeated game of discounted pay-outs. Thus, the concepts of game solution are implemented, which is the best strategy that a coherent defender can take against a coherent aggressor. This way, optimal puzzlebased defense strategies are industrialized. This mechanism is applicable to defeat attacks distributed by a distinct source, but cannot claim that higher payments are realistic in the game. This defense is based on the hypothesis that the defender knows the dimensions of the attack federation. The mechanisms of the project can provide such defenses. On the other hand, estimates made by a reactive mechanism can be used to regulate the mechanisms.

#### REFERENCE

- Yongdong Wu,g Zhao;Feng Bao; Deng,"Software Puzzle: A Countermeasure to Resource-Inflated Denial of Service Attacks", IEEE Transactions on Information Forensics and Security, Vol.10, No.1, October 2015.
- [2] D. Moore, C. Shannon, D.J. Brown, G.M. Voelker, and S. Savage, "Inferring Internet Denial-of-Service Activity", ACM Trans.Computer Systems, Vol. 24, No.2, May 2013, pp. 115-139.
- [3] K.Iwai, N.Nishikawa and T.kurokawa, "Acceleration of AES Encryption on CUDA GPU", Int. J. Netw.Comput., Vol. 2, No.1, 2012, pp.131-145.
- [4] Vancha Maheshwar Reddy, B Sandhya Rani, J Vedika, Ch. Veera Reddy, "International Journal of Engineering Research and Applications (IJERA)", ISSN: 2248-9622 www.ijera.com Vol. 3, No.1, January -February 2013, pp.751-757.
- [5] R.Shankesi, O.Fatemieh and C.A. Gunter, "Resource Inflation Threats to Denial of Services Countermeasures", Dept. Computer Science, UIUC, Champaign, IL, USA, Tech Rep., Oct 2010.
- [6] Mehran S. Fallah, "A Puzzle-Based Defense Strategy Against Flooding Attacks Using Game Theory", IEEE Transactions on Dependable and Secure Computing, Vol.7, No.1, January- March 2010.

- [7] Mirkovic and P. Reiher, "A Taxonomy of DDoS Attacks and DDoS Defense Mechanisms", ACM SIGCOMM Computer Communication Rev., Vol. 34, No. 2, Apr. 2004, pp. 39-53.
- [8] A.Hussain, J. Heidemann and C. Papadopoulos, "A Framework for Classifying Denial of Service Attacks", Proc. ACM SIGCOMM '03, 2003, pp. 99-110.
- [9] A.R. Sharafat and M.S. Fallah, "A Framework for the Analysis of Denial of Service Attacks", The Computer J., Vol. 47, No. 2, Mar. 2004, pp. 179-192.
- [10] A. Juels and J. Brainard, "Client Puzzles: A Cryptographic Countermeasure Against Connection Depletion Attacks", in Proc. Netw. Distrib. Syst. Secur. Symp., 1999, pp.151-165.
- [11] X. Wang and M. K. Reiter, "Mitigating Bandwidth-Exhaustion Attacks using Congestion Puzzles", in Proc. 11<sup>th</sup> ACM Conf. Comput. Commun. Secur., 2004, pp. 257-267.
- [12] R. L. Rivest, A. Shamir and D. A. Wagner, "Time-Lock Puzzles and Timed-release Crypto", Dept. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. MIT/LCS/TR-684, Feb. 1996. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.110.5709.
- [13] NVIDIA CUDA. (Apr. 4, 2012). NVIDIA CUDA C Programming Guide, Version 4.2. [Online]. Available: http://developer.download.nvidia.com/
- [14] A. Juels and J. Brainard, "Client Puzzles: A Cryptographic Defence against Connection Depletion Attacks", In Proceedings of NDSS '99 (Networks and Distributed Systems Security), 1999, pp.151-165.
- [15] T. Aura, P. Nikander and J. Leiwo. "DoS-Resistant Authentication with Client Puzzles", Lecture Notes in Computer Science, Vol. 2133, 2011.
- [16] Q. Wu, S. Shiva, S. Roy, C. Ellis, V. Datla and D. Dasgupta, "On Modeling and Simulation of Game Theory-based Defense Mechanisms against DoS and DDoS Attacks", 43<sup>rd</sup> Annual Simulation Symposium (ANSS10), part of the 2010 Spring Simulation MultiConference, April 11-15, 2010.
- [17] W. Feng, E. Kaiser and A. Luu, "The Design and Implementation of Network Puzzles", Proc. 24<sup>th</sup> Ann. Joint Conf.IEEE Computer and Comm. Societies, 2005, pp. 2372-2382.

# Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques

V. Kaviya Shree, R.Kalaivani and K.Gowri Nayaki

Department of Computer Science and Engineering Kongu Engineering College, Erode - 638 060, Tamil Nadu

#### Abstract

Globalization has affected many countries [1]. There has been a drastic increase in the economic activities and consumption level, leading to expansion of travel and transportation. The increase in the vehicles, traffic lead to road accidents. Considering the importance of the road safety, government is trying to identify the causes of road accidents to reduce the accidents level [1]. The exponential increase in the accidents data is making it difficult to analyse the constraints causing the road accidents. The paper describes how to mine frequent patterns causing road accidents from collected data set [1]. We find associations among road accidents and predict the type of accidents for existing as well as for new roads. We make use of association and classification rules to discover the patterns between road accidents and as well as predict road accidents for new roads.

#### **1. INTRODUCTION**

There are a lot of vehicles driving on the roadway every day, and traffic accidents could happen at any time anywhere. Some accident involves fatality, means people die in that accident[2].As a human being we all want to avoid all these accidents. So it is necessary to find the root cause for all the accidents. Data mining technique could be used to find the reason on trained dataset to find out some valuable information.

Descriptive or predictive mining applied on previous road accidents data in combination with other important information as weather, speed limit or roadconditions creates an interesting alternative with potentially useful and helpful outcome for all involved stakeholders.

Association rule mining is used to analyse the previous data and obtain the patterns between road accidents [2]. The two criterion used for association rule mining are support and confidence. Apriori algorithm is one of the techniques to implement association rule mining [2]. In the proposed system, we use apriori algorithm to predict the patterns of road accidents by analyzing previous road accidents data [2].

To find out the association between independent variables in a huge data, there are number of data mining algorithms are available. Association rule mining algorithm is the most popular methodologies to detect the significant associations between the data stored in the large database [2]. There are a number of association rule mining algorithms available. Apriori and FP-growth

algorithm are the most common association rule mining methods to find out the association between various road traffic accident severity factors that influencing the traffic accident severity levels in India [2].

# 2. RELATED WORKS [2]

Jayasudha [4] analyzed the traffic accident using data mining technique that could possibly reduce the fatality rate. Using a road safety database enables to reduce the fatality by implementing road safety programs at local and national levels. Those database scheme which describes the road accident via roadway condition, person involved and other data would be useful for case evaluation, collecting additional evidences, settlement decision and subrogation. The International Road Traffic and Accident Database (IRTAD), GLOBESAFE, website for ARC networks are the best resources to collect accident data. Using web data a self-organizing map for pattern analysis was generated. It could classify information and provide warning as an audio or video. It was also identified that accident rates highest inintersections then other portion of road. [4].

The effect on road speed on accident in the state of Washington was investigated by Eric. [7] Some researches claim that those states which increased speed limit from 55mph to 65mph after 1974 had the fatality rate go up by 27% compared to increase in 10% in the states that did not increase the speed limit. It is claimed that as the effect of change in maximum speed is varies between urban and rural areas. After 1987 accident in rural areas increased while urban areas stayed relatively

IJEST Vol.12 No.2 July - December 2018

constant but clash rate in urban intersection is twice as high as in rural intersection. Accident is dependent on area (urban/rural), type of street (intersection, highway). [7]It is assumed that the fatality rate of an accidentmight be reduced with the introduction of an express emergency system. Reducing the time of delivery for emergency medical services (EMS) accident victimscould be treated in time saving their lives. Accident notification time, the difference between crash and EMS notification time is the most crucial. Trauma is time dependent disease and trauma victims could be saved if treated on time. Trauma victims could stabilized if treated soon. The first 10 minutes is called golden hour. The time was 12.3 minutes for the one who died and 8.4 minutes for the one who survived. The fatality rate was also much higher in rural areas because of unavailability of rapid EMS response in those areas. There are also several other factors affecting the fatality such as vehicle kilometers traveled, alcohol consumption, driver age distribution, accident notification time, personal income per capita and so on. [3] Solaiman et. al. [8] describes various ways accidentdata could be collected, placed in a centralized database server and visualized the accident. Data could be collected via different sources and the more the number of sources the better the result. This is because the data could be validate with respect to one another few could be discarded thus helping to clean up the data.

Different parameters such as junction type, collision type, location, month, time of occurrence, vehicle type could be visualized in a certain time strap to see thehow those parameters change and behave with respect to time. Based on those attributes one could also classify the type of accident. Using map API the system could be made more flexible such that it could find the safestand dangerous roads. [8] Partition based clustering and density based clusteringwere performed by Kumar [6] to group similar accidents together. Based on the categorical nature of most of the data K-modes algorithm was used. To find the correlation among various sets of attributes association rule miningwas performed. First the data set is classified into 6 clusters and each of them are studied to predict some patterns. Among the various rules that are generated those which seemed interesting were considered based on support count and confidence. The experiments showed that the accidents were dependent of location and most of the accident occurred in populated areas such as markets, hospitals, local colonies. Type of vehicle was also a factor to determine the nature of accident; two wheelers met with an accident

the most in intersections and involved two or more victims. Blind turn on road was the most crucial action responsible for those accidents and main duration of accidents were on morning time 4.00 a.m. to 6 a.m. on hills and 8 p.m. to 4 a.m. on other roads. [6] Krishnaveni and Hemalatha [5] worked with some classification models to predict the severity of injurythat occurred during traffic accidents. Naive Bayes Bayesian classifier, AdaBoostM1 Meta classifier, PARTRule classifier, J48 Decision Tree classifier, and Random Forest Tree classifier are compared for classifying the type of injury severity of various traffic accidents. The final result shows that the Random Forest outperforms the other four algorithms. [5] Amira, Pareek, [2] and Araar applied association rules mining algorithm on a dataset about traffic accidents which was gathered from Dubai Traffic Office, UAE. After information preprocessing, Apriori and Predictive Apriori association rules algorithms were applied to the dataset to investigate the connection between recorded accidents and factors to accident severity in Dubai. Two sets of class association rules were generated using the two algorithms and summarized to get the most interesting rules using technical measures. Exact results demonstrated that the class association rules created by Apriori algorithm were more viable than those created by Predictive Apriori algorithm. More relationship between accident factors and accident severity level were investigated while applying Apriori algorithm [2].

# 3. METHODOLOGY 3.1 Data Preparation

It was performed before each model construction.we have to find the dataset for this paper.All records with missing values in the chosen attributes were removed [2].Then we should convert the numeric values into nominal values according to the dataset.

Fatal rates were calculated to two categories. They are high and low. Formula for finding fatal rate is:

Fatal rate = Fatal / persons [2]

Where fatal is the number of fatalities and persons is the number of persons involved in the accident [2].

# 3.2 Data Modeling

To find the relationship among the attributes and the patterns we should applied Association rule mining or Naïve Bayes classification or k-means clustering [2].



# 3.2.1 Apriorl Algorithm

To generate the frequent itemset in the data set we use association rule mining algorithm.

# **Steps for Apriori Algorithm** [1]:

- 1 Scan the dataset and find the support of each itemset.
- 2 Generate the frequent itemset and set of candidate itemset.
- 3 Scan the candidate itemset and generate the support of each itemset.
- 4 Add to frequent itemset until it becomes null.
- 5 For each item in the frequent item set generate all non empty subsets.
- 6 For each non empty subset determine the confidence. If confidence is greater than or equal to specified confidence, then add to strong association rule.

# **Example for Strong Association Rule:**

Sl.No.	Items
1	A,C,D
2	A,C,E
3	A,B,C,E
4	B,E

Item set:A,B,C,D AND E. Minimum support=50% Minimum confidence=80%

This is the result obtained:

 ${B} > {E} {CE} > {A}$ 

IJEST Vol.12 No.2 July - December 2018

 ${AE} > {C}$  ${A} > {C}$  ${C}$  ${C} > {A}$ 

# 3.2.2 Naive Bayes Classification

Main purpose of Naïve Bayes is to classify the dataset and find the probability of each attributes. First we should scan the dataset. Then calculate the probability of each attribute value by classification method. Finally compare the values and classify the attribute values to one of the predefined set of class.

# 3.2.3 K-Means Clustering

Clustering algorithm was performed on the fatal accident dataset to find out which states are similar to each other and which states are safer or more risky to drive. To perform clustering total number of fatality rate was calculated.clustering should be done in three various steps. They are:

**Cluster 1:** It represents the safer state with relatively low fatal rate per million people.

Cluster 2: It represents the higher fatal rate.

**Cluster 3:** Here it has relatively large population but lower fatal rate. These are considered as safe region.

# **3.3 RESULT ANALYSIS**

The result of our analysis include association rule among the variables, clustering of states on their population and number of fatal accidents, and classification of the region as high or low risk of fatal accident [2]. Finally it shows the percentage of fatal accidents in these four attributes.

# 4. CONCLUSION

Current system is manual where government sector make use of ledger data and analyze the data manually, based on the analysis they will take precautionary measures to reduce the number of accidents [1].Proposed system uses road accident data to mine frequent patterns and important factors causing different type of accidents. It discovers the associations among road accidents using Apriori algorithm. It also predicts the common accidents that may cause for new roads with the help of Naïve Bayes algorithm and k-means clustering. From the clustering result we could see that some states/regions have higher fatal rate, while some others lower [2]. We may pay more attention when driving within those risky states/regions. Through the task performed, we realized that data seems never to be enough to make a strong decision. If more data, like non-fatal accident data, weather data, mileage data, and so on, are available, more test could be performed thus more suggestion could be made from the data [2].

# REFERENCES

- [1] Poojitha Shetty, P.C. Sachin, Supreeth V Kashyap and Venkatesh Madi, "Analysis of Road Accidents Using Data Mining Techniques", Dept. of Computer Science and Engineering, National Institute of Engineering, Karnataka, India.
- [2] Liling Li, Sharad Shrestha and Gongzhu Hu, "Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques", Department of Computer Science, Central Michigan University, USA.
- [3] William M Evanco, "The Potential Impact of Rural Mayday Systems on Vehicular Crash Fatalities", Accident Analysis &Prevention, Vol.31, No.5, September 1999, pp.455-462.
- [4] K Jayasudha and C Chandrasekar, "An Overview of Data Mining In Road Traffic and Accident Analysis", Journal of ComputerApplications, Vol.2, No.4, 2009, pp.32–37.
- [5] S. Krishnaveni and M. Hemalatha, "A Perspective Analysis of Traffic Accident Using Data Mining Techniques", InternationalJournal of Computer Applications, Vol.23, No.7, June 2011, pp.40-48.

- [6] Sachin Kumar and Durga Toshniwal, "Analysing Road Accident Data Using Association Rule Mining", In Proceedings of International Conference on Computing, Communication and Security, 2015, pp.1-6.
- [7] Eric M Ossiander and Peter Cummings. Freeway Speed Limits and Traffic Fatalities in Washington State. Accident Analysis &Prevention, Vol.34, No.1, 2002, pp.13-18.
- [8] KMA Solaiman, Md Mustafizur Rahman, and Nashid Shahriar, Avra Bangladesh Collection, Analysis & Visualization of Road Accident Data in Bangladesh", In Proceedings of International Conference on Informatics, Electronics & Vision, IEEE, 2013, pp.1-6.

# Precise Inspection of Mechanical Systems Based Android APP

# <sup>1</sup>B.Gokulnath<sup>1</sup>, G.Angusamy<sup>1</sup>, P.Prakash Kumar<sup>1</sup>, K.Hirushi Kesan<sup>1</sup>, R.Naveen Kumar<sup>1</sup>, S.Rahul<sup>1</sup>and V.Vignesh<sup>2</sup>

<sup>1</sup>Department of Mechanical Engineering,<sup>2</sup>Department of Information Technology, Sri Ramakrishna Engineering College, Coimbatore - 641 022, Tamil Nadu. E-mail: gokulnath.1504054@srec.ac.in, vignesh.venkataraman@srec.ac.in

#### Abstract

Inspection of mechanical system involves huge time and requires some special skilled personnel to inspect the data evaluated by the operators. Because the data in the mechanical system plays a vital role in quality and productivity of the product and ultimately results in the profit of an organisation. To make an inspection of mechanical systems precise and simpler, we have developed a android based process chart so that the data is evaluated by operator can be inspected by higher personnel at any time without the knowledge of the operator. So, the transfer of data from the machine to the personnel server takes place either by Wifi module / in terms of Bluetooth module. All data will be stored in firebase and also in file manager or SQL when there is no availability of Wifi or presence of Wifi Jammer. This saves the time and makes the inspection more precise.

# **1. PROCESS FLOWCHART**



#### 2. EXISTING SYSTEM

Existing system does not contain any kind of technique to read the data collected for inspection. It requires a special inspector to inspect the data to be evaluated by an operator and consumes huge time of the official and also inspection will not be precise as there is a chance of human error. This will affect the quality of the product, productivity, and customer satisfaction and workers loyalty towards organisation.

# **3. PROPOSED SYSTEM DESCRIPTION 3.1 Reading of Data**

Data of the mechanical systems are processed by either microcontroller, microprocessor, Arduino or other processing systems. The output of the processing systems is given to the LED display to know the data of the process. This output can also be given as input to the Wifi module which transfers the data simultaneously to the receiver.

# 3.2 APP Design

App will be designed in such a manner to act as an intermediate between Wifi module and data collecting system. The preferred tool is used to design app is Android studio because it is better than eclipse while considering its characteristics. Front End will be designed by XML and Back end will be designed by Java.

#### 3.3 Offline Storage System

The datas which got red by app and stored in mobile by creating a folder in File manager or SQL. This part is not mandatory and acts as an optional when the internet is available. In case if internet is not available or Jammer presence in organisation, this system gives the support.

#### 3.4 Cloud Storage System

Initially, personnel mail id is logged in and machine gets integrated with this mail id and works on the firebase mode. Datas from Wifi module of machine are collected and sends to this firebase. When the personnel login his mail id, he/she can able to view the data collected/ performed by the operator and precise inspection can be made by this data collection.

# 4. PROS OF THE PROCESS CHART

- Simultaneous Process
- Low time consumption
- High process speed
- Adaptive to any scale of industires/organisation
- Highly precise leads to high productivity and high quality
- No time delays
- Detailed data collection information
- Less expensive
- Highly secured

# 5. CONCLUSION

Highly precise inspection can be made in Mechanical systems with use of simple technique. Datas collected are highly secured and does not affect the confidentiality of the organisation. Adaptive measures can also be taken.

# **Predictive Analytics for Diabetes Mellitus Using Classification Models**

M. Munirathinam and K. Premalatha

Department of Computer Science and Engineering, Bannari Amman Institute of Technology, Sathyamangalam - 638 401, Erode District, Tamil Nadu Email : munirathinammrm187@gmail.com, kpl\_barath@yahoo.co.in

#### Abstract

Diabetes is one of the common and growing diseases in several countries and all of them are working to prevent this disease at early stage by predicting the symptoms of diabetes using several methods. The main aim of this study is to compare the performance of algorithms those are used to predict diabetes using data mining techniques. In this paper we compare machine learning classifiers (Random Forest, Boosting and, Support Vector Machines) to classify patients with diabetes mellitus. These approaches have been tested with data samples downloaded from UCI machine learning data repository. The performances of the algorithms have been measured in both the cases i.e dataset with noisy data (before pre-processing) and dataset set without noisy data (after pre-processing) and compared in terms of Accuracy, Sensitivity, and Specificity.

Keywords: Boosting, Data mining, Diabetes mellitus, Machine learning, Random Forest, SVM

# **1. INTRODUCTION**

Diabetes is one of the major health problems of all over the world. Diabetes mellitus is classified into four broad categories: type 1, type 2, gestational diabetes and other specific types. All forms of diabetes increase the risk of long-term complications. These typically develop after many years (10–20), but may be the first symptom in those who have otherwise not received a diagnosis before that time. The criteria for diagnosing diabetes in pregnancy have been given the World Health Organization (WHO) in 2006. The criteria are as follows,

- fasting plasma glucose e" 7.0 mmol/l (126 mg/ dl)
- 2-hour plasma glucose e" 11.1 mmol/l (200 mg/dl) following a 75g oral
- glucose load
- random plasma glucose e" 11.1 mmol/l (200 mg/ dl) in the presence of diabetes symptoms.

Diagnostic criteria for diabetes in non-pregnant individuals are based on the relationship between plasmaglucose values and the risk of diabetes-specific Micro vascular complications. People with diabetes have an increased risk of developing a number of serious health problems. Consistently high blood glucoselevels can lead to serious diseases affecting the heart and blood vessels, eyes, kidneys, nerves and teeth. Inaddition, people with diabetes also have a higher risk of developing infections. In almost all developedcountries, diabetes is a leading cause of cardiovascular disease, blindness, kidney failure, and lower limbamputation. Now it is very important to develop predictive models using the risk factors for thedevelopment of diabetes. Many studies have suggested traditional methods (statistical) as predictors. Data mining predicts the future by modelling. Predictive modelling is the process by which a model is created to predict an outcome. The data mining process for diagnosis of diabetes can be divided into five steps, though the underlying principles and techniques used for data mining diabetic data bases may differ for different projects in different countries. Data mining is one of the "Knowledge Discoveryin Databases" processes. The overall goal of the data mining process is to extract information from a dataset and transform it into an understandable structure for further use.

This process has become anincreasingly pervasive activity in all areas of medical science research. Data mining problems are oftensolved using different approaches from both computer sciences, such as multidimensional databases,machine learning, soft computing and data visualization; and statistics, including hypothesis testing, clustering, classification, and regression techniques. In recent years, data mining has been used widely in the areas of science and engineering, such as bioinformatics, genetics, medicine, and education.

# 2. PROPOSED WORK

The following algorithms are considered for our comparison analysis for prediction of diabetes. a. Random Forest

IJEST Vol.12 No.2 July - December 2018

# b. Boosting

c. Support Vector Machine

- a. Tree bagging
- b. From tree bagging to random forest

# 2.1 Random Forest

Random forest algorithm is the statistical and machine learning algorithm which uses multiple learning algorithms to obtain better predictive performance than others. This algorithm has twoparts



Fig.1 Random Forest Tree

Each tree is grown as follows:

- If the number of cases in the training set is N, sample N cases at random but withreplacement, from the original data. This sample will be the training set for growing thetree.
- If there are M input variables, a random number of attributes are selected and the bestsplit used to split the node. The value of M is held constant during the forest growing.
- Each tree is grown to the largest extent possible. There is no pruning.

# 2.2 Boosting

The term 'Boosting' refers to a family of algorithms which converts weak learner to strong learners. Boosting is a machine learning ensemblemeta-algorithm for primarily reducing bias, and also variance in supervised learning, and a family of machine learning algorithms which convert weak learners to strong ones.

# 2.2.1 Types of Boosting Algorithms

Underlying engine used for boosting algorithms can be anything. It can be decision stamp, marginmaximizing classification algorithm etc. There are many boosting algorithms which use other types of engine such as:

- AdaBoost (Adaptive Boosting)
- Gradient Tree Boosting
- XGBoost

# 2.3 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In this algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyperplanethat differentiates the two classes very well. Figure2 shows sample example.



Fig.2 Sample Example for finding the hyper-plane

In SVM, maximizing the distances between nearest data point (either class) and hyper-plane will help us to decide the right hyper-plane. This distance is called as Margin. In Figure 3, The margin for hyper-plane C is high as compared to both A and B. Hence, the right hyper-plane as C. Another lightning reason for selecting the hyper-plane with higher margin is robustness. If we select a hyper-plane having low margin, then there is high chance of miss-classification.



Fig.3 SVM hyper-plane identification

To construct an optimal hyperplane, SVM employs an iterative training algorithm, which is used to minimize an error function. According to the form of the error function, SVM models can be classified into four distinct groups:

- Classification SVM Type 1 (also known as C-SVM classification)
- Classification SVM Type 2 (also known as nu-SVM classification)
- Regression SVM Type 1 (also known as epsilon-SVM regression)
- Regression SVM Type 2 (also known as nu-SVM regression)

# 3. Comparison For Predictive Analytics Using Classifier Models

Figures 4-7 show the Precision/Recall Curve, Sensitivity/Specificity and Predicted vs Observed, Cost Curve Model for DM dataset.





Fig.5 Sensitivity/Specificity Curve









Fig. 7 Cost Curve

# 4. CONCLUSION

DM is rapidly emerging as one of the greatest global health challenges of the 21<sup>st</sup> century. To date, there is a significant work carried out in almost all aspects of DM research and especially biomarker identification and prediction-diagnosis. In summary, we have compared four prediction models for predicting diabetes mellitus using 8 important attributes under two different situations. One is before pre-processing the dataset. Random Forest performance much better than the other two classifiers and they provide 73% accuracy. From this we can come to know that after removing the noisy data from our dataset it will provide good result for our problems. This study can be used to select best classifier for predicting diabetes. In future we can use this type of study for any other diseases with their suitable data sets.

#### REFERENCES

- [1] U. Fayyad, G.Piatetsky-Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", AI Mag, Vol.17, 1996, pp.37-54.
- [2] J. Han, M. Kamber and J.Pei, "Data Mining: Concepts and Techniques", The Morgan Kaufmann Series in Data Management Systems, 2011.
- [3] CA. Mattmann, "Computing: A Vision for Data Science", http://dx.doi.org/10.1038/493473a, Nature 493, 24 Jan 2013, pp.473-475.
- [4] T. Mitchell, "Machine Learning", McGraw Hill 0-07-042807-7, 1997.
- [5] V. Marx, "Biology:the Big Challenges of Big Data". http://dx.doi.org/10.1038/498255a. Nature 498, 13 Jun 2013, pp.255-260.
- [6] Russell, Stuart and Norvig Peter, "Artificial Intelligence: A Modern Approach (2nd Ed.)", Prentice Hall. ISBN 978-0137903955, 2003, 1995.
- [7] RA. Wilson and FC. Keil, "The MIT Encyclopaedia of the Cognitive Sciences".
- [8] Diagnostic Criteria and Classification of Hyper Glycaemia First Detected in Pregnancy-WHO Publications,2013
- [9] The International Diabetes Federation (IDF) [Internet]. http://www.idf.org/complicationsdiabetes.
- [10] Encyclopaedia of Data Warehousing and Mining, Edited by John Wang- Idea Group Publishing, PCK Edition 2005.

# A Survey on Data Mining in Healthcare

#### T. GitanjaliSimran, D. Sujeetha and M.Sushmitha

Department of Computer Science and Engineering,

Bannari Amman Institute of Technology, Sathyamangalam - 638 401, Erode District, Tamil Nadu E-mail: gitanjalisimran.cs17@bitsathy.ac.in,sujeetha.cs17@bitsathy.ac.in,sushmitha.cs17@bitsathy.ac.in

#### Abstract

The knowledge discovery in database (KDD) is alarmed with development of methods and techniques for making use of data. One of the most important step of the KDD is the data mining. Data mining is the process of pattern discovery and extraction where huge amount of data is involved. Both the data mining and healthcare industry have emerged some of reliable early detection systems and other various healthcare related systems from the clinical and diagnosis data. In regard to this emerge, we have reviewed the various paper involved in this field in terms of method, algorithms and results. This review paper has consolidated the papers reviewed inline to the disciplines, model, tasks and methods. Results and evaluation methods are discussed for selected papers and a summary of the finding is presented to conclude the paper.

#### **1. INTRODUCTION**

Across all the fields, data are being collected and accumulated at a vivid pace. There is an urgent need for a new generation of computational theories and tools to assist humans in extracting useful information (knowledge) from the rapidly growing volumes of digital data. At the core of the process is the application of specific data mining methods for pattern discovery and extraction[1]. Among the data mining techniques developed in recent years, the data mining methods are including generalization, characterization, classification, clustering, association, evolution, pattern matching, data visualization and meta-rule guided mining. [2]. As an element of data mining technique research, this paper surveys the development of data mining technique, through a literature review and the classification of articles from 2005 until 2015 are reviewed. The period is important because, during the time period there is a newly widespread of data mining techniques being used in the healthcare industry where technology has played a significant role especially in the development of methodologies for the collection of data from online databases. The review interest for this literature review. started in the March 2015 with searches made of the keyword indices on the ScienceDirect, Springerlink and IEEE Xplore online databases, for full article containing the phrase "application of data mining techniques in healthcare". For the period from 2005 to 2015, 3840 articles were found. Topic filtering reduced this number to 205 articles, which were related to the keyword. From the 205 articles, 50 articles is used for this review, the papers are collected based on the phrase "application of data mining techniques in healthcare" in no specific categorization. The remaining part of the paper is organized as follows.

# 2. DATA MINING ALGORITHMS IN HEALTHCARE

Healthcare covers a detailed processes of the diagnosis, treatment and prevention of disease, injury and other physical and mental impairments in humans [15]. The healthcare industry in most countries are evolving at a rapid pace. The healthcare industry can be regarded as place with rich data as they generate massive amounts of data including electronic medical records, administrative reports and other benchmarking finding [16]. These healthcare data are however being underutilized. As discussed in 2.0 data mining is able to search for new and valuable information from these large volumes of data. Data mining in healthcare are being used mainly for predicting various diseases as well as in assisting for diagnosis for the doctors in making their clinical decision. The discussion on the various methods used in the healthcare industry are discussed as follows.

#### 2.1. Anomaly Detection

Anomaly detection is used in discovering the most significant changes in the data set [17]. Bo Lie et al [18] had used three different anomaly detection method, standard support vector data description, density-induced support vector data description and Gaussian mixture to evaluate the accuracy of the anomaly detection on uncertain dataset of liver disorder dataset which is obtained from UCI. The method is evaluated using the AUC accuracy. The results obtained for a balanced dataset by average was 93.59%. While the average standard deviation obtained from the same dataset is 2.63. The uncertain dataset are prone to be available in all datasets, the anomaly detection would be a good way to resolved this matter, however since there is only one paper discussing this method, we cannot comment much on the effectiveness of the method.

# 2.2. Clustering

The clustering is a common descriptive task in which one seeks to identify a finite set of categories or clusters to describe the data [17]. RuiVelosoa [19] had used the vector quantization method in clustering approach in predicting the readmissions in intensive medicine. The algorithms used in the vector quantization method are kmeans, k-mediods and x-means. The datasets used in this study were collected from patient's clinical process and laboratory results. The evaluation for each of the algorithms are conducted using the Davies-Bouldin Index. The k-means obtained the best results while x-means obtained a fair results while the k-mediods obtained the worst results. From the results the work by these researchers provide a useful result in helping to characterize the different types of patients having a higher probability to be readmitted. A more significant comparison on the method cannot be made since this is the only one paper in my review discussing on the vector quantization.

# 2.3. Classification

Classification is the discovery of a predictive learning function that classifies a data item into one of several predefined classes [17]. The related work in classification will be discussed in the following subsections.

# 2.3.1. Statistical

The MTS algorithm is being extensively applied in multivariable statistical analysis. The Mahalanobis distance (MD) is used to build statistical judgements to distinguish one group from another and the Mahalanobis space (MS) is used to represent the degree of abnormality of observations from the known reference group. In the statistical classifiers, the authors Su et al.[20], have used the Mahalanobis Taguchi System (MTS) to design the prediction model for pressure ulcers. The class imbalance problems are very much prevalent in the healthcare datasets. Usage of the data mining algorithms are often affected with skewed distribution when using skewed or imbalanced data sets. This problem often leads to the tendency of producting highly predictive classification accuracy over the majority class and poor accuracy over the minority class. Having such a nature to distinguish the degree of abnormality of observations, this method would be a good method to test on the real data set pressure ulcers. This method is also used since the MD is suitably scaled. The test conducted using this algorithms were done in four phases with scaled datasets ranging from 14 to 8, 5, and 2 accordingly. The results obtained in the paper [20] shows that the measurement scale for this algorithm has good a performance based on the huge difference between the normal and abnormal examples. Being an algorithm which is suitable for scaling the MTS proves to have better sensitivity and g-means values in the testing stage. The MTS has enhanced performance in terms of sensitivity.

# 2.3.2. Discriminant Analysis

Linear discriminant analysis (LDA) is widely used in discriminant analysis to predict the class based on a given set of measurements on new unlabeled observations [17]. AuthorsArma anzas et al. [21] and Jen et al [22] have used the linear discriminant analysis in their respective work. Jen et al [21] had the algorithm in predicting the severity staging of Parkinson's disease patient using scores of non-motor symptoms. Their study is intended to quantitatively analyze the inner relationships between both motor and non-motor symptoms. The linear discriminant analysis is the conditional probability density function of the predictors follows a normal distribution based on the given class value. The algorithm's ability to capture statistical dependencies among the predictor variables indicates that this algorithm would be suitable to explore the linear constraint of this study to discovery the synergy between motor and non- motor symptoms. The proposed model obtained an accuracy estimation of 69% compared to other algorithms since the algorithm's performance increases significantly when the dependencies are in linear form. Based on the same nature of the algorithm the authors Arma anzas et al. [21], used the algorithm to evaluate the classification accuracy to seek the most substantial risk factor and establish the initial set of substantial risk factors for chronic illness early warning. From the results of the two works we can safely say that the algorithm has good results and it is suitable to be utilized to identify significant accuracy if the relationships of the healthcare data are in linear form.

IJEST Vol.12 No.2 July - December 2018

#### 2.3.3. Decision Tree

Several study have explored the decision tree method to analyze clinical data. The authors Sharma & Om [23], Wang et al. [24] and Zolbanin et al.[25] have used the decision tree algorithm in their respective work. Having the nature to examine data and make the tree and its rules are used to make a prediction. All the three works have used the decision tree to the data set to improve the prognostic performance, in terms of accuracy. The nature of the data set used in this research are rather balanced set of data set. From the comparative of the works, we conclude that decision tree as cannot be used in proposing prognostic decision to solve imbalanced problems because the decision tree recursively separate observations into branches to construct a tree.

#### 2.3.4. Swarm Intelligence

The authors Yeh et al. [26], Fei 2010 [27], and Abdi&Giveki [28] have used the swarm intelligence method to designed their diagnosis model. The algorithm particle swarm optimization (PSO) is able to efficiently find the optimal or near optimal solutions in large search spaces. All the three authors tried to resolve optimzation problem which often involves features in the classification problems. The classification process will be faster and more accurate if less number of features are used. From the work studied, the PSO based approach proves to improve the overall classification results since PSO is being used to select suitable parameters in the involved classifiers.

#### 2.3.5. K-Nearest Neighbor

Authors Garc a-Laencina et al. [29], Arma anzas et al.[21], Jen et al.[22], Bagui et al.[30], and a ahan et al. [31] have used the k-nearest neighbour in their respective predictive models. Thek-nearest neighbour is an instance based classifier method. The parameter units consists of samples that are used in the method and this algorithm then assumes that all instances relate to the points in the *n*-dimensional space  $R^N$ . The algorithm is very expedient as the information in the training data is never lost. However, this algorithm would be suitable if the training data set is large as this algorithm is very time consuming when each of the sample in training set is processed while classifying a new data and this process requires a longer classification time. From the work by the mentioned authors, the classification accuracy is what they would like to attain instead of classification time as

the classification accuracy is more important in the medical diagnosis.

#### 2.3.6. Logistic Regression

Logistic regression (LR) is a method that would use the given set of features either continuos, discrete, or a mixture of both types and the binary target, the LR then computes a linear combination fo the inputs and passes through the logistic function [29]. This method is commonly used because it is easy to implementation and it provides competitive results. Authors Garc a-Laencina et al. [29], Mamiya et al. [32], Su et al. [20], Wang et al.[24], Zolbanin et al. [25], Thompson et al. [33], and Samanta et al. [34] have adopted the LR in their respective research work. The results obtained from all the authors are not very significant, due to the signifiant decrement in the size of the input data sets. The results would have been more significant if the datasets were large in quantity as the boundaries of accuracy would be larger. The LR works well for larger datasets.

#### 2.3.7. Bayesian Classifier

Authors Arma anzas et al. [21], and Bandyopadhyay et al.[35] have used the Bayesian classifier method in their respective predictive model. The Bayesian classifiers is well known for its computational efficient and ability to handle missing data naturally and efficiently. Having this advantage both the authors have recorded a good prediction accuracy from the models designed respectively. By having the models implemented the Bayesian classifier also proves that the model is suitable since the averaging approach has led to improved prediction accuracy and allows authors to extract more features from the data without being overfitting. This method would be a good approach if there data sets are suffering from missing data.

#### 2.3.8. Support Vector

The support vector method (SVM) is proven to be advantageous in handling classification tasks with execellent generalization performance. The method seeks to minimize the upper bound of the generalization error based on the structural risk minimization principle. The SVM training is equivalent to solve a linear constrained quadratic programming problem [36]. The method isvery commonly used in medical diagnosis. Authors Garc a-Laencina et al. [29], Zheng et al. [36], Kang et al. [37], and Su et al.[20] have used the method in their model in medical diagnoses. Some of the authors have used the SVM method for comparative study purpose. The SVM method generalization ability is controlled by two different factors, that is the training error and the capacity of the learning machine measured. The training error rate can be controlled by changing the features in the classifiers. From the results obtained from the studies, it clearly shows that the SVM showed greater performance since it maps the features to higher dimensional space.

# 3. DISCUSSION

From the papers reviewed and discussed, the data mining methods accuracy varies depending on the features of the data sets and the size of data set between the training and testing sets. The common characteristics among the healthcare data sets are highly imbalanced data sets, where by the majority and the minority classifier are not balanced resulting prediction erroneous when run by the classifiers. Another characteristics of healthcare data sets are the missing values. The sample size of the data is often seen as another characteristics as the data available are usually in small scale. There is no one suitable data mining method to resolve all this issues.

# 4. CONCLUSION

The data mining has played in an important role in healthcare industry, especially in predicting various types of diseases. The diagnosis is widely being used in predicting diseases, they are extensively used in medical diagnosing. In conclusion, there is no one data mining method to resolve the issues in the healthcare data sets. In order to obtain the highest accuracy among classifiers which is important in medical diagnosing with the characteristics of data being taken care, we need to design a hybrid model which could resolve the mentioned issues. Our future directions is to enhance the predictions using hybrid models.

### REFERENCES

- [1] R. Agrawal and G. Psaila, "Active Data Mining", Current, 1995, pp.3-8.
- [2] S. H. Liao, P. H. Chu and P. Y. Hsiao, "Data Mining Techniques and Applications-A Decade Review from 2000 to 2011", Expert Syst. Appl., Vol.39, No. 12, 2012, pp.11303-11311.
- [3] G. E. Vlahos, T. W. Ferratt and G. Knoepfle, "The Use of Computer-based Information Systems by

German Managers to Support Decision Making", Inf. Manag., Vol. 41, No.6, 2004, pp.763-779.

- [4] I. H. Witten, E. Frank and M. Hall, "Data Mining: Practical Machine Learning Tools and Techniques", (Google eBook). 2011.
- [5] D. K. Bhattacharyya and S.M. Hazarika, "Networks, Data Mining And Artificial Intelligence: Trends And Future Directions", 1<sup>st</sup> ed. Narosa Pub House, 2006.
- [6] M. Karegar, A. Isazadeh, F. Fartash, T. Saderi and A. H. Navin, "Data-Mining by Probability-Based Patterns", 2008, pp. 353-360.
- [7] H. Thomas and L. Paul, "Statistics: Methods and Applications", 1<sup>st</sup>ed. StatSoft, Inc, 2005.
- [8] M. Kantardzic, "Data Mining: Concepts, Models, Methods, and Algorithms", 2<sup>nd</sup> ed. Wiley-IEEE Press, 2011.
- [8] P. Berkhin, "A Survey of Clustering Data Mining", Group. Multidimens. Data, 2006, pp. 25-71.
- [9] T. P. Hong, K. Y. Lin and S. L. Wang, "Fuzzy Data Mining for Interesting Generalized Association Rules", Fuzzy Sets Syst., Vol. 138, No.2, 2003, pp. 255-269.
- [10] D. R. Hardoon, S. Sandor and S. John, "Canonical Correlation Analysis: An Overview with Application to Learning Methods", Fuzzy Data Mining for Interesting Generalized Association Rules J. Neural Comput., Vol.16, No. 12, 2004, pp. 2639-2664.
- [11] M. Chau, R. Cheng, B. Kao, and J. Ng, "UUncertain Data Mining: An Example in Clustering Location Data", Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect.Notes Bioinformatics), Vol. 3918 LNAI, 2006, pp. 199-204.
- [12] Z. Wu and C. Li, "L0-Constrained Regression for Data Mining", 2007, pp.981-988.
- [13] A. Genkin, D. D. Lewis and D. Madigan, "Large-Scale Bayesian Logistic Regression for Text Categorization", Technometrics, Vol. 49, No. 3, 2007, pp. 291-304.
- [14] J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang and H. Pan, "Emerging Information Technologies for Enhanced Healthcare", Comput. Ind., Vol. 69, 2015, pp.3-11.
- [15] N. Wickramasinghe, S. K. Sharma and J. N. D. Gupta, "Knowledge Management in Healthcare", Vol. 63, 2005, pp. 5-18.
- [16] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", AI Mag., 1996, pp.37-54.

- [17] B. Liu, Y. Xiao, L. Cao, Z. Hao and F. Deng, "SVDD-based Outlier Detection on Uncertain Data", Knowl. Inf. Syst., Vol.34, No.3, 2013, pp. 597-618.
- [18] R. Veloso, F. Portela, M. F. Santos, A. Silva, F. Rua, A. Abelha and J. Machado, "A Clustering Approach for Predicting Readmissions in Intensive Medicine", ProcediaTechnology, Vol.16, 2014, pp. 1307-1316.
- [19] C. T. Su, P. C. Wang, Y. C. Chen and L. F. Chen, "Data Mining Techniques for Assisting the Diagnosis of Pressure Ulcer Development in Surgical Patients", J. Med. Syst., Vol.36, No.4, 2012, pp.2387-2399.
- [20] R. Arma anzas, C. Bielza, K. R. Chaudhuri, P. Martinez-Martin and P. Larra aga, "Unveiling Relevant Non-Motor Parkinson's Disease Severity Symptoms Using A Machinelearning Approach", Artif. Intell. Med., Vol.58, No.3, 2013, pp.195-202.
- [21] C.H. Jen, C.C. Wang, B. C. Jiang, Y.-H. Chu and M.S. Chen, "Application of Classification Techniques on Development an Early-Warning System for Chronic Illnesses", Expert Syst. Appl., Vol.39, No.10, 2012, pp. 8852-8858.
- [22] N. Sharma and H. Om, "Data Mining Models for Predicting Oral Cancer Survivability", Netw. Model. Anal. Heal. Informatics Bioinforma., Vol. 2, No.4, 2013, pp. 285-295.
- [23] K.J.Wang, B.Makond and K.M. Wang, "An Improved Survivability Prognosis of Breast Cancer by Using Sampling and Feature Selection Technique To Solve Imbalanced Patient Classification Data", BMC Med. Inform. Decis. Mak., Vol.13, 2013, pp.124.
- [24] H. M. Zolbanin, D. Delen, and A. Hassan Zadeh, "Predicting Overall Survivability in Comorbidity of Cancers: A Data Mining Approach", Decis.Support Syst., Vol. 74, 2015, pp.150-161.
- [25] W.C. Yeh, W.W. Chang and Y. Y. Chung, "A New Hybrid Approach For Mining Breast Cancer Pattern Using Discrete Particle Swarm Optimization And Statistical Method", ExpertSyst. Appl., Vol. 36, No.4, 2009, pp. 8204-8211.
- [26] S. W. Fei, "Diagnostic Study On Arrhythmia Cordis Based On Particle Swarm Optimization-Based Support Vector Machine", Expert Syst. Appl., Vol. 37, No.10, 2010, pp.6748-6752.
- [27] M. J. Abdi and D. Giveki, "Automatic Detection Of Erythemato-Squamous Diseases Using PSO-

SVM based on Association Rules", Eng. Appl. Artif. Intell., Vol.26, No.1, 2013, pp. 603-608.

- [28] P. J. Garc a-Laencina, P. H. Abreu, M. H. Abreu, and N. Afonoso, "Missing Data Imputation on the 5-year Survival Prediction of Breast Cancer Patients With Unknown Discrete Values", Comput. Biol. Med., 2015, Vol.59, pp. 125-133.
- [29] S. C. Bagui, S. Bagui, K. Pal, and N. R. Pal, "Breast Cancer Detection Using Rank Nearest Neighbor Classi cation Rules", Vol.36, 2003, pp. 25-34.
- [30] S. Sahan, K. Polat, H. Kodaz, and S. Gunes, "A New Hybrid Method Based On Fuzzy-Artificial Immune System And K-Nn Algorithm for Breast Cancer Diagnosis", Comput. Biol.Med., Vol. 37, No.3, 2007, pp.415-423.
- [31] H. Mamiya, K. Schwartzman, A. Verma, C. Jauvin, M. Behr, and D. Buckeridge, "Towards Probabilistic Decision Support in Public Health Practice: Predictingrecent Transmission of Tuberculosis from Patient Attributes", J. Biomed. Inform., Vol.53, 2015, pp. 237-242.
- [32] V. L. S. Thompson, S. Lander, S. Xu, and C. Shyu, "Identifying Key Variables in African American Adherence to Colorectal Cancer Screening: The Application of Data Mining", 2014, pp. 1-10.
- [33] B. Samanta, G. L. Bird, M. Kuijpers, R. A. Zimmerman, G. P. Jarvik, G. Wernovsky, R.Clancy, D. J. Licht, J. W.Gaynor and C. Nataraj, "Prediction of Periventricular Leukomalacia. Part I: Selection of Hemodynamic Features Using Logistic Regression and Decision Tree Algorithms", Artif. Intell. Med., Vol.46, No.3, 2009, pp.201-215.
- [34] S. Bandyopadhyay, J. Wolfson, D. M. Vock, G. Vazquez-Benitez, G. Adomavicius, M. Elidrisi, P. E. Johnson and P. J.O'Connor, "Data Mining for Censored Time-To-Event Data: A Bayesian Network Model for Predicting Cardiovascular Risk from Electronic Health Record Data", 2014.
- [35] B. Zheng, S. W. Yoon and S. S. Lam, "Breast Cancer Diagnosis Based on Feature Extraction Using A Hybrid of K-means and Support Vector Machine Algorithms", Expert Syst. Appl., Vol.41, No. 4, Part 1, 2014, pp.1476-1482.

# Survey on Application of Image Processing

J.Gayathri and S.Ramya

Department of Computer Science and Engineering Bannari Amman Institute of Technology,Sathyamangalam - 638 401, Erode District, Tamil Nadu E-mail: gayathri.cs17@bitsathy.ac.in, ramya.cs17@bitsathy.ac.in

#### Abstract

The growth in the use of image processing software applications has been accompanied by a parallel increase in their use in criminal activities. Image processing tools have been associated with a variety of crimes, including counterfeiting of currency notes, cheques, as well as manipulation of important government documents, wills, ûnancial deeds or educational certiûcates. Thus, it is important for the Document Examiner to keep up to date with latest technological and scientiûc advances in the ûeld. The present research focuses on the use of image processing tools for the examination of computer-manipulated documents. The altered documents were examined using a suite of currently available image processing tools. The results demonstrate that a number of tools are capable of detecting computer-based manipulations of written documents.

# **1. INTRODUCTION**

The development of digital technology has drastically changed the concept of written document and, along with it, the nature and perpetration of white-collar crimes. Owing to revolution- ary changes made possible by new technology, forgers are becoming tech-savvy and are capable of fabricating documents with or without signatures using the skills and tools these mod- ern technologies afford. Forgery of documents, e.g., important government documents, wills and educational certiûcates have become easy to perfect. As a result of these advances in science and technology, it is becoming increasingly difúcult todetermine the authenticity of a ûle transferred through a net- work. Therefore, it is essential for the Forensic Document Examiners to keep abreast of the latest technological and scientiûc advances in the ûeld.

In most cases, forgers apply cut and paste techniques to manipulate ûgures, letters, or words that consequently alter the meaning of the document. Workers have reported the use of image processing techniques in forensic document exam- inations.1–5 The present study aims to explore the uses of image processing techniques for the examination of computer manipulated documents. The primary goal of this research is to study and characterize the various forms of alterations that have been found in a written document.

#### 2. MATERIALS AND METHODS

The study on the image processing techniques in forensic document examinations, this research was divided into two phases. Inphase-I, documents (in particular, those typically bused by for- gers) were selected. The original documents were ûrst scanned (Figs. A1–H1) and then, alterations were made using image processing software (Figs. A2–G2). In phase-II, image process- ing software was used to detect alterations in the manipulated documents.

#### 2.1. Collection of Samples

The documents collected for the study, included detail mark sheets, identity cards, bank cheques and documents consisting text on plain paper. Original samples were scanned using a col- our digital scanner.

#### 2.2. Manipulation of Scanned Samples

Forty manipulated samples were created out of four scanned copies of original samples. The manipulations included both addition and deletion of text from the original documents. The details of which can be found in Tables 1 and 2, respec- tively. The manipulations were done in multiple ways, as described below.

#### 2.2.1. Deleting Text using Paint Software

A scanned copy of the document was opened in the standard Paint software application. A portion of the

text was erased using the Erasure tool. To make the erased surface appear con- sistent with the rest of the surface, the Color Picker tool was used. The manipulated ûle was saved in JPEG format from the Paint software application.

# 2.2.2. Adding text using Paint and Adobe Photoshop Software

A scanned copy of the document was opened in the Paint soft- ware application. The Rectangular Selection tool was selected. Next, the Text Tool was selected in order to add new text. A Rectangular Selection appeared on the document where the desired ûgure or text was inserted and in the ûnal step, the altered document was saved in JPEG format. Alternatively, text could be added to the document using, Adobe Photoshop software. The document was edited using the copy and paste tool. Various words and letters in the document were copied and pasted at new locations within the samedocument. The Rectangular Marquee tool was selected from the editing window. The portion of the document to be copiedwas selected using Rectangular Marquee tool. Next, ' layer viacopy was selected and the Move tool was used to place the copied portion of the document at the desired location. After alladditions were made, the document was saved in JPEGformat. Using these procedures, various alterations were made on anumber of samples at different locations in the document. Then manipulations were carefully made so that they were difûcult to detect by the naked eye. All of the altered documents were saved in both PSD and JPEG format and then examined. The layers of the documents were merged using Merge Visiblecommand from the layer option in the main panel. As theresult of this process the background surfaces were typically disturbed. This was verified by converting all of the documents with printed backgrounds into Negative Image using Picasa editing software.

# 3. EXAMINATION OF THE ALTERED DOCUMENTS

The documents were thoroughly examined to check whether there were any irregularities such as:

(a) Irregular spacing between letters and words.

(b) Discrepancies in font and design of inserted words and letters.

- (c) Discrepancies in size of inserted letters or words.
- (d) Crowding of various letters and words.



Figure A (A-1) The original scanned document of detail mark sheet I. (A-2): The addition of text done on scanned document of detail mark sheet I. (A-3): The background disturbance in altered scanned copies of detail mark sheet I. (A-4) The background disturbance in document changed into negative form of detail mark sheet I.



Figure B (B-1): The original scanned document of bank cheque. (B-2): The removal of text done on scanned copies of bank cheque. (B-3): The background disturbance in altered scanned copies of bank cheque.



Figure C (C-1): The original scanned document of detail mark sheet II. (C-2): The addition of text done on scanned copies of detail mark sheet II. (C-3): The change in text, size and space in altered document of detail mark sheet II.



Figure D (D-1): The original scanned document of detail mark sheet III. (D-2): The addition of text done on scanned copies of detail mark sheet III. (D-3): The

presence of dots in the background surface of altered documents of detail mark sheet III.

The documents were examined for these irregularities using image processing tools, described below

#### 3.1 Use of Picasa Photo Editing Software

A scanned document was opened in the Picasa application. The resolution was increased using the Cropping tool to mag-nify the image approximately 100•. The Highlighting tool was applied twice in order to highlight the background and make it more prominent and visible. By doing so, the disturbance in the background could be observed. The background surfaces were examined carefully; in general, no matter how digital document manipulations are made, the background surfaces always get disturbed.

#### 3.2. Use of Adobe Photoshop

The number of layers in the document was easy to determine if PSD, as opposed to JPEG, ûles were opened in Adobe Photoshop. On the other hand, if the altered ûle was saved as a JPEG from Adobe Photoshop, information about the number of layers applied was lost and no layer detection could be performed.

# 4. RESULTS

The altered documents were examined using both Picasa andAdobe Photoshop software. The following observations have been made:

(a) In the case of text additions, disturbances in the background surface of the document were observed for documents with printed backgrounds such as detailed mark sheets, cheques, and identity cards (Figs. A-3, A-4, B-3). Moreover, discrepancies in the text and font of altered words were observed (Fig. C-3).

(b) Irregular spacing between letters and words was observed both in the case of additions and deletions of text in the scanned document's body (Fig. C-3).

(c) In the case of text additions, discrepancies in the size of various words and letters were observed in the newly added text (Fig. C-3).

(d) Disturbances in the form of small dots on the background surface of the document were additionally observed (Fig. D-3). For example, using the copy and paste technique in Adobe Photoshop to add text resulted.

Document manipulation. Two types of document manipulation, or forgory, implicate com Bocumants stored in electronic form can be forged by changing or copying the original dox Technological developments. Rehigh-quality home laser printers, also make forgery more computer expertise is required to create counterfelt checks, involces, letterhead and other with or without originals.

> Fig - E -1 The original scanned document.

Document manipulation. Two types of document manipulation, or forgery, implicate comp Bocuments stored in electronic form can be forged by changing or copying the original doc Technological developments. It high-quality home laser printers, also make forgery more computer expertise is required to create counterfelt checks, involces, latterhead and other t with or without originals, manipulation, home taser printers

Document manipulation. Two types of document manipulation, or forgery, implicate com Bocuments stored in electronic form can be forged by changing or copying the original do Technological developments. Ike high-quality home laser printers, also make forgery mon computer expertise is required to create counterfelt checks, involces, latterhead and other with or without originals, manipulation, home laser printers

The circle wit le pattern of the Dots which donot join with the above line when copied

Figure E (E-1): The original scanned document of plain paper. (E-2): The addition of words in scanned document of plain paper. (E-3): The circle with the arrow shows the unique pattern of dots which do not join with the above line when are copied in the scanned document of plain paper.

Figure F (F-1): The original scanned document of identity card. (F-2): The removal of text done on scanned copies of identity cards. (F-3): Removal of some part of the background surface in white colour of altered documents of identity cardsin disturbances in the patterns of the joined text and the background surface. These disturbances arose because the background surface was copied and pasted along with the added text. This feature, in particular, proved to be very useful as forgers are possibly not familiar with it. These observations were made on both types of documents, having either printed or plain backgrounds (Fig. E-3).

(e) In the case of certificates, the watermarks present in the background of the certificate can be changed or forged; however, it was observed that text written on thewatermark cannot be removed. Thus, when the



#### J.Gayathri and S. Ramya

document-bearing text on the watermark was examined in Adobe Photoshop, it always showed a single layer. The forger cannot paste new text onto the background of the watermark. When such an alteration was made, some loss in background color was observed. Moreover, distur- bances in the sequence of pixels were observed around the edges of the alterations (Figs. F-3, G-3, G-4).

(f) In altered documents, copied letters exhibited different pixel colors as compared to other letter forms of the same document when examined at 100• magniûcation

Figure G (G-1): The original scanned document of detail mark sheet II. (G-2): The addition of text on scanned document of detail mark sheet II. (G-3): Removal of some part of the background surface in white colour of altered document of detail mark sheet II. (G-4): Removal of some part of the background surface in white colour of altered document of detail mark sheet II. If an altered document was printed out and rescanned, it was observed that all letters (both original and altered) had the same pixel color.



(g) Plain or white backgrounds exhibited pale colors follow- ing deletion. (Figs.F-3, G-3, G-4). Furthermore, it was observed that pasting text resulted in irregularities in the background.

(h) If the forger saves the altered ûle in PSD format, without ûrst merging the layers then the layers are visible when the ûle is reopened in Adobe Photoshop. Manipulateddocuments that have printed backgrounds can be chan-ged from positive to negative form in Adobe Photoshop. The area containing additions or deletions typically appears highlighted and is easily visible. On the other hand, when a ûle is altered in Adobe Photoshop and saved in any format other than PSD, or as a PSD ûle after merging the layers, then the layer information is lost and it would not be possible to detect manipulations by examining layers.

# 5. DISCUSSION

Here, it was shown that there is utility in using image processing software to analyze digitally manipulated documents to detect manipulations. The results demonstrate that many features were associated with manipulations in digital documents, such as disturbances of the surface of the document, size and spacing discrepancies of words and letters, the presence of dots, irregularities in the pixel arrangements, changes in pixel colors of altered letters and words, as well as the presence of unique sequence of dots around each inserted letter or word. Earlier studies3,5 reported the use of image processing tools for the analysis of questionable documents. In the present study, the results obtained after examination of all of the docu- ments were encouraging and suggest ways to help document examiners in detecting manipulations of digital documents.

# 6. CONCLUSION

In this era of cybercrime, computers are used most frequently to commit certain crimes, such as altering documents to change their meaning. In this paper, we studied alternations in system-generated documents, including documents with printed backgrounds as well as plain backgrounds. The manipulation of the documents was done with the help of image processing software applications including Adobe Photoshop and Paint. After manipulating the documents, they were exam- ined for alterations; encouraging results were obtained. The results demonstrated that a number of features were associated with image manipulation and could easily be detected using standard image processing applications.

# REFERENCES

- [1] RE Gibson, "Obliteration Decipherment Using Computer Scanner Technology", Int. J.Forensic Doc Examiners, Vol.2, 1996, pp.326-332.
- [2] MOGaudreau, DC Purdy, JS Harris, "Where Document Imaging and Scientific Image Analysis Meet: Document Evidence", Int.J Forensic Doc Examiners, Vol.3, 1997, pp.261-4.
- [3] AF Hicks, "Computer Imaging for Questioned Document Examiners II: the Potential for Abuse. J Forensic SciJFSCA, Vol.40, 1995, pp.1052-4
- [4] OP Jasuja, VK Garg, "Deciphering Obliterated Writings: A Computer Based Simple Method", Int. J Forensic Doc Examiners, Vol.5, 1993, pp.270-9.
- [5] MC Joshi, A. Kumar and S. Thakur, "Examination of Digitally Manipulated-Machine Generated Document, A Case Study Elucidating the issue of Such Unwanted Progenies of Modern Technology", ProbForensic Sci., Vol.56, 2011, pp.162-73.

# **Efficient Content Based Image Retrieval**

S.Ramya, R.Prathiba, S.Raja Ganesh and PCD.Kalaivaani

Department of Computer Science and Engineering Kongu Engineering CollegeErode,Tamilnadu E-mail: ramyacs30@gmail.com,rprathiprathiba@gmail.com,rajaganesh1357@gmail.com, kalaivaani.ramesh@gmail.com

#### Abstract

The paper presents an efficient Content Based Image Retrieval (CBIR) system using color, shape and texture. Image search is a specialized data search used to find images. Image retrieval based on color, texture and shape is a wide area of research scope.Color Texture and shape are the low level image descriptor in Content Based Image Retrieval. These low level image descriptors are used for image representation and retrieval in CBIR. Combining the color, texture and shape feature leads to a higher retrieval efficiency.Here K-means algorithm is used to cluster the images into different clusters and the similarity measure is taken only with the images in the same cluster. The retrieval is more accurate and time taken is less.

#### **1. INTRODUCTION**

Digital Image Processing involves the modification of digital data for improving the image qualities with the aid of computer. The processing helps in maximizing clarity, sharpness and details of features of interest towards information extraction and further analysis. The digital image sources range from commercial earth resources satellites, airborne scanner, airborne solid-state camera, scanning micro-densitometer to high-resolution video camera. Digital image processing is a broad subject and often involves procedures which can be mathematically complex, but central idea behind digital image processing is quite simple. The digital image is fed into a computer and computer is programmed to manipulate these data using an equation, or series of equations and then store the results of the computation for each pixel (picture element). These results form a new digital image that may be displayed or recorded in pictorial format or may itself be further manipulated by additional computer programs. The possible forms of the digital image manipulation is literally infinite. The challenge in content based retrieval is to represent each image in a unique way to make accurate identification of the image. Therefore, the key to successful retrieval system depends on choosing the right features to accurately represent the image and the size of feature vector. Then distance measure between the query image and the images in that class is computed to retrieve the most similar images. This two tier approach reduces retrieval time and increase the accuracy.

#### 2. IMAGE RETRIEVAL

Here, the query image is selected from the database and its features are extracted. Similarly the features of database images are extracted. A comparison is made by using the K-means clustering algorithm where the distance measure between the query image and the images in the database is calculated and most similar images are retrieved and displayed to the user. Accuracy depends on the characteristics of the object. There are two types of features: global features and local features. Global features include color, texture and shape features of the whole image. Global representation does not give information about spatial location of the pixels. To overcome the limitation of the global representation the images are segmented into set of regions or the features are extracted from the sub images, which is called local image representation. The presented paper includes global and local feature extraction methods.

#### **3. RELATED WORK**

Most of the search engines (ex. google, yahoo, etc.,) are based on a semantic search, i.e., the user types in a series of keywords and the images are also annotated using keywords. Thus the match is done primarily through these keywords. In the recent years CBIR system have been developed to handle the large image database effectively. Basically color, texture and shape have been used for extracting similar images from an image database. Different CBIR techniques have used global color and texture features where as some have used local colors and texture features .After that the method of segmentation is proposed, where the image is segmented into regions based on color and texture features. And then region to region similarity is done. The regions are closed to human perception and are used as basic building blocks of computation of feature and similarity measurement. These systems are known as region based image retrieval. But image segmentation algorithm is difficult in the human perception point of view. To ensure robustness against such inaccurate segmentation region maintaining algorithm is used. Color space is divided into small ranges. Each interval is regarded as a bin. Then matching techniques are carried out.

Shape is another important feature for perceptual object recognition. Various applications proposed techniques including the chain codes, characteristic, circumference, area and circular degree. High rate of computation is observed here. Combination of edge histogram and Fourier transforms for computing edge image is proposed in literature. Gradient method for shape feature extraction and retrieval of similar image from image database is reported in literature. The literature proposed the gradient method for extraction of shape features from the image and to retrieve the similar image from the database. This paper uses color, texture and shape information for image retrieval to enhance the image retrieval results to a better efficiency.

# 4. K-MEANS CLUSTERING ALGORITHM

Clustering is a method to divide a set of data into specific number of groups. In k-means clustering, it partitions a collection of data into a k number group of data. It gives a given set of data into k number of disjoint cluster. K-means algorithm consists of two separate phases. In the first phase it calculates k centroid and in second phase it takes each point to cluster which has nearest centroid from respective data point. Distance is calculated using Euclidean distance. Once the grouping is done it recalculates the new centroid of each cluster and based on that centroid, a new Euclidean distance is calculated between each cluster and each data point and assigns the points in cluster which have minimum distance. So K-means is an iterative algorithm in which it minimizes sum of distances from each object to its cluster centroid, over all clusters.

Let us consider an image with resolution of  $x^*y$  and image has to be clustered into k number of cluster centers.

#### 5. ALGORITHM

- Initialize number of cluster k and Centre.
- For each pixel of an image, calculate Euclidian distance d, between centre and each pixel of an

image using

$$D=p(x,y)-ck$$

- Assign all the pixels to the nearest centre based on distance d.
- After all pixels have been assigned, recalculate new position of the center.

#### 6. RESULTS

This work was implemented in Microsoft Visual Studio 2008 on Intel(R) Core(TM) i5 @1.90GHz, 4GB RAM, 64 bit Windows 8 Operating System. Wang image database of 1000 images spread across 10 classes, each class with 100 images in the same class is used for the study of the CBIR system. Images are of resolution 256\*256 and they are converted into grayscale and binary image before processing.

# 7. CONCLUSION

On the basis of previous researches, the paper explored low-level features of color, texture and shape extraction of CBIR. After comparing the CBIR based on color, texture and shape features with that of the color ,texture and shape fused features, it is observed that CBIR based on color, texture and shape fused features provides better results i.e.results of color, texture and shape fused features are robust than the color, texture and shape features based image retrieval. The main aim of the project to enhance the CBIR is satisfied in this project. S.Ramya, R.Prathiba, S.Raja Ganesh and PCD.Kalaivaani

# REFERENCES

- [1] Datta, Ritendra, et al. "Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Computing Surveys (CSUR), Vol.40, No.2, 2008. pp-5.
- [2] ElAlami, M. Esmel. "A New Matching Strategy For Content Based Image Retrieval System", Applied Soft Computing, Vol.14, 2014, pp.407-418.
- [3] Koheri Arai and Ali RidhoBarakbah, Heirarchical K-means: An Algorithm for Centroid Initialization for K-means", Saga University, 2007.
- [4] PallaviPurohit and Ritesh Joshi, "A New Efficient Approach towards k-means Clustering Algorithm", International Journal of Computer Applications, (0975-8887), Vol. 65, No.11, March 2013.
- [5] Jun Yue, Zhenbo Li, Lu Liu, Zetian Fu., "Content based Image Retrieval Using Color and Texture Fused Features", 1121-1127, 2011, Elsevier.
- [6] Alan Jose, S. Ravi and M. Sambath, "Brain Tumor Segmentation using K-means Clustering and Fuzzy C-means Algorithm and its Area Calculation", International Journal of Innovative Research in Computer and Communication Engineering, Vol. 2, No.2, March 2014.

# **Question Paper Generation System**

# S.C. Lavanya, K.K. Aarthi, R. Devaprithika, S. Durgapiriyadharshini and K. Dinesh kumar

Department of Computer Science and Engineering,

Dr.Mahalingam College of Engineering and Technology, Pollachi, Coimbatore - 642 003, Tamil Nadu E-mail: lavanyack@drmcet.ac.in, aarthikoventhan@gmail.com,devaprithika@gmail.com

#### Abstract

Examination process is an important activity for educational institutions to assessstudent performance. The nature of the exam questions would determine the quality of the students produced by the institutions. Preparing the exam questions is time consuming. Thus with the help of Automatic question Paper Generator System which makes use of a shuffling algorithm and string matching techniques. The user needs to specify the Semester or CCET or unit test and from the entered input, the examination paper will be generated automatically. Keywords- Question paper generator, Randomization, shuffling algorithm, Course outcomes, Blooms Taxonomy.

# **1. INTRODUCTION**

Examination is a tool to qualify the ability and understanding level of leaners. Evaluation may be done on question and answering system based on different taxonomy. Generating an effective question paper is a task of great importance for any educational institution. In traditional method, faculty manually prepare question paper, which is very tedious and challenging. This Automated Question Paper Generator System (AQPGS) make use of Shuffling algorithm and string matching techniques.

In this proposed system, an automated process of Question Paper Generation is fast, streamlined, randomized and secure. Every task performed by this system is automated, so that storage space, bias and security are not a concern anymore. In this Automatic question Paper Generator System, a new algorithm which ensures total randomization of questions and avoids repetitions is used. The proposed system can be helpful to many educational institutes and NGO based institutes.

#### 2. RELATED WORK

The question paper is generated automatically using two algorithms where the questions are stored in the database with limited and normalised row in a table using apriori algorithm which can fetch data in large database in bottom-up approach.

Role Based Access Control Model (RBAC) is suitable for proposed system. RBAC involves implementation of user hierarchy by assigning the access control based on senior and junior roles. Activity log keep track of small changes done with the record and can be retrieved. Auto compute list will be helpful in finding questions in Database management system which avoid duplication. JSP, JAVA Servlet are implemented in front end, Oracle/Mysql is used in back end and Apache Tomcat server used for connectivity. After signing up, it will store parameter like time, date, staffId.

The question are selected randomly according to specific requirement and conditions. In the combination of manual and intelligent paper combination. After selecting the question, number of question and difficulty level the percentage of marks should be calculated by Score Distribution Algorithm. To generate question automatically, first select number of question set, secondly the number of questions in each set.

AQPG is based on fuzzy model which will be implemented using Microsoft visual studio Microsoft SQL server as a database management system. The system will have registration and login functionalities. After registration, the credentials provided by the user will be verified by the admin. If user is authenticate then admin will make his username and password valid and access will be granted. While registration user can select maximum four subjects for which he wants to set the paper. Admin has rights to change the subjects of a particular user.

The main part of the shuffling algorithms is to provide randomization technique in question paper generation system, thus different sets of question could be generated without repetition and duplication. Shuffling algorithms is a suitable and very effective way to implement for IJEST Vol.12 No.2 July-December 2018 describes grammatical rules to classify and analyze written programming exam questions through natural language processing and randomization of stored questions. The nature of this algorithm is as followed, for a set of N (the total number of question in the database) elements for generating a random alteration of the numbers 1–N.

Programming Exam Questions Classification Based on Bloom's Taxonomy describes grammatical rules to classify and analyse written programming exam questions through natural language processing. In which question are pre-processed to simpler form based on rule-based system.

# **3. EXISTING SYSTEM**

The automated question generated system generates question based on the randomization technique using shuffling algorithm. It keep track of generated question to avoid repetition of questions from one paper to another and also provide multiple sets of question paper based on user need. The initial database is created with question and their weightage along with unit details. Once database created, question can be generated automatically based on user suggestion such as for CCET, Semester or Quiz paper. The system can be modified for other exams also. This system enables the generation of QP automatically. QP generation includes selection of several attributes like CO, questions and marks. QB or material repository is maintained where QP is generated out of existing DB.

#### 4. PROPOSED SYSTEM

The purpose of test and question is to evaluate leaner performance. It is based on Blooms taxonomy using rule based system. There are six proposed blooms level. They are Knowledge, Comprehension, Application, Analysis, synthesis and Evaluation. The Bloom's level for the questions is generated automatically by string matching. The question is pre-processed and transform to simpler form and given as input. By using POS tagging, each word is identified grammatically. Using shallow phrasing the phrase is identified using POS tagging can be grouped and matched with the keywords in the bloom's taxonomy. To simplify the job instead of shallow phrasing use partial phrase syntax to identify the questions level easily.



Block diagram for question paper generator

#### 5. RESULT AND DISCUSSION

The Proposed System enables to automate question paper generation process with less time and effort. Question paper is generated with great ease, accuracy and keeps record of generated question paper. Question paper can be generated and viewed only by admin and password security is provided. Full customization of test paper institute name, time, instructions and maximum marks are provided. This software provides unique user authentication facilities. Database backup and restore backup facilities are also provided. Faculty login with password and selects marks, CO's and number of questions in each division. With the selection of these criteria, QP is generated automatically using Shuffling algorithm without repetition of questions.

The efficiency of the system is computed using the evaluation measure called Accuracy. Accuracy of the system should be high enough to make a correct prediction on the prediction system.

# 6. CONCLUSION

The creation of question paper is based on the various criteria by using shuffling algorithm will generate question based on the specified needs automatically. The QP is generated without repetition of questions, number of question paper set required, blooms level for generated questions and will be generated in a required format. The DB can be included with the string/verb matching technique and NLP technique for the automatic generation of Bloom's level in the DB for the respective questions. This system overcomes the complexity in generation of question paper with time saver.

### REFERENCES

- [1] Ashok Immanuel and B. Tulasi, "Framework for Automatic Examination Paper Generator System", IJCST, Vol.6, No.1, March 2015.
- [2] Prof. AlkaLeekha, "Automatic Question Paper Generator System", IJSRET, Vol.4, No.6, April 2017.
- [3] Mr. Amit Sanjay Khairnar, Mr. Bhagwat Chintaman Jadhav, Mr. Rahul Birhade and Mr. Pramod Patil "Automatic Question Paper Generator System", IJTRE, Vol.4, No.9, May 2017.
- [4] Amrutaumardand and Ashwinigaikwed, "A Survey on Automatic Question Paper Generator System", (IARJSET) Vol.4, No.4, Jan 2017.
- [5] Mrs. AshaRawat, UrvashiBhadarka and Manish Patil, "Question Paper Generator System", (IJTRA), Vol.5, No.3, May2017.

- [6] DivyaKarekar and AishwaryaChavan, "Automated Question Paper Generator System using Apriori Algorithm and Fuzzy Logic", IJIRST, Vol.2, No.11, April 2016.
- [7] FenilkiranGangar and WitalGopalGori, "Automatic Question Paper Generator System", IJCA, Vol.166, May 2017.
- [8] KapilNaik, Shreyassule and Shrutijadhav, "Automatic Question Paper Generation System using Randomization Algorithm", IJETR, Vol.2, No.12, Dec 2014.
- [9] Mayur Joshi, SiddheshDoiphode, "A Survey on Question Paper Generation System", IJCA, Vol.2, April 2016.
- [10] Pratik Gadge and Ravikntvishwakarma, "Advanced Question Paper Generator using Fuzzy Logic", IRJET, Vol.1, May 2016.
- [11] RohanBhirangi and SmitaBhoir, "Automated Question Paper Generation System", IJERMT, 2016.

# Indian Journal of Engineering, Science, and Technology (IJEST)

(ISSN: 0973-6255)

(A half-yearly refereed research journal)

#### **Information for Authors**

- All papers should be addressed to The Editor-in-Chief, Indian Journal of Engineering, Science, and Technology (IJEST), Bannari Amman Institute of Technology, Sathyamangalam - 638 401, Erode District, Tamil Nadu, India.
- 2. Two copies of manuscript along with soft copy are to be sent.
- A CD-ROM containing the text, figures and tables should separately be sent along with the hard copies.
- 4. Submission of a manuscript implies that : (i) The work described has not been published before;(ii) It is not under consideration for publication elsewhere.
- 5. Manuscript will be reviewed by experts in the corresponding research area, and their recommendations will be communicated to the authors.

# **Guidelines for submission**

#### Manuscript Formats

The manuscript should be about 8 pages in length, typed in double space with Times New Roman font, size 12, Double column on A4 size paper with one inch margin on all sides and should include 75-200 words abstract, 5-10 relevant key words, and a short (50-100 words) biography statement. The pages should be consecutively numbered, starting with the title page and through the text, references, tables, figure and legends. The title should be brief, specific and amenable to indexing. The article should include an abstract, introduction, body of paper containing headings, sub-headings, illustrations and conclusions.

#### References

A numbered list of references must be provided at the end of the paper. The list should be arranged in the order of citation in text, not in alphabetical order. List only one reference per reference number. Each reference number should be enclosed by square brackets.

In text, citations of references may be given simply as "[1]". Similarly, it is not necessary to mention the authors of a reference unless the mention is relevant to the text.

#### Example

- M.Demic, "Optimization of Characteristics of the Elasto-Damping Elements of Cars from the Aspect of Comfort and Handling", International Journal of Vehicle Design, Vol.13, No.1, 1992, pp. 29-46.
- [2] S.A.Austin, "The Vibration Damping Effect of an Electro-Rheological Fluid", ASME Journal of Vibration and Acoustics, Vol.115, No.1, 1993, pp. 136-140.

# SUBSCRIPTION

The annual subscription for IJEST is Rs.600/- which includes postal charges. To subscribe for IJEST a Demand Draft may be sent in favour of IJEST, payable at Sathyamangalam and addressed to IJEST. Subscription order form can be downloaded from the following link http:// www.bitsathy.ac.in./ jjest.html.

For subscription / further details please contact:

#### IJEST

#### Bannari Amman Institute of Technology Sathyamangalam - 638 401, Erode District, Tamil Nadu Ph: 04295 - 226340 - 44

Fax: 04295 - 226666 E-mail: ijest@bitsathy.ac.in Web:www.bitsathy.ac.in

# Indian Journal of Engineering, Science, and Technology

Volume 12, Number 2

July - December 2018

# CONTENTS

V. Alamelu, Priyadharshini, Suganthi and Suriya Varshini	
Secured Energy Efficient Wireless Data Transfer Using Li-Fi Technology 01 C. Manipriya, A.S.Deepika, S. Indhumathi, S. DivyaPratheeba and S.Nithya	7
Denial of Service Using Software Puzzle 1: E.Anjana Devi, S.Raj Kumar, G.M.Madhu Mitha and S.Ponmutharthan	1
Analysis of Road Traffic Fatal Accidents Using Data Mining Techniques 10 V. Kaviya Shree, R.Kalaivani, K.Gowri Nayaki and B.Gokulnath	6
Precise Inspection of Mechanical Systems Based Android APP 20 G.Angusamy, P.Prakash Kumar, K.Hirushi Kesan, R.Naveen Kumar, S.Rahul and V.Vignesh	)
Predictive Analytics for Diabetes Mellitus Using Classification Models 22 M. Munirathinam and K. Premalatha	2
A Survey on Data Mining in Healthcare 27 T. GitanjaliSimran, D. Sujeetha and M.Sushmitha	7
Survey on Application of Image Processing 32 J Gayathri and S Ramya	2
Efficient Content Based Image Retrieval 38 S.Ramya, R.Prathiba, S.Raja Ganesh and PCD.Kalaivaani	3
Question Paper Generation System S.C. Lavanya, K.K. Aarthi, R. Devaprithika, S. Durgapiriyadharshini and K. Dinesh kumar	_