ISER

Institute for Scientific and Engineering Research

International Conference on Computer Science and Engineering (ICCSE-17)

Mumbai
17th September'17

Mumbai

17th September'17

Institute for Scientific and Engineering Research

4A, Girija Apartment, MMDA,

Arumbakkam, Chennai-600106, India

www.iser.org.in

Editorial:

We cordially invite you to attend the International Conference on Computer Science and Engineering (ICCSE-17), which will be held in Grand Hyatt Mumbai, Off Western Express Highway, Santa Cruz East, Mumbai on September 17th, 2017. The main objective of ICCSE-17 is to provide a platform for researchers, engineers, academicians as well as industrial professionals from all over the world to present their research results and development activities in Computer Science Engineering and information Technology. This conference provides opportunities for the delegates to exchange new ideas and experience face to face, to establish business or research relations and to find global partners for future collaboration.

These proceedings collect the up-to-date, comprehensive and worldwide state-of-art knowledge on software engineering, computational sciences and computational science application. All accepted papers were subjected to strict peer-reviewing by 2-4 expert referees. The papers have been selected for these proceedings because of their quality and the relevance to the conference. We hope these proceedings will not only provide the readers a broad overview of the latest research results on Electrical, Electronics, Mechanical, Computer Science and Information Technology but also provide the readers a valuable summary and reference in these fields.

The conference is supported by many universities and research institutes. Many professors plaid an important role in the successful holding of the conference, so we would like to take this opportunity to express our sincere gratitude and highest respects to them. They have worked very hard in reviewing papers and making valuable suggestions for the authors to improve their work. We also would like to express our gratitude to the external reviewers, for providing extra help in the review process, and to the authors for contributing their research result to the conference.

Since July 2017, the Organizing Committees have received more than 40 manuscript papers, and the papers cover all the aspects in Computer Science Engineering and information Technology. Finally, after review, about 11 papers were included to the proceedings of ICCSE-17

We would like to extend our appreciation to all participants in the conference for their great contribution to the success of International Conference 2017. We would like to thank the keynote and individual speakers and all participating authors for their hard work and time. We also sincerely appreciate the work by the technical program committee and all reviewers, whose contributions make this conference possible. We would like to extend our thanks to all the referees for their constructive comments on all papers; especially, we would like to thank to organizing committee for their hard work.

Editor-In-Chief
Dr. Nalini Chidambaram
Professor
Bharth University

Acknowledgement

ISER is hosting the International Conference on Computer Science and Engineering this year in month of September. Technical advantage is the backbone of development and Nano electronics has become the platform behind all the sustainable International Conference on Computer Science and Engineering will provide a forum for students, professional engineers, academician, and scientist engaged in research and development to convene and present their latest scholarly work and application in the industry. The primary goal of the conference is to promote research and developmental activities Computer Science Engineering and information Technology and to promote scientific information interchange between researchers, developers, engineers, students, and practitioners working in and around the world. The aim of the Conference is to provide a platform to the researchers and practitioners from both academia as well as industry to meet the share cutting-edge development in the field.

I express my hearty gratitude to all my Colleagues, staffs, Professors, reviewers and members of organizing committee for their hearty and dedicated support to make this conference successful. I am also thankful to all our delegates for their pain staking effort to travel such a long distance to attain this conference.

Er. R. B. Satpathy
Secretary
Institute for Scientific and Engineering Research (ISER)

CONTENTS

CONTENTS

ORGANIZATION COMMITTEE

MANAGING DIRECTOR
Dr. P. C. Srikanth
Head of the Department,
Department of ECE,
Malnad College of Engineering

PRESIDENT:
Dr. P A Vijaya
Professor, Department of ECE,
BNM Institute of Technology

ORGANIZING SECRETARY
Asst. Prof. Bonia Mohan,
Department of CSE,
Dayananda sagar College of Engineering,
Bangalore,

PUBLICATIONS COMMITTEE
Dr. Shankar Narayanan
Dr.S. Sangeetha

PROGRAM CHAIR
Dr. Nalini Chidambaram
Professor
Bharth University

Automatic Scraper of Celebrity Images from Heterogeneous Websites Based On Face Recognition and Sorting For Profiling

[1] Sneha, [2] N Lalithamani
[1] PG Student, [2] Assistant Professor (SG),
[1][2] Dept of Computer Science and Engineering,
[1][2] Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidhyapeetham,
Amrita University, India.
[1] snehakandacharam@gmail.com, [2] n_lalitha@cb.amrita.edu

Abstract: Now days it has become trend to follow all the celebrities as we consider them as our role models. So instead of searching the images of various celebrities in different websites we can find them in a single website by sorting all the images. Reliable database of images is essential for any image recognition system. Through Internet we find all the required images. These images will serve as samples for automatic recognition system. With these images we do face detection, face recognition, face sorting using various techniques like local binary patterns, haar cascades. We make an overall analysis of the detector. Using opencv we detect and recognize images. Similarity matching is done to check how the images are related to each other. Collection of images is based on user defined templates, which are in web browser environment. With the help of this system one can give their requirement and the image of celebrity is displayed based on that.

Index Terms— *Celebrity Images, Image recognition, Image sorting*.

I. INTRODUCTION

The World Wide Web gives the information of any kind. It acts as a source of data. There is every topic available in Internet. It is termed as web of everything. Apart from sharing information among people it has huge data which helps in processing and gives us new product or services. Since majority of websites is designed for human browsing, it can be challenging to extract data computationally. Though data on web are structured a diversity of the structure is very high. Every page is different and the structure is very complex. An internet bot which systematically browses the World Wide Web is called and Web crawler[9]. It downloads all the links and webpages that point to a website as it recursively processes it. Extracting the specified data from web pages is called Web Scrapping. In some kind of websites there is some specific template in which they have been build. Example of this kind is a web shop, where a page of product details has always same structure, only differs in content loaded from a database. Data of similar category are displayed in same way. This fact is used to avoid complexity of general web page structure.

After the crawling of images these images are taken for face detection, face recognition and sorting them which is termed as similarity matching. The algorithms used for face detection, face recognition are being discussed in detail. After the face recognition these images are being sorted to check the similarity among images.

II. RELATED WORK

There have been enormous papers about extracting specific data from web pages [14]. One important element is extraction rule which has certain pattern and helps us to locate and extract content [15]. When using web pages it can be path of required HTML element [7]. A web wrapper downloads web page located at specified URL and according to the path it gets data from elements. A standard for selecting nodes from an XML (HTML) document is XPath (XML Path Language)[1][8]. Many works use this. Another standard-based approach to locate required data is using CSS selectors.

This can also be done using beautiful soup which is an python library for pulling data from HTML and XML [11][12]. It works with your favorite parser to provide idiomatic ways of navigating, searching, and modifying the parse tree. It commonly saves programmers hours or days of work.

Face recognition has attracted much attention and its research has rapidly expanded by not only engineers but

also neuroscientists, since it has many potential applications in computer vision communication and automatic access control system [19]. Especially, face detection is an important part of face recognition as the first step of automatic face recognition. However, face detection is not straightforward because it has lots of variations of image appearance, such as pose variation (front, non-front), occlusion, image orientation, illuminating condition and facial expression[20].

Digital images and video are becoming more and more important in the multimedia information era [2]. The human face is one of the most important objects in an image or video. Detecting the location of human faces and then extracting the facial features in an image is an important ability with a wide range of applications, such as human face recognition, surveillance systems, human computer interfacing, video-conferencing, etc.

A general statement of the problem of machine recognition of faces can be formulated as follows: given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces. Available collateral information such as race, age, gender, facial expression, or speech may be used in narrowing the search (enhancing recognition). The solution to the problem involves segmentation of faces (face detection) from cluttered scenes, feature extraction from the face regions, recognition, or verification (Figure 1). In identification problems, the input to the system is an unknown face, and the system reports back the determined identity from a database of known individuals, whereas in verification problems, the system needs to confirm or reject the claimed identity of the input face.
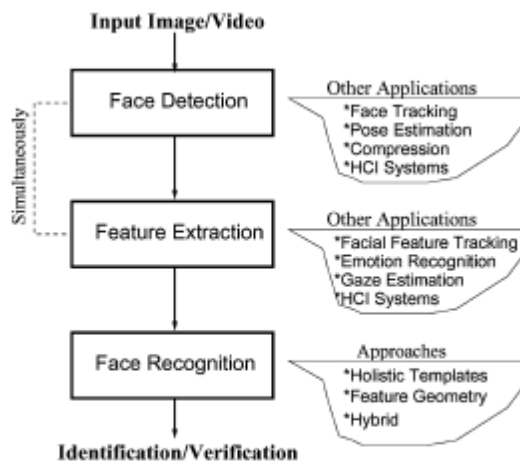
pixels in normalized training images). The classes of face and non-face images are decidedly characterized by multimodal distribution functions and effective decision boundaries are likely to be nonlinear in the image space. To be effective, either classifier must be able to extrapolate from a modest number of training samples or be efficient when dealing with a very large number of these high-dimensional training samples.

Determining similarity between visual data is necessary in many computer vision tasks, including object detection and recognition, action recognition, texture classification, data retrieval, tracking, image alignment, etc. Methods for performing these tasks are usually based on representing an image using some global or local image properties, and comparing them using some similarity measure.

III. AUTOMATIC SCRAPPING AND PROFILING

The main motivation for this work is development of website which consists of all the celebrity images by doing face detection, face recognition and sorting them.
This system will be more useful for people who want to follow present trend as well as they can follow them as per their requirements. First of all there are two Modules. First Module is Admin module and another is User Module.

In Admin Module there are websites or URLs present. Along with them we give celebrity names as inputs. We develop a crawler which stores all the images of celebrities and stores it in a temporary database. Then the preprocessing of the images takes place where it consists of another three sub modules. Face detection and recognition, face sorting and similarity matching.

Fig. 1. Configuration of a generic face recognition system.

Fig 1 [16] Face detection also provides interesting challenges to the underlying pattern classification and learning techniques. When a raw or filtered image is considered as input to a pattern classifier, the dimension of the feature space is extremely large (i.e., the number of
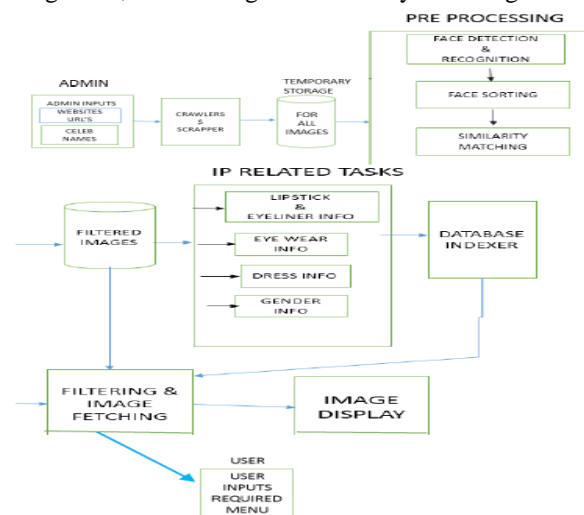

Fig 2. Architecture of the Automatic scrapper and profiling system

After the pre-processing the filtered images are taken and tested for image processing related tasks like lipstick and eye liner , eye wear information , dress information and

gender information. After all this processing it is given to the database indexer. All the filtered images and database indexer combined together and form a Filtering and Image fetching module where all the images will be stored with the required content. Then the user can request for any kind of image data and can find the image in image display. The work flow is as follows

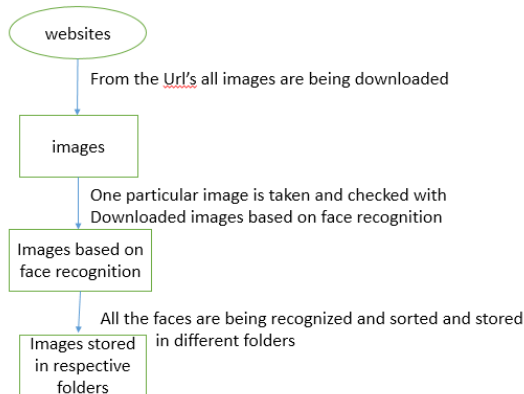Fig 3. Flow diagram of the Automatic scrapper and profiling.

Fig 3. Flow diagram of the Automatic scrapper and profiling.

From Fig 3 we get that Images are being downloaded using python library called Beautiful Soup. After downloading all the images from different websites one particular image is taken and check with all the other images. These images are sorted according to face recognition. This face recognition is done using opencv python. One particular image is saved in one folder and next checks for same image and saves into that folder. It removes all the junk images as well.

IV. FACE DETECTION AND RECOGNITION

Face recognition has attracted much attention and its research has rapidly expanded by engineers, since it has many potential applications in computer vision and automatic access control system. Especially, face detection is an important part of face recognition as the first step of automatic face recognition[26]. However, face detection is not straightforward because it has lots of variations of image appearance, such as pose variation (front, non-front), occlusion, image orientation, illuminating condition and facial expression.

Automatic facial expression recognition involves two vital aspects: facial feature representation and classifier design[27]. Facial feature representation is to derive a set of features from original face images which minimizes within class variations of expressions whilst maximizes between class variations[25]. There are two main types of approaches to extract facial features: geometric feature-based methods and appearance-based methods.

Local Binary Patterns (LBP) has been introduced as novel low-cost discriminative features for facial expression recognition[4]. LBP was proposed originally for texture analysis. A facial image is divided into a set of small regions from which LBP histograms are extracted and concatenated into a single, spatially enhanced feature histogram[21][23]. The simple LBP features can be fast derived in a single scan through the raw image, whilst still retaining enough facial information in a compact representation [6].

The original LBP operator was introduced by Ojala et al . The operator labels the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number (see left of Fig 4 for an illustration). Then the histogram of the labels can be used as a texture descriptor.
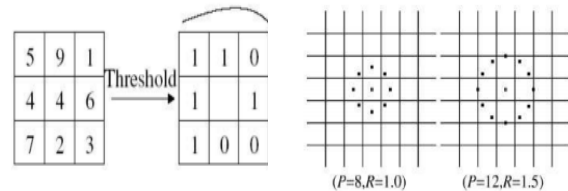
Binary – 11010011 (P=8,R=1) (P=12,R=1.5)
Decimal - 211
Figure 4 Texture descriptor

Fig. 4.[4] Left: The basic LBP operator . Right: Two examples of the extended LBP : a circular (8, 1) neighborhood, and a circular (12, 1.5) neighborhood. The limitation of the basic LBP operator is its small 3×3 neighborhood cannot capture dominant features with large scale structures[24]. Hence the operator was extended to use neighborhood of different sizes. Using circular neighborhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighborhood. Examples of the extended LBP are shown in right of Fig 4, where (P, R) denotes P sampling points on a circle of radius of R. Further extension of LBP is to use uniform patterns. A Local Binary Pattern is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. For example, 00000000, 001110000 and 11100001 are uniform patterns. It is observed that uniform patterns account for nearly 90% of all patterns in the (8, 1) neighborhood and for about 70% in the (16, 2) neighborhood in texture images [18].

Local binary patterns can also found by using OpenCV in python. With help of this face detection has been done. It can detect more than one person in the image.

```python
from sklearn.svm import LinearSVC
from imutils import paths
import argparse
import cv2
import os

ap = argparse.ArgumentParser()
ap.add_argument("-t", "--training", required=True,
        help="path to the training images")
ap.add_argument("-e", "--testing", required=True,
        help="path to the tesitng images")
args = vars(ap.parse_args())

desc = LocalBinaryPatterns(24, 8)
data = []
labels = []

for imagePath in paths.list_images(args["training"]):
        image = cv2.imread(imagePath)
        gray = cv2.cvtColor(image, cv2.COLOR_BGR2GRAY)
        hist = desc.describe(gray)

        labels.append(os.path.split(os.path.dirname(imagePath))[-1])
        data.append(hist)
```
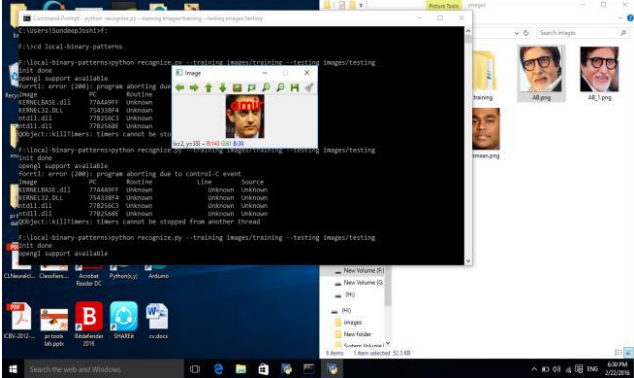

Fig 5. Here face recognition is being done using opencv

V. HAAR CASCADE DETECTION

For combining successively more complex classifiers in a cascade structure which dramatically increases the speed of the detector by focusing attention on promising regions of the image[3]. The notion behind focus of attention approaches is that it is often possible to rapidly determine where in an image an object might occur [17, 8, 1]. More complex processing is reserved only for these promising regions. The key measure of such an approach is the "false negative" rate of the attentional process. It must be the case that all, or almost all, object instance.

The simple features used are reminiscent of Haar basis functions which have been used by Papageorgiou et al. . More specifically, we use three kinds of features. The value of a two-rectangle feature is the difference between the sum of the pixels within two rectangular regions. The regions have the same size and shape and are horizontally or vertically adjacent (see Figure 6). A three-rectangle feature computes the sum within two outside rectangles subtracted from the sum in a center rectangle. Finally a four-rectangle feature computes the difference between diagonal pairs of rectangles.es is selected by the intentional filter.
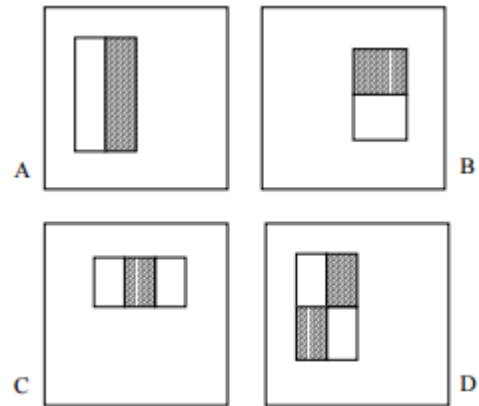
Figure 6 An detection window

In the Figure.6 [13] shown above, : Example rectangle features shown relative to the enclosing detection window. The sum of the pixels which lie within the white rectangles are subtracted from the sum of pixels in the grey rectangles. Two-rectangle features are shown in (A) and (B). Figure (C) shows a three-rectangle feature, and (D) a four-rectangle feature.

1.Integral Image

Rectangle features can be computed very rapidly using an intermediate representation for the image which we call the integral image.2 The integral image at location x,y contains the sum of the pixels above and to the left of x,y , inclusive[4]:

$$(x, y) \quad = \sum_{x' \leq x, y' \leq y} i(x', y')$$
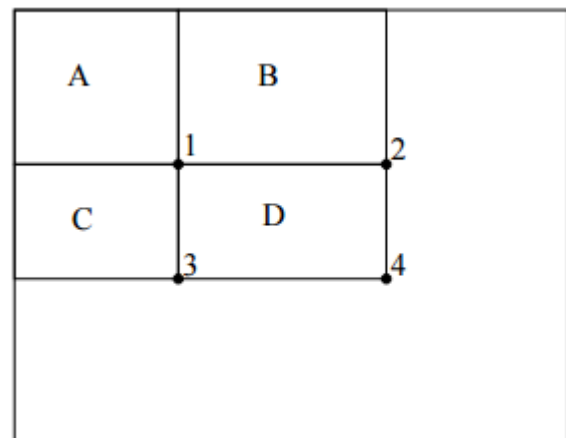
Figure 7 .To compute sum of pixels

Fig 7. [4].The sum of the pixels within rectangle can be computed with four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is A+B, at location 3 is

A+C, and at location 4 is A+B+C+D . The sum within can be computed as 4 + 1 − (2 +3).

where ii(x,y) is the integral image and i(x,y) is the original image. Using the following pair of recurrences:

s(x,y) = s(x, y-1) + i(x,y) (1)

ii(x,y) =ii(x-1,y) + s(x,y) (2)

(where s(x,y) is the cumulative row sum,s(x,-1)=0 and ii(-1,y)=0) the integral image can be computed in one pass over the original image. Using the integral image any rectangular sum can be computed in four array references (see Figure 7). Clearly the difference between two rectangular sums can be computed in eight references. Since the two-rectangle features defined above involve adjacent rectangular sums they can be computed in six array references, eight in the case of the three-rectangle features, and nine for four-rectangle features.

VI.   SIMILARITY MATCHING

Comparing two images is the fundamental operation for any retrieval systems. The similarity matching of two images can reside in the hierarchical levels from pixel-by-pixel level, feature space level, object level, and semantic level. In most systems of interest, a simple pixel-by-pixel comparison will not suffice: the difference that we determine must bear some correlation with the perceptual difference of the two images or with the difference between two adequate semantics associated to the two images[5]. Similarity matching techniques are developed mostly for recognition of objects under several conditions of the distortion while similarity measures, on the other hand, are used in applications like image databases. Matching and dissimilarity measurement are not seldom based on the same techniques, but they differ in emphasis and applications.

Similarity judgments play a central role in theories of human knowledge representation, behavior, and problem solving and as such they are considered to be a valuable tool in the study of human perception and cognition. Tversky describes the similarity concept as ''an organizing principle by which individuals classify objects, form concepts, and make generalizations.'' Retrieval by similarity[10].

One can identify three components that typically every system for retrieval by similarity needs to have:

- ❖ Extraction of features or image signatures from the images, and an efficient representation and storage strategy for this precomputed data.
- ❖ A set of similarity measures, each of which captures some perceptively meaningful definition of similarity, and which should be efficiently computable when matching an example with the whole database.
- ❖ A user interface for the choice of which definition of similarity should be applied for retrieval,

presentation of retrieved images, and for supporting relevance feedback.

Similarity Matching can be found using open cv where we can find the similarity between images and an histogram which shows the similarit

Figure 8 Similarity Matching
Figure 8. shows the similarity between two same images but from different website.
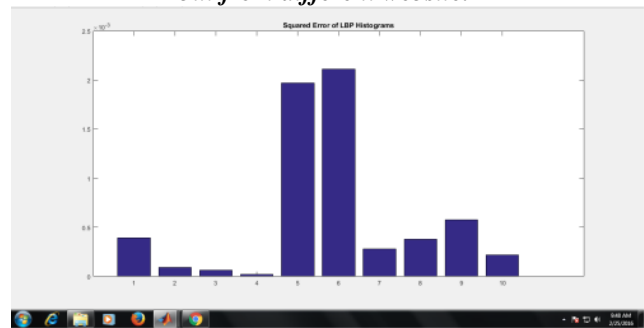
Figure 9. shows the histogram of the images [22]

VII.   DATABASES

For experimental purposes , we have taken the below websites for downloading the images and have performed face detection , face recognition and similarity matching and found significant results.

http://www.behindwoods.com/index.html

http://www.missmalini.com/

http://www.harpersbazaar.com

http://www.highheelconfidential.com/

http://fashionista.com/2015/02/most-influential-style-bloggers-2015

Images of all the celebrities collected in the database

VIII. CONCLUSION AND FUTURE ENHANCEMENT

Face recognition has got its importance on a wide range in Image Processing. It is an part of biometrics which helps in recognizing human faces in digital images. Similarity matching also plays a vital role where we can know how similar those images are when they are matched.

The proposed system helps to crawl the images and perform face detection, recognition and similarity matching which helps to develop any facial recognition system using the algorithms Local Binary Patterns and Haar cascades. Using these we will be able to recognize any face in digital images and check similarity.

ACKNOWLEDGMENT

REFERENCES

[1] Michal vagac , Miroslav Melichercik , Matus Marko (2015). Crawling images with web browser support : IEEE 13th International Scientific Conference on Informatics. Informatics'2015 . November 18-20 . poprad . Slovakia

[2] Junghoo Cho , Hector Garcia-Molina, Lawrence Page (2012). Reprint of : Efficient crawling through URL ordering : Elsevier Journal , Computer Networks 26(2012) 3849-3858.

[3] Paul Voila , Michael Jones . Rapid Object Detection using a Boosted Cascade of Simple Features . Accepted Conference on Computer Vision and Pattern Recognition 2001.

[4] Caifeng Chan , Shogang Hong , Robust Facial Recognition using Local Binary Patterns . Image Processing, 2005. ICIP 2005. IEEE International Conference on (Volume:2)

[5] Eli shechtman , Michal Irani. Matching Self local similarities across Image and Videos. http://www.wisdom.weizmann.ac.il/~vision/VideoAnalysis/Demos/SelfSimilarities

[6] V. Ferrari, T. Tuytelaars, and L. V. Gool. Object detection by contour segment networks. In ECCV, May 2006.

[7] R. Baumgartner , S. Flesca and G. Gottlob, Visual web information extraction with lixto , In Proceedings of the 27th International Conference on Very Large Data Bases, VLDB 01, pages 119-128, San Francisco, CA, USA: Morgan Kaufmann publishers Inc, 2001.

[8] V. Crescenzi , P. Merialdo , and D. Qui, Alfred : Crowd assisted data extraction, In Proceedings of the 22nd International Conference on World Wide Web Companion , WWW 13 Companion, pages 297-300, Republic and Canton of Geneva, Switzerland, 2013.

[9] T.Furche , G. Gottlob, G.Grasso , C. Schallhart and A. Sellers, Oxpath: A language for scalable data extraction , automation, and crawling on the deep web, The VDLB Journal , 22(1):47-72, Feb. 2013.

[10] R. Brooks, T. Arbel, D. Precup, Anytime similarity measures for faster alignment, Computer Vision and Image Understanding 110 (3) (2008) 378–389.

[11] T.Grigalis, Towards web-scale structured web data extraction , In proceedings of the Sixth ACM International Conference on Web Search and Data Mining , WSDM 13, pages 753-758, New York, NY, USA:ACM, 2013.

[12] K. Kanaoka , Y. Fujii , M. Toyama , Ducky : a data extraction system for various structured web documents , In Proceedings of the 18th International Database Engineering & Applications Symposium, IDEAS ' 14. Pages 342-347, New York , NY, USA: ACM, 2014

[13] N. Kushmerick , Wrapper induction : Efficiency and expressiveness, Artificial Intelligence , Vol 118, Issue 1-2 , pages 15-68. Essex, UK: Elsevier Science Publishers Ltd., 2000.

[14] M. Tlo and M. Suzuki, Design and implementation of a facility for wandering and manipulating the structure of

on-line data, In Information Science and Applications (ICISA), 2011 International Conference on, pages 1-8, April 2011.

[15] M. Geel , T. Church and M.C . Norrie, Sift : An end – user tool for gathering web content on the go , In Proceedings of the 2012 ACM Symposium on Document Engineering , DocEng 12, pages 181-190, New York, NY, USA: ACM, 2012.

[16] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In International Conference on Computer Vision, 1998.

[17] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. IEEE Patt. Anal. Mach. Intell., 20(11):1254–1259, November 1998

[18] H. Greenspan, S. Belongie, R. Gooodman, P. Perona, S. Rakshit, and C. Anderson. Overcomplete steerable pyramid filters and rotation invariance. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1994.

[19] Edgar Osuna, Robert Freund, and Federico Girosi. Training support vector machines: an application to face detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997.

[20] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, , and T. Poggio. Robust object recognition with cortex-like mechanisms. to appear in PAMI, 2006.

[21] Di Huang , Caifeng Shan, Mohsen Ardabilian. Local Binary Patterns and Its Application to Facial Image Analysis: A Survey. IEEE transactions on systems, man, and cybernetics—part c: applications and reviews, vol. 41, no. 6, november 2011 .

[22] Chi-Ho Chan, Josef Kittler, Kieron Messer, Multi scale local Pattern Histograms for Face Recognition. Advances in Biometrics Volume 4642 of the series Lecture Notes in Computer Science pp 809-818.

[23] G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu. Boosting Local Binary Pattern (LBP)-based face recognition, volume 3338. Springer Berlin / Heidelberg, 2004.

[24] S. Yan, S. Shan, X. Chen, and W. Gao. Locally assembled binary (lab) feature with feature-centric cascade for fast and accurate face detection. 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2008.

[25] B. Fröba and A. Ernst. Face detection with the modified census transform. In Sixth IEEE Int. Conference on Automatic Face and Gesture Recognition, pages 91–96, 2004

[26] H. Jin, Q. Liu, H. Lu, and X. Tong. Face detection using improved lbp under bayesian framework. In Third Int. Conference on Image and Graphics, pages 306–309, 2004.

[27] M. Heikkilä, M. Pietikäinen, and C. Schmid. Description of interest regions with local binary patterns. Pattern Recognition, 42(3):425– 436, 2009.

An Intelligent Robotic Home Security System

[1]A.R.Revathi, [2]B.Sathyanarayanan [3]M.Raja, [4]J.Sathish, [5]D.Suresh,
[1] Assistant Professor (Sl.G), [2][3][4][5] UG Student
[1][2][3][4][5] Department of Information Technology,
SRM- Valliammai Engineering College
[1] revathi96@gmail.com

Abstract:-- **The security needed for humans and their belongings is more important in the day-to-day life. In the growing technological world many automated systems are being developed to alert the user about any hazards taken place in the home environment. Normally this kind of automated system will alert the user within a particular area, this can be made through using Extended Infrastructure Network. Even if user needs to be communicated wirelessly it's more expensive. Considering all these factors we design the smart home security system which can be controlled by Android Application through the Internet. This robot gives live surveillance streaming about the home environment to the user Android Application. Live surveillance streaming is made possible with the help of IP Camera controlled by Arduino board. However, this robot can functioned by two modes either user mode or automatic mode. In addition to Home Security System, the robot can detect LPG gas leakage and fire accident happen in the home environment with minimum cost. Home Security System can alert the user through email and Android App notification. User address will be send to the nearest fire station, if fire accident is occurred.**

Index Terms - **Android, Live surveillance, Trigger, Arduino.**

I. INTRODUCTION

In the growing world, electronic components are being popular due to comfort and convenience. In obvious to which, many people pay a lot of attention on the electronic things in recent years. Automation plays major role in many fields but security is the more important concern. Many experts and companies are trying to provide good automated solution to the user and even various researches are going on in the security aspect. Therefore, we suggest best solution as the robotic technology with the security system, because the issue of the safe living environment for human beings is more important. The robots can be designed in such a way to identify potential hazards to warn human beings in advance.

Now a day's smart phones are being popular among the mobile handset users. According to data from the International Data Corporation (IDC) in 2015,

Android conquered the smart phone market with a share of 82.8%, where IOS has 13.9%, where Blackberry has 0.3%, where windows has 2.6% and where other phones has 0.4% share. Android users are more in number [19].

Considering all these aspects, this paper discusses an approach where a robot is operated with Android Application to provide security in the home environment. Home Security System can monitor private, public and even dangerous areas through the Internet anytime, anywhere remotely. Android Applications are developed using Object Oriented Programming language. The android app developed with internet support. A new type of Home Security System developed in which there is a moving wireless robot in home which has IP Camera and various sensors like ultrasonic sensor, gas sensor and temperature sensor are connected with it and the robot needs Internet for remote accessing. The internet for robot is provided with the help of Wi-Fi chip. The robot is fixed with an IP Camera that can be used for live surveillance in home environment. The robot operated over two modes, either User mode or Automatic mode. During the live surveillance process, if any LPG Gas leakage or fire accident happens, it can send the alert to the user through E-mail and Android App notification and as well in case of fire accident, it sends the user address to the fire station. In this paper section II provides information literature survey and section III explains about the proposed methodology. It describes about the live video surveillance, hazards detection and notification process. The section IV deals with technical aspects needed and section V is about experimental result. Finally, section VI will conclude the work.

II. LITERATURE SURVEY

Many experts and organizations where concentrating in the development of Home Security System. In smart way, different methodologies proposed at different fields. However, Home Security System using Smart phones is still ongoing research field. Some of existing projects are discussed below. Rupam Kumar

Sharma [1] looks into the development of an Android application which alerts (message) the mobile device on possible intrusion and subsequently a reply message needs to be send through SMS to trigger alarm/buzzer. The AuthorRen C. Luo[2] looks the multiple remote interface security system can senseunusual and risky situation and notify userover internet, or send the message to mobile phone overGSM module.The appliance module can feedback response result to the user through mobile phone. In the existing work [1] [2] they works with GSM module to send notification, whereas Ch.ManoharRaju [3] uses Bluetooth module to send notification; they look into the development of a prototype of an autonomous android based mobile robot for detecting gas leakage in large industrial facilities. The robot may be used for continuous inspections of facilities or for scheduled inspections of specific system components. Lim[4] explains the development of, the PC hosts a Java program that connects wirelessly with a mobile robot using the ZigBee module and shows a live video stream from a wireless camera on the robot. The robot gets commands from the PC and it then sends instructions to switches wirelessly to turn on or off. It also deals with switches controlled by robot through video stream. Dr.Shantanu K. Dixit[5] deals with the video feed for human detection;the robot will help in rescue operation and user can access the video feed from the remote location such as the high sensitive areas or areas which are beyond our reach. The total system comprises mobile robot, controlled over the Internet. The robot has camera and PIR sensor for living body detection. Mr.Vishnoo.S.K[6]providesthe location information of remote zones and even helps the military personnel in fighting terrorism. The robot used in the borders where terrorist presence may be high and human being cannot be risked in such areas. The author[7] Wen- Chung Chang, mainly focused into a novel approach to develop mobile based localization robot and navigation in a huge unknown workspace with a set of vision sensors based on an effective on-line calibration strategy. They control the mobile robot to aggressively perform calibration with any static IP camera. The above existing technology were using GSM and Bluetooth module which cost more and even it cannot access remotely.

III. PROPOSED METHODOLOGY

A Multi-purpose intelligent robot is designed with multiple sensors (ultrasonic sensor, Gas sensor and Temperature sensor) connected with Arduino circuit board. Ultrasonic Sensor (HC-SR04) is used to find the distance of the targeting object. The major benefit of this sensor is transparent to sunlight and black material but remains affected to clothes. Gas sensor detects the gas leakage in the home environment. Here MQ-6 gas sensor is used to detect the LPG gas. It has heater coil made up of NI-CR alloy which detect LPG quickly in the atmospheric air.
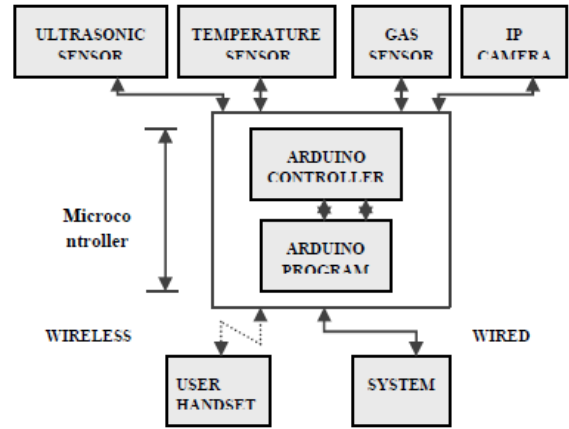
Fig. No.1 Block Diagram

The temperature in the house is calculated using Temperature sensor (LM35), it has benefit over linear temperature sensors generated in kelvin as it provides calibrated centigrade scaling. If the detected value exceeds 120 degree centigrade [16], fire alert will be send to the user. The notification is send by Email and also through Android App. Code which is used for the functioning of the robot is stored in the Microcontroller (i.e.Arduino mega 2560). The entire robot system can be controlled remotely, (through internet) either by User mode or Automatic mode.IP camera is used to give the live video feed to Android App. Block diagram is shown in the Fig. no.1
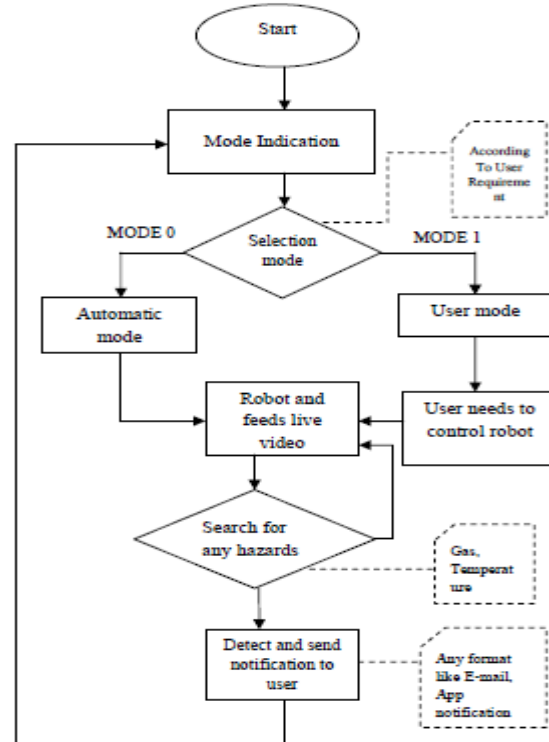
Fig. no. 2 Flow Diagram

a). User Mode: In the user mode, user needs to control the robot through user handset with help of

Android Application. As per the user direction robot give live video feed to the user application. According to user requirement robot navigate the home environment. It described in flow diagram (Fig.no.2).

b). Automatic Mode: In this mode, user don't have access to control the robot. Using ultrasonic sensor the robot automatically avoid the obstacles interfered in its path. So, the live video stream feed to the user application.

IV. TECHNICAL BACKGROUND

4.1 Arduino Mega

In this work implemented by microcontroller board(Arduino Mega 2560).The Arduino Mega 2560 is a microcontroller Board and it is based on ATmega 2560 [8].

Fig. no 3 Arduino Microcontroller Board

Arduino Microcontroller Board is shown in Fig no 3. It has 54 input and output pins of which includes 16 analog input, 4 UARTs, a 16MHz crystal oscillator, a power jack, an ICSP header, a USB connection, and a reset button. Datasheet for Arduino MEGA is shown in Table no.1.

4.2 MQ6 Gas Sensor

This implementation hasMQ6 gas sensor. MQ6 gas sensor has a 6 pin device and it requires 5 voltage DC supply. Whenever the sensor senses gas leakage among 100 to 1000ppm, in the atmospheric air, its output turns high and triggers the buzzer[10].

Fig. no.4 MQ-6 Gas Sensor

Table no.1 Arduino Datasheet [9]

Microcontroller	ATmega2560
Operating Voltage	5V
Input Voltage	7-12V
Digital I/O Pins	54
Analog Input Pins	16
Length	101.52 mm

Table no.2 MQ6 Datasheet [10]

| Sensor Type	Semiconductor
Detection Gas	Isobutane, LPG, Butane
Concentration	100-10000ppm
Heater Voltage(VH)	5.0V±0.2V AC (or) DC
Heater Resistance(RH)	31Ω±3Ω Room Tem.
Heater consumption(PH)	≤900Mw
Preheat time	Over 48 hours

It is widely used in domestic gas leakage alarm and portable gas detector. MQ-6 gas sensor is shown in fig. no.4. Datasheet for gas sensor is shown in Table no.2.

4.3 Temperature Sensor LM35

We require LM35 temperature sensor for this work. LM35 is a precision integrated circuit temperature device. It has an advantage over linear temperature sensors generated in kelvin as it provides calibrated centigrade scaling [12].
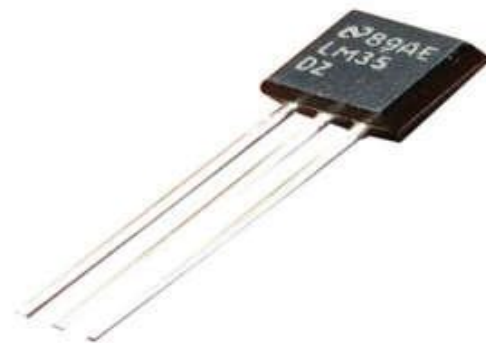

Fig. no 5 Temperature Sensor LM35

The main feature of LM35 is it calibrates directly to Celsius. It is mainly applicable in remote applications. Temperature Sensor LM35 is shown in fig. no.5. Datasheet for LM35 is shown in Table no.3.

Table no.3 Temperature Sensor LM35 Datasheet [13]

Parameter	Condition	Lm35		
		Typical	Tested Limit	Design Limit
Accuracy	TA=+25°c	±0.2	±0.5	
	TA=-10°c	±0.3		
	TA=TMAX	±0.4	±1.0	
	TA=TMIN	±0.4	±1.0	
Sensor gain	TMIN≤ TA≤ TMAX	±10.0	±9.9 ±10.1	
Minimum Temperature For rated accuracy	IL=0	±1.5		±2.0
Long term stability	Tj=TMAX, for 1000 hours	±0.06		

4.4 Ultrasonic Sensor

In this implementation HC-SR04 ultrasonic sensor is used. The HC-SR04 ultrasonic sensor that makes use of sonar device.

Fig. no 6 Ultrasonic Sensor

It is used to calculate the distance to an object. It offers non-contact range detection with high accuracy and it produces stable readings [11].It is transparent to sunlight and black materials but remains affected to clothes.
Ultrasonic Sensor is shown in fig. no. 6. Datasheet for Ultrasonic sensor is shown in Table no.4.

Table no. 4 Ultrasonic Datasheet

Electrical Parameters	HC-SR04 Ultrasonic module
Operating Voltage	DC-5V
Operating Current	15mA
Operating Frequency	40KHZ
Farthest Range	4m
Nearest Range	2cm
Input Trigger Signal	10μs TTL pulse
Output Echo Signal	Output TTL level signal, proportional with range
Dimensions	45*20*15mm

4.5 IP Camera

For live surveillance IP camera is used. IP camera is a high definition, full functionality video endpoint device with high image quality and better processing power.

Fig. no 7 IP camera

The camera can produce full HD 1080p resolution by processing 30 frames per second [14]. It optimizes network usage with MJPEG compression. IP Camera is shown in fig. no. 7. Datasheet for IP Camera is shown in Table no.5.

Table no. 5 IP Camera Datasheet [14]

Item	Specification
Lens selection	P-Iris lens
Shutter speed	1/5 to 1/32,000 sec
Video compression	H.264 and MJPEG
Video Resolution	2MP HD
Video streaming	Single-stream H.264 or MJPEG up to 1080p Dual-stream H.264 and MJPEG
Power consumption (max)	DC: 4.7W AC: 6.3W

4.6 ESP8266

To interact with internet we used ESP8266. ESP8266 offers a complete Wi-Fi networking solutions which either act as a host application or from another application processor.

Fig. no 8 ESP8266 Circuit Board

Whenever ESP8266 acts as a host application it boots up directly from the external flash memory. It improves the performance of the system and minimize the memory requirements.[15] It also serves as a Wi-Fi adapter. It is mainly applicable in smart power plugs and IP cameras. ESP8266 is shown in fig. no.8. Datasheet for ESP8266 is shown in Table no.6.

Table no.6 ESP8266 Datasheet [15]

Wi-Fi Chip/Module	ESP8266
Wi-Fi Standards	802.11b/g/n
Packets	TCP and UDP
Modes	Client and Server
Size	25 *38 mm
Digital Pins	9
Analog Pins	1
Programmable Microcontroller	Yes

V. EXPERIMENTAL RESULT

5.1 Component & Software Specification

Arduino microcontroller is integrated with many sensors, MQ6 (Gas Sensor) operates normally in 5V and maximum of 5V±0.2V.The possible component for MQ6 sensor is MQ2 sensor. HC-SR04 (ultrasonic sensor) works normally in 5V and other alternate component is IR sensor. LM35 (temperature sensor) takes input of 4V normally and it can operates upto 20V. It can be replaced with IR Flame sensor. ESP6682 (Wi-Fi chip) and IP Camera inputs with 5V and 4.7V respectively. Arduino MEGA normally operates with 5V and it works upto 12V and other substitute microcontroller is RASPBERRY PI.

This entire work was implemented with Arduino 1.6.5 and Android Studio 1.5.1 software. The program is coded in Arduino 1.6.5 using Arduino 'C' Programming.Sensor inputs detected by using analog Read function which is coded as, sensor Value = analog Read(sensor); The digital Read function reads the digital value outputted from LM35 sensor which is programmed below, intsensorReading = digitalRead(2); The above given, is a sample lines which is used in the gas and flame detection module.

```
servo.write(90); scan();
FrontDistance = distance;
if(FrontDistance> 40 || FrontDistance == 0)
{ moveForward(); }
Else
{ CollisionCounter = CollisionCounter + 1;
moveStop(); navigate(); }
```

The given code is used for the robot movement during live surveillance operation. The result for gas detection value is shown in Fig. no. 9. The reading has taken in both normal and leakage conditions.
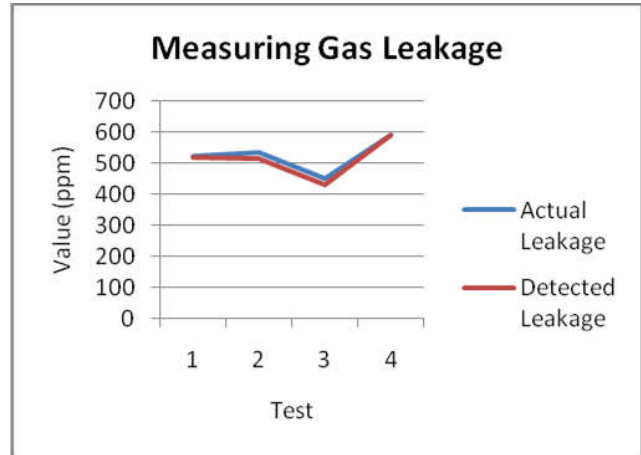
Fig. no. 9 Measuring Gas Leakage

VI. CONCLUSION

A Multipurpose Intelligent Robot is used in home security system, which is controlled by Android Application through the Internet. It is used for the purpose of remote surveillance system and for the detection of gas and temperature. This robot can functioned by two modes either user mode or automatic mode. When intrusion occurs, it can alert the user by sending an E-mail and Android App notification. Finally, we implement the Android based multipurpose intelligent robot using android phone which has been tested with real data and it performs with 90% efficiency in minimal cost. In future, we are going to include more functions like voice, image recognition and heartbeat reading.

REFERENCES

[1].Rupam Kumar Sharma,Ayub Mohammad, HimankaKalita, DhirajKalita,"ANDROID INTERFACE BASED GSM HOME SECURITY SYSTEM", IEEE Conference Publication 2014.

[2]. Ren C. Luo, KuoL.Su,"THE DEVELOPMENT OF INTELLIGENTHOME SECURITY ROBOT", IEEE Conference July-2005 .

[3].Ch.ManoharRaju and N.Sushma Rani, "AN ANDROID BASED AUTOMATIC GAS DETECTION AND INDICATION ROBOT", International Journal of Computer Engineering and Applications, Volume VIII, Issue I, Oct 2014, ISSN 2321-3469.

[4].ZakiUd Din, Wee-Gin (David) Lim, "HOME AUTOMATION WITH SMART ROBOT FEATURING LIVE VIDEO FEED AND REMOTELY CONTROLLED SWITCHES", IEEE Conference Publication 2013.

[5].Dr.Shantanu K. Dixit, Mr. S. B. Dhayagonde, "DESIGN AND IMPLEMENTATION OF ESURVEILLANCE ROBOT FOR VIDEO MONITORING AND LIVING BODY DETECTION", International Journal of Computer Engineering and Applications, Volume 4, Issue 4, April 2014 ISSN 2250-3153.

[6].Mr.Vishnoo.S.K, Mrs.Aarthy.S.T, "ROBOT FOR SURVEILLANCE, COMBATING TERRORISM AND RESCUE OPERATIONS", International Journal Of Engineering Development And Research, Volume 3, Issue 2 ISSN: 2321-9939.

[7].Wen-Chung Chang, Ping-Rung Chu, Taiwan, "AN INTELLIGENT SPACE FOR MOBILE ROBOT NAVIGATION WITH ON-LINE CALIBRATED VISION SENSORS", IEEE Conference Publication 2010.

[8].https://www.arduino.cc/en/Main/ArduinoBoardMega2560

[9].https://www.arduino.cc/en/Main/arduinoBoardMega

[10].https://www.parallax.com/sites/default/files/downloads/28015-PING-Sensor-Product-Guide-v2.0

[16].https://www.quora.com/At-what-temperature willcotton- ignite

[17].Professional Android Application Development, Reto Meier, WROX publication, May 2012.

[18].Head First Android Development, Dawn Griffiths & David Griffiths, O`Reilly Media, June 2015.

[19].http://www.idc.com/prodserv/smartphone-os marketshare. Jsp

[11].http://www.elecfreaks.com/store/download/product/Sensor/HC-SR04/HCSR04_ Ultrasonic_Module_User_Guide

[12].https://www.facstaff.bucknell.edu/mastascu/eLessons HTML/Sensors/LM35

[13].http://www.engineersgarage.com/electroniccomponents/lm35-sensor-datasheet

[14].http://www.cisco.com/c/en/us/products/collateral/physical-security/video-surveillance-6000-series-ipcameras/ datasheet-c78-728703.html

[15].https://www.adafruit.com/images/productfiles/ 2471/0A-ESP8266__Datasheet__EN_v4.3

Employment Exchange Application for Android Platform

[1] B Brindha, [2] S Dhanya [3] Mr.K.Elaiyaraja (Sr.G),
[1] Assistant Professor (Sl.G), [2][3][4][5] UG Student
[1][2][3][4][5] Department of Information Technology,
Valliammai Engineering College, Kancheepuram District, India.
[1] brindhavecit@gmail.com [2] k.elaiyaraja@rediffmail.com [3] dhanyait2012@gmail.com

Abstract: -- **The prime objective of "Employment Exchange Application" is to create a full-fledged Android Application which would allow the user to register their qualifications and experience with the valid proof and provide employment assistance on the basis of that. Job seekers could search suitable jobs based on their qualification. After registration, a unique user ID will be generated as a reference. So that next time the user could login with the user ID and could make any updates if available. Job seekers could also check their status by the option Status. By choosing SMS option, whenever government vacancies are announced, the registered member with the suitable education profile will be intimated through an SMS/Mail.**

This project was done using Eclipse. It also uses Android Software Development kit (SDK) which includes a comprehensive set of development tools that help us to develop mobile Applications on the Android platform.

Keywords- **Android SDK, Eclipse-IDE, java, SQLite**

I. INTRODUCTION

Employment Exchange Application is a new Technology for job seekers to register their qualifications through this application rather than online. One of the main disadvantages in the present online system is that, it provides potential candidates with a facility to save their work as they do, but Internet service drop-offs and system timeouts can result in the loss of application content at any point throughout the process, forcing potential candidates to go back to the beginning and enter their details.

In other case, One need to visit nearest employment exchange office in the area of residence and wait for long an hour in Employment office to register. And also they occupy an enormous amount of space and create a noise pollution over there it would be very simple and easy if it is carried over using this application. To provide an integrated solution for the above said problems is the motto of our project.

II. OBJECTIVES

- ❖ Job seekers can easily search for suitable jobs based on their qualification.
- ❖ To store the data about the users in a secure way.
- ❖ To make Employment Exchange System more easy and user friendly.
- ❖ To reduce the manual work during registration and updating details.
- ❖ Usage of ICT infrastructures in minimum amount.
- ❖ To reduce the electricity and power backup facility.

III. LITERATURE REVIEW

This article [1] gives detailed information about how to start an android application and to overcome the hurdles while developing an application. It provides the basic steps for application development and in which it has four phases as setup, Development, Debugging and Testing, Publishing. It also provides detailed information about the installation of software.

This article [2] gives detailed information about the job search systems. It uses an Algorithm in which it provides secure login. Unauthorized users cannot login until they are provided with unique identification number. It provides Un-registered users to view jobs available but the user should be logged-in to apply for job. It causes fuzzy preference rules to make proper decisions. It uses ANDROID 2.2, JADE-LEAP and the Google API to provide a robust and user friendly solution.

The authors in [2] have developed three kinds of application on Java and Android SDK- Weibo client, video player, Audio player. Test hardware environment is Lenovo Y460 laptop and millet MI phone; software environment is windows 7 and phone system environment

is Android 4.0.3. Network environment is 10M broadband, WIFI LAN and chine Mobile GPRS network. The audio player uses Content Resolver and curor to obtain music files and play the music. The video player uses Media player class provided by Android SDK. Sina weibo client can complete OAuth 2.0 authentication method for user to complete the login process.

The authors in [3] discuss about the challenges and issues faced during Android application development. Android is an open source operating system with its powerful function enables the user friendly atmosphere. This paper gives a detailed reference of a new app PikDish developed for restaurants. It consists of main components like Users which allows a people to rate individual food items, Entities like food items, Dimensions which describes about the categories that are formed to rate the food items, Belief system which is personal to each user. It says that experience of developing this app is quite challenging, motivating as well as satisfying.

IV.   METHODOLOGY

A. Users
After installing the application the user can login using the unique ID and password. If he/she is a new user, options are available to register. Once the user logs in, details must be given. Thus the details are uploaded in the database. The registration will be successful. The user can login whenever he/she needs any information and also can update status periodically.

B. Administrator
Once the user submits his details the administrator verifies whether they are valid. On confirmation, intimation is provided to the user. After this the administrator keeps the user updated about the job opportunities. This facility can be provided through automated SMS and E-mail. Therefore this will be helpful for the user to access.

C. Implementation
Initially, applications are designed to satisfy multiple user needs. On registration, a unique id(mail ID) will be generated for the user. No two registrations are possible for a single user. The user logs in using his / her unique id to access the details. To view the details, the user id and password is just enough.

Any information in this database is notified to the user using mobile communication. Administrator gathers information from various government offices and updates the database for each user. User can render the information in the database by accessing the application.

V.   ARCHITECTURE

1. Android Architecture
Android System is a Linux-based system. Android architecture (fig. 1) consists of four layers such as Linux kernel, Libraries and Android Runtime, Application Framework, Application and Widgets.

A. Application and Widgets
Android application gives a set of applications including browser, contacts, maps, calendar which is been developed in Java.

B. Application Framework
The developer can access all the API framework of the programs. It simplifies the reuse of its components. It has to follow the security of the framework when all other apps can access and be able to use this component.
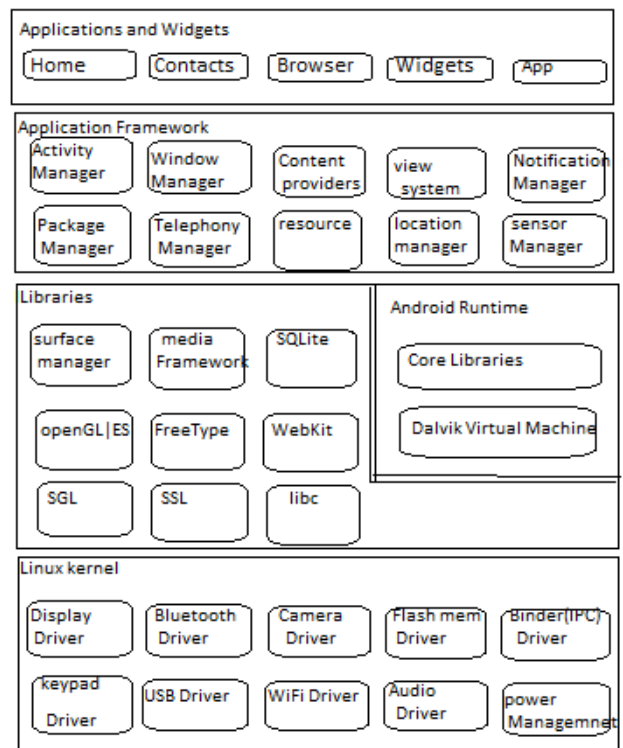
Fig. 1- Android Architecture

C. Libraries and Android Runtime
The library is dividing into two components; They are Android Runtime and Android Library. Android Runtime consists of Core libraries and Dalvik Virtual Machine. Core Library provides Java Core Library with some functions. Dalvik Virtual Machine is a register based and runs classes compiled by the Java language compiler that has been transformed intodex format. Android system library that connects between application framework and Linux Kernel. It is developed in C or C++ language. They

provide service for the developers through the application framework.

D. Linux Kernel

Linux gives the Android developers a pre-built, already maintained operating system kernel to start. The kernel system service provided by Android inner nuclear layer is based on Linux 2.6 kernel. Operations like internal storage, process management and other core service are all based on Linux Kernel.

2. SYSTEM ARCHITECTURE

The application being developed is located in the workspace which contains the necessary packages, manifest file, .xml and .java files. Each page in the application is called an activity which has to be recorded in the manifest file. The event handling is done in .java file. When the application is developed and compiled, it is run on an android emulator during which a .apk file is created which is transferred to the mobile and the application is installed.
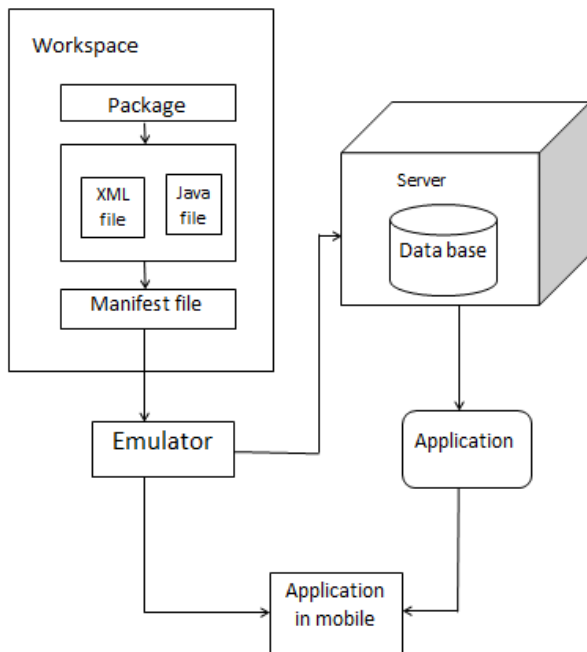
Fig. 2-System Architecture

3. PROPOSED SYSTEM

Android is an open source operating system used for smart phones and tablet computers from a host of manufacturers offering user access to Google's own services like search, YouTube, Maps and more. Dramatic breakthrough in processing power along with the number of extra features included in these devices has opened the doors to a wide range of commercial possibilities. One Application that falls into this category is our 'Employment Exchange

Application' developed for Google Android Phones.
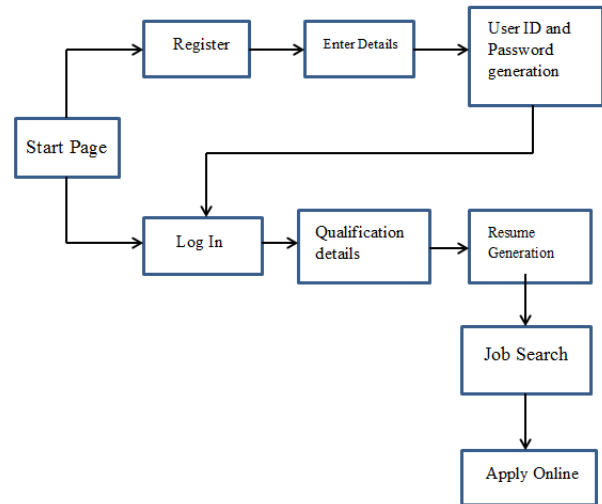
Fig. 3-Process Flow

A. Tools Required
- ❖ Android SDK Tools.
- ❖ SDK platform Android 4.4(API 19). API level shout be set to higher. So that it can run in all environments.
- ❖ Eclipse IDE In this paper we are using Eclipse 4.2(Juno).
- ❖ ADT plug-in Android Development Toolkit is a plug-in for the Eclipse IDE that is designed to give powerful, integrated environment to build Android Applications.
- ❖ AVD Manager We create Android Virtual Device and connect hardware devices, on which we can run our applications.

B. Process Flow
1. New User

When we get into this application as shown in Fig.2, we would start up with login page where there could be two options like login and register. If it is a new user he/she should click on to the register button. Therefore user can enter their own personal and contact details. After completing the process user will be provided with a unique ID and password.

2. Registered User

In the case of registered user, he/she can login with his/her own unique ID. Also user can view the details else they can update their details if needed. In case the user forgets the password an OTP will be sent to the mobile with which it can be recovered.

3. Resume Generation

The application has an additional feature to auto generate a resume for the user. In this process the user must provide all the qualification details and skills with

which a resume is formatted and forwarded to their mail in a .pdf format.

4. VIA Mail/SMS

This application also has the possibility of viewing their jobs not only by viewing this app every time, but also the information can be passed via Mail and through SMS (Short Message Service) regarding the Job fairs and interviews for jobs that match their qualification.

5. Additional Features

The application also contributes to provide continuous notifications regarding the government aided social events like, medical camps, awareness campaigns, and other public related service camps. Using the information users can also volunteer for blood donation, fund raising and other social events.

6. Server Connectivity

An important process involved during the development of the application is the server connectivity to the mobile user. The android app consists of SQLite database driver which can be used to store and fetch data.

VI. CONCLUSION

Our objective behind this paper is to provide Employment Exchange System more user friendly. The Job search system provides the users with the ease of registering for jobs without having to wait in long queues in government employment office. And the application also invokes people to try for government jobs and not alone prefer private jobs.

ACKNOWLEDGEMENT

REFERENCES

[1] "*Android Mobile Application Build on Eclipse*", volume 4, Issue 2, Feb 2014,by Garima Pandey, Diksha Dani.

[2] "*Job search system in Android environment-Application of Intelligent Agents*", volume 2,No.3,May 2012,by Salathiel Bogle and Suresh Sankaranarayanan.

[3] "*Research and Development of Mobile Application for Android platform*", volume 9, No.4, 2014, by Li Ma, Lei Gu and Jin Wang.

[4] "*Challenges and Issues in Android App Development-An Overview*", volume 5,Issue 1,Jan 2015,by Deepa v.Jose, Lakshmi Priya C, G.Priyadharshini, Monisha Singh.

"Academia" - An Android Application

[1] Janani.K [2] Kiran Berlene Marita.J.J [3] Sandhya.S,
[1][2] UG Student [[3] Assistant Professor (O.G),
[1][2][3][4][5] Department of Information Technology,
Valliammai Engineering College, Kancheepuram District, India.
[1] jananikannan995@gmail.com [2] kiranberlene09@gmail.com [3] sanrsn@gmail.com

Abstract:-- **An Android is a mobile operating system (OS) based on the Linux kernel. In the current scenario, the students are the larger users of Android Applications. The proposed system is an "Android Application for Engineering Institutions" which allows the students to view and download their Question Banks and get notified about both the technical and non-technical events at the college. The Application is developed using Android Studio with the help of Android SDK tools. The MySQL Server is used for storing the data. It helps the users to know about the most possible questions on each chapter for the Assessments as well as the University Examinations. Key Links are provided, which directs the users to the website where the study materials are available. It also helps in knowing the performance of the students, once the assessment details of each student has been uploaded in the Institution's Academic Portal. This app facilitates the student community to have an easy access to the study materials and it can be extended by various academic institutions in future.**

Keywords—Android Application, Android SDK Tools, SQLiteDatabase.

I. INTRODUCTION

Android is an operating system that powers all the android devices. It is the heart of all the android mobile devices. It is an Open source technology based on Linux platform. There are numerous android apps developed. It could be downloaded for free of cost. The proposed work mainly uses the android platform to develop an application compatible with all android devices. This application enables the users of various departments of engineering disciplines to know the unit-wise important questions. This would help them to have a quick glance during their last minute preparation. In addition to this, the users are made aware of the upcoming news and events of the various departments in college. The users can also view the assessment details, once the performance of students in various assessment periods has been uploaded in the Institution's Academic Portal.

The major goal of the proposed system is to develop an android application that would retrieve and display the question banks from the database as per user's request. The question banks for various Departments are made available from which students can select according to their requirements. Students are free to download the question banks for their future use. This is achieved by
(a) Creating a backend application of the system which would be a web application that allows the authorized admin people to upload the necessary data items
(b) Implementing and deploying a database that would host the captured data

(c) Developing a frontend application which would be an Android application that would interact with the database with the web service and displays the data retrieved on the Android device.

II. LITERATURE REVIEW

Google Play store has one million Android Applications published, among which 50 billion applications were downloaded. In the corporate world, managers face a lot of difficulties in managerial decision making. The decisions taken by management are vital as it could make a company either overtake or stay within reach of its competitors. At times managers would tend to have voluminous information which would be tedious for them to extract the summarized corporate data for decision making. Delayed decisions would lead to organizational loss. In order to overcome that, the author developed an android app to facilitate quick managerial decisions

[1].The workers working at offices capture the data and store it in a central database. The managers access the data from the central database using android app. Android based ERP system helps to maintain the data of ERP at centralized location. The author had developed an android app that allows the users to access the data from anywhere in the world.

[2].In this app the access includes both storage and retrieval of data for the authenticated users. In addition to this the author provides the facility of tracking the office personal (Sales Person) using GPS. The location of the

personal is sent to the server who can be tracked on Google Maps. Manager can easily retrieve data in the form of report which helps them in decision making. It provides live updates to Director and helps them in monitoring the business and undertaking the strategic decision making.

"I am in ASEAN" is a learning media app that enhances the knowledge about the ASEAN community. The author mainly focuses on the children and the youth to get a vast knowledge of the member countries of the ASEAN community.

[3]. The app is an alternative to learn about ASEAN community. It comprises 5 topics – ASEAN Community, their countries, flags, costumes and the way they greet. It attracted the children with the colorful images which urges them feel joyful and sounds interesting for them to repeat with the app. The author had developed an android application.

[4] Which manage student details like Personal, Academic and Placement details and keeps them updated with the latest events in the college. The app focuses on staffs, students and parents. The app automates the procedure of managing and controlling the student details, thereby reducing the overhead of managing paper documents. The users need to register and login with the proper credentials in order to upload their details. The students need to register at the beginning of every semester. There is a discussion forum available where the teachers can upload notes, respond to the queries of the students. The students are also given access to upload the notes. The notices are uploaded and broadcasted to the respective classes alone. Attendance of the students are generated as reports and sent to the students and their parents. To avoid the difficulty of searching for a good college and to clear the entrance examination the author has proposed an android app – "My College Finder"

[5] This app facilitates the search and helps the users to find their appropriate college. It reduces the effort, and saves time and money spent on the search for the appropriate college. The list of colleges are displayed to the user according to the higher secondary cutoff marks. The students have to register and login, after which they are mandated to fill in their entrance examination's score along with their desired engineering stream. With the help of standard SQL Queries the list of colleges are fetched from the SQL DATABASE. On selecting a particular college, its location is provided to the user. "Scribd"

[6] Is an online library app that enables the students to find millions of documents and books relatively important for their studies. It's a reader's paradise. The world wide data are organized according to their topic. They could find their favorite books in all genre. Millions of books and comics are available for endless reading. Audio books are also available. Data are easily shared by people from all over the world. Monthly membership is provided for the users and it is of automatic renewal unless and until the users cancel the membership. "EasyBib"

[7] Is an app for searching books. The majority of the app users are the college students. In this app, they could get the right citation just by entering the title of the book. The citations are added by logging in with the Easy Bib. com's account from the mobile devices. Those citations are perfect and accurate as it is checked by Librarians and Teachers from various Universities.The "EasyBib" app includes various citations styles of Harvard, ASA, AMA, CSE and ACS "Todoist"

[8] Is a to-do list app that helps the users to stay organized. It features a very simple, day-based interface so it helps to wake up, check details, and see what the users need to do that day. This is great for keeping assignments in order and users can tag tasks. Users can know exactly what it's for. There is also a desktop application so users can sync the tasks between platforms. It's simple, effective, and it works. "Udemy"

[9] Wants to be a university replacement app. And it's a valuable resource for college students. It has classes that can freshen up on topics and also helps to learn new ones. In college, students may need a little extra help in learning some concepts in some areas and "Udemy" is perfect for that sort of learning. The sessions can get a bit expensive but there are plenty of cheaper and free ones to choose from. It's worth a shot if the users need a little supplemental learning."Feed.ly"

[10] Allows the users to be updated with the news and events all around. College students always want to know everything, but it may be quite difficult to do because of constant deadlines, part-time jobs, overdue essays, exams, etc. Feed.ly is here to help the students in dealing with the above problems. This is one of the best RSS aggregators. It lets the users/students to consolidate all news into one feed. Helps to check the news, and get notified every time something interesting appears. On the search for the appropriate college. The list of colleges are displayed to the user according to the higher secondary cutoff marks. The students have to register and login, after which they are mandated to fill in their entrance examination's score along with their desired engineering stream. With the help of standard SQL Queries the list of colleges are fetched from the SQL DATABASE. On selecting a particular college, its location is provided to the user. "Scribd"

III. PROPOSED SYSTEM

A. System Components

The system architecture of the proposed system is shown in *Fig.1*. It includes the college source for the app, and the server. A web service is used to communicate between the server and the app. A PHP page is created to feed the college data into the database which is present in the server. The page contains registration facility which enables only the authenticated users to login. Only the authorized users have access to the data in the server.
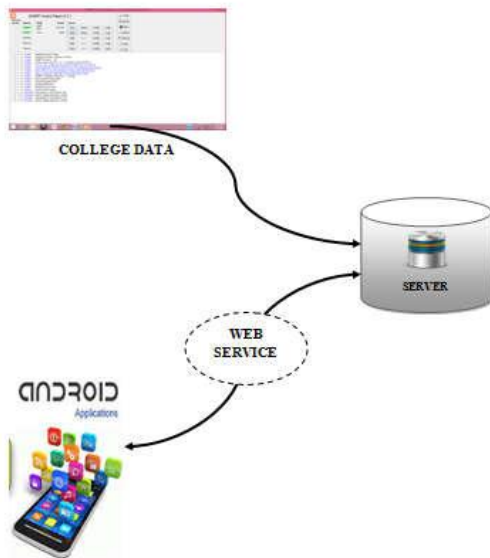
Fig.1. System Architecture

The notification of news and events of the college remain public where as the question banks of various departments allow only the authorized users to have access.

B. Module Description

The modules of the proposed system are represented in *Fig.2*. There are 3 different modules namely - Question Banks, Key Links and Notification. The home page allows the users to select their graduation (U.G / P.G). According to which they will be directed to the question banks page.
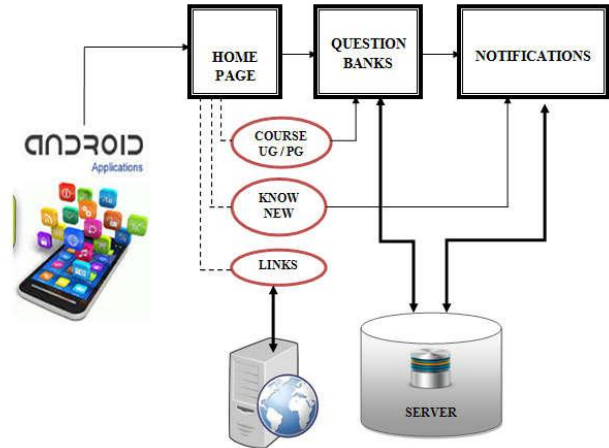
Fig. 2. Modules of Proposed System

In *Fig.3,* the HOME Page of the app is shown. It contains the list of modules as menus. On clicking those menus, the users will be directed to the corresponding module's activity screen.
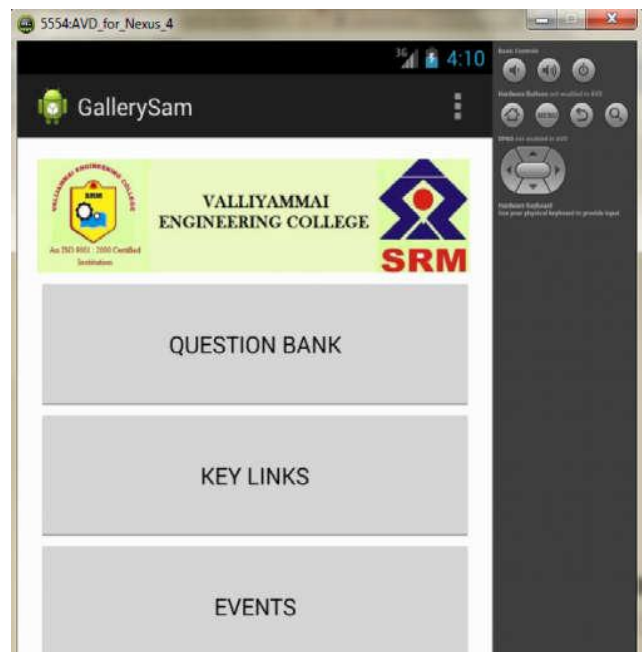
Fig.3. Home Page

The question banks are uploaded by the authorized staff to the database present in the server. To get authenticated, staffs need to register themselves in PHP page at the backend of the system. They are requested for their name, email-id and a password. Once they register, their credentials are saved at the SQLite Database. The staff need to login with the registered credentials, otherwise they would not be given access for uploading data. It has been considered that the question banks should be of PDF format. The question banks are provided for the regulations R2008 and R2013 as per Anna University

syllabus. The view of one of the question banks is shown in *Fig.4*. The concerned college students can download the question banks through proper credentials.
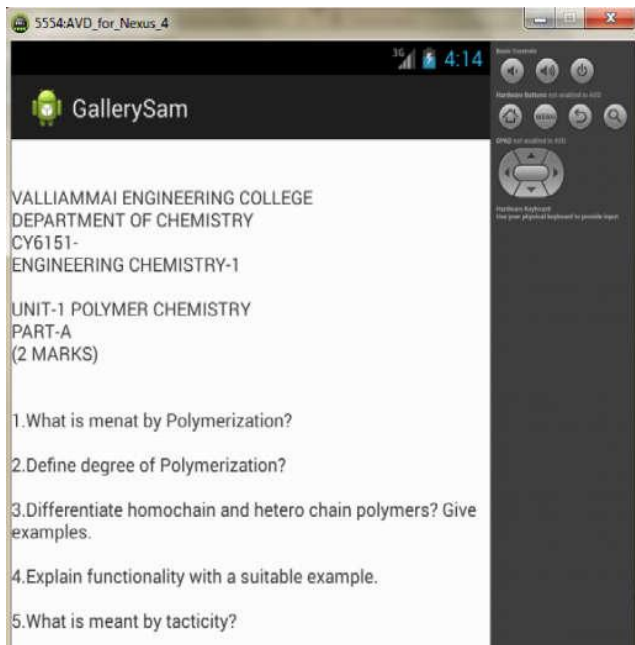
Fig.4. Question Bank

In *Fig.5*, the key links provided leads to the websites having study materials related to the question banks uploaded. There are number of websites to refer with and we have analyzed and chosen the best among them. The key links which are given in the app includes both the study materials and university links. Links of study materials are directly given with which it is the website's responsibility to filter the request and give the required materials. The website includes both the Anna University's official website and their Web Portal. The official website gives the information about the date of examination, date of results, conferences, journals and other important details related to academics. The Web Portal in general has the complete details of the students. Their assessment marks and attendance details. At the time of the University examination the University calculates the internal marks according to the assessment marks and attendance of each individual and uploads in the Web Portal. The students can login with their register number and date of birth and get to know about their internal marks.
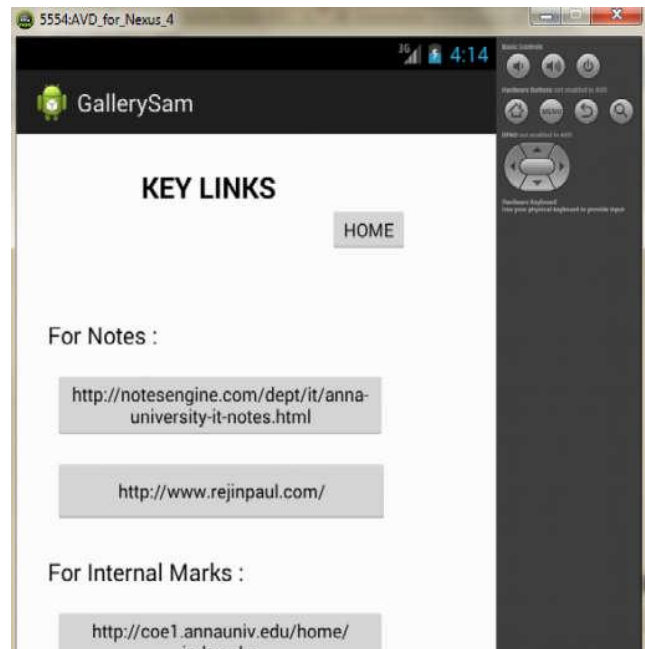
Fig.5. Key Links

Notification module tends to have the latest news and events of the college. It includes both the inter and intra events. Those events may possibly include Symposium, Conferences, Seminars, Webinars, Guest Lectures, Workshops, Value Added Courses, Cultural programs, etc.,

IV. SYSTEM IMPLEMENTATION

The major goal of the proposed system is to ease the use of college sources like question banks and to provide easy updating of the upcoming news and events. The proposed system facilitates these requirements. The Table I is a brief detail about the implementation specifications.

TABLE I. IMPLEMENTATION DETAILS

Software Used	API Level	Version	Application
Eclipse	17	Jellybean	
Android Studio	23	Lollipop	Native and HTML

The backend is created with the help of PHP so that the college staff can easily update the information to the students. In this page only the authenticated staffs are allowed to update the information. The staffs need to register with proper credentials and use the same to log-in the PHP Page. Once they are properly authenticated, they can upload the subject materials and notification messages to the database. They also have access to remove the

outdated notifications. The frontend is created with the help of android studio in conjunction with java programming language. The frontend communicates with the database using web service. This web service retrieves the data from the database and displays it to the user. The user pertaining to the college that has created the application can alone access the study materials. The notifications remain common to all the users. A part of manifest code is given below,

Manifest File –

```
packagecom.example.jak;
importandroid.os.Bundle;
importandroid.app.Activity;
importandroid.content.Intent;
importandroid.view.Menu;
importandroid.view.View;
importandroid.widget.Button;
public class MainActivity extends Activity
{
Button b,b1,b2,b3;
@Override
protected void onCreate(Bundle
savedInstanceState) {
super.onCreate(savedInstanceState);
setContentView(R.layout.activity_main);
b=(Button)findViewById(R.id.button1);
b1=(Button)findViewById(R.id.button2);
b2=(Button)findViewById(R.id.button3);
b3=(Button)findViewById(R.id.button4);
}
public void qb(View v){
Intent i=new
Intent(MainActivity.this,One.class);
startActivity(i);
}
public void not(View v) {
Intent i=new
Intent(MainActivity.this,Two.class);
startActivity(i);
}
public void kl(View v) {
Intent i=new
Intent(MainActivity.this,Three.class);
startActivity(i);
}
public void gal(View v) {
Intent i=new
Intent(MainActivity.this,Four.class);
startActivity(i);
}
@Override
publicbooleanonCreateOptionsMenu(Menu menu)
{
// Inflate the menu; this adds items to the
action bar if it is present.
getMenuInflater().inflate(R.menu.main, menu);
return true;
}
}
```

V.   CONCLUSION

This app enables the users to could have an easy access to the question banks instead of getting into the college website and searching for the question banks of their desire. It also narrows down the search for the notes and materials required for the subjects. The notifications that are given through the app are very much advantageous. It helps the users to be updated with the college events. The development of this application can be extended with additional features that are more useful for the students.

REFERENCES

[1] Seth Y.Fiawoo andRobert A.Sowah, "Design and Development of an Android Application to Process And Display Summarized Corporate Data", IEEE 4th International Conference on Adaptive Science & Technology, 2012.

[2] JulieePanse, Monika Memane and Monika Bagul, "Android Based ERP System, International Journal of Computer Science and Mobile Computing", Vol. 4 - Issue 1, January 2015.

[3]   YupadeeSiboopimpa,   ThonthanChaipranitan, SirapatChiewchanwattana andKhamronSunat, "I am in ASEAN : A Learning Media Application On Android Operating System", Third ICT International Student Project Conference,2014.

[4]ManasiKawathekar,   KirtiK.Bhate   and PankajBelgoankar,"An Android Application for Student InformationSystem", International Journal of Advanced Research in Computer Engineering & Technology, Volume 4, Issue 9, September 2015.

[5] AnkitKarandikar, Sameer Kalantre, AjinkyaKohok, Deepak Ignale and RutujaKulkarni, "My College Finder-An Android Based Application", International Journal of Advanced Research in Computer Engineering & Technology, Volume 2, Issue 1, Januray 2013.

[6]https://play.google.com/store/apps/details?id=com .easybib.easybibandroid&hl=en

[7]   https://play.google.com/store/apps/details?id=com .scribd.app.reader0&hl=en

[8]   https://play.google.com/store/apps/details?id=com .todoist&hl=en

[9]https://play.google.com/store/apps/details?id=com
.udemy.android&hl=en

[10]https://play.google.com/store/apps/details?id=com
.devhd.feedly&hl=en

"Complex Event Processing Application for Smart City"

[1] Poonam Hajare, [2] Rajeshri Kachole, [3] Ankush Deshmukh, [4] V. S. Mahalle
[1][2][3] Third Year B. E., [4] Assistant Professor
[1][2][3][4] Department of CSE, Shri Sant Gahanna Maharaj College of Engineering
(SSGMCE) Shegaon-444203
[1] hajarepv123@gmail.com, [2] rajeshrikachole2012@gmail.com, [3] deshmukhank@gmail.com,
[4] vsmahalle@gmail.com

Abstract:-- **Smart city is defined as the ability to incorporate multiple technological solutions in secure fashion to manage the city assets. Emerging needs to make cities smarter, as proposed by IMB in Smarter planet program which mention the collaboration between the different city agencies such as health care agencies, transport agencies, various govt. agencies etc. This collaboration in smart cities will generate a huge data set. By applying complex event processing on these data set we can solve the various real time problems related to above mentioned agencies. Event processing is nothing but processing of the past or real time dataset to generate the new conclusion. These conclusions are helpful to find the opportunities or threats about any particular event. Complex event processing uses data mining to process the given dataset and give the essential event pattern as output. There is very less research work in India for calculating required conclusion from multiple dataset. This proposed system will help to propagate the optimum conclusion.**

Index Terms: **Event Processing, data mining, smart city, ID3 Algorithm.**

I. INTRODUCTION

Data Mining is process of extracting an event pattern from large database. Event is nothing but changing from one state to another. Suppose there is a car in showroom at that time the state of that car is 'SELL'. After purchasing that car by customer, the state of that car will be 'SOLD'. The main goal of Data Mining is to extract relevant event pattern for performing various operations. In data mining term event processing refers to any type of meaningful computation on event for getting efficient output. The incoming source data can be come from any type of data streaming sources. To store such a large amount of data with various type is not possible by DBMS(Data Base Management System). For that there are two type of system:

1) Data Stream Management Systems (DSMS)
2) Complex Event Processing (CEP) systems.

In this paper we are focusing on Complex Event Processing as an Application for 'Smart City'. To get quick response Complex Event Processing is new paradigm to process on data stream. CEP is used to analyze event pattern and it will also help in business side for better communication with IT and service department.

II. CONSEQUENCES FOR SMART CITY:

The goal of smart city building is to improve quality of life by using various technologies to improve efficiency of services. Smart city uses information and communication technology [ICT] to enhance quality, performance and interactivity between citizen and government. Smart city criteria depend on six different aspects, and these are as follows.

❖ Economy
❖ People
❖ Governance
❖ Mobility
❖ Environment
❖ Living.

Out of above mentioned six criteria, for complex event processing we are focusing on mobility i.e. referring to accessibility of modern and fast transport system.

III. LONDON CASE STUDY:

In London, the case study has been done to examine and analyze the data coming from different data sources like traffic data, data about event happening in city, weather condition etc. And to find the reasons and

solutions about traffic disturbances by applying complex event processing mechanism. As per the case study held in London it is seen that we can correlate data between traffic and various events happing in city (like cultural event). The step in above case study was to identify incoming data sources.

*Traffic data source***:** Bing map, Traffic for London.

*Weather conditions***. Sources:** Weather Underground, Yahoo! Weather, AccuWeather, etc.

*Events happening in the city***:** Eventful.com, upcoming.org, last.fm, zvents.com, socialevents.com.

All the data coming from these resources is stored in single database. The dataset obtained by these resources are dynamic in nature. That means they are updated after specified time interval. By analyzing the dataset obtained from the above data sources various problematic decisions like arranging the events, time and place about event and decision about rout of travelling etc.
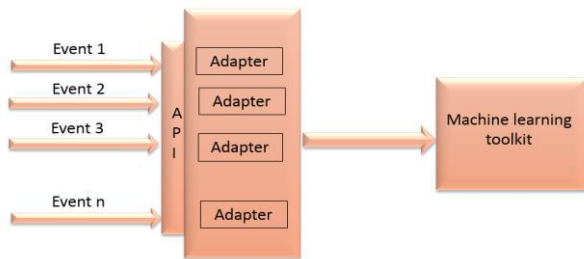
FIGURE (1) Data Sources

IV.   PROPOSED METHODOLOGY:

Road traffic conditions in India are getting worse day by day. The average number of vehicles in India is growing at the rate 10.16 percent annually, since last five years. Spending hours in traffic jam has become part and parcel of metropolitan life style, leading to health and environmental hazards. The vehicle penetration in metropolitan cities like Mumbai is suffering from about 590 vehicles per Km of road stretch and Bangalore with around 5 million of vehicle ply over a network that extends barely up to 3000kms. There could be an approaches to solve this problem which is nothing but our 'Smart City' project will be based on these 3 scenario.

- ❖ Traffic Data
- ❖ Weather Report
- ❖ Cultural Event
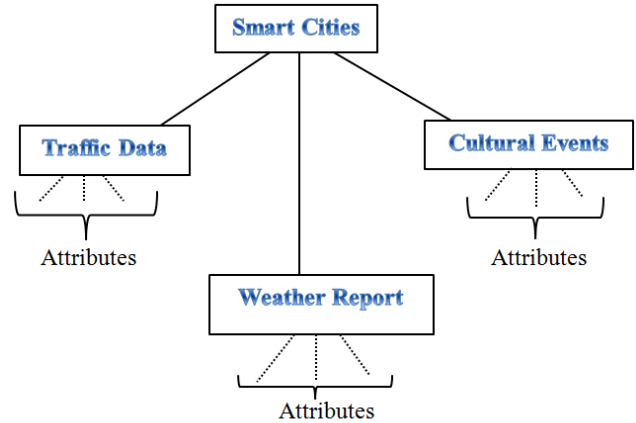
Figure. Decision Tree Of Complex Event Processing.

FIGURE (2) Decision tree for CEP

In case of traffic data, let's take an example, when we are driving a car then, we may need information regarding speed breaker, turning, traffic on rout etc. In case of weather report, we may need information about today's weather, forecast of next two days weather etc. In case of cultural event, we need information about functional hall. By using complex event processing we can combine all events from dataset obtain from given resources and generate conclusion for current and future incident. Our proposed methodology has four basic steps. And they are as follows.

- ❖ Dataset Collection
- ❖ Data Processing
- ❖ Data Mining
- ❖ Result Set Generation

V.   DATASET COLLECTION:

In dataset collection we need to collect datasets about traffic, which we can collect from various sources like Google map, Bing map, various traffic related app databases etc.

Datasets related weather can be collected from meteorology department forecast, news, and various mobile apps AccuWeather etc. Dataset related to cultural events can be obtained from many cultural websites, some events are fix events, and information related events can be obtained from social media and newspapers. All datasets collected from various data sources is stored in a single database to maintain correlation between them and to make final conclusion from them.

VI.   DATA PROCESSING:

The data collected from various sources is not structured data. To make that data structured we need to apply various mechanisms like data selection, data

clearing, data transformation etc. As all the data collected in data collection step is not fully useful so we need to delete useless data. Data cleaning involve error removing from dataset. Errors like missing values or inconsistent data etc. We need to apply various techniques to remove these anomalies. Only data cleaning will not allow us to apply data mining techniques on datasets. We need to transform our dataset. We have to apply smoothing aggregation and normalization to transform our dataset.

VII. DATA MINING:

After removing all the anomalies from the collected dataset it's time to apply data mining techniques on it. Data mining can be done on any type of data i.e. text or media, only algorithm or approach varies. There are many algorithms like c4.5, k-means, Apriori and many more. In 1 clustering and association analysis are done for data mining. We perform data mining to discover the interested pattern. Only discovering interested Pattern is not sufficient, we need to remove the redundant patterns from the extracted data.

A. *Result set generation*:
Result set generation is the last phase of data mining. In this phase the virtual representation of discovered result is done. Data set which is extracted by using interested pattern is called as result set. This result set is used to generate the final conclusion.
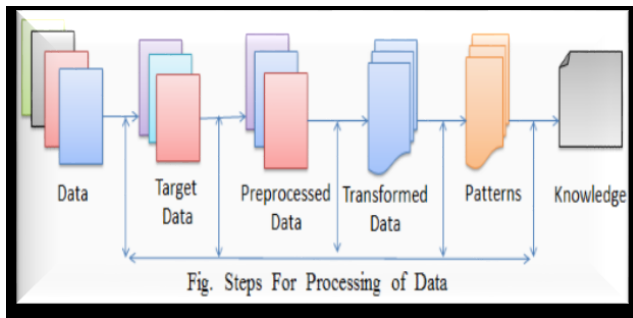
FIGURE (3) Steps For Processing of Data

VIII. ALGORITHM:

Step 1: Create a root node for the tree
Step 2: If all examples are positive, Return the single-node tree Root, with label = +.
Step 3: If all examples are negative, Return the single-node tree Root, with label = -.
Step 4: If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
Else
A = The Attribute that best classifies examples.
Decision Tree attribute for Root = A.
For each possible value, *vi*, of A,

Add a new tree branch below Root, corresponding to the test A = *vi*.
Let Examples(*vi*), be the subset of examples that have the value *vi* for A
If Examples (*vi*) is empty
Then below this new branch add a leaf node with label = most common target value in the examples
Else below this new branch add the sub tree ID3 (Examples (*vi*), Target Attribute, Attributes – {A})
End
Step 5: Return Root.
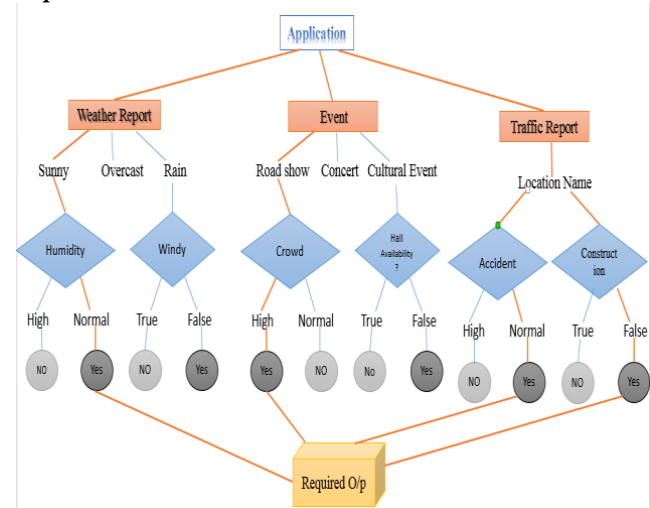
FIGURE (4) Decision Tree with Example.

IX. CONCLUSION:

By analyzing the today's various problems related to day to day life, like traffic problems, weather problems and problems like finding the place to arrange different events need to solve using today's technology. Today's need to upgrade our metropolitan cities to smart city for renovation of our standard of living can be solved by using latest technology. This paper emphasis a solution to our different problems mentioned above. The proposed technology help us to find real time and optimum solution to our problems. And also to enhance the standard of living.

FUTURE WORK:

We can increase our data sources to large number to focus on many more problems like mentioned in this paper. We can extend our work to various government departments, educational departments, public sector and many more to improve the standard of city and make it "SMARTER CITY".

Consider an example of Mumbai as a 'Smart City', we can collect traffic data of Mumbai fromGoogle and Bing map etc. and weather report from meteorology department, AccuWeather app etc. and data of cultural

event from the Social Networking Sites and other website. By collecting above data we can bind it into single unit. After that we can extract target data that is useful event pattern then apply some pre-processing on that data. For storing multiple event patterns we can useevent cloud. For getting appropriate event pattern, we can calculate entropy and information gain by using some formula. There may be common attributes in above mentioned events like location, time etc. by calculating entropy we can combine those attributes and store complete data in less memory.

Suppose, there is a medical emergency and there is more than one rout to reach the hospital then ambulance driver can search the shortest route with less traffic Suppose user want to search functional hall with less traffic and know about weather of next day then also this proposed methodology can be used.

REFERENCE

[1] Stonebraker, M., Çetintemel, U., Zdonik, S."The 8 requirements of real-time stream processing" ACM SIGMOD Record. 34, 42-47 (2005).

[2] Cugola, G., Margara, A."Processing Flows of Information" From Data Stream to Complex Event Processing. ACM Computing Surveys.

[3] Niblett, P.: Event Processing In Action. (2010).

[4] Wang, Q., Meegan, J., Freund, T., Li, F.T., Cosgrove, M.: Smarter City: The Event Driven Realization of City-Wide Collaboration. 2010 International Conference on Management of e-Commerce and eGovernment. 195-199 (2010).

[5] Giffinger, R., Fertner, C., Kramar, H., Kalasek, R., Pichler-Milanovic, N., Meijers, E.: Smart cities Ranking of European medium-sized cities. , Vienna, Austria (2007).

[6] Transport for London, Live Traffic Disruptions – Data Dictionary, (last accessed 1st September 2012), http://www.tfl.gov.uk/assets/downloads/businessandpar tners/data-dictionary-live-traffic-disruptions.pdf

[7] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten. The WEKA Data Mining Software: An Update.SIGKDD Explorations, Volume

Smart Education for Smart City: A Solution to School Dropout

[1] V. S. Mahalle, [2] Shyam Dhanuka, [3] Ashish Sharma, [4] Abhishek Karanjkar
[1] Assistant Professor Dept of CSE [2][4] Final Year UG Student,
[1][2][4] Department of CSE, Shri Sant Gahanna Maharaj College of Engineering
(SSGMCE) Shegaon Maharashtra, India [3] Hindustan Institute of Technology and Management, Agra, India
[1] vsmahalle@gmail.com, [2] shyamdhanuka21@gmail.com, [3] ashish2311sharma@gmail.com
[4] abhishekkaranjkar944@gmail.com

Abstract: -- School dropout is a major factor responsible for low quality education and lack of relevant skills among the students. The findings reveal that both the school and family related factors were responsible for the increase number of dropouts. In order to reduce the dropout rate of the students make education system more efficient the use of modern technology can definitely help using low cost solution that can be even deployed in the Government schools also for Smart City Application. The present study is analysis of dropout rate and the proposed solution to reduce the same.

Index Terms:— **Quality of education, low cost solution, school dropout, smart city application.**

I. INTRODUCTION

School Dropout is the most pervasive problem in the Indian Education System, multiple factors like unawareness of importance of education among the students and their parents, financial instability, abuse of girl child, irresponsible behavior of teachers, lack of accountability, and unavailability of transport facility leading to students dropping out of the school. Domestic and financial problems create a negative environment that affects the quality of education. Other factors like migration of parents from one city to another in search of new jobs and opportunities, discrimination of the children and complete absence of extra-curricular activities accounts for the high dropout rate.

When students experience failure in school, they become frustrated up alienated with lack of achievement and experience exclusion leading to eventual dropout. Further, girl student could dropout as a result of lack of sanitation facilities, etc. It is important to design a sustainable solution and strategies that could be adopted in order to help the dropouts. For this study dropout is defined as the children who takes admission in the school but leaves the school before the proper completion of the secondary education without transferring to another school. Dropout rate is the percentage of dropouts in a given year out of the total number of those enrolled in a program in the same year. Dropout can also be count as the total number of children who start a program and do not finish it.

Taking this into consideration the present paper includes all the major and minor factors responsible for the dropouts and provides an innovative solution that can be easily implemented to overcome this problem for Smart City. In a study in 2010, Reddy Anugula and Sinha found that more than 27 million children in India, who joined in Class I in 1993, only 10 million of them reached Class X, which is only about 37% of those who entered the school system [1]. Out of 100 children enrolled in Class I only 47 reached Class VIII, putting the dropout rate in primary and elementary schools at 52.79% [2]. In a study in 2011 Chugh found that about 33 percent of children were interested to study further if given an opportunity to continue the studies, provided additional coaching and special attention by the teachers of the school. Few were keen to complete secondary education [3]. Providing right type of education at right time will surely help the right people to become more productive and to the potential in them, build the necessary condition requires high school retention and the quality education. Therefore the quality education and reducing the number of dropout will prove as master key for the development of the smart city that how we utilize and develop this section of people.

II. DROPOUT RATE

In 2011, Chugh found that very few children complete elementary education and even fewer transits to secondary education. Despite a small proportion of children actually reach secondary education the dropout rates at secondary level are found to be very high. Though,

the available data does not give precise figures on how many enroll and how many actually dropout at secondary stage but the difference between children who dropout at upper primary and secondary stage may hint at the volume of dropouts. The dropout rate for Delhi stands at 36.04 percent compared to 56.71 percent at the all-India levels between classes I-X in 2007-08. It may further be noted that the dropout rate fell drastically by 10 percentage points in one single year i.e. 2005-06 and 2006-07 whereas between 2007-08 and 2006-07 it remained constant. Surprisingly in case of SCs, the dropout rates are lower than the general population in 2005-06 and 2006-07 whereas in 2004-05 and 2007-08 the dropout rates of SCs are higher than general population. This implies that the data on dropouts is very inconsistent and more so in case of Scheduled Castes. This raises questions about the quality of official data on enrolment and dropouts. But this does not hide the fact that the dropout rate is still very high.

Though, the dropout rates at secondary level are lower than elementary education but soon it may increase consequent upon persistent efforts to reduce dropout at elementary education. This necessitates studying the dropout phenomenon at micro level on who drops out and what factors contribute for dropout at secondary level. This paper presents results from a micro study of dropout from secondary schools that serve children of slums in Delhi. Before discussing the solution, an overview of a few root causes of dropout is given in the next section [3].

III. ROOT CAUSES OF DROPOUT

A. Children as Economic Potential

Children act as bread earners, are the economic support of the family. They are forced to stay as caretaker of their younger siblings if both the parents are working. Some type of social as well as parental pressure is also responsible that students have to act as economic potential of family and can't attend the school. Considering the importance of the family decision to send the children to school or to discontinue their studies, information should be collected from the families of the drop out children to look at their environmental, social and economic compulsions which may act against their continuation in school. Besides household income, the education level of the parents is also expected to influence the continuation of children in school

B. Migration of Parents

Parents those who are contract labors or farm labors or seasonal labors have to migrate occasionally due to financial instability, in search of new jobs, tribal festivals. Some illiterate parents are unaware of the importance of the education and the process to be followed to transfer student from one school to another one. In 2005, Ersado observed that parental education is the most

consistent determinant of child education. Higher parental education is associated with increased access to education, higher attendance rates and lower dropout rates [5]. Parents, who have attained a certain educational level, might want their children to achieve at least the same level.

C. Absence of efficient tracking of dropout children

There is complete absence for monitoring of the dropout students. Student has to leave the school according to their parents' jobs, migration; there is no record of the migrated students and no transfer certificate availability that restricts them from taking admission in another school. This obsolete and inefficient tracking should be eliminated.

D. Non Interactive Teaching System

Lack of activity based learning, old teaching methodology, language barrier, unskilled and untrained teachers, resistance to implement new technology, old teaching methodology, not updated modules, no incentive to teachers for implementing interactive system are some of the root causes of the dropout of students.

Proper environment should be created so that we can attract the students to the school. It has been repeatedly observed that low-achievers and students from low socio-economic backgrounds are at much higher risk of dropping out which could be due to several reasons such as inadequate parenting, inability to afford the educational expenditure, poor schooling infrastructure, de-motivated teachers, pressures to augment family income, accompanied by a view that schooling has limited economic returns, peers with low aspirations, poor nutrition and health, and too few role models in the community. Dropout decision also depends upon the academic performance of the student. In 2012, Rupon found that poor school performance, low attendance and late enrolment are likely to be act as signal for teachers that the student is more likely to dropout [4]. However, one of the key problems is that students don't usually dropout for a single reason. Multiple factors are at play and no single risk factor can accurately predict who will dropout.

IV. A REVIEW OF RELATED STUDIES

Over the past years appreciable research has been undertaken to find the root causes for the dropout and by using the relevant data from the Census of India, the NSSO and other sources. Multi-level statistics model that focused on the school efforts and individual efforts to address the dropout problems. Over the decades we have not yet been able to address the poor attendance of both teachers and students which has a direct correlation to the drop out ratios in schools across India. The studies are based on experimental studies examining particular aspect of the dropout. These studies proved useful in determining at risk students and the developing estimates of the extent of the dropout, symptoms of dropping out much before they actually dropout.

Researchers have consistently found out that the socio-economic pressure that is mostly measured by the parental education and economic condition, is a powerful predictor of dropout behavior [6] [7]. Very less number of children completes the elementary education and even fewer transits to the higher secondary education. Despite a small proportion of children actually reach secondary education the dropout rates at secondary level are found to be very high. Though, the available data does not give precise figures on how many enroll and how many actually dropout at secondary stage but the difference between children who dropout at upper primary and secondary stage may hint at the volume of dropouts. The dropout rate for Delhi stands at 36.04 percent compared to 56.71 percent at the all-India levels between classes I-X in 2007-08[3].These studies help in identification of the factors related to the child and household and those related to their respective schools and is influenced by both the social and academic experiences of students. In 2002, Brown and Park found that good family income makes it convenient to provide more resources to support children's education like access to better quality schools, private tuitions and more support for learning within home. Therefore, poverty still remains as one of the significant causes of children dropping out of school [6].

V. PROPOSED SOLUTION

One of the solutions for tackling transportation problem is either providing movable schools to them like School-On-Wheels, an amazing mobile-van platform for reaching out to far off communities' children founded by Sachin Joshi and its implemented in Nashik City of Maharashtra[8], young kids of around 7-8 years who cannot reach to school and hence such platform can foster a little hope for them in their dark lives, children taught about maintaining health hygiene and made aware about daily life needs it will also create awareness about the importance of education among the student. Another option is to promote, provide the 'Bicycle Sharing' concept in Smart City so that they can reach to the nearest school. To tackle with lack of awareness and caretaking of younger siblings Self Help Groups can help to a good extent.

Out of these we propose to address solution to major problems directly related to dropouts of the student also it includes solution to teachers' and students' attendance and accountability. The detailed solution to these problems is mentioned here

We could design an Android or iOS application that helps to track the dropout children, this will be for the citizens and college students and volunteers, as it is a social issue we should take involve the society to help us. Application will verify the volunteers, citizens those who are interested to help the society and we will have complete database of such volunteers, citizens and it is

proposed to divide the volunteers' area wise and allot them specific area of the City. The complete process will include following points.

A. Identify

Starting from the month of June every year, verified volunteers, citizens, college students with the help of NGOs will search for the 'out-of-school' children. In the application approach we will include some basic question related to children that help to collect the information and prepare the databases and following questions will be asked to every child like i) Child name ii) Age iii) Address iv) Why you are out-of school v) How many days since have not attended the school and why? Along with these we will click a snap of student. We will further share these data with the NGOs.

B. Connect

In this stage the collected data goes to the City Administration, who shares this information with the local NGOs of the specific area based on GPS/location of these students. Now it is responsibility of the NGO to analyze the relevant data, to convince, consult the child and his parents and to make a reliable network.

C. Counsel

The local NGO will visit to the specified location of the children and NGOs will counsel with the child and their parents regarding the importance of education, how education will help their child to fulfill his dreams, government facilities available and some other benefits related to the education.

D. Enroll

After counseling these students will be enrolled in the nearest public school by the NGOs. Students will be provided with all the essential facilities.

E. Track

Now it becomes important to keep track of the enrolled students, a simple mobile application for teachers can be developed to capture the attendance everyday Properly maintaining the children information and keeping the track of enrolled students will surely solve this major problem of dropout.

Fig. 1 Proposed Solution
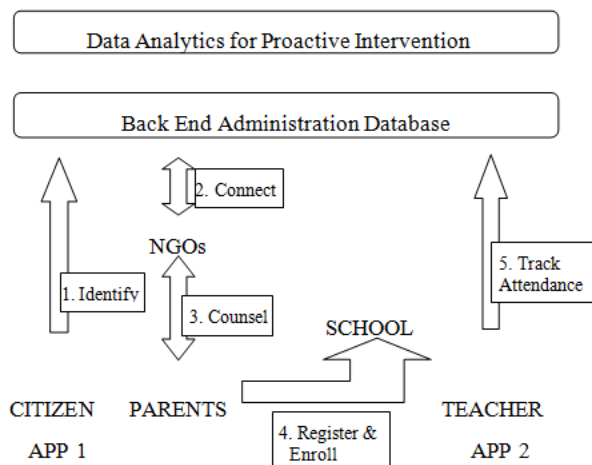
REFERENCES

[1] Reddy Anugula N and Sinha Shantha, School Dropouts or Pushouts? Overcoming Barriers for the Right to Education, NUEPA, 3 (2010)

[2] InfochangeEducation,http://infochangeindia.org/education/news/52-dropout-rate-in-indias-primary-and-elementary-schools.html, accessed on February 15, 2016

[3] Sunita Chugh, "Dropout in Secondary Education: A Study of Children Living in Slums of Delhi" NUEPA Occasional Paper series, February 2011

[4] Rupon Basumatary , "School Dropout across Indian States and UTs: An Econometric Study" International Research Journal of Social Sciences, vol. 1(4), pp. 28-35, December 2012

[5] Ersado, L. "Child labour and schooling decision in urban and rural areas: comparative evidence from Nepal, Peru and Zimbabwe. World Development", 33(3), pp. 455-480, 2005

[6] Brown, P. H., & Park, "Education and Poverty in Rural China. Economics of Education Review", vol. 21, pp. 523-541, 2002

[7] Pong, S.-L., & Ju, "D.B. Effects of change in family structure and income on dropping out of middle and high school" Journal of Family Issues, vol. 21(2), pp. 147-169, 2000

[8] DNA,http://www.iamin.in/en/nashik/news/educator-nashik-wants-all-children-above-six-school-60527,May 24, 2015 accessed on January 29, 2016

F. Awareness

Society is involved to find out the children not pursuing the education, so most of the people in the city will be aware of the need to enroll students into schools and will actively participate in this campaign. Parents will be aware of various government schemes and incentives, advantages of enrolling their wards into school and it will prove helpful for them to emphasize the need for their child's education

G. Real time measures

Teacher and student attendance will be available on real-time basis. Real-time availability of attendance data will help track various parameters such as school and teacher quality, mid day meal expense tracking, health tracking during outbreak of diseases to ensure availability of medicines at Public Health Centers. Incentivization programs can be implemented with more transparency and outcome-focus.

H. Quality Improvement

Push model can move to a pull model over time with better data availability and program effectiveness. The improvement in quality of both educations as well as of teacher and student is likely to create a major impact while moving towards smart education.

VI. CONCLUSION

Using the above mentioned methodology to solve the dropout problem will surely going to help the government as well as the citizens of the city. Improving our elementary education system for the betterment of society will surely help us to build Smart City by the providing Smart Education to every citizen of society.

Survey on Privacy Preserving Data Mining: Techniques and Application

[1] Vivek Jawanjal, [2] Shubham Pawar, [3] Akshay Kamble, [4] Prof. V.S. Mahalle
[1][2][3] BE. Final Year CSE. [4] Computer Science & Engineering,
[1][2][3][4] Department of CSE, Shri Sant Gahanna Maharaj College of Engineering
(SSGMCE) Shegaon Maharashtra, India
[1] vivekjawanjal3@gmail.com [2] shubhampawar006@gmail.com, [3] akshaykamble247@gmail.com
[4] vsmahalle@gmail.com

Abstract:-- **Data mining is the process of extraction of data from large amount of database. One of the most important topics is nowadays in research community is Privacy preserving data mining (PPDM). The goal of privacy preserving data mining is to extract relevant knowledge from large amounts of data while protecting at the same time sensitive information. To solve such problems there are number of methods and techniques have been proposed for protecting sensitive information. This paper provides a wide survey of different privacy preserving data mining algorithms and A tabular comparison of different technique is presented.**

Keywords—**Data Mining; Privacy Preserving; Sensitive Data; Privacy Preserving Techniques;**

I. INTRODUCTION

Data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories. By performing data mining, interesting knowledge, high-level information can be extracted from database.

The main goal of data mining is to extract knowledge and new patterns from large data sets in human understandable structure. For data mining computations we have to first collect data without much concern about privacy of data. Because of privacy concerns some people are not giving right information. Therefore Privacy preserving data mining has becoming important field of research. In order to make a publicly system secure, we must ensure that not only private sensitive data have been trimmed out, but also that certain inference channels should be blocked with respect to privacy.

The paper is organized as follows. In Section 1, we give the basic concept of data mining and privacy. In Section 2, we describe Privacy Preserving data mining with its framework. Section 3 provides some of the Application in this field. Section 4 contains different techniques with their limitations. A tabular comparison of different techniques of PPDM given by different authors is shown in section 5. And finally we conclude in Section 6.

II. PRIVACY PRESERVING DATA MINING

Privacy preserving data mining (PPDM) deals with protecting the privacy of individual data or sensitive knowledge without sacrificing the utility of the data. People have become well aware of the privacy intrusions on their personal data and are very reluctant to share their sensitive information.

In figure 1, framework for privacy preserving Data Mining is shown. Data from different data sources or operational systems are collected and are preprocessed using ETL tools. This transformed and clean data from Level 1 is stored in the data warehouse. Data in data warehouse is used for mining. In level 2, data mining algorithms are used to find patterns and discover knowledge from the historical data. After mining privacy preservation techniques are used to protect data from unauthorized access. Sensitive data of an individual can be prevented from being misused.
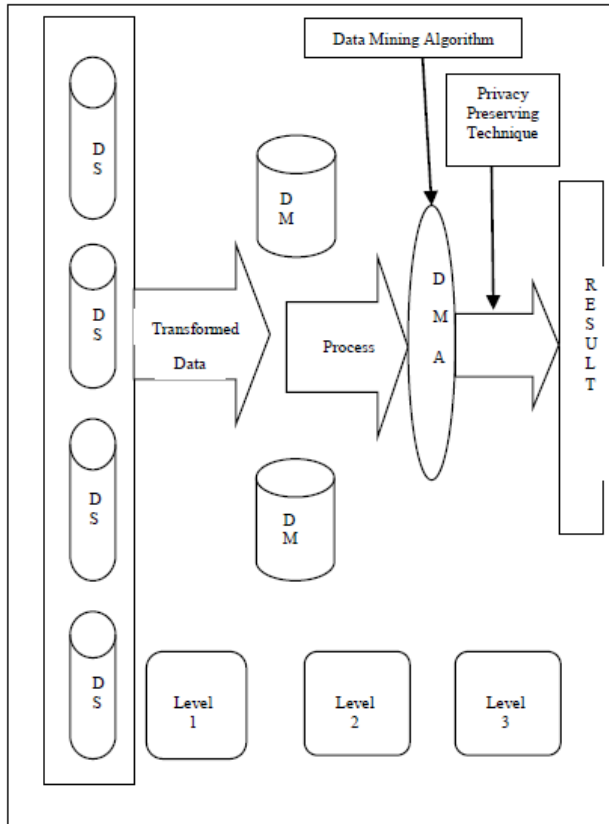
Fig. 1 Framework of privacy preserving data mining

III. APPLICATION

A. Medical Database:-Scrub & Data fly System. The scrub System design for identification of medical record which in the form of text data. Medical records are typically in the form of text which contains information about patient. Data fly was design to avoid identification of the subject of medical record.

B. Bioterrorism Application: - In Order to identify disease affected area it is necessary to track incidences of those common diseases . That Medical data would need to be report to public health agency. However there are some diseases that are not reportable disease by law. E.g. Common Respiratory Diseases.

C. Homeland Security Applications:-

- ❖ Credential validation problem
- ❖ Identity theft
- ❖ Web camera surveillance
- ❖ Vide-surveillance
- ❖ The watch –list problem

IV. PRIVACY PRESERVING DATA MINING TECHNIQUES

In this section, we focus on different privacy preserving data mining techniques such as a cryptographic technique, blocking based, hybrid technique etc.

A. Data Perturbation

Data Perturbation is a technique for modifying data using random process. In this technique by changing the sensitive data by adding, subtracting or any mathematical formula. Different data types are used: character type, Boolean type, classification type and integer. In this preprocess the original data set required. According to the preprocessing of data it classified into attribute coding and obtains coded data sets. Discrete formula prescribed is: A(max) – A(min)/n = length. Where A is continuous attribute, n is number of discrete and length is discrete interval length. Data perturbation cannot reconstruct the original data; it can reconstruct only distribution data.

It is important to secure the sensitive information therefore data perturbation play an role for preserving sensitive information. Distortion can done by different method applying such as a by adding unknown values, adding noise etc. In some technique it is difficult to protect the sensitive information, to overcome this problem new algorithm are developed which is able to reconstruct the distributed data.

An new approach is develop, It is based on singular value decomposition (SVD) and sparsified singular value distribution (SSVD) technique. In this matrices is introduced to compare the original datasets and distorted dataset. SSVD used for data utility, SVD method add noise to make perturbed data. Perturbation has drawback is that each data dimension is reconstructed independently.

B. Blocking based technique

In blocking based technique there is a sensitive classification rule, it is used to hiding the sensitive data from outsiders. In this technique two steps are used for protecting the privacy. First is to identify transaction of sensitive rule and second is to replace the known values to the unknown values (?). In this, there is a scanning of original database and then identifying the transactions supporting sensitive rule, and then for each record or transaction, algorithm replaces the sensitive information with an a unknown values. This techniques is used those applications that can save unknown values to the some kind of attribute. Generally we can hide the actual values by just replacing the '0' by '1' or '1' by '0' or with unknown values (?) in transaction. Blocking based techniques aim is that to preserve the sensitive data from

unauthorized access. There are different sensitive rule for according requirement. For every sensitive information scanning of original database is needed. When the left side pair of rule is a subset of attribute values pair of transaction and the right hand side of the rule should be same as the attribute class of the transaction then only transaction supports any rule. These steps are continuing till the sensitive data are not hidden by the unknown values.

C. Cryptographic Technique

Cryptographic is a technique through which we can encrypt the sensitive data. It is used for preserving the data. This method is very popular because is provide security and safety for sensitive attributes. There are many algorithm are available for cryptographic technique. But this method has may drawbacks such as, it fails to protect the output of computation. It prevents privacy leakage of computation. This algorithm does not provide accurate results when two or more parties talks. It is very difficult to technique for large amount of database. Final result may break the privacy of individual record.

D. Randomization Technique

The Randomization method is a popular method in current privacy preserving data mining. In which noise is added to the data in order to mask the attribute values of records, this technique provides simple and effective way. This can be easily implemented at data collection phase for privacy preserving data mining by resisting user from learning sensitive data.

Randomization techniques consists the following steps:
1. Data providers randomize their data and transmit the randomized data to the data receiver.
2. Data receiver estimates the original distribution of the data by employing a distribution reconstruction algorithm.
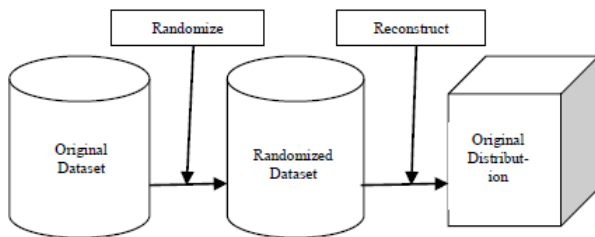

Figure 2. The Model of Randomization

E. Hybrid Technique

Hybrid is new technique in which we combine two or more privacy preserving data mining technique to preserve the data. A hybrid technique in which used randomization and generalization. In this approach first they randomize the original data and then generalized the modified or randomized data. This technique protect private data with better accuracy, also it can reconstruct original data and provide data with no information loss,

makes usability of data and also data is reconstructed. Many other technique are used to make hybrid technique such as cryptographic technique, data perturbation etc.

V. COMPARISON TABLE
TABLE 1
COMPARISON OF DIFFERENT TECHNIQUES

S. No	Author	Year of Publication	Technique Used For PPDM	Approach	Result And Accuracy
1	Y.Lindell, B.Pinks[11]	2000	Cryptographic Technique	A technique through which sensitive data can be encrypted. There is also a proper toolset for algorithm of cryptography.	This approach is especially difficult to scale when more than a few parties are involved. Also it does to hold good for large database.
2	L.Sweeney[22]	2002	K-Anonymity	A record from a dataset cannot distinguished from at least k-1 records whose data is also in the data set.	K-Anonymity Approach is able to preserve privacy.
3	J.Vaidya, C Clifton[20]	2002	Association Rule	Distribution of data vertically into segment.	Distribution based Association Rule Data Mining provide privacy.
4	Hillol Kargupta, Souptick Datta, Qi Wing And Krishnamoorthy Sivakumar[7]	2003	Data Perturbation	They tried to preserve the privacy by adding random noise while making sure that the random noise still preserve the signal from the data so that the pattern can still be accurately estimated.	Randomization based technique are used to generate random matrices.
5	CharuC.Aggrwa, Philip S. Yu[12]	2004	Condensation Approach	This approach works with pseudo-data rather than with modification of original data ,this helps in better preservation of privacy than techniques which simply use modification of the original data.	The use of pseudo data no longer necessitates the redesign of data mining algorithm, since they have the same format as the original data.
6	A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramaniam[24]	2006	L-Diversity Algorithm	If there are well represented values for sensitive attribute than that class is said to have L-Diversity.	It is better than K-Anonymity.
7	Slava Kisilevich,Lior Rokach, Yuval Elovici, Bracha Shapira[21]	2010	Anonymization	Anonymization is technique for hiding individual, sensitive data from owner records. K-anonymity is used for generalization and suppression for data hiding.	Background knowledge and homogenuity attacks of K-Anonymity Algorithm do not preserve sensitivity of an individual.
8	P. Deivanai,J. Jesu Vedha Nayahi and V. Kavitha[3]	2011	Hybrid Approach	Hybrid approach is combination of different technique which combine to give an integrated result.	It uses Anonymization and Suppression to preserve data.
9	George Mathew, Zoran Obradovic[25]	2011	Decision Tree	An approach which is technical methodology and should give judgmental knowledge.	A graph based framework for preserving patient's sensitive information.
10	Anita parmar, Udai Pratap Rao , Dhiren R. Patel[10]	2011	Blocking Based Technique	The sanitized dataset is generated from which sensitive classification rule are no longer mined.	Unknown value helps in preserving privacy but reconstruction of original dataset is quite difficult.
11	Sara Mumtaz, Azhar Rauf And Shah Khusro[16]	2011	Distortion Based Perturbation Technique in OLAP Data Cube	Data perturbation technique is also called which is also called uniformly adjusted distortion is proposed which initially distorts one cell of a cube and then distortion occurs in whole cube.	This distribution of distortion technique not only preserve ,but also provide utmost accuracy with range sum queries and high availability.
12	Hsiang chah Huang Wei –chi Fang[17]	2011	Histogram Based Reversible Data Hiding	A concept of reversibility which states that an original data can easily be hidden data can also be recovered perfectly , sensitive data is embedded into medical images which is very good technique for hiding secret data.	Histogram technique is basically used for X-ray or CT medical images and it has the potential to be integrated into database for managing the medical images in the hospital.
13	Jinfei Liu, Jun Luo and Joshua Zhexue Huange [5]	2011	Rating Based Privacy Preservation	A novel algorithm which overcomes the curse of dimensionality and provides privacy.	It is better than K-Anonymity and L-Diversity.
14	Khaled Alotaibi, V. j. Rayward-Smith,Wenjin Wang and Beatriz de la Iglesia[6]	2012	Multidimensional Scaling	A non linear dimensionality reduction technique used to project data on lower dimensional space.	The application of non metric MDS transformation works efficiently and hence produces better result.
15	Elahe Ghasemi Komishani and Mahdi Abadi[8]	2012	Trajectory data	Approach for privacy preservation in trajectory data publishing in which trajectories and sensitive attributes are generalized with respect to different privacy requirements of moving objects.	It is able to provide personalized privacy preservation trajectory data publishing, but also it is resistant to all three identity linkage, attribute linkage and similarity attacks.
16	Thanveer Jahan, Dr G. Narsimha and Dr. C. V. Guru Rao[15]	2012	Data Perturbation Using SSVD	An analyzing system used to transfer original dataset into distorted data set using sparsified singular value Decomposition.	Use of Sparsified SVD than SVD is more successful.
17	D.Karthikeswarant , V.M.Sudha, V.M.Suresh and A.J.Sultan[19]	2012	Association Rule	Sanitize datasets using sliding window algorithm and preserves data.	A novel approach that modifies the database to hide sensitive rules.
18	M.N.Kumbhar and R.Kharat[18]	2012	Association Rule By Horizontal and Vertical Distribution.	Different approach in the field of Association rule are reviewed.	The Performance of all models is analyzed in terms of privacy, security and communication.
19	Savita Lohiya and Lata Raghra[9]	2012	Hybrid Approach	A combination of K-Anonymity and Randomization.	It has better accuracy and original data can be reconstructed.
20	Martin Beck And Michel March Ofer[26]	2012	Anonymizing Demonstrator	Making a demonstrator with user friendly interface and performs Anonymization.	Swapping and Recording can be applied to enhance the utility.
21	Mi Wen, Rongxing Lu,Jingshen Lei,Xiaohui Liang[27]	2013	(ECQ) Efficient Conjunctive Query	ECQ can protect the data confidentiality and integrity, as well as data and query privacy	Achieve Conjunctive query without data and query privacy leakage. ECQ is more efficient in terms users computation cost.
22	Mittal D., Kaur D., Aggarwal A[28]	2014	K-Means Algorithm	It Prevents Intermediate data leakage in the process of Computation while maintaining the correctness and validity of data mining process.	This method is used to solve the privacy issue of the cloud.

Algorithms are classified on the basis of performance, utility, cost, complexity, tolerance against data mining

algorithms etc.

VI. CONCLUSION

In today's world, privacy is major concern to protect the sensitive data. People are concerned about that they do not share the sensitive information. Our survey is focuses on the on existing literature present in field of the privacy preserving data mining. From our analysis, we found that there is no single technique in all domain consistent. All method perform in a different way depending on the type of data and a type of application or domain. But still from analysis, we conclude that Cryptography and Random Data Perturbation method better than existing method and Cryptography technique is best for encryption of sensitive data. Perturbation helps to preserve data hence sensitivity is maintained.

REFERENCES

[1] J. Han and M. Kamber, "*Data Mining: Concepts and Techniques*", 2nd ed.,The Morgan Kaufmann Series in Data Management Systems, Jim Gray, Series Editor 2006.

[2] M. B. Malik, M. A. Ghazi and R. Ali, "Privacy Preserving Data Mining Techniques: Current Scenario and Future Prospects", in *proceedings of Third International Conference on Computer and Communication Technology*, IEEE 2012.

[3] P.Deivanai, J. Jesu Vedha Nayahi and V.Kavitha," A Hybrid Data Anonymization integrated with Suppression for Preserving Privacy in mining multi party data" in *proceedings of International Conference on Recent Trends in Information Technology*, IEEE 2011.

[4] M. Prakash, G. Singaravel, "A New Model for Privacy Preserving Sensitive Data Mining", in *proceedings of ICCCNT Coimbatore, India*, IEEE 2012.

[5] J. Liu, J. Luo and J. Z. Huang, "Rating: Privacy Preservation for Multiple Attributes with Different Sensitivity requirements", in *proceedings of 11th IEEE International Conference on Data Mining Workshops*, IEEE 2011.

[6] K. Alotaibi, V. J. Rayward-Smith, W. Wang and Beatriz de la Iglesia, "Non-linear Dimensionality Reduction for Privacy-Preserving Data Classification" in *proceedings of 2012 ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security,Risk and Trust*, IEEE 2012.

[7] H. Kargupta and S. Datta, Q. Wang and K. Sivakumar, "On the Privacy Preserving Properties of Random Data Perturbation Techniques", in *proceedings of the Third*

IEEE International Conference on Data Mining, IEEE 2003.

[8] E. G. Komishani and M. Abadi,"A Generalization-Based Approach for Personalized Privacy Preservation in Trajectory Data Publishing", *in proceedings of 6'th International Symposium on Telecommunications (IST'2012),* IEEE 2012.

[9] S. Lohiya and L. Ragha, "Privacy Preserving in Data Mining Using Hybrid Approach", in *proceedings of 2012 Fourth International Conference on Computational Intelligence and Communication Networks*, IEEE 2012.

[10] A. Parmar, U. P. Rao, D. R. Patel, "Blocking based approach for classification Rule hiding to Preserve the Privacy in Database" , in *proceedings of International Symposium on Computer Science and Society*, IEEE 2011.

[11] Y. Lindell, B.Pinkas, "Privacy preserving data mining", in *proceedings of Journal of Cryptology*, 5(3), 2000.

[12] C. Aggarwal , P.S. Yu, "A condensation approach to privacy preserving data mining", in *proceedings of International Conference on Extending Database Technology (EDBT),* pp. 183–199, 2004. 746

[13] R. Agrawal and A. Srikant, " Privacy-preserving data mining", in *proceedings of SIGMOD00*, pp. 439-450.

[14] Evfimievski, A.Srikant, R.Agrawal, and Gehrke , "Privacy preserving mining of association rules", in proceedings of KDD02, pp. 217-228.

[15] T. Jahan, G.Narsimha and C.V Guru Rao, "Data Perturbation and Features Selection in Preserving Privacy" in *proceedings of 978-1-4673-1989-8/12*, IEEE 2012.

[16] S. Mumtaz, A. Rauf and S. Khusro, "*A Distortion Based Technique for Preserving Privacy in OLAP Data Cube*", in proceedings of 978-1-61284-941-6/11/$26.00, IEEE 2011.

[17] H.C. Huang, W.C. Fang, "*Integrity Preservation and Privacy Protection for Medical Images with Histogram-Based Reversible Data Hiding*", in proceedings of 978-1-4577-0422-2/11/$26.00_c, IEEE 2011.

[18] M. N. Kumbhar and R. Kharat, "*Privacy Preserving Mining of Association Rules on horizontally and Vertically Partitioned Data: A Review Paper*", in proceedings of 978-1-4673-5116-4/12/$31.00_c, IEEE 2012.

[19] D.Karthikeswarant, V.M.Sudha, V.M.Suresh and A.J. Sultan, "A Pattern based framework for privacy

preservation through Association rule Mining" in *proceedings of International Conference On Advances In Engineering, Science And Management (ICAESM -2012),* IEEE 2012.

[20] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data", in *The Eighth ACM SIGKDD International conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, CA, July 2002*, IEEE 2002.

[21] Slava Kisilevich, Lior Rokach, Yuval Elovici, Bracha Shapira, "Efficient Multi-Dimensional Suppression for K-Anonymity", *in proceedings of IEEE Transactions on Knowledge and Data Engineering*, Vol. 22, No. 3. (March 2010), pp. 334-347, IEEE 2010.

[22] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy", in *proceedings of Int'l Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 2002.

[23] The free dictionary.Homepage on Privacy [Online]. Available:http://www.thefreedictionary.com/privacy.

[24] A. Machanavajjhala, J.Gehrke, D. Kifer and M. Venkitasubramaniam, "I-Diversity: Privacy Beyond k-Anonymity", *Proc. Int'l Con! Data Eng. (ICDE)*, p. 24, 2006.

[25] G. Mathew, Z. Obradovic," A Privacy-Preserving Framework for Distributed Clinical Decision Support", *in proceedings of 978-1-61284-852-5/11/$26.00 ©2011 IEEE*.

[26] Martin Beck and Michael Marh¨ofer," Privacy-Preserving Data Mining Demonstrator", in *proceedings of 16th International Conference on Intelligence in Next Generation Networks*, IEEE 2012.

[27] Mi Wen, Rongxing Lu ; Jingshen Lei ; Xiaohui Liang , 2013,ECQ: An Efficient Conjunctive Query scheme over encrypted multidimensional data in smart grid, Global Communications Conference (GLOBECOM), 2013 IEEE, 796 – 801

[28] Mittal, D. ; Kaur, D. ; Aggarwal, A., 2014 , Secure Data Mining In Cloud Using Homomorphic Encryption IEEE International Conference on Cloud Computing in Emerging Markets CCEM),2014, pp : 1 – 7

Data Mining Using Matrix Factorization for Enhancing a Patient's HealthCare

[1] Manigandan. J [2] Dr. Soundararajan. S, M.E., PhD,
[1] M.E. Computer Science and Engineering, [2] Vice Principal
[1][2] Velammal Institute of Technology,
Panchetti, Chennai.
[1] rjsmanigandan23@gmail.com. [2] viceprincipalvit@gmail.com.

Abstract: **Web mining is used to discover as well as extract data from web documents and service. Social networking sites are used to discuss the current topics and reactions to current happening on the internet. The discussion which reflects the opinion of people, thoughts and their innovative ideas. Detection of current topic and tracking valid data from offline articles is quite difficult. Detection of topic from social networking sites will helps to gather and analyses the huge volume of up-to-the minute. Topics are detected based on vigorously and provides path to various treatments to cure the diseases. The techniques are called as Formal Concept Analysis [3] based on Matrix Factorization are intended to pick up the evolution and issues of current topic in unstructured content which are present in a social media. Extraction and analyses of data based on the user-needed data content. Self organizing maps [16] are used to correlate the data based on positive and negative words present in the user's status. Scores of text will give as numerical value of each user forums. The pictorial representation can be viewed based on the scored values and for easy understanding. It helps to determine the better treatments and least cost medicine to cure incurable diseases can be identified and try to cure by early stage as soon as possible.**

Index Terms: Cluster Analsyis, FCA, Grid factorization, Neural Networks, NMF, R Console, SOM, Twitter.

I. INTRODUCTION

The Internet has become the major part in day to day life of Kinsfolk. Nobody will imagine a world without internet. Were each and everyone getting to grasp things that happen around the world through internet? This project is to detect disease and various diseases from the foremost noted social networking web site known as Twitter that exhibits the views of individuals on certain person, event, and organization. Web Usage Mining[11] which helps out to explore interesting web usage facts in order to understand the needs of the user in which tends help or provide the needy to users who use the web applications. Web usage mining is a mechanism of extracting useful facts from server activity reports.

The count of social networking sites users will reach 2.55 billion by the end of the year 2017 this proves that how people show their interest in sharing their views and opinions in social networking sites such as Orkut, Twitter, Face book, and Google plus. This project aims out to detect various treatments for incurable disease like lung cancer, blood caner, etc., from Twitter. Multiple treatments for diseases detection is a seminal event or activity which is directly related to disease. Multiple treatments for diseases detection helps out in knowing the most talked about and important happenings that has been spoken all over the social networking sites. All these social contents are analyzed to detect the multiple treatments for diseases from the posts and links such as link mining [15], classification through links [15], predictions based on objects as well as links [8], existence [10], estimation [11], object [7], group and subgroup detection [12], and mining the data [4], [8]. Link prediction, viral marketing [13], and online discussion [4][14] groups (and rankings) allow for the development of solutions based on user feedback. When the social contents are analyzed with appropriate statistical and computational tools, social media contents can be turned into invaluable future Insights.

Detection of incurable disease for multiple treatments which helps out to know how public reacted to a particular cured from various disease and recover their lives. And also it helps us to know how new medicine and current booming medicine get evolved and who has been influenced because of the news can also be retrieved. Healthcare providers could use patient opinion to improve their services. Physicians could collect feedback from other doctors and patients to improve their treatment recommendations and results.

II. PROBLEM DEFINITION

A. Existing System

In the web mining scenario, the records to match are highly query-dependent, since they can only be obtained through online forums. Moreover, the dictionary has capability to store only 110 words i.e., 55 positive and 55 negative words generally. And also, the dictionary is not efficient to analyses the medical terms and their semantic definitions.

B. Problem Statement

The existing research does not sufficiently assess whether sentiments can be analyzed using sentiment function. And by scoring the tweets using positive and negative scores within the limited number of 110 words are stored in dictionary. The dictionary is not efficient to analyses a medicinal terms and semantically words of medicines.

C. Objectives of the Proposed System

The aim of the work is more clearly to retrieve and analyze the sentiments of a cancer and incurable disease from twitter posts. In order, to add more medical termed words to dictionary. Efficient to retrieve the medical terms and dictionary words. Determine the least cost treatments to cure the cancer diseases.

III. EXTRACTION OF TWEETS

A. Create an Application on Twitter

Twitter helps out the students and research scholars to do something useful with tweets. There are certain steps and procedures to fetch tweets from the twitter. The first step is to create an application [4] in twitter which asks for username and password of the creators twitter account. Once the application is created successfully twitter generates a consumer key and a consumer secret key which is like a basic gateway in fetching out tweets from twitter.

Fig.1. Functional Architecture for detection of diseases And their multiple treatments from Twitter.

B. Execution of R console

The main purpose of using R console is R is an open source environment for statistical computing and graphics. Another major reason of using R is it provides a wide variety of statistical and graphical techniques. R can retrieve tweets from twitter by running its chunk of code one and only if twitter supporting packages are installed such as twitter, ROAuth, plyr, stringr, ggplot2. Once these packages are installed the consumer key and the consumer secret is entered in the code which was given by the twitter when the new application was created. After the execution of these lines of code a API link will be generated which was shown in Fig.1. After the link is copied and pasted in twitter a 7 – digit pin is generated which will be source to retrieve tweets. Pin generation and execution of that pin will create handshake based authentication which acts as a pathway between R and Twitter. In twitter, the nodes are interconnected and they form a clustering of nodes.

Similar nodes are interconnected with each other to retrieve a user needs data. Self-Organizing Map (SOM) [16] will helps to organize the data based on the links of each other from twitter. Using Apirori algorithm we can capable to retrieve the data and subsets of data from social network.

Apirori Algorithm:
C_k: Candidate itemsets of size k
L_k : frequent itemsets of size k
L_1 = {frequent 1-itemsets};
for (k = 2; L_k !=Æ; k++)
C_{k+1} = GenerateCandidates(L_k)
For each transaction **t** in database do increment count of candidates in
C_{k+1} that are contained in **t**
endfor
L_{k+1} = candidates in C_{k+1} with support
≥**min_sup**
endfor
return È$_k$ L_k;

C. Hash Based Tweets

Hash based Tweets is nothing but the type or it can called as calling tweets like for example if tweets related to cancer is to be downloaded from twitter it should be called as #blood cancer were a hash(#) symbol must be priory added before the word were the tweets about is to be retrieved from twitter.
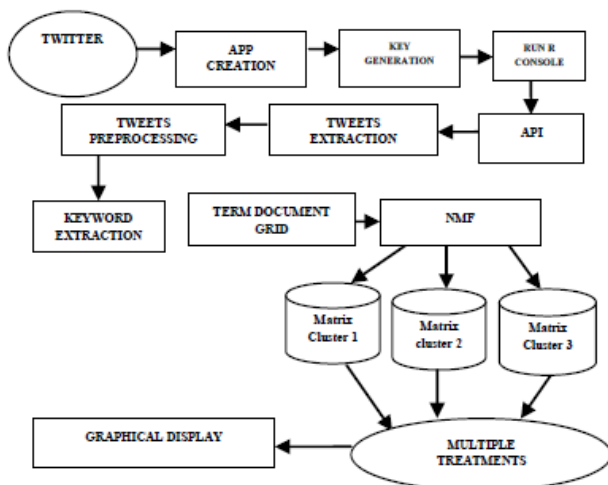
Fig. 2. Cluster based web mining of data from Twitter

```
requestURL = "https://api.twitter.com/oauth/request_token"
accessURL = "https://api.twitter.com/oauth/access_token"
authURL = "https://api.twitter.com/oauth/authorize"
consumerKey = "xxxxxxxxxxxxxxxxxxxxxxxx"
consumerSecret = "xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx"
Cred <- OAuthFactory$new(consumerKey=consumerKey,
consumerSecret=consumerSecret,requestURL=requestURL,
accessURL=accessURL,   authURL=authURL)
Cred$handshake(cainfo   =   system.file("CurlSSL",   "cacert.pem",
package = "RCurl") )
XXXXXX
save(Cred, file="twitter authentication.Rdata")
registerTwitterOAuth(Cred)
Hashtag.list          <-          searchTwitter('#Hashtag',         n=1000,
cainfo="cacert.pem")
Hashtag.df = twListToDF(Hashtag.list)
write.csv(Hashtag.df,          file='C:/temp/HashtagTweets.csv',
row.names=F)
```
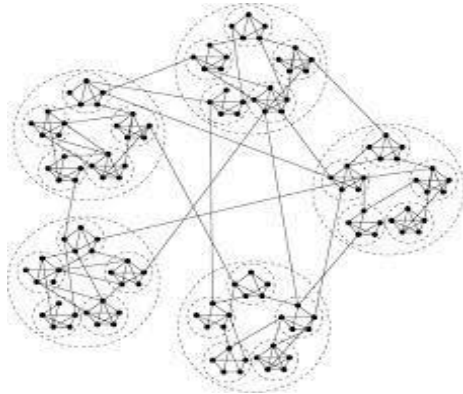
Fig. 3. R Chunk to Download Tweets from Twitter

IV.   PREPROCESSING OF TWEETS

As explained in the Fig.1. the functional architecture of this paper goes like this were after the API link is generated in R console by running its R chunks it generates a 7 digit pin by entering that pin and by executing the code the tweets can be downloaded based their hash tag(#). Once the tweets are downloaded and saved as CSV files and immediately if the tweets are analyzed several noise data and unwanted symbols appear that are not necessary. Once the tweets are downloaded the symbols are first removed from the twitter using java code and the process goes like this by removing special symbols and filling space in the places were symbols previously existed.

Once the symbols are removed from the tweets and filled with the space they are ready for further smoothing process.

We then assigned weights for each of the words found in the user posts using with the following formula:

$$\text{Weight}_{i,d} \begin{cases} \text{Log } (tf_{i,d} + 1)\log n\, x/x_i & \text{if } tf_{i,d} > 1 \\ 0 & \text{otherwise} \end{cases}$$

In which tf_i, d represents the word frequency (t) in the document (d), n represents the number of documents within the entire collection, and xt represents the number of documents where t occurs. Once the symbols are removed from the tweets and filled Times New Roman with the space they are ready for further smoothing process.

A. Text Transformation

Once the symbols are removed the tweets are first transformed to a data frame which is used for storing data tables. They are nothing but list of common size vectors.

The built in data frames are used to create or convert data frames from tweets. The data frames are like primary data structure in R which is used to exhibit data. A data frame is a table which can be also called as two-dimensional array-like structure, were each column contains measurements on one variable, and each row contains a case.

There are no constraints in R that all columns in a table should have same features one column may be of numerical other may be of character. After converting the tweets into data frames then they are converted to data corpus mainly for the purpose of specifying the source to be of character vectors. Once the corpus is created there are certain transformations need to be made they are changing upper case letters into lower case letters and removing numbers, stop words, and hyperlinks are also will be removed.

B. Stemming of Words and text

To perform the above mentioned operations such as transforming text and to stem words from the downloaded tweets tm package from library must be installed without that no operations can be performed. In most of the applications the words needed to be stemmed to retrieve their radicals, so that several forms can be retrieved from a stem that can be taken as same word while determining their frequency in occurrence.

Hence the tweets are preprocessed finally by the completion of stop words removal and stemming of words.

A term document grid [9] represents the relationship between terms and documents where each row stands for terms and each column stands for documents.

Positive	Negative	Positive	Negative
Appreciate	Cannot	Agree	Bad
Beneficial	Concern	Benefit	Loss
Importance	Damage	Ease	Don't
Effective	Didn't	Easier	Died
Greatly	Difficult	Good	Isn' t
Favorably	Depression	feasible	Doubt
Favorable	Error	Great	Failure
Grateful	Impossible	Help	Hasn't
Greatest	Hasn't	Enjoy	Hate
Greater	Hurt	Helped	Fear
Helpful	Discomfort	Hoped	Lack
Hopeful	Limited	Honest	Lose
Honestly	Concerns	Hope	Lost
Helping	Miss	Helps	Nasty
Hopefully	Nausea	Hopes	Negative
Hoping	No	Comfort	No
Importantly	Painful	Like	Poor
Improve	Problem	Improved	Sad
Improvement	problems	Improves	scared
Improving	Scary	Inspiration	severe
Impresses	Sorry	Love	hate

Table I. General stored dictionary words.

V.   BUILDING A TERM DOCUMENT GRID

The term document grid[9] is formed from the above processed data corpus. Term document grid can be build in R using Term Document Grid[9] function. Since the multiple treatments for diseases is being detected from the tweets and the csv file consists of about 1000 tweets.

Tweets	Tweet 1	Tweet 2	Tweet 3	Tweet 4	Tweet 5
Term 1	1	3	7	4	2
Term 2	4	8	5	9	6
Term 3	3	5	2	1	4
Term 4	1	3	6	2	1
Term 5	2	6	7	3	4

Table II: Term Document Grid

So each and every single tweet is to be considered as a single document and the term document grid will be formed in such a manner that all keywords found in all 1000 tweets will be mentioned in the rows respectively as shown in Table I, the tweets are the documents and terms are the rows. Table I. illustrates that term 1 has been appeared 1 time in document 1 i.e. in tweet 1 and similarly term 2 has been appeared 8 times in tweet 2 respectively. All the numerical value signifies that the frequency of all terms in each and every document i.e. each and every

tweet. And hence finally a perfect term document grid[9] is build successfully.

A. Mining of medical termed text and Pre-processing

A Rapid miner (www.rapidminer.com) data collection [8] and processing tree was developed to look for the most common positive and negative words and their term-frequency-inverse document frequency (TFIDF) [13] [1]scores within each post.

Acne	Headache	Itchin	Cachexia	Pneumonia
Throat cancer	Stomach cancer	Zika virus	Viral latency	Chickenpox
vomiting	Genetic disease	Blood cancer	Lung cancer	weakness
Diabetes	bacteria	Rash	Heart disease	Viral disease

Table III. Dictionary stored medical termed words

To this goal, we used the National Library of Medicine's Medical Sub-ject Heading (MeSH), which is controlled vocabulary. (http://www.nlm.nih.gov/mesh/) that consists of a hierarchy of descriptors and qualifiers that are used to annotate medical terms.

A custom designed program was used to map words in the forum to the MeSH database. A list of words present in forum posts that were associated to treatment side effects was thus compiled. This was done by selecting the words simultaneously annotated with a specific list of qualifiers in MeSH (CI– chemically induced; CO – complications; DI – diagnosis; PA– pathology, and PP – physiopathology)

B. Data score function

The score function of tweet based on the medical terms and semantics words of dictionary stored in it. TFIDF [1], [13] are used to correlation of user posts and forums found in each modules. They help us to analysis the positive and negative terms of words are known as Module Average Opinion (MAO)[1] by equation written as

$$MAO= \frac{Sum_+ - Sum_-}{Sum_{all}}$$

Sum$_+$ = x$_{ij}$ is the total sum of the TF-IDF [1]
, [13] scores will compare the positive words in the Wordlist vectors within the module containing medical Termed stored words in dictionary. The units i represent post index. The unit j represents the wordlist index (matching the positive words in the module list). Sum$_-$ = x$_{ij}$ is the total sum of the TF-IDF[13],[1] scores will compare the negative words in the wordlist vectors within the module containing medical termed stored words in dictionary. The units i represent post index. The unit j represents the wordlist index (matching the negative words in the module list).

$$\text{Sum}_{\text{all}} = \begin{matrix} N \\ i=1 \end{matrix} \quad \begin{matrix} M \\ k=1 \end{matrix} x_{ik}$$

Is the sum of the total words in the wordlist vectors within the module containing medical termed stored words in the dictionary?

VI.   GRID FACTORIZATION

In linear algebra, a grid factorization [6] is deriving products of matrices from single grid. To reduce the dimensionality of a larger complexity grid into a lesser complexity grid and reduction of dimensionality is the main aim of grid factorization [9].

A.   Formal Concept Analysis

Formal Concept Analysis (FCA) [10] is the most effective data analysis methodology which was based on ordered lattice theory. It is also called as concept hierarchical structure representing relationships and attributes in to a particular domain. The main aim of FCA [3] is to define the unit of two parts extension and intension. FCA methodology [7] is used to mine association rules from web usage lattice that has been constructed from web lattice and it has been discovered from those web lattices is used to detect multiple treatments for diseases using FCA[10] techniques called Singular Value Decomposition (SVD) and Nonnegative Grid Factorization (NMF)[9].

Terms	NMF VALUE
Term 1	8.2574
Term 2	14.3030
Term 3	6.4969
Term 4	6.4236
Term 5	10.4298

Table IV. Illustrates the NMF values for the Term Document Grid [9] Table I.

Clearly signifies that the Term 2 has the highest value 14.3030 since the terms occurrence or frequency is comparatively higher than the other terms this may look easy for a 5*5 grid and guess which term has the highest value but if the size of the term document grid is 500*1000 then it cannot be guessed just by seeing here comes the job of NMF [9] to find the multiple treatments for diseases. Hence the multiple treatments for diseases are concept related to Term 2 is the multiple treatments for diseases.

VII. CLUSTERING THE MATRIX

Matrix is being clustered using a formal concept analysis approach called NMF [9] which is the most efficient matrix factorization algorithm. What it does is

reduce the dimensionality of the matrix and further produce several matrixes from the term document matrix that has been generated in the earlier phase. There are various types of matrix factorization techniques for clustering they are,

A. NMF

When the input data is non negative, and it restricts F and G to be nonnegative. The standard NMF [6] can be written as, NMF: $X+ \approx F+G+$ Using an intuitive notation for X, F, $G \geq 0$

B. SEMI-NMF

When the input data has mixed signs, it can restrict G to be nonnegative while placing no restriction on the signs of F. This is called SEMI- NMF. SEMI-NMF: $X\pm \approx F\pm G+$ Semi-NMF [9] can be motivated by K-means clustering. Let F = ($f1, \ldots\ldots ,fk$) be the cluster centroids obtained via K-means clustering. Let G be the cluster indicators: i.e., $g_{ki} = 1$ if x_i belongs to cluster ck; $g_{ki} = 0$ otherwise.

C. CONVEX-NMF

In general, the basis vectors F= ($f1, \cdots ,fk$) can be anything in a large space, in particular, a space that contains the space spanned by the columns of X= ($x1, \cdot \cdot xn$). Inorder for the vectors F to capture the notion of cluster centroids. It restricts them to lie within the space spanned by the columns of X, i.e., $Fl = w1lx1 + \cdot \cdot \cdot + wnlxn = Xwl$, or F= XW. Furthermore, fl is restricted as a convex combination of the data points. Hence they are called as restricted form of factorization as Convex-NMF [6] and it applies to both nonnegative and mixed-sign input data.

D. TRI-FACTORIZATION

This technique is used to simultaneously cluster the rows and the columns of the input data matrix X, They are considered as the following nonnegative 3 factor decomposition, $X \approx FSGT$ Note that provides additional degrees of freedom such that the low-rank matrix representation remains accurate while row clusters and *G* gives column clusters. An important special case is that the input X contains a matrix of pair wise similarities: X = XT =W. In this case, F = G= H. The optimization of the symmetric NMF is done.

E. KERNEL NMF

Consider a mapping $xi \rightarrow f(xi)$, or $X \rightarrow f(X) = (f(x1), \cdot \cdot \cdot ,f(xn))$. A standard NMF or Semi-NMF like $f(X) \approx FGT$ would be difficult since F,G will depends explicitly on the mapping function f(·). However, Convex-NMF provides a nice possibility: $\varphi(X) \approx \varphi(X)WG$ Depends only on the kernel K= f T(X)f(X). This kernel extension of NMF [9] is similar to kernel-PCA and kernel K-means.

F. Word Cloud and Graphical display

A word cloud depicts the visual representation of keywords based on their importance as well as values in this paper the highest NMF [9] value holding term will acquire the center and larger part of word cloud respectively the lesser terms with lesser NMF value will hold the remaining left parts with respect to their NMF[9] values. Fig.3.Represents an report about Cancer disease and their causes

Fig. 3. Word Cloud of CANCER.

From 2010 to 2025 describes cancer statistics derived from tweets and their respective scores. One of the reasons for cancer diseases are hypertension prevalence is shown as bar chat statics of men and women Fig. 4 respectively. As a consequence, the estimated number of yearly cancer deaths is expected to increase slightly for men and more for women. With respect to prevalence of hypertension, female mortality is expected to be highest for established market economies and former socialist economies. Female is quite larger compared to males as per derived tweets of cancer statistics.
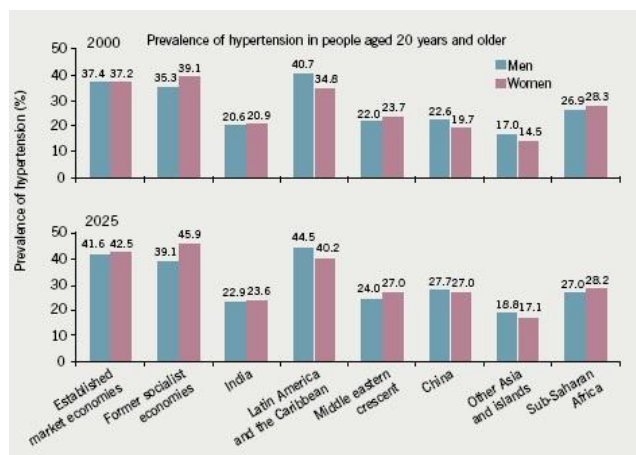

Fig. 4. Bar Chart for CANCER survey from 2000 to 2025.

VIII. CONCLUSION

The main aim of this paper is to detect multiple treatments for diseases from the social networking site twitter using a formal concept analysis based Grid factorization i.e. Non-negative Grid Factorization[9] a best grid factorization technique that can used to determine the current multiple treatments for diseases or new treatments or it can be also called as event detection. Hence this paper proposed a most effective solution and accurate solution in detecting current multiple treatments for diseases. The multiple treatments for diseases are detected on the basis of closest similarity based distance calculation using Nonnegative grid factorization [9] and the multiple treatments for diseases that have been detected will be of perfect accuracy in medical field.

This paper deals with only textual information and now a day's most of the posts in social networking sites are images and videos rather than textual posts. Since the main purpose of inventing NMF [6] is to work with images. So, in future image processing concepts can be built in here to find out multiple treatments for diseases from the images that are being posted and shared in social networking sites [4][12]. Social media can open the door for the health care sector in address cost reduction, product and service optimization, and patient care.

REFERENCES

[1] Altug Akay, Member, IEEE, Andrei Dragomir, Bjorn Erik Erlandsson, Senior Member, *IEEE,* "Network-Based Modeling and Intelligent Dapta Mining of Social Media for Improving Care" 2015.

[2] Ping Li, Jiajun Bu, Yi Yang and RongrongJi, "Discriminative Orthogonal Nonnegative Grid Factorization", *Journal on Expert Systems with Applications*, 2013 pp. 01–11.

[3] Elena Nenova and Dmitry I. Ignatov, "An FCAbased Boolean Grid Factorization for Collaborative Filtering", *Conference on Formal Concept in information retrieval,* 2012 pp. 57–73.

[4] Alan Ritter, Mausam and Oren Etzioni, "Open Domain Event Extraction from Twitter", *Conference on Knowledge discovery and data mining*, 2012, pp. 1104–1112.

[5] Anish Das Sarma and Alpa Jain, "Dynamic Relationship and Event Discovery",*Conference on Web search and data mining,* 2011, pp. 207–216.

[6] Deng Cai and Jiawei Han, "Graph Regularized Nonnegative Grid Factorization for Data Representation",

Transactions on Pattern analysis and Machine Intelligence, 2011, Vol. 33, No. 8, pp. 1548–1560.

[7] Abderrahim El Qadi, DrissAboutajdine and YassineEnnouary, „Formal Concept Analysis for Information Retrieval‟*, Journal of computer science and information security*, 2010, vol.7, No. 2, pp. 119–125.

[8] C. Corley, D. Cook, A. Mikler, and K. Singh, "Text and structural data mining of influenza mentions in web and social media," *Int. J. Environ. Res. Public Health*, , Feb. 2010, vol. 7, pp. 596–615

[9] Fei Wang, "Community discovery using nonnegative grid factorization", *Conference on data mining and data representation,* 2010, pp. 01–29.

[10] Jonas Poelmans, Paul Elzinga and Stijn Viaene, "Formal Concept Analysis in knowledge discovery: a survey",*Conference on conceptual structures,* 2010, pp. 139–153.

[11] Vasumathi, D. and Govardhan, 'Efficient Web usage Mining Based on Formal Concept Analysis', Journal on Theoretical and Applied Information Technology, , 2009, pp. 99–109.

[12] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the Web," *Manag. Sci.*, Sep. 2007, vol. 53, pp. 1375–1388.

[13] P. Soucy and G. W. Mineau, "Beyond TFIDF weighting for text catego-rization in the vector space model," *in Proc. 19th Int. Joint Conf. Artificial Intell., Edinburgh, U.K.,* 2005, pp. 1130–1135.

[14] W. Yih, P. H. Chang, and W. Kim, "Mining online deal forums for hot deals," *in Proc. IEEE/WIC/ACM Int. Conf. Web Intell., Beijing, China,* 2004, pp. 384– 390.

[15] B. Taskar, M. Wong, P. Abbeel, and D. Koller, "Link prediction in relational data," *in Proc. Adv. Neural Inform. Process. Syst., Vancouver, B.C. Canada*, 2003.

[16] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "Self-Organizing Map in MATLAB: The SOM Toolbox," *in Proc. Matlab DSP Conf., Espoo*, Finland, 1999, pp. 35–40.

A Health-IoT Platform Based On the Bio-Sensor and Mobile Application

[1]K.V.S.N.Rama Rao, [2] D.Sai pushkar, [3]T.Bhargavi, [4]Satya Sena Reddy , [5] K.Priya Ranjith

[1][2] [3],[4], [5] Department of CSE , MLR Institute of Technology, Dundigal, Hyderabad

[1] kvsnramarao@yahoo.co.in, [2]saipushkar07@gmail.com, [3]bhargavichoudary95@gmail.com, [4]gvsr320@gmail.com , [5] ranjith.yuvraaj12@gmail.com

Abstract—**In-home healthcare services based on the Internet-of-Things (IoT) have great business potential; however, a comprehensive platform is still missing. In this paper, an intelligent home-based platform, the iHome Health-IoT, is proposed and implemented. In particular, the platform involves an open-platform-based intelligent health analysis system with enhanced connectivity and interchangeability for the integration of devices and services, flexible and wearable bio-medical sensor device (Bio-Patch) enabled by the state-of-the-art inkjet printing technology and system-on-chip. The proposed platform seamlessly fuses IoT devices (e.g., wearable sensors.) with in-home healthcare services (e.g., telemedicine) for an improved user experience and service efficiency.**

Index Terms—**Internet-of-Things, Health-IoT, Bio-Patch, mobile application (e-health application)..**

I. INTRODUCTION

Now-a-days, global ageing and the prevalence of chronic diseases have become a common concern . Many countries are undergoing hospital restructuring by reducing the number of hospital beds and increasing the proportion of home healthcare. A promising trend in healthcare is to move routine medical checks and other healthcare services from hospital (Hospital-Centric) to the home environment (Home-Centric) . By doing so, firstly, the patients can get seamless healthcare at anytime in a comfortable home environment; secondly, society's financial burden could be Greatly reduced by remote treatment; thirdly, limited hospital

Resources can be released for people in need of emergency care. In-home healthcare and services can drastically reduce the total expenditure on medical care or treatment. Therefore, it is urgent in the near future for the healthcare industry to develop advanced and practical health-related technologies and services by leveraging information and communication technology (ICT), and apply them directly in the home environment. In order to track the physical status of the elderly and in the meanwhile keep them healthy, the following two daily tasks are essential: 1) real-time monitoring and analyzing vital signs to early-detect or predict life-threatening adverse events, 2) checking whether they are following their prescribed treatment, including taking their prescribed medicine on time. However, with rapidly aging populations, these daily tasks have brought great pressure and challenges to global healthcare systems. One review

estimates that about 25% of the adult population does not adhere to their prescribed medication, which may lead to poor health outcomes and increased mortality. Poor medication adherence is a major problem for both individuals and healthcare providers. Technology improvements in healthcare facilities and services are highly desirable to meet the requirements of this giant group.

In the meantime, Internet-of-Things (IoT) has been recognized as a revolution in ICT since it started at the beginning of the 21st century . IoT technology provides the possibility to connect sensors, actuators or other devices to the Internet and is conceived as an enabling technology to realize the vision of a global infrastructure of networked physical objects . IoT extends the Internet into our everyday lives by wirelessly connecting various smart objects and will bring significant changes in the way we live and interact with smart devices . Various companies are exploring this domain as it can potentially unlock the door to new business opportunities. As part of IoT – intelligent components, radio-frequency identification (RFID) , embedded sensors and actuators, etc. – have been rapidly developed and significantly expanded in scope. As a consequence, the number of IoT-based applications has boomed as well. All these technologies facilitate the deployment of IoT devices in the home environment for 24/7 healthcare . Some researchers attempt to integrate wearable devices and systems in IoT scenario to achieve better e-health services . As a result, the physical size, rigid nature and short battery life become limiting factors for potential long-term use. Some research

groups focus on the user-comfort issues, by leveraging advanced materials to develop user-friendly sensors. Functional textiles are utilized to manufacture a sensorized garment for physiological monitoring, where electrocardiogram (ECG) signals are successfully recorded using fabric sensing elements. A desirable system should be capable of taking care of the patients from all aspects, covering personalized medication, vital signs monitoring, on-site diagnosis and interaction with remote physicians. In addition, the existing systems seldom integrate new materials or apply new manufacturing approaches, which are always the key elements for bringing new devices or solutions into healthcare fields. By taking the forementioned issues into consideration, an intelligent healthcare IoT system and mobile application proposed in this paper. IoT devices (e.g., wearable sensors) are seamlessly connected to the e-health application via a heterogeneous network which is compatible with multiple existing wireless standards. Considering the present and future importance of IHIS and IoT in e-health field, we developed the Health-IoT platform which can well find its applications in patients' home and nursing home scenarios. The proposed system takes the advantages of System-on-Chip (SoC) technology, material technology, and advanced printing technology, to build a patient-centric, self-assisted, fully-automatic intelligent in-home healthcare solution. The functions developed can be applied in various health-related scenarios, including environmental monitoring, vital signs acquisition, medication management, and healthcare services.

II. PROPOSED HEALTH CARE APPLICATION
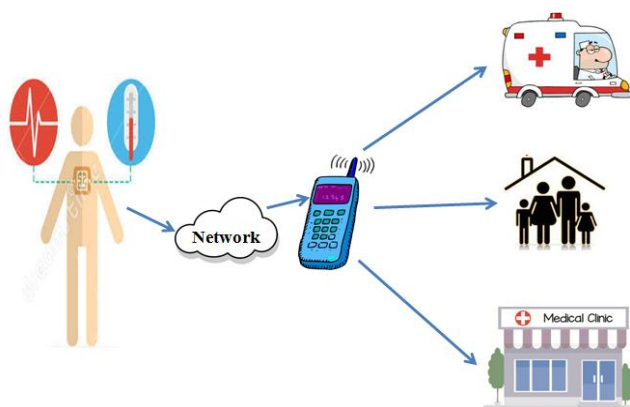
A. HOME HEALTH-IOT SYSTEM

Fig. 1 illustrates the concept of the iHome Health-IoT System

The figure-1 is discussed below:
- ❖ The body-worn Bio-Patch can detect and transmit the user's bio-signals to the mobile app in real time.

- ❖ All the collected information is interpreted, stored and displayed locally on the mobile application.
- ❖ By using e-health mobile application, it automatically intimates to the doctor, emergency services and the family.

One major contribution of the proposed iHome Health-IoT system is that it dramatically expands the scope and coverage of traditional Healthcare Information Systems (HIS), extending from a confined hospital environment to a patient's home. By doing so, the overall healthcare system could be optimized at the top level, turning from the conventional Enterprise Resource Planning into the Entire Resource Planning.
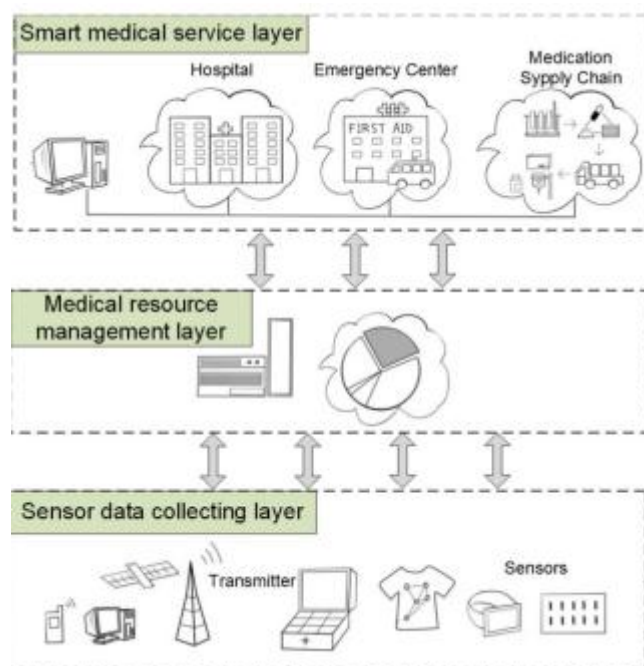
B. DATA FLOW AMONG DIFFERENT LAYERS

Fig. 2 illustrates the concept of data flow among different layers

I t consists of three different layers i.e , smart medical service layer , medical resource management layer ,sensor data collecting layer.

- ❖ A smart medical service layer is directly linked to professional medical facilities such as hospitals, emergency centers, and medicine supply chain. For example, doctors can efficiently manage a large group of patients. They can inspect the medication history as well as the physiological status history of a specific patient, make further analysis of a suspicious portion of patient's bio-signals (e.g., ECG) and based on the analysis, doctor makes a new e-prescription accordingly. The doctors can easily identify the patient group whose health conditions have improved, and

make them aware of their progress. Both patients and their family may feel reassured which helps build positive loops into rehabilitation and self-care.

❖ The medical resource management layer works as a transition auxiliary layer, which involves the administration and management of medical resources in an efficient manner and facilitates the smooth operation of the iHome system. In this layer, cloud computing services are available to health and life science provides an efficient way for data security and patient privacy protection.

❖ The sensor data collecting layer is the basis of the entire network. It consists of data sensing and acquisition devices, local computing and processing units, data storage devices, and wired/wireless transmitting modules. It is a multi-standard wireless sensor platform, compatible with different wired/wireless protocols, such as Ethernet, RFID, Zigbee, Wi-Fi, Bluetooth, and 3G/4G network. With this three-layer iHome Health-IoT system, interaction between clinical professionals and home-stay patients can easily take place on demand or on a regular basis. In this present work, we have used wireless zigbee to transmit the data. And we have used micro chip sensor to sense temperature, blood pressure and heart beat. The details of the sensors used are explained in the next section.

C. BIO-SENSORS

Biological signals, such as ECG and electroencephalograms (EEG), are the most commonly used vital tools for monitoring patients' physical condition and diagnosing diseases. In particular for premature heart attacks, a very high proportion of health attacks happen during sleep or daily activities. The sooner the symptom is detected, the earlier medical treatment and the better prognoses can be made for the patients. However heart diseases, many chronic diseases are asymptomatic which lead to difficulties in accurate detection during a short visit to a hospital. Therefore, long-term continuous health monitoring is essential for detecting and treating diseases. Existing continuous monitoring systems (e.g., Holter system, etc.) are usually uncomfortable and inconvenient for long-term use, due to their physical limitations, e.g., large size, rigid package and twisted wires, etc. In previous work, a wearable ECG sensor node was developed using off-the-shelf components, which successfully detected ECG signals.
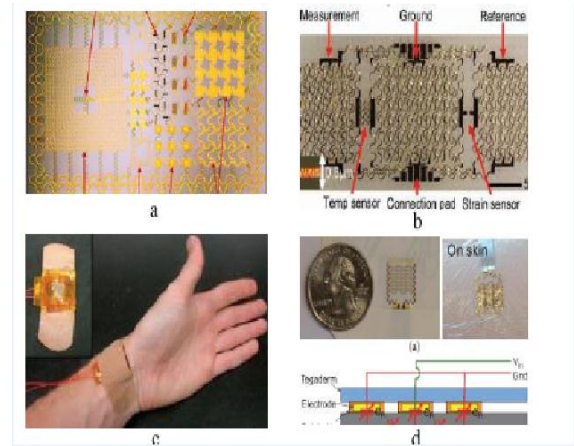

Fig. 3 Bio-chips and micro-sensors

Fig-3 illustrates about bio-chips and micro sensors. However, this sensor node was strap-based. In order to ensure good physical contact between embedded electrodes and human body, the strap had to be tightly pressed against the user's chest, which make it uncomfortable for long-term use. This present study used a new generation of healthcare devices with features such as: 1) low power consumption therefore long battery life, and 2) an affordable price.

Bio-sensors transmit the data to mobile application using wireless zig bee device. The mobile applications and its features are discussed in the next section.

D. Mobile applications

The elderly are more likely to forget to take or fail to properly take the medicine as prescribed. The following cases may be common: the user takes a wrong medicine, takes too much or too little of a specific medicine, takes the medicine at the wrong time, or drug reactions happen with the possibility of causing death. According the severity in each case, various levels of alarms should be triggered. Keeping these in view our mobile application provides the following features.

❖ *Body analysis* – shows the entire analysis of heart rate, BP and temperature in different blogs.

❖ *Alert* – Whenever an user is in abnormal condition, an alert message is sent to both doctor and patient's family.

❖ *Prescription-* Based on the analysis the doctor sends the prescription to the patient. The message includes user information and the doctor can log into the database and make an estimation to decide whether it is necessary to contact the user immediately or deliver the information to the emergency center.

The doctor can take it as a reference for the next prescription.

❖ If the patient condition becomes normal within 15 min then the alarm turn off. By this the doctor will decide whether or not to contact the patient's relatives or delivery the case to an emergency center.

III. RELATED WORK

Li Da Xu[6] has proposed that the emerging compressed sensing (CS) theory can significantly reduce the number of sampling points that directly corresponds to the volume of data collected. They discussed how CS can provide new insights into data sampling and acquisition in wireless sensor networks and IoT. He briefly introduced the CS theory with respect to the sampling and transmission coordination during the network lifetime through providing a compressed sampling process with low computation costs. Then, a CS-based framework is proposed for IoT, in which the end nodes measure, transmit, and store the sampled data in the framework. Then, an efficient cluster-sparse reconstruction algorithm is proposed for network compression aiming at more accurate data reconstruction and lower energy efficiency. Performance is evaluated with respect to network size using datasets acquired by a real-life deployment.

In this, he proposed a CS framework for WSNs and IoT and introduced how the framework could be utilized to reconstruct the compressible information data into a variety of information systems involving WSNs and IoT. This framework provides a promising approach for compressible signal and data in information systems. It makes an effective new information and data gathering paradigm in networks and information systems.

Antonio J. Jara[10] has proposed that the Communication and information access is the basis to reach a personalized health end to end framework. Personalized health Capability is limited to the available data from the patient. The data is usually dynamic and incomplete. Therefore, it presents a critical issue for mining, analysis and trending. For this reason, he presents an interconnection framework for mobile Health (mHealth) based

on the Internet of Things. It makes continuous and remote vital sign monitoring feasible and introduces technological innovations for empowering health monitors and patient devices with Internet capabilities. It also allows patient monitoring and supervision by remote centers, and personal platforms such as tablets. In terms of hardware it offers a gateway. He presents the architecture and evaluates its capability to provide continuous monitoring, ubiquitous connectivity, extended device integration, reliability, and security and privacy support. The proposed interconnection frame work and the proposed protocol for the sensors have been exhaustively evaluated in the framework of the project, which is focused on patients.

G. Kortuem[8], has proposed that the term Internet Of Things refers to networked interconnection of objects of diverse nature, such as electronic devices, sensors, but also physical objects and beings as well as virtual data and environments. He tried to resolve the existing restrictions of current architectural models by integrating both RFID(Radio Frequency Identification) and smart object-based infrastructures. He proposed the architecture that is based on a layered lightweight and open middle-ware solution following the Service Oriented Architecture and the Semantic Model Driven Approach, which is realized at both design-time and deployment–time covering the whole service lifecycle for the corresponding services and applications provided. According to him, one could track the entire existence of an object, from the time before it was made (its virtual representation), through its manufacture, its ownership history. He says IoT as a composition of smart objects that can understand and react to their environments. Through practical experimentation and by prototyping some generations of smart objects. It offers easy-to-use web service interfaces for controlling any type of physical sensor devices irrespective of its network technology. It also incorporates means of device and service discovery, semantic model driven architecture and security.

[1] Sonam V. Maju proposed An IoT application in the health platform which involves sensors for reading the human heart rate in digital format and an intelligent medicine box with a light sensor to indicate the variations in the medicine slots like counting the number of tablets a patient is consuming, alarms are there for consuming wrong medicine and more than that this medicine box will act as medication reminders. An IoT-based intelligent home-centric healthcare IOT platform, which flawlessly connects smart sensors attached to human body for biological monitoring and intelligent medical packaging for daily medication management. It includes the scenario of assisted living for people with physically and mentally

disabled, where users can intermingle with smart objects deployed in a home environment to ensure their health and well-being. The Med Box serves as a home healthcare station providing strong interoperability and network connectivity. The Internet of Things will change our society, and will bring seamless 'anytime, anywhere' personalized healthcare and monitoring over fast reliable and secure networks. This implies that we are approaching the end of the divide present between digital, virtual and physical worlds.

[2] Z. PanG proposed that "Technologies and Architectures of theThe emerging technology breakthrough of the Internet-of-Things (IoT) is expected to offer promising solutions for food supply chain (FSC) and in-home healthcare (IHH), which may significantly contribute to human health and well-being. In that thesis, he investigated the technologies of the IoT for these two applications called Food-IoT and Health-IoT respectively. He intended to resolve a series of research problems and integration architectures. In order to reduce the time-to-market and risk of failure, he took the business aspects into account. The challenges about enabling devices that he have addressed include: the WSN mobility and wide area deployment, efficient data compression in resource-limited wireless sensor devices, reliable communication protocol stack architecture, and integration of acting capacity to the low cost intelligent and interactive packaging (I2Pack). At the system level, he had addressed the challenges about effective integration of scattered devices and technologies, including EIS and information integration architectures such as shelf-life prediction and real-time supply chain re-planning for the Food-IoT, and device and service integration architectures for the Health-IoT. The emerging technology breakthrough of the Internet-of-Things (IoT) is expected to offer promising solutions for food supply chain (FSC) and in-home healthcare (IHH), which may directly contribute to human health and well-being. To reduce the time-to market and risk of failure, business aspects should be taken into account more than before in the early stage of IoT technology development because the technologies and applications are both immature.

[5] K. Ashton Personalized Health Assistants have gained popularity over the last few years. Such technologies allow users to monitor their health information in real time and often integrate with their smart devices, especially smart phones. Augmented Quick Health (AQH) is such an intelligent health monitoring system, which uses multiple sensors to read heartbeat, body temperature and sweat rate information. The sensor readings are used together to determine the health condition of a subject. With the resulting output, it interacts with a smart device and data is pushed in to a robust cloud-based infrastructure via the device (e.g. Smartphone). The system is designed to be extensible and

flexible, so adding new sensors and/or use-cases is straightforward. While there are other personalized health monitoring systems, AQH incorporates location based search, presented using augmented reality, which has proven to be an effective tool in emergency situations.

[7] E. Welbourne work objective is the discussion of challenges and requirements for pervasive healthcare computing towards an integrated service-oriented platform for the management of mission critical Healthcare Environments (HEs), such as Intensive Care Units (ICUs). Due to the large amount of equipment and technologies for patient monitoring, especially for those patients in intensive care as well as the vast amount of available information about patients' health to deal with, there is a great motivation related to the environment integration on an intelligent computing platform capable of processing messages, providing healthcare services, and eventually, making decisions autonomously and safely, to ensure the health and well-being of patients. The challenge of integrating medical equipment for monitoring patients' health goes beyond the economic and social aspects, i.e. deals with aspects related to technology, infrastructure and even technology acceptance by institutions, physicians and society, in general. Regarding to the technological aspects is possible to physically integrate medical devices through different communication networks, since equipment is available in each ICU, and specific communication protocols are defined. Moreover, several issues related to the HE requirements, and specifically towards to patients monitoring, suffer an impact from barriers imposed by proprietary medical equipment specifications that need to be prepared to be integrated on the PHE.

IV.   CONCLUSION AND FUTURE WORK

Health Related IoT applications is the need of the present smart city concept. Since the diseases are chronic and severe, these types of applications will be more helpful. In our present study, we have developed a prototype. Our future work focuses on reducing cost and dropping the inconvenience while using the strap.

REFERENCES

[1] Department of Economic and Social Affairs of the United Nations. 2009. World Population Ageing 2009, New York: United Nations.

[2] Z. Pang, "Technologies and Architectures of the Internet-of-Things (IoT) for Health and Well-being," PhD Thesis, Royal Institute of Technology (KTH), Stockholm, Sweden, 2013.

[3] C. E. Koop, et al., "Future delivery of health care: Cybercare," IEEE Engineering in Medicine and Biology Magazine, vol.27, no.6, pp.29-38, Nov. 2008.

[4] B. Schuz, et al., "Medication beliefs predict medication adherence in older adults with multiple illnesses," Journal of Psychosomatic Research, vol. 70, no. 2, pp. 179-187. 2011. [5] K. Ashton "That 'Internet of Things' Thing," RFID Journal, Jun. 2009.

[6] S. Li, L. Xu, and X. Wang, "Compressed Sensing Signal and Data Acquisition in Wireless Sensor Networks and Internet of Things, " IEEE Transactions on Industrial Informatics vol.9, no.4, pp. 2177-2186, Nov. 2013.

[7] E. Welbourne, et al., "Building the Internet of Things Using RFID: The RFID Ecosystem Experience," IEEE Internet Computing vol.13, no.3, pp.48-55, Jun. 2009.

[8] G. Kortuem, F. Kawsar, D. Fitton, and V. Sundramoorthy, "Smart objects as building blocks for the Internet of things," IEEE Internet Computing, vol.14, no.1, pp.44-51, Feb. 2010.

[9] S. Tozlu, M. Senel, W. Mao, and A. Keshavarzian, "Wi-Fi enabled sensors for internet of things: A practical approach," IEEE Communications Magazine, vol.50, no.6, pp.134-143, Jun. 2012.

[10] A. J. Jara, M. A. Zamora-Izquierdo, and A. F. Skarmeta, "Interconnection Framework for mHealth and Remote Monitoring Based on the Internet of Things," IEEE Journal on Selected Areas in Communications, vol. 31, no.9, pp.47-65, Sep. 2013.

[11] S.-Y. Lee, L.-H. Wang, and Q. Fang, "A Low-Power RFID Integrated Circuits for Intelligent Healthcare Systems," IEEE Transactions on Information Technology in Biomedicine, vol.14, no.6, pp.1387-1396, Nov. 2010.
[12] X. Li, R. Lu, X. Liang, X. Shen, J. Chen, and X. Lin, "Smart community: an internet of things application," IEEE Communications Magazine, vol.49, no.11, pp.68-75, Nov. 2011.

[13] G. Broll, E. Rukzio, M. Paolucci, M. Wagner, A. Schmidt, and H. Hussmann, "Perci: Pervasive Service Interaction with the Internet of Things," IEEE Internet Computing, vol.13, no.6, pp.74-81, Nov. 2009.

[14] European Commission Information Society. 2008. "Internet of Things in 2020: A Roadmap for the Future." http://www.iot-visitthefuture.eu.

[15] A. Hande, and E. Cem, "Wireless sensor networks for healthcare: A survey," Computer Networks, vol. 54, no. 15, pp. 2688-2710, Oct. 2010. [16] National Information.

[16] National Information Council. 2008. "Global Trends 2025: A Transformed World. US Government Printing Office."http://www.acus.org/publication/global-trends-2025-transformed-world.

[17] S. Li, L. Xu, and X. Wang, "A Continuous Biomedical Signal Acquisition System Based on Compressed Sensing in Body Sensor Networks," IEEE Transactions on Industrial Informatics, vol.9, no.3, pp.1764-1771, Aug. 2013.

[18] P. Castillejo, J.-F. Martinez, J. Rodriguez-Molina, and A. Cuerva, "Integration of wearable devices in a wireless sensor network for an E-health application," IEEE Wireless Communications, vol.20, no.4, pp.38-49, Aug. 2013.

[19] X.-F. Teng, Y.-T. Zhang, C. Poon, and P. Bonato, "Wearable Medical Systems for p-Health," IEEE Reviews in Biomedical Engineering, vol.1, no., pp.62-74, and 2008.

[20] J. Morak, H. Kumpusch, D. Hayn, R. Modre-Osprian, and G. Schreier, "Design and Evaluation of a Telemonitoring Concept Based on NFC-Enabled Mobile Phones and Sensor Devices," IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 1, pp. 17–23, Jan. 2012.

[21] H. Baek, G. Chung, K. Kim, and K. Park, "A Smart Health Monitoring

[22] R. Paradiso, G. Loriga, and N. Taccini, "A wearable health care system 10.1109/TII.2014.2307795, IEEE Transactions on Industrial Informatics.

[23] E. Xu, M. Wermus, and D. B. Bauman, "Development of an integrated medical supply information system," Enterprise Information Systems, vol.5, no.3, pp. 385-399, May. 2011.

[24] S. Coyle, et al., "BIOTEX—Biosensing Textiles for Personalised

[25] C. R. Merritt, H. T. Nagle, and E. Grant, "Fabric-based active electrode

[26] M. Kaltenbrunner, et al., "An ultra-lightweight design for imperceptible

[27] H. Fuketa, et al., "1μm-thickness 64-channel surface electromyogram

[28] T. Someya, "Building bionic skin," IEEE Spectrum, vol.50, no.9,

[29] H. Liu, "Exploring Human Hand Capabilities Into Embedded

[30] L. Xu, "Enterprise Systems: State-of-the-Art and Future Trends," IEEE Operators: The ProeTEX Project," IEEE Transactions on Information vol.7, no.4, pp.630-640, Nov.2011

An Approach to the Development of Smart Village under Virtual Environment

[1] Laveti.Lakshmideepak [2] M.Kesava [3] M.Suswara [4] Dr. Challa Narasimham
[1][2][3] Undergraduate Student, [4] Professor,
Department of CSE, VIIT, Visakhapatnam, A.P, India
[1] lakshmideepak.laveti@gmail.com [2] mekakesava@gmail.com, [3] narasimham_c@yahoo.com
[4] suswaramupparajum@gmail.com

Abstract— **Smart village development is the main objective of this paper. Smartness can be achieved by Wireless Sensor Networks (WSN) and pervasive computing. People in the villages will be able to communicate with the government from any device, any location, anytime and in any format. This paper describes a mobile application by which one can communicate with the mobile server, cloud, every sensor that is connected to another sensor wirelessly to form a wireless sensor network to sense the surrounding environment. The rapid increase in the development of these devices in communicating-actuating network develops the Internet-of-Things (IOT). Sensors and actuators in this environment sense the information and is shared across platforms in order to create a Common Operating Picture (COP). Cloud centric vision is one of the application of this paper.**

Index Terms— **Actuators, Common Operating picture (COP), Internet of Things (IOT), Pervasive computing, Sensors, Wireless Sensor Networks (WSN).**

I. INTRODUCTION

India is a country of villages. It needs to provide the villages with the latest technologies to facilitate their better livelihood.. It is necessary to build smart villages and smart wards. Smartness can be achieved by using smart devices which can sense the data and can do operations on the data acquired from the surrounding environment. The main theme of this system is to develop a free basic application. This system has to sense the data that is one part and this system has to communicate with the people to facilitate and to solve their problems. This system should help the villages to develop the smarter environment. There are some points this paper to develop smart villages are zero dropouts from the schools and farmers subsidiary systems, grievance cell, health of the citizens and to prevent malnutrition for the children. Skill development centers are to be established to improve the skills of the youth in the villages. People can be able to communicate with government and their officials from any place, time, and format without any data charges to make use of the application. The data or the information collected through this system is very huge. The data are maintained in the cloud platform. This cloud platform should share the data equally at any instance of time to all the resources. Now the www is upgraded to web3, since the data usage is more and the format of using the data is formatted. As we move on from the static web pages to mobile web pages. Skill development should be done as to shift the people from the primary sector, i.e. from agriculture to secondary sector i.e. industries. More people have to shift in the industrial sector than the agriculture sector in the villages for the development of the nation.

II. EXISTING SYSTEM

Now, everything today is manually done without the involvement of technology. So it is difficult to maintain with the paper work. Some places like district headquarters and mandala's got the fiber optic internet connectivity. This connectivity is to mainly reside with government sector only. But in a few cases the fiber optic internet connectivity is given to the common man. The government is maintaining some data records of the public as like aadhar details, bank details etc. Even most of the schools don't have internet connectivity.
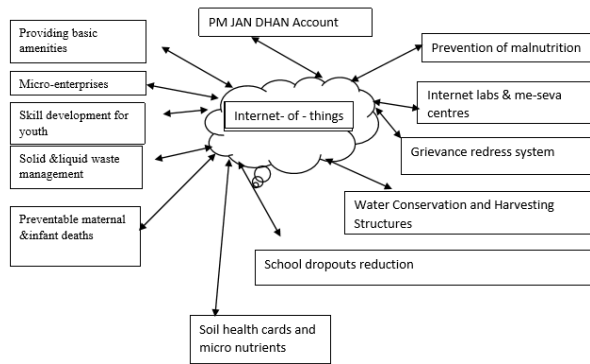
III. PROPOSED SYSTEM

Fig 1 overall diagram for smart villages

The above figure shows the areas to be worked for the development of the smart villages. The people in the villages have to connect to the internet to facilitate them. The main areas to be focused are to provide basic amenities to all the people, to establish the micro enterprises with women in the village and they are to be connected with the internet, skill development for youth and for students, solid and waste management should be monitored digitally, prevention of maternal and infant death rates, providing of soil health cards and micro-nutrients to farmers for the lands they have, reduction of school dropouts up to 12th class, development of the water conservation and harvesting structures, grievance redressal system have to be improved, establishment of internet centers to educated the people, pm jan dhan account has to be digitally monitored to prevent the money loss due to corruption, prevention of mal-nutrition to the children and to pregnant ladies and finally health services are to be developed in the villages. These are the main areas which are identified for the development of smart villages.
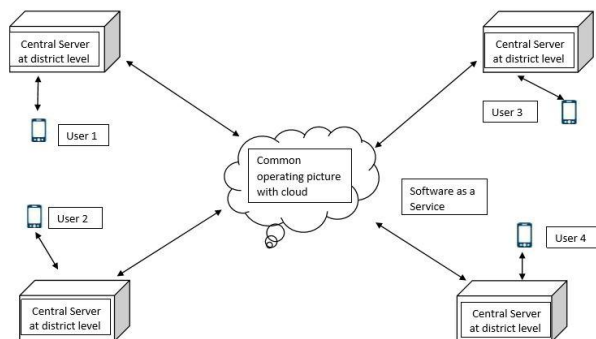

Fig2 connectivity of the local server with cloud and to users

This central idea is to be to develop a cloud based internet connectivity to all the users with common operating picture. Cloud is one of the best solutions for huge data and with more user population. Users can know this system by using the free basics application which runs in the cloud as Software as a service (Saas). People can access to the data from any place anytime in any format. Everything in the system has to be digitalized and partially automated to prevent the corruption and for the development of the better society. People can utilize this system and its services even from their handheld devices and smart phones. Common operating picture (cop) is located centrally in the cloud to provide information efficiently to all the users equally. There are some preventive measures to protect this system. Mobile application comes into the picture to communicate effectively with people. This system will provide some efficient access to the users without any breakdown of the system or overload to the server as the population is more. Population is the main constrain for bring up any idea with the proper utilization of the resources. Information or data is passed to the nodes or to server with shortest routing algorithms for the efficient access and the retrieval of the data from the end user and even from the server end also.
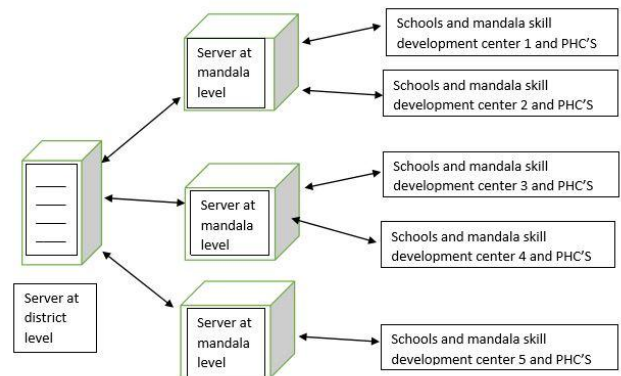

Fig 3 connectivity of PHC's and schools and skill development centers

There will be a server at the district level which is directly connected to the cloud. The servers at the district level is connected internally with the servers in the mandala level. Every school in the each mandala is connected with the server in that mandala. Schools are to be first connected to the internet for the development of the society. As the children are the future citizens of the nation. Every skill development center is connected to the internet. This connectivity is required is to learn new things from the outside world to develop themselves. There should be some activities to common people to actively participate in the course to develop the nation or else they don't show the interest in the development of the smart nation. Main thing is the health issue of the public. So it is the main consideration to the development and all the PHC's are connected to the servers to know the public health issues and the diseases in that area. This prevents

the maximum deaths of children in the initial stages. Important thing is that people are connected through mobile application to know the actions to be taken in the emergency times. They can easily get the first aid and medical prescription for health problems or to unexpected accidental conditions. Even for the uneducated people they can upload a one minute video to express their problem to the government officials. There they can conduct camps for the development of the villages. Mainly in the villages many people are working in the agriculture sector, so it does not improve the society much. So they have to be transformed to the industrial sectors to create and develop new products. If the new products are improved then there will be in improvement in the exports business from our country with other countries. Thus there will be increase in the value of our countries money. Automation of the sector is essential for the improvement but it should not be fully automated. Schools are to be connected with the internet to learn knowledge as it is a huge data source. They can learn new techniques to improve the performance of the system. Fiber optic connectivity is to be established as it is important like a water to a village. Water bring liveliness to the lives of the people and internet connectivity will improve their standard of living. Common operating picture has to identify the system faults and it has to be rectified easily. Unexpected incidents might happen so that information is to be intimated all the departments of the government to take necessary actions.
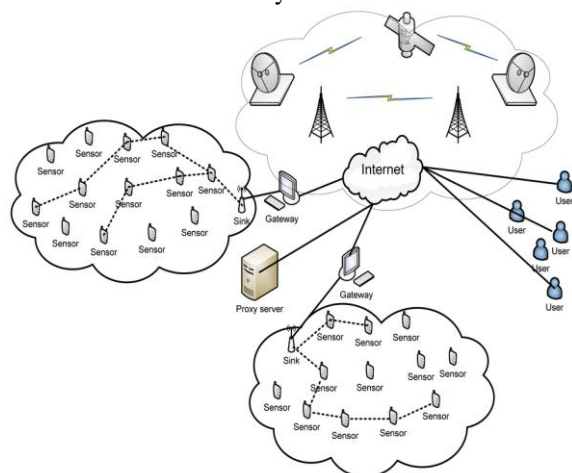

Fig 4 wireless sensor network (WSN) connectivity

Sensors place a major role in protecting in the people from natural calamities. Main purpose is to identify the temperature conditions to suggest the people to take precautions to prevent deaths. We can provide emergency services to the people in the effective way. They can know temperature levels at different times like in summer they can take precautions to protect themselves from the heat and even in the winter times also. With these sensors we can know the information about the ground water levels

and to know the water level which can improve the crop yield. This information can be sent to the public with the help of free basics mobile application. People can know the information and can the severity of the issues in the natural calamities.

Fig 5 Main location of the centers in the state.

These are the predictable locations to establish the internet connectivity centers in the state.

IV.   CONCLUSION

The main output of this system is to build smart villages. The free basic application has to serve the people to improve their connectivity with their people's government. Sensors and actuator have to work for the development of smarter cities. The cloud will be the basic data source and internet of things is the major concept in developing the smart cities and villages.

REFERENCES

[1] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, Marimuthu Palaniswami, "Internet of Things (IoT): A Vision, Architectural Elements and Future Directions", Elsevier, Page No 1 to 28, cited by 893.

[2] J.Srinivas1, K.Venkata Subba Reddy, Dr.A.Moiz Qyser, "CLOUD COMPUTING BASICS," International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 5, July 2012.

[3] W.Bajwa,J.Haupt, A.Sayeed,R.Nowak, "Compressive wireless sensing, in: ACM, 2006.

[4] Andhra Pradesh State Government Gazette.

[5] K.Ashton, That "Internet Of Things" Thing Rfid Journal. (2009).

Prediction of Cancer Risk in Perspective of Symptoms using Naïve Bayes Classifier

[1] Pallavi Mirajkar [2] Dr. G. Prasanna Lakshmi

[1]Research Scholar, Faculty of Computer Science [2] Guide, Associate Professor

[1] Pacific Academy of Higher Education and Research University, Udaipur.

[2] Andhra University

Abstract— **Health Care professionals face the complex task of predicting the type of cancer in patient. This earlier prediction of cancer helps to the practitioner to recommend cancer treatment. Various studies have been carried out that examine patient to improve help practitioner's prognostic accuracy. Here we worked on major factor that is symptoms. In this thesis paper we have used Naïve Bayes Classification algorithm of data mining to predict the type of cancer. Proposed method of risk prediction aims to predict probability of cancer. Based on the classification algorithm, symptoms of the cancer are classified using Navie Bayes algorithm to recognize the risk of cancer such as lung, breast, ovarian, stomach and oral. Therefore accurate prediction of cancer in patient is important for good clinical decision making in health care strategies.**

Index Terms— **Data Mining, Naïve Bayes Classifier, Probability, Cancer, Symptoms.**

INTRODUCTION

Cancer is a collection of diseases concerning anomalous cell growth with the potential to attack or spread to other parts of the body. Cancer becomes the most life threatening disease now days. Accurate prediction of cancer in patient is important for good clinical decision making and for recommending the treatment, at present, accurate detection remains challenge even for experienced clinicians. The most effective way to decrease cancer deaths is its early recognition. Early diagnosis requires an accurate and reliable diagnosis procedure that allows physicians to diagnose symptoms of cancer.

Data mining plays a crucial role in the discovery of knowledge from huge databases. Data mining has created its important hold in every field including healthcare. Data mining has its most important role in extracting the concealed information. Data mining process is more than the data analysis which includes classification, clustering, association rule mining and prediction. A patient affected with any Cancer may suffer symptoms in the body. Cancer symptoms are used to forecast the risk level of the cancer disease. The main aim of this study is to predict the risk of type of cancer using Naïve Bayes classification algorithm. Naive Bayes is one of the most effective statistical and probabilistic classification algorithms. Naïve Bayes algorithm is called "Naïve" because it makes the assumption that the occurrence of certain feature is independent of the occurrence of other features. Naive

Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. Naive Classifiers assumes that the effect of an attribute value on the given class is independent of the values of other attributes. Naive Bayesian Classifiers depends upon the Baye's theorem. An advantage of the Naive Baye's classifier is that it requires a small amount of data to estimate the parameters needed for classification. It performs superior in numerous complex real world situations like weather forecasting, Medical Diagnosis, spam classification. It is suitable when dimensionality of input is high.

RELATED WORK

Kritharth Pendyala et.al (2017) proposed a system that detects and predicts breast cancer. They have applied classification algorithm to data using Naïve Bayes and SVM algorithm. The proposed hybrid algorithm gives more accuracy and reduces the time complexity.

Subrata Kumar Mandal (2017) has applied various techniques such as data cleaning, feature selection, feature extraction, data discretization and classification for predicting breast cancer as accurately as possible. They states that Logistic Regression Classifier has given maximum accuracy.

N.V. Ramana Murty and Prof. M.S. Prasad Babu (2017) analyzed lung cancer data set using WEKA tool with various data mining classification techniques. It is

recognized that the Navie Bayes classification algorithms.

Keerti Yeulkar et. al. (2017) classified the SEER dataset in to malignant tumor and benign tumor using C4.5 and Naïve Bayes algorithm. They have also applied decision tree classification technique.

Younus Ahmad Malla et.al (2017) performed analytical evaluation of machine learning algorithm such as Naïve Bayes, Logistic Regression and Random Forest. WEKA tool is used for running different data mining algorithms for breast cancer data set.

Rashmi M, Usha K Patil(2016) have used Navie Bayes algorithm to diagnose cancer using Gene data set. They stated that data mining can be used to classify cancer patient by analyzing the data*.*

Animesh Hazra et.al. (2016) have studied various classification techniques and compared them in terms of accuracy and time complexity. They have concluded that Navie Bayes gives better accuracy and it is best classification technique.

Shweta Kharya , Sunita Soni (2016) have developed a predictive model using Navie Bayes algorithm. They have applied weighted concept on Navie Bayes Classification for breast cancer diagnosis. They developed model is more readable, modifiable, efficient and can be handled easily.

Dr. Vani Prumal et.al. (2016) have proposed methodology that uses Naïve Bayes classifier to predict stomach cancer. They stated that Naïve Bayes algorithm is the most appropriate algorithm for the prediction of stomach cancer. It gives more accurate result.

Dr. P. Indra Muthu Meena (2016) implemented C4.5 and Naïve Bayes algorithm for prediction of stomach cancer. They have also analyzed and compared the performance of both algorithms.

Kathija, Shajun Nisha (2016) proposed an algorithm by using support vector machine and Naïve Bayes algorithm. They evaluated performance of the proposed algorithm using confusion matrix. After comparing both algorithms, it is found that Naïve Bayes algorithm gives highest accuracy.

Samuel Giftson Durai et.al (2016) have done survey and shown that decision tree is the best suitable classification algorithm for prediction of breast cancer and experiments have been conducted using MATLAB software.

B. M. Gayathri, C. P. Sumathi (2016) applied Naïve Bayes algorithm and designed GUI to enter details of the patient for predicting whether given data is benign or malignant. Proposed work has shown that Naïve Bayes classifier predicts breast cancer with less attributes.

P. Saranya, B. Satheeskumar (2016) analyzed that data mining techniques helps to recognize data to predict the cancer at an early stage. They have studied various data mining algorithms such as Naïve Bayes, ANN, Support Vector Machine and decision tree*.*

Monika Hedawoo et. al. (2016) proposed method to classify mammogram using different classification algorithms for predicting tumor cells in breasts. It has been revealed that Naïve Bayes algorithm is more sensitive.

Animesh Hazra et. al. (2016) applied data cleaning, feature selection, feature extraction, data discretization and classification techniques to predict breast cancer correctly. In proposed method Naïve Bayes algorithm gives maximum accuracy.

V. Kirubha, S. Manju Priya (2015) proposed method to analyze risk factors of lung cancer. They have used classification algorithm like Naïve Bayes, Random Forest, REP Tree and Random tree. It is proved that REP Tree provides better result.

S. Kalaivani et.al (2015) studied three classifiers Naïve Bayes, Random Tree and Support Vector Machine. They have calculated the performance of these algorithms and it is observed that Naïve Bayes algorithm gives more accurate results.

Mandeep Rana et.al (2015) has applied different algorithms on dataset such as KNN, logistic regression and Naïve Bayes algorithm. It is observed that Naïve Bayes and logistic regression gives accurate results for diagnosing breast cancer.

PROPOSED WORK

Introduction
For better perceptive of approach, we take the paradigm which is shown below in the following Tables:

Symptoms of Lung Cancer
Cough that won't quit
Pain in the chest area
Wheezing
Raspy, hoarse voice
Drop in weight
dyspnea (shortness of breath with activity)
dysphasia(difficulty Swallowing)
Shortness of breath

Table No. 1 Lung Cancer

Symptoms of Breast Cancer
Hard Lump
Changes in the shape or size of the breast
Changes to the nipple(inverted nipple)
Discharge that comes out of the nipple without squeezing
Breast pain

Table No. 2 Breast Cancer

Symptoms of Oral Cancer
Lump in the neck
Unexplained bleeding in the mouth
Losing in the teeth
Pain in chewing, swallowing or moving tongue

Table N. 3 Oral Cancer

Symptoms of Ovarian Cancer
Irregular menstrual
Painful intercourse
Increase in urination
Changes in bowel movement
Abnormal bleeding
Pain in the abdomen
Bloating

Table No. 4 Ovarian Cancer

Symptoms of Stomach Cancer
Indigestion or heartburn
Vomiting and nausea
Pain in the abdomen
Constipation or diarrhea
Bloating Stomach

Table No. 5 Stomach Cancer

Naïve Bayes classifiers are a group of basic probabilistic classifiers based on applying Bayes theorem with independence assumptions between the features. Using Bayesian classifiers, the method can determine the hidden facts related with diseases from symptoms of the patients. According to naïve Bayesian classifier the occurrence (or unoccurrence of a particular feature of a class is considered as independent to the presence (or absence) of any other feature. When the dimension of the inputs is high and more efficient result is expected. Naïve Bayes model identifies the physical characteristics and symptoms of patients suffering from cancer disease.Naive Bayes algorithm is based on Bayes theorm. There are two types of probabilities

- Posterior Probability [P(H/X)]
- Prior Probability [P(H)]

where X is data tuple and H is some hypothesis. According to Bayes' Theorem, $P(H/X)= P(X/H)P(H) / P(X)$.
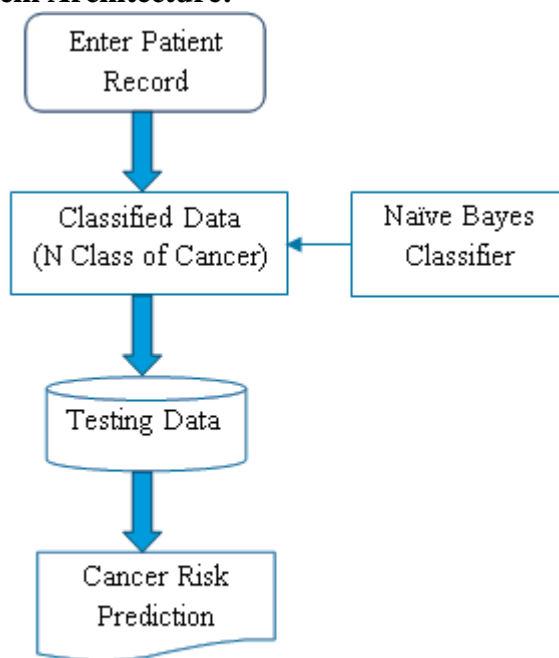
System Architecture:

Figure 1 System Architecture

Implementation of an Algorithm is as given below:
Suppose that there are n classes of cancer, C1,C2,...,Cn and each data sample is represented by n dimensional symptoms of cancer, S= {S1 ,S2 ,…Sn}.

Step 1: To calculate the probability of type of cancer. First recognize the symptoms of the specific cancer. P(C1|S1,S2,...,Sn).

Step 2: Calculate the probability of each symptoms.

Probability of each Symptom = P(Si|C1)

Where i= {1,2,…,n}

Step 3: Multiply each probability of symptoms together

Probability of C1=

P(S1|C1) * P(S2|C1) * … * P(Sn |C1)

Step 4: Repeat step 1 to Step 3 for each cancer class.

Step 5: Predict that S belongs to the type of cancer which is having the highest posterior probability.

P(Ci|S) > P(Cj>S) for all 1< = j < = n and j != i

Display the risk of one of the type of cancer from predefined set of class.

The Naïve Base algorithm is playing a vital role in mining the necessary information. Based on the above mentioned algorithm and computed the risk of cancer, some tests will be prescribed to confirm the presence of cancer.

CONCLUSION

In thesis paper, we have used Naïve Bayes classifier for predicting cancer by taking into consideration the related symptoms. It is essential to work on different symptoms and discover the probability of cancer in patient. Prediction of cancer at an early stage can be helpful for better treatment and survival of the patient. In thesis paper, we compute the probability of risk of the cancer based on symptoms. And then gives a warning of which type of cancer patient may have. The goal of thesis paper is to give the prior warning to the patient. This will eventually spare the time and decrease the cost of treatment and conjointly will increase the possibility of survivability. Future work can be directed on the prediction of cancer by integrating the various dimensions such as cancer causing factors, contaminated food items and symptoms.

REFERENCES

1. Rashmi M, Usha K Patil "Cancer Diagnosis Using Navie Bayes Algorithm" International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 05; May - 2016 [ISSN: 2455-1457].

2. Animesh Hazra, Subrata Kumar Mandal, Amit Gupta "Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 145 – No.2, July 2016.

3. Shweta Kharya, Sunita Soni "Weighted Naive Bayes Classifier: A Predictive Model for Breast Cancer Detection" International Journal of Computer Applications (0975 – 8887) Volume 133 – No.9, January 2016.

4. Dr. Vani Prumal, Shibu Samuel, Dr. P. Indra Muthu Meena "Application of Training Dataset using Naïve Bayes Classifier for Prediction of Stomach Cancer in Female Population" International Journal of Scientific Engineering and Technology Research Volume.05, IssueNo.45, November-2016, Pages: 9230-9234

5. Dr. P. Indra Muthu Meena, Dr. Vani Perumal "Performance of C4.5 and Naïve Bayes Algorithm to Predict Stomach Cancer - An analysis" International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Vol. 5, Issue 11, November 2016

6. Kathija, Shajun Nisha "Breast Cancer Data Classification Using SVM and Naïve Bayes Techniques" International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Website: www.ijircce.com Vol. 4, Issue 12, December 2016

7. Subrata Kumar Mandal "Performance Analysis Of Data Mining Algorithms For Breast Cancer Cell Detection Using Naïve Bayes, Logistic Regression and Decision Tree" International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 6 Issue 2 Feb. 2017, Page No. 20388-20391

8. Kritharth Pendyala., Divyesh Darjee., Vaishnavi Adhyapak and Vaishali Gaikwad "Cancer Detection & Prediction Using Dual Hybrid Algorithm" International Journal of Recent Scientific Research Vol. 8, Issue, 3, pp. 15777-15789, March, 2017.

9. Younus Ahmad Malla, Mohammad Ubaidullah Bokari "A Machine Learning Approach for Early Prediction of Breast Cancer" International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 6 Issue 5 May 2017, Page No. 21371-21377

10. V. Kirubha, S. Manju Priya "Comparison of Classification Algorithms in Lung Cancer Risk Factor Analysis" International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

11. Samuel Giftson Durai, S. Hari Ganesh and A. Joy Christy "Prediction of Breast Cancer through Classification Algorithms: A Survey "IJCTA,9(27),2016,pp.359-365© International Science Press.

12.   B. M. Gayathri, C. P. Sumathi "An Automated Technique using Gaussian Naïve Bayes Classifier to Classify Breast Cancer" International Journal of Computer Applications (0975 – 8887) Volume 148 – No.6, August 2016.

13.   P. Saranya, B. Satheeskumar "A Survey on Feature Selection of Cancer Disease Using Data Mining Techniques" International Journal of Computer Science and Mobile Computing, Vol.5 Issue.5, May- 2016, pg. 713-719.

14.   Animesh Hazra, Subrata Kumar Mandal, Amit Gupta "Study and Analysis of Breast Cancer Cell Detection using Naïve Bayes, SVM and Ensemble Algorithms" International Journal of Computer Applications (0975 – 8887) Volume 145 – No.2, July 2016.

15.   Keerti Yeulkar, Dr. Rahila Sheikh "Utilization of Data Mining Techniques for Analysis of Breast Cancer Dataset Using R" www.ijraset.com,Volume 5 Issue III, March 2017 IC Value: 45.98 ISSN: 2321-9653.

16.   Monika Hedawoo, Abhinandan Jaiswal, Nishita Mehta "Comparison of Data Mining Algorithms For Mammogram Classification" International Journal Of Electrical, Electronics And Data Communication, ISSN: 2320-2084 Volume-4, Issue-7, Jul.-2016.

17.   S. Kalaivani, S. Gandhimathi "An Efficient Bayes Classification Algorithm for analysis of Breast Cancer Dataset using Cross Validation Parameter" International Journal of Advanced Research in Computer Science and Software Engineering 5(10), October- 2015, pp. 512-515.

18.   Mandeep Rana, Pooja Chandorkar, Alishiba Dsouza, Nikahat Kazi "Breast Cancer Diagnosis and Recurrence Prediction Using Machine Learning Techniques" International Journal of Research in Engineering and Technology eISSN: 2319-1163 | pISSN: 2321-7308 Volume: 04 Issue: 04 | Apr-2015, http://www.ijret.org

19.   N.V. Ramana Murty and Prof. M.S. Prasad Babu "A Critical Study of Classification Algorithms for Lung Cancer Disease Detection and Diagnosis" International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 13, Number 5 (2017), pp. 1041-1048

Biography:

Mrs. Pallavi Mirajkar passed M.Sc. (Computer Science) from North Maharashtra University, Jalgaon. Currently she is pursuing her Ph.D. from Pacific University, Udaipur.

Dr. G.Prasanna Lakshmi, Guide.
Associate Professor, Andhra University.