ISSN#1973-9013 (Print)

Vol. : 8, January 2019

ABVU-MTMI's International Journal of Decision Science & Information Technology (IJDSIT)

(A Peer Reviewed Journal)

EDITOR-IN-CHIEF :

DR. H.S. HOTA

ATAL BIHARI VAJPAYEE UNIVERSITY,

BILASPUR (C.G.)

PUBLISHED BY :

DEPARTMENT OF COMPUTER SCIENCE & APPLICATION

ATAL BIHARI VAJPAYEE UNIVERSITY, BILASPUR (C.G.), INDIA

IN ASSOCIATION WITH :

MODERN TECHNOLOGY & MANAGEMENT INSTITUTION

(MTIMI) Inc., USA

ABVU - MTMI's

ISSN#1973-9013 (Print) Vol. : 8, January 2019

INTERNATIONAL JOURNAL OF DECISION SCIENCE AND INFORMATION TECHNOLOGY

(IJDSIT)

(A Peer Reviewed Journal)

Editor-in-chief

DR. H.S. HOTA

Atal Bihari Vajpayee University, Bilaspur, (C.G.), India

:: Published by ::

DEPARTMENT OF COMPUTER SCIENCE AND APPLICATION
ATAL BIHARI VAJPAYEE UNIVERSITY, BILASPUR, (C.G.), INDIA

In Association with

MODERN TECHNOLOGY AND MANAGEMENT INSTITUTION
(MTMI) Inc., USA

INTERNATIONAL JOURNAL OF
DECISION SCIENCE AND INFORMATION TECHNOLOGY
(IJDSIT)

ISSN# 1973-9013(Print) **Vol. : 8, January 2019**

Editorial Board:

Editorial Notes

The International Journal of Decision Science and Information Technology (IJDSIT) is an international peer reviewed journal dedicated to the latest advancements in the field of Decision Science and Information Technology. The current issue of journal is being published by Department of Computer Science and Application, Atal Bihari Vajpayee University, Bilaspur, Chhattisgarh, India in association with Modern Technology and Management Institution (MTMI) Inc. USA under the MoU in between Atal Bihari Vajpayee University and MTMI, USA.. The goal of this journal is to promote authentic and original research in many areas of achievements in Decision Science and information technology. Its broad nature allows for a wide dissemination of knowledge amongst researchers both from academe and industry, to initiate, cultivate, share, and discuss various new issues and developments in different areas of theories and practices. The journal will hopefully benefit scientists, business, industry and government leaders and managers relating particularly to Information Technology.

Topics that would be covered in IJDSIT include, but not limited are: Artificial Intelligence and Expert System, Decision Science, Information Technology, Big Data, Data Analytics, Machine Learning, Internet of Things (IoT), Computer Networking, Information Communication and Technology (ICT), Block chain and Crypto currency, Soft Computing, Information security, Cryptography and Network security, Graph theory, Data Mining, Distributed Computing, Decision support system, Cloud Computing, Robotics and Augmented reality, Pattern Recognition, Business Intelligence, Business Analytics, Smart Governance, E-Commerce and M-Commerce, E-Learning and M-Learning, Fuzzy logic and Fuzzy mathematics, Biotechnology and Bioinformatics, Optimization techniques or any other related topics.

I am happy to release the current issue of IJDSIT with six research articles received from the various parts of the word. These research articles covers the broad topics of the journal like Decision support System, Business Intelligence.

In the end I would like to thank Prof. G.D. Sharma, Hon'ble vice chancellor, Dr. Sudhir Sharma, Registrar, Atal Bihari Vajpayee University, Bilaspur, Chhattisgarh, India and Prof. Kamal Nayan Agarwal, Vice-Chairman, MTMI, USA and Professor, School of Business, department of Information System and Supply chain Management, Howard University, Washington DC, USA along with Prof. Dinesh Sharma, School of Business and Technology, Department of Business, Management and Accounting, Maryland University, Eastern Shore, USA to make this journal as collaborative journal of Atal Bihari Vajpayee University, Bilaspur, Chhattisgarh, India and MTMI, USA. I also thank all the members of editorial board for their untiring supports to see this journal in reality. I would also like to thank all authors to publish their research articles in IJDSIT. It is worth mentioning here the important feedbacks provided by the reviewers to authors to improve the quality of their manuscripts. Finally, I hope that readers will find the research articles published in IJDSIT useful and thought provoking.

Dr. H.S. Hota
Editor-in-Chief

INTERNATIONAL JOURNAL OF
DECISION SCIENCE AND INFORMATION TECHNOLOGY
(IJDSIT)

INDEX

GENERATING MAXIMUM ORDER LINEAR FEEDBACK SHIFT REGISTERS ON COMMODITY MULTICORES COMPUTING SYSTEMS

Daniel Okunbor, Fayetteville State University Fayetteville, NC 28301, U.S.A. (diokunbor@uncfsu.edu)

ABSTRACT

Linear Feedback Shift Registers (LFSRs) are used as pseudo random number generators and are they are obtained using primitive polynomials. These primitive polynomials provide maximum length LFSRs. Constructing primitive polynomials adopting LFSRs is computationally expensive both in terms of processor time and memory usage. In this paper, we demonstrate how to generate LFSRs using message passing single program multiple data (SPMD) constructs on commodity computing systems with 4 to 8 cores. In addition, we analyze the linear complexity of a few of the LFSRs generated.

Keywords: Linear Feedback shift Register (LFSR), message passing single program multiple data (SPMD).

INTRODUCTION

In cryptography, an encryption algorithm or cipher is used to systematically scrabble a message to ensure confidentiality and, in some cases, integrity of the message. While there are a variety of ciphers, this paper focuses on symmetric stream ciphers which encrypt and transmit message bits one at a time. It is symmetric because is based on a single secure key, also called private key. The stream cipher that is utilized in this paper is constructed adopting Linear Feedback Shift Registers (LFSRs). The LFSRs are the most popular for generating pseudo random sequences (keystream, $s = s_0, s_1, s_2, \ldots,$) due to their ease of implementation in hardware.The keystream is used to encrypt the message $m = m_0, m_1, m_2, \ldots,$ by XORing the keystream bits and the message bits, $e_i = s_i \oplus m_i$,where $e = e_0, e_1, e_2, \ldots,$ is the encrypted message.

An LFSR, as the name indicates is made of a register consisting of a linear array of circuitry flip flops that are clock controlled. The usual definition is that1 flip flop represents 1 bit of memory cell, however, when the bit is not stored but shifted to the next flip flop (either to the right or left), we have what is called a shift register. Aclock is used to determine the pulse (high or low) for each flip flop.

An LFSR is one in which its input is determined by some of the flip flop output bits called taps and these bits are used for clocking (called regular clocking). The associated clock values of these taps are 1's and other flip flops have clock values of 0's.The output bits of these taps are XORed to produce the next input bit for the LFSR (see Figure 1 below), hence the feedback mechanism. The output bit of the current state of the LFSR is the bit of the last flip flop. This is regarded as the Fibonacci LFSR, the other type but related LFSR is the Galois LFSR and it uses zero initial input and updates bits of some of the flip flops using the output bit, which is the most significant bit. Irregular clocking occurs only in multiple LFSRsthat are combined to generate the output bit (Okunbor, Amado, and Sharma, 2017; Okunbor, Sarami, Bhattacharya, and Sharma, 2017; Okunbor, 2018)

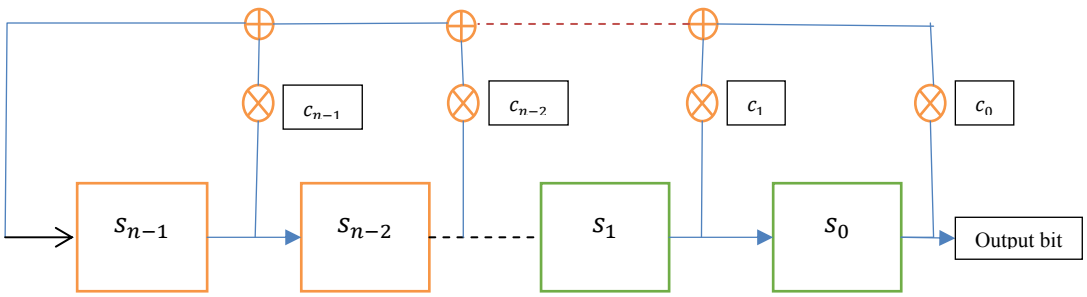

Figure 1: Linear Feedback Shift Register

As described previously, theelements of the connection vector are used to define the LFSR and are denoted by

$c = \{c_i | c_i \in \{0,1\}, i = 0,1, \dots, n-1\}$ of length n.

The initial state vector of an LFSR is a bitstreamdenoted by

$$s^0 = \{s_i^0 | s_i^0 \in \{0,1\}, i = 0,1, \dots, n-1\}.$$

The current state vector is obtained using the formula:

$$s^k = \{s_i^k | s_i^k = s_{i-1}^{k-1}, i = 1,2, n-2, s_{n-1}^k = \sum_{i=0}^{n-1} c_i s_i^{k-1} (mod\ 2)\}.$$

The output bit of Fibonacci LFSR after a shift cycle is the least significant bit (LSB). For example, suppose the initial state vector of an LFSR is $s_4^0 s_3^0 s_2^0 s_1^0 s_0^0 = 00001$ and the connection vector is $c_4 c_3 c_2 c_1 c_0 = 00101$, after first clocking the state of the LFSR is 10000 with an output bit of 1; after the second clocking the state of the LFSR is 01000 with output bit of 0; after the third clocking, the state of the LFSR is 00100 with output bit of 0; after the fourth clocking, the state is 10010 with output bit of 0; and so on. The bit sequence generated will be 1000010010110011…(see Table 1 below for state vectors after clocking63 times.)

Table 1: LFSR State vectors

00001	00111	00010	00011
10000	00011	00001	10001
01000	10001	10000	11000
00100	11000	01000	01100
10010	01100	01001	10110
01001	10110	10100	11011
10100	11011	11010	11101
11010	11101	01101	01110
01101	01110	00110	10111
00110	10111	10011	01011
10011	01011	11001	10101
11001	10101	11100	01010
11100	01010	11110	00101
11110	00101	11111	00010
11111	00100	01111	00001
01111	10010	00111	10000

Clearly, an LFSR is defined by its connection vector and this paper will discuss how to derive the connection vector, but first, the relationship amongst LFSRs, primitive polynomials and Galois fields will be discussed. Generating primitive polynomials and therefore, LFSRs can be computationally expensive in terms of storage and processing time, the use of distributed computing systems based on the message passage on commodity computing systems will be described. We will present results of some of the generated LFSRs.

LFSR, PRIMITIVE POLYNOMIAL AND GALOIS FIELD

As discussed in the preceding section, an LFSR is of length n and it is defined by the connection vector c. The bit sequence generated by the LFSR and vector of all states are:

$$s = s_0, s_1, s_2, \dots, s_{n-1}, s_n, s_{n+1}, \dots$$

and

$$s^* = \{s^k | k = 0,1,2, \dots\}.$$

An LFSR for which there exists a positive integer $N > 0$ such that $s^N = s^0$is periodic with period N.For all LFSRs, $N \leq 2^n - 1 = N^*$. An LFSR with period N^* is called a maximum-length LFSR or maximum period LFSR. Maximum-length LFSRs are of interest in cryptography, since they generate pseudo-random bit sequences that are large enough for them to be computationally secure using exhaustive search methods. A maximum-length LFSR has better linear complexity and satisfies the Golomb randomness postulates – the number of 1 and 0 bits is at least 1; the bit patterns of all 1's and all 0's have unique properties; and the autocorrelation function has two values only.

The associated or characteristic polynomial of any maximum-length LFSR is a primitive polynomial in Galois field modulus 2: $GF(2^n)$.The characteristic polynomial is represented as

$$p(x) = x^n + \sum_{i=0}^{n-1} c_i x^i \ (mod\ 2)$$

A primitive characteristic polynomial has an associated reciprocal primitive polynomial that is given by

$$p^*(x) = x^n p\left(\frac{1}{x}\right).$$

For n-degree LFSRs, the number of primitive polynomials in $GF(2^n)$ is $\varphi(2^n - 1)/n$, where $\phi(n)$ is the Euler phi function that represents positive integer that is $\leq n$ and relatively prime to n [Weisstein(a), undated; Weisstein(b), undated]. For 10-degree LFSRs, the number of primitive polynomials is 60, for 20-degree LFSRs, the number is 24000, and for 25-degree LFSRs, the number is 1296000. Clearly, the number of maximum-length LFSRs of degree n grows super exponentially and generating primitive polynomials using LFSR will require multiprocessor computing systems. Due to this computational complexity, the literature is replete with techniques for generating large trinomial primitive polynomials (Brent, Larvala and Zimmermann, 2003; Zivkovic, 1994). These are primitive polynomials of three terms only (Brent and Zimmermann, 2016):

$$p(x) = x^n + x^d + 1.$$

CONSTRUCTING PRIMITIVE POLYNOMIALS USING LFSRs

To determine whether a polynomial is primitive, the exhaustive search technique is used to test the associated LFSR for maximum-period. That is, starting with a non-zero initial state vector, generate subsequent state vectors to obtain the period N and if this is $2^n - 1$, we conclude that the associated characteristic polynomial is primitive, otherwise it is discarded. There are two known techniques for simulation LFSR, namely, the Fibonacci LFSRand the Galois LFSR. The Fibonacci LFSR uses taps and state to determine the input, while the Galois uses the output bit as the new input bit and XORing the clocked flip flop with output bit to form the input bit of the next flip flop. For simplicity, the Fibonacci LFSR will be used in this paper.

Given initial vector S0, connection vector c and degree n.
Set count:=0, complete:=false, S:=S0
while not complete:
    update LFSR S:= (S,c)
    count:= count +1
    if S is equal S0: complete:=true
if count is equal $2^n - 1$*: LFSR is maximum-length*

From Table 1 above, he LFSR connection vector is $(c_0 c_1 c_2 c_3 c_4) = 10100$ and the initial state vector is $(s_0 s_1 s_2 s_3 s_4) = 10000$. The 5-degree LFSR repeats after every 31 iterations as highlighted in Table 1 and it is therefore, maximum-length and the associated primitive polynomial is

$$p(x) = x^5 + x^2 + 1.$$

Now, for LFSRs of degree n the number of possible connection vectors is 2^n. It is easily determined that many of these connection vectors do not give primitive polynomials. Clearly, for all maximum-length LFSRs, $c_0 = 1, c = \{1, c_1, c_2, \dots, c_{n-1}\}$. It has been shown that connection vectors that will give maximum-length LFSR must have even number of 1's in its bitstream, i.e., odd number of terms in the primitive polynomial (Benvenuto, 2012).These two conditions of consistency and parity of terms eliminate ¾ of the possible connection vectors as candidates for primitive polynomials. The symmetry of polynomial reciprocity reduces the domain of possible search candidates to 2^{n-3}. This is still a large number for large$n$, e.g., this number will be over 2 quintillions with $n = 64$.

SINGLE PROGRAM MULTIPLE DATA (SPMD) MESSAGE PASSING PROGRAM

As indicated in the previous section, determining primitive polynomials is a very computationally expensive task. If all LFSR states are stored, the simulation for maximum-period will require $k(2^n - 1)2^n = O(2^{2n})$ storage, where k is the memory storage size of a bit. However, only two states are needed to test for maximum period.

To determine all primitive polynomials of degree n, a total of 2^{n-1} connector vectors will be tested without accounting for polynomial reciprocity and parity of terms. On a single core processor (laptop), the compute time for $n = 16$ was 204 seconds, which is comparable to the performance reported by Saxena and McCluskey (2004). However, on 2-core processor, the compute time was 130 seconds yielding 78.5% efficiency and on 4-core processor, the compute was 87 seconds with 58.6% efficiency.This encouraging performance was the motivation to consider an in-depthapplication ofcommodity-based multicore processor for the determining primitive polynomials. The single program multiple data (SPMD) model isutilized. In SPMD model, a single program runs on all processors, however, each processor is responsible for a different data, which in this case are the polynomials. The message passing library used is the C++ version of MPICH on Visual C++ of Microsoft Visual Studio. All simulations were done on a commodity computing system with the following configuration: Dell Ultrabook/Tablet with 2 physical Intel cores (corresponding to 4 logical processors), 8GB RAM, 1.80GHz CPU, and 256GB hard drive. The SPMD program is

Set n:= degree of LFSR
Set m:= maximum primitive polynomial candidates (2^n)
Set max_p:= maximum period ($2^n - 1$)
Set np:= number of processes
Set first:= first polynomial candidate after starting with $c_0 = 1$ ($\frac{m}{2} + 1$)
Set step:= number of polynomial candidates per process $\frac{m}{2+np}$
Set start:= start polynomial candidate for process with rank k ($first + k * step$)
Set end:= start + step
For p:=start to end-1
    Generate the connector c
    If oddTerm(c) and checkMaxPeriod(c) then
        c is a primitive polynomial
    End If
End For

The result shows that the partitioning above allocates equal number of polynomials to a process and produces a very good load balance. The number of primitive polynomials determined was almost the same for each processor with very minimal differences. The CPU times for LFSRs of degrees 10 to 20 are shown in Table 2 below.

Table 2: Primitive Polynomials Per Processor and Efficiency (calculated as speed divided by the number of processors)						
LFSR Degrees	**Processor 0**	**Processor 1**	**Processor 2**	**Processor 3**	**CPU Time (Secs)**	**Efficiency(%)**
10	13	15	16	16	0.003	65
11	43	45	46	42	0.018	63
12	38	38	34	34	0.043	61
13	155	160	155	160	0.323	59
14	184	190	191	191	0.987	61
15	454	446	452	448	4.122	61
16	522	513	509	504	17.753	60
17	1935	1920	1935	1920	91.786	
18	1944	1946	1955	1931	270.934	
19	6912	6885	6885	6912	1568.526	
20	6005	5995	6047	5953	4497.279	

For LFSR of degree 19, each processor will executecheckMaxPeriod function 2^{18} times and each time, the floating operations count is $O(4n * 2^n)$, where each operation is comparison, logical XOR, logical AND or an assignment. Clearly,exhaustive method is exorbitantly slow. The future research will examine other fast techniques for determining primitive polynomials.

REFERENCES

Benvenuto, C.J. 2012.Galois Field in Cryptography.http://sites.math.washington.edu/~morrow/336 12/papers/juan.pdf.

Weisstein, E. W. (a)"Totient Function." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/TotientFunction.html

Weisstein, E. W. (b) "Primitive Polynomial." From MathWorld--A Wolfram Web Resource. http://mathworld.wolfram.com/PrimitivePolynomial.html

Okunbor, D., Sarami, C., Bhattacharya, S., and Sharma, R. (2018). Python Implementation of A5-Based GSM Cryptosystems. *Journal of Global Information Technology*, in press.

Okunbor, D., Amado, C. E., and Sharma, R. (2017). A5-Based GSM Cryptosystem Implementation and Analysis. Review of Business and Technology Research, 14(2), 46-52. ISSN 1941-9414

Saxena, N and McCluskey, E.J. (2004). Primitive Polynomial Generation Algorithms Implementation and Performance Analysis. Center for Reliable Computing, *Stanford University.Technical Report*. CRC TR 04-03.

Brent, P.B., Larvala, S., Zimmermann, P. (2003). A Fast Algorithm for Testing Reducibility of Triomials mod 2 and Some New Primitive Trinomials of Degree 3021377. *Mathematics and Computation.* 72(243):1443-1452.

Zivkovi, M. 1994. A Table of Primitive Binary Polynomials. *Mathematics and Computation.* 63(205):285-286.

Brent, P.B., Zimmermann, P. (2016). Twelve new primitive binary trinomials. https://hal.inria.fr/hal-01378493/document.

CURRENT TRENDS OF BIG DATA IN BIOINFORMATICS AND COMPUTATIONAL BIOLOGY

S.Jha, MCA Dept,V.B.S. Purvanchal University, Jaunpur (UP), India (sonamjha1980@gmail.com)
S.Kumar, MCA Dept,V.B.S. Purvanchal University, Jaunpur (UP) (India, surjeet_k_yadav@yahoo.co.in)
A.Gupta , Dept Of Pharmacy,Kunwar Haribansh Singh College of Pharmacy, Jaunpur (UP) India, (81.akgupta@gmail.com)

ABSTRACT

Big Data explains a huge amount of data in bioinformatics and computational biology. It performs a new model that converts the studies to a large scale research. The act and appearance of big data in bioinformatics is to produce storage place of data, better computing facilities and data manipulation equipments to examine or determine data. Parallel computing is one of the central frameworks that control big data tasks. It permits executing algorithms concurrently on a cluster of machines or supercomputers. The present work is aimed to represent key concept in the analysis of big data to extract data from increasing data source for bioinformatics by explaining all tools, resources and architectures.

Keywords: Big Data, Bioinformatics, Machine Learning, Genomics, DNA, Proteomics, Big Data Analytics, Parallel Computing, Grid Computing, Cloud Computing, Hadoop .

INTRODUCTION

Big Data means large amount of data. It is powerful data sets that may be analyzed computation to display figure, mode or style. It is not only used in specific area but also every field. Bioinformatics is one of those fields. Big Data is resources of information. Similarly, it also provides golden opportunities for quick and fast growth of related data. We can see the annual growth of data generation may reach 44 trillion zettabyte by the year 2020. Big Data produces huge amount of data related to exhaustive areas such as atmospheric science, genomics, research, astronomical studies and network traffic monitor .Data is created in wide way day to day by the inter-connection of trillions of people using computer, GPRS and sensors etc. Big Data is a mode of an expression for large or complex data sets that are conventional data processing; application software is insufficient to deal with them. The task of big data is an extract data, store data, data analysis, search sharing, reproduce and transfer, catching visualization, querying, updating and information privacy**.**

BIG DATA ANALYSIS IN BIOINFORMATICS

Big Data Analytics is the process of exploring extensive or huge and diverse data sets to discover secret patterns, unspecified correlations, market trends, customer desire and other functional information(Huang, W.2012).Big data always includes large data but they also frequently added online data and data varieties.Machine learning and bioinformatics proceed towards have been established for research arithmetical connection from big data in medicine and behavioural science that usually include clinical, genomic and proteomics (Kashyap, H. 2015).The task of big data in bioinformatics is to produce computing data, storage facilities and data manipulation equipment to evaluate data (Bellazzi, R. 2014).

APPLICATIONS OF BIG DATA USED IN BIOINFORMATICS

Biology: Now Biologists never used common laboratories to search out the results so they depend on large and continual growing of technologies for captured sequences therefore making it economical and more effective. Data like automated genome sequencers, DNA computing and many more growing up to new era of big data in bioinformatics. Saying about Bioinformatics in big data means it has become a dynamic site of research area which is cable of collecting the benefits from big data. It produces to handle large amount of information into a fundamental biological mechanisms and how the results can be applied in. Big data is one of the largest features of biological studies. Researchers are capable of producing petabyte of data within hours and it have a great impact on the bioinformatics such as sequencing data, genomic sequence, protein sequence, DNA computing etc (Hussain, S. 2014).

Personalized Medicine: At present day, researchers are trying to hard work to meet the present demands in the field of personalized medicine by studying several ways to collaborate and coordinate medicine with new technologies. Big data in personalized medicine is one the biggest game changer for pharmaceutical companies (Al-Shorbaji, N. 2016).Manufacturers are assuming to develop high value, cost effective and targeted drugs .It has long been maintained that personalized medicine offers the chance of better health with less harm. Making assumptions using a single mind is no longer enough. So there is an urgent emerging of combining both the IT and biology together to find an innovative answer .The quantity of data is available today is an unbelievable. These changes have lead an effective mining data on genomics proteomics and metabolomics alongside clinical trial data and real world clinical data allowed researchers to more completely understand the particular structure of a disease shown in figure 1.

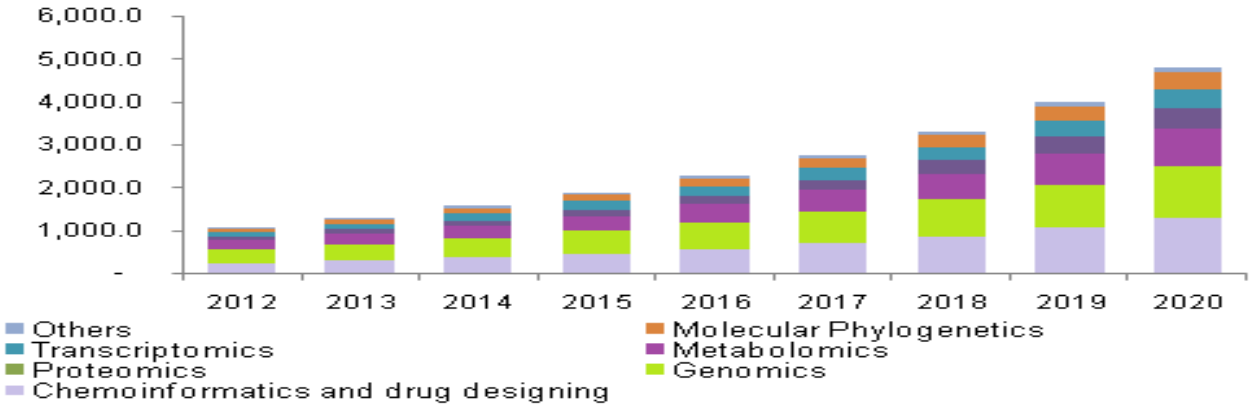

Figure 1 : Industry outlook report, 2012:2020

Gene Sequences: Researchers are getting attracted by new outlooks on the human genome and the processed of developing made in big data analytics. For many years genes have been studied, mapped and experimented in gene sequencing. The highest achievement was fulfilment of the Human Genomes project in the early 2000s.According to biologist by 2025,100 million to 2 billion human genomes could have been sequenced where concurrently data storage demands also increases from 2-40 Petabyte, because the number of data that must be stored for a single genome are 30 times larger than the size of the genome itself. Recently some revolution has taken place to handle quantity of data and the speed of analysis. Now scientists will be able to look more closely at human genes and much of this progress comes as they apply big data analytics to these issues.

Preventative Medicine: Recently to make a great deal in healthcare and medicine with the help of digital revolution. The primary aim is to provide better foundation to produce or develop protective and predictive field of medical science .Now a day's many advance computers support systems which helps us to improve protect, serve, recommend and maintain health to prevent disease, disability and death. Using big data in the field of preventative medicine, we can improve the health and give a better diagnose while treating the disease. As the role of big data comes more and more information from all around the world can be balanced. As the prototypes are being made with the help of large collection of data using big data techniques, it is easy to measure the outcome (Martin-Sanchez, F.J.).

Healthcare: Due to the huge collection of information, the healthcare industries has entered into the new era to generate large quantity of data by keeping records and maintain requirement for patient's health and care. Today's healthcare has been debated topic in developing healthcare information system. Using big data it is easy to find out fraud claims. It helps in transformation of medical claims payment system. Using this technique it is easy to treat a wide range of conditions such as diabetes, chronical disorders, heart diseases and various cancer conditions. Big data analytical technique also helps in cost reducing treatment for better health of patient. Even electronic records and tools are available to identify the condition of patients which helps to give the best care available (Kumari, D. 2014).

Industry: Bioinformatics involves the application of data, rich computational and informatics methods to support the scientific study of complex biological problems. Recently, Technological and intellectual advances in bioinformatics are transforming the way of biological research and currently explosive growth in academic, industry and government sectors. The prominent field of the bioinformatics visualization indicates the design of visual image and the execution of an effective software tools that provide an accurate and deep understanding into complex biological data (Kumari, D. 2014; Wilhelm, M. 2014).

TOOLS AND TECHNIQUES OF BID DATA

There are a several techniques used in Big Data analytics such as distributed storage, tiered storage and parallel processing. Some of the many big data analytics tools available include Hadoop, NoSQL and Google Analytics. Some of open source data tools are:

1. Knime helps to discover the potential hidden in our data, mine for fresh insights or predict new futures. KNIME analytics platforms are the perfect toolbox for any data scientists.
2. Open Refine, also called Google Refine, it is a powerful tool for working with disorganised data. Open Refine are mostly used for scrubbing than converting data from one format into another and added with web services and external data.
3. R-Programming is a free software programming language and software environment for mathematical calculating and graphics. Data miners uses R-Programming for developing statistical software and data analysis (Ernst, J. 2012).
4. Orange is open source data tools for analyse and visualize data.
5. Rapid Miner operates by visual programming for manipulating, analysing and modelling data. It is an open source platform that provides productive data science teams, machine learning and model employment (Karlic, R. 2010).

BIG CHALLENGE OF BIG DATA FOR BIOINFORMARICS

The Tools that are being used currently is not very efficient in dealing the complexity of every biological aspect. Most of the tools are powerful but they still possess some flaws that are not encouraged by biologists. With the increase in the volume of data, there is a steep rise in the need for higher computational speed, better software packages and pipelines to process, retrieve or submit the genomic and proteomics data. One of the major challenge that bioinformatics face today is that biological data is generated at a faster pace that the Moore's law of computational power does not meet the requirements of molecular biology (Bakken, S. 2016; Marx, V. 2013). The block diagram of big data extracts in bioinformatics specifications is shown in figure 2.

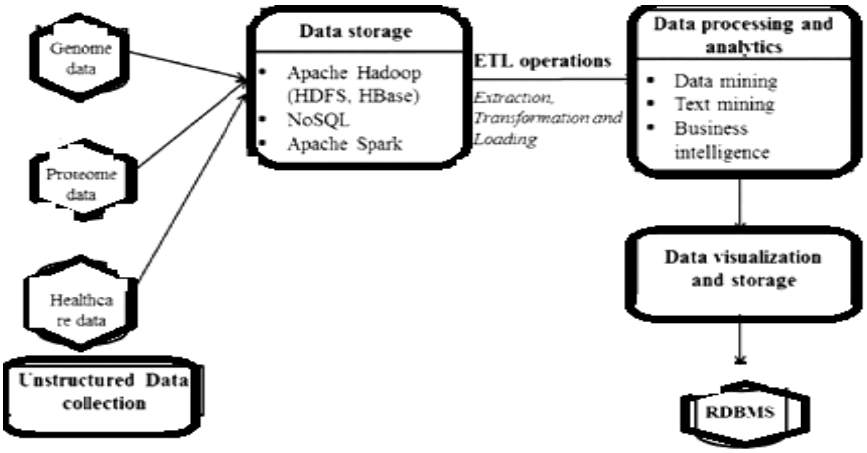

Figure 2 : Big data extracts in bioinformatics

PROBLEMS ASSOCIATED WITH BIG DATA IN BIOINFORMATICS

Storage: When the data is collected from multiple sources and this data being heterogeneous, it is very important to filter and select the useful data to avoid redundancy. When several researchers from all over world might sequence the genome of a particular organism and submit their data to a database, if the difference in the data is small or negligible, where if the difference do not hold any biological significance then the data can cause redundancy if stored. Database design is an art today and is carefully executed in the enterprise by highly paid professionals.However there are quite number of hypothetical theories and successful experiments conducted by data scientists to overcome the storage problem of big data in bioinformatics (Karlic R 2010).

Data Transfer: A single human DNA comprises around 3 billion base pairs (bp) representing approximately 100 GB of data. The transfer of such huge data from a database to local server or a personal computer is time consuming and inefficient process. When a study is conducted which involves the comparison of quite large number of organisms and the transfer and analysis of each set of data is difficult and time consuming process. For overcome these problems researchers have need advanced and updated tools to analyse and handle large volume of data because being to handle very huge data is more time consuming and not very economical method to every biologists. For overcome these problems researchers have need advanced and updated tools to analyse and handle large volume of data because being to handle very huge data is more time consuming and not very economical method to every biologists (Sacchi, L. 2015).

Security: With the very large amount of data flowing in and out of the databases and converted into large information. It is an extremely challenging task to maintain the confidentiality and security of entire system. Without a proper encryption of data being transferred is a great risk of loss or damage data due to several reasons. In a cloud computing system the privacy and security of data depends on the quality of service providers depending on their terms and conditions. Bioinformatics especially in the field of personalised medicine is very sensitive and must be achieved in secure and safe database.

Computational Power: Computational power is a need that grows parallel to the data volume in Big Data. Parallel computing still requires new paradigms in order to harness the additional processing power for bioinformatics. There are various sophisticated algorithms developed to manage the storage and processing of biological big data, it still does not meet the requirement which the big data in omics place in front of the computer science. As long as sequencing takes place and more comparative and analytical research and carried out in the field of biology, a need for the new and updated tools to compute the statistical data in biology is never ending (Kumari, D. 2014).

Solutions for Problems Associated With Big Data In Bioinformatics: To solve problems of big data, it is necessary to look at the possibilities in a broader way by a better understanding of biology and computer science.

Cloud Computing: Since we are facing the problems of data transfer and storage in personal storage system, Cloud computing is gaining a wider acceptance these days. Cloud computing have a number of advantages due to its scalability, extensibility and provision capability. Cloud computing works in a way where individual can transfer large volume of data from databases. Cloud computing provides a great solution to the problems of data transfer because of its higher data transfer speed and computational power in comparison to peer to peer data transfer or server to client (Dillon, T. 2011).

Storage of data in DNA: Digital data storage in DNA is progressing at a rate of 10 times every year in comparison to Moore's law on computational power and electronics. 1 gram of dry DNA has a storage potential of 455 Exabyte of information, but at the current situation the cost of storage and retrieval of one megabyte of data in DNA is 12,500 USD and 220 USD respectively. But the high durability of DNA molecule for several million years in addition to their confidentiality and security in archival storage of data. With the current rate of development in the field of bioinformatics, the possibility of storage of digital data in biological molecules such as DNA is inevitable with respect to time (Martin-Sanchez, F.J. 2013).

Parallel Computing: A parallel computer uses a set of processors that are able to cooperate in solving computational problems. Parallel computing is an example for the use of high computational power in data

analysis. Here, a large problem is broken down into small tasks and distributed to multiple processors, and then each bit of solved data is collected and submitted as a single solution. Two or more microprocessors can be used simultaneously, in parallel processing to divide and conquer tasks that would overwhelm a single, sequential processor. Database searching is the most heavily used bioinformatics application (Ware, A. 2017).

Hadoop Open Source Framework: Hadoop, a software framework, is said to be the most efficient tool in parallel computing for handling biological data. There are several tools already existing in Hadoop that deals with bioinformatics problems. Parallel computing using software frameworks like Hadoop can get the result out with less cost in comparison with the use of supercomputers. Hadoop was developed by Google's Map Reduce that is a software framework where an application breaks down into various parts (Schumacher, A. 2014).The Hadoop distributed file system architecture of hadoop allows the data storage without loss or damage of data.

Grid Computing: Grid computing holds an advantage of being cost effective method of analysing large problems. It has potential capabilities to apply supercomputing power to address a vast range of Bioinformatics problems. Grid computing is a technology that enables the division of large problems.

CONCLUSION

Bioinformatics is a branch of biology. It is a new field under which acquisition of bioinformatics, storage, processing, analysis, distribution, interpretation etc. In this work, the techniques of biology, information technology and mathematics are used. It is also combination of computer science and Biology. Big data is one of the general attribute of biological studies, and today, researchers are capable of generating terabytes of data in hours. Over the last decade, biological datasets have been grown massively in size, mostly because of advances in technologies for collection and recording of data. Therefore, big data posses a great impact on the bioinformatics field and a researcher in this field faces many difficulties in using biological big data. Thus, it is essential that bioinformatics develop tools and techniques for big data analysis so as to keep pace with our ability to extract valuable information from the data easily thereby enhancing further advancement in the decision-making process related to diverse biological process, diseases and disorders.

REFERENCES

Al-Shorbaji, N. (2016), Discussion of the new role of biomedical informatics in the age of digital medicine, Methods Inf Med, 55, 403-421.

Bakken, S. (2016), The promise and potential perils of big data for advancing symptom management research in populations at risk for health disparities, Ann Rev Nurs Res, 34, 247-260.

Bellazzi, R. (2014), Big Data and Biomedical Informatics: A challenging opportunity, Yearb Med Inform, 9, 8-13.

Dillon, T. (2011), Cloud Computing:Issue and Challenges, *Int Con Adv Info Net*, 12, 27-33.

Ernst, J. (2012), Automating Chromatin state discovery and characterization, Nat Methods, 9, 215-216.

Huang, W. (2012), ART: a next generation sequencing read simulator, *Bioinformatics,* 28 (4), 593-594.

Hussain, S. (2014) Behaviour life style analysis for mobile sensory data in cloud computing through MapReduce Sensors, Basel, 14(11), 22001-22020.

Karlic, R. (2010) Histone modification levels are predictive for gene expression, Proc Natl A cadSci, USA, 107, 2926-2931.

Kashyap, H. (2015), Big Data Analytics in bioinformatics:A machine learning perspective. *Computing Research Repository*, 15, 1506-05101.

Kumari, D. (2014), Impact of Biological Big data in Bioinformatics, *International Journal of Computer applications*, 101, 0975-8887.

Martin-Sanchez F.J. (2016), The new role of biomedical informatics in age of digital medicine methods, Inf Med, 55, 392-402.

Marx, V. (2013), Biology: The big challenges of big data, Nature, 498(7453), 255-60.

Sacchi, L. (2015), Analyzing Complex patients,temporal histories : new frontiers in temporal data mining, Methods Mol Biol, 1246, 89-105.

Schumacher, A. (2014), Simple and scalable scripting for large sequencing data sets in Hadoop, Bioinformatics, 30(1) , 119-120.

Ware, A. (2017), HADOOP: Solution for big data challenges in bioinformatics and its prospective in India, *IOSR Journal of computer engineering*, 51, 2278-0661.

Wilhelm, M. (2014), Mass spectrometry based draft of the human proteome, Nature, 509(7502), 582-587.

VISUAL GESTURE RECOGNITION FOR DRAWING APPLICATIONS

Aaron Rababaah, American University of Kuwait, Kuwait (arababaah@auk,edu.kw)
Aya Kandil, American University of Kuwait, Kuwait (s00024803@auk,edu.kw)
Yossif Gharib, American University of Kuwait, Kuwait (s00028610@auk,edu.kw)
Ahmed Helwa, American University of Kuwait, Kuwait (s00039084@auk,edu.kw)

ABSTRACT

Painting and drawing software are rapidly produced by the software industry. Each of them has its own pros and cons. One of the main disadvantages of these drawing software's is the lack of accurate pointing devices. As gesture recognition technology has improved it has been applied to various consumers appliances to provide natural interfaces. Drawing annotations with 3D hand gestures in augmented reality is useful for creating visual and spatial references in the real world, especially when these gestures can be issued from a distance. Here we introduce the VIDS (Virtual Intelligent Drawing System), which is a gesture-based painting software for professional and nonprofessional use. It uses the hand movements for drawing. User does not need an external hardware-pointing device to draw. Instead, the different gestures of the fingertips are used to activate the drawing functionalities such as pen-up, pen-down, copy, paste and scaling drawn shape etc. We designed this system as a primary level software product, which aims to provide entertainment to the user. The software can be further upgraded by adding extra drawing features. VIDS was tested by number of subjects and found useful and effective and promising to enhance users' experiences and has potentials to improve children educational environment.

Keywords:Machine Vision, Gesture Recognition, Virtual Drawing, Image Processing.

INTRODUCTION

Since the beginning of time and man has tried to express himself through different means for many purposes. Usually the means are used to either help the person create an idea or make it clearer to the audience. In the past, several basic instruments were used to express such ideas. Some examples would be using material to create paint in order to draw basic shapes on surfaces that were usually flat, using papyrus and ink, etc. Things starting evolving until the electronic age was reached. Other physical instruments were provided to people so they can communicate, such as basic input devices like mouse and keyboard, or even complex drawing pads that can be drawn upon using special input devices. What all those methods have in common are the need for physical tools that will have to touch the surfaces on which the drawing will appear.

This leads to a very important point, which is that as time passes technologies evolve and as humans evolve, new means of expressing oneself is created that are much more suitable to the age in which one exists. When it comes to expressing oneself, one has focused on the drawing side of Art since Art has many fields and methods of expression. In order to add something relatively new to this form of art, one has decided to create a software designed for drawing and will not have to use any physical tools that have to touch the device on which the software is stored in order to create a drawing. As a result, the input method will be remote based on machine vision algorithms.

RELATED LITERATURE

Gestures are widely defined due to its wide variety of applications and uses in particular domains. Gestures are considered the most primary and expensive form of human communication and the reason behind is the importance of gestures for computer interaction with human. (Ghotkar et al., 2012) have defined the gestural taxonomy that is the motion of body intended for communication. Fingers, hands, arms, head, face or body movement produces these gestures. Since hand gestures will be used for computer interaction then to have a successful communication, both the sender and the receiver must have the same set of information for

particular gesture. Successful communication generates meaningful information to interact with the environment. Gestures provides more flexibility for interacting with environment compared to graphical user interface like mouse and keyboards. Gestures are categorized by the involvement of the human body in to: hand and arm gestures, head and face gestures, body gestures. Among all the gestures stated previously hand gestures are the most expressive and used more frequently. The use of hand gesture for computer interaction communication generated the motivation for a research for hand gesture recognition. Understanding the static and dynamic gestures over period is the process of hand gesture recognition.

Through the decades, algorithms developed and many algorithms have been proposed for hand gesture recognition. Different attempts used to solve the hand gesture recognition problem using gloves-based devices, vision-based techniques and specific types of cameras and sensors.

Many projects, prototypes and algorithms have been developed to translate the gestures-based interactions between human and machine. Specifically, in the area of sign language interpretation the user's used to have hard time communicating with the machines around them, but by the rapidly increasing advanced electronic devices these devices are built to recognize the user's gesture and respond to their commands. By this era, there are a lot of increasing interest in this field due to the efficient and effective solutions developed. In this section, and the coming sections several attempts in previous work and their experiences in their studies are presented. We are highly interested by information revolution, which includes the automation in many aspects of our life. Productivity, accuracy, efficiency, quality, consistency and safety are main factors when considering automaton devices for human environment interaction.

Authors in (Rababaah, 2013) stated that there are two main ways to produce sign language recognition system by, data gloves based and vision based. The data glove approach that includes censored-gloves that detects the position of the finger establishing a data vector that uniquely represents a gesture. The vision-based approach uses image processing to segment and extract the hand gesture from the environment to precisely represent a gesture pattern. In (Rababaah et al., 2013) it is stated that the approach provided a multi-colour based glove used to establish uniquely patterns of the different hand signs. The approach used image processing to recognition, transform a visual hand gesture into spoken letters, and displayed text. The testing included RGB-filter in the segmentation stage by collecting sample image frames of individual signs then sampling each fingertip multiple times. For each of the sample the mean and sigma tolerance were computed then for each sample the calibration accuracy was computed to test if it is acceptable in the range. The system generated accuracy greater than 93 percent.

Other attempt was produced by Jaya et al. (Shukla & Dwivedi, 2014) proposed a method for hand gesture recognition system using Microsoft Kinect sensor. Their system includes scans of and object in real time using three-dimensional scans. Kinect depth feature was used for background segmentation of hand gesture images captured with Kinect. Segmented hand images processed using image processing techniques to find contour of the segmented parts. Contour area and convexity defects as features for classification. Five hand gestures classes were considered using one, two, three, four, and five fingers one by one. In terms of implementation and testing this algorithm was applied for 15 images of each class. It gives a correct classification rate of 100%.

The work of (Chang et al., 2017) produced a "Gesture based augmented reality annotation" that presented different annotation drawing techniques for issuing arrow, circle, and free form annotations in mid-air or at real world surfaces at a distance. Shapes considered are all made up of strokes are drawn as free-form annotations. The proposed system designed three different drawing methods: surface drawing, air drawing, and mid-air drawing. Surface drawing is applied directly on detected surface. Both air-drawing and mid-air

drawing is applied directly at the user fingertip. Different types of gestures are included in the system, pinch and drag gestures are used to perform the drawings by both applying and releasing the gestures.

Natural interfaces are having a great interest of various consumers and gesture recognition technology has improved to provide interfaces that are more natural for users. Other project was proposed by (Jeong et al. 2012) that is "Camera dedicated control system using gesture drawing" it proposes a gesture-based method for easy and quick control interaction with the TV. The user starts by drawing different symbols in the air and these hand gestures are recognized by capturing the motion path when the user draws. This system is based on real time parallel processing structure that used a single-camera for hardware. Some dynamic gestures in early projects used stereo visions but these stereo vision does not provide image resolution sufficient for hand shape analysis due to the lack of the texture resulting from the image. The system contains a selection of gestures that are used in manipulation and control each gesture based on a function or command. It is therefore important to differentiate the intended behavior from the unintended behavior. This project was tested on 15 participants of both genders. Each participant is allowed to perform the gesture as efficient as possible while minimizing the errors. It is hard to standardize the proposed system due to the different environments the testing was carried out. One of the main disadvantages of this system is the low-cost because it allows more users to use it and increase the human machine interaction.

Hand-gesture implementation involves significant usability challenges, including fast response time, high recognition accuracy, quick to learn, and user satisfaction. Many factors affect the efficiency and effectiveness of human interaction activities. No single algorithm for hand-gesture recognition favors every application. The suitability of each approach depends on application, domain, and physical environment. In the previously stated paragraph by (Jeong et al. 2012) some limitations were related to physical fatigue for the action to be performed after the applied gesture. Since gesture shapes are more diverse and will have to be taught for the user in order to use the interface, some gestures might be uncomfortable. Considering the response time taken to take the action applied has to be quick to gain more user interest and loyalty to use the proposed system. Environment is a huge factor to be considered especially when considering a simple hardware for the proposed system as stated project. Other issue is multiple methods lends robustness to hand-tracking algorithms because once the hand tracking is lost due to collision it will take more time to re-preform both the gestures and actions.

THE PROCESS OF THE SYSTEM

In this section we are going to present in details the stages of the process of the proposed system. We also provide a process model in Figure 1.

Functionality 1: Translating hand gestures into commands

This functionality will require MATLAB to read hand gestures from the user through the camera that will be used and using different algorithms to analyse the gesture, compare it to its match in the dictionary and display or compute the command associated with the hand gesture on the screen. In our software, we will be using algorithms to detect the colour of the glove then we will be later associating the gestures for the commands.

Functionality 2: Detecting hand motion for drawing

This functionality will be the main aim and concept of our project. As soon as a certain gesture is r ead, the user will have to move his/her hand to draw whatever they want by hovering over certain positions on screen. The diagram drawn will then be processed by the software, converted into an image type file, and then shown on the screen for the user to see what is drawn.

Functionality 3: Starting and pausing software

This part of the software where user have total freedom to start and pause or end the drawing experience under any circumstances for any reasons that fit the user needs. Having this feature provides user with more flexible experience.

Functionality 4: Editing options to the shape drawn (cutting, copying and pasting)

Once the gesture is identified then the user will be able to edit the shape he/she is drawing by performing gestures. Editing features include, erasing the shape drawn, clear the shape drawn, copy the shape drawn, paste the shape drawn, and finally they will be able to cut the shape drawn. The user will be introduced to the editing gestures by the handbook that comes with the software. Previously drawn shapes once retrieved that user will be able to draw, copy/paste whole shape and erase the shape. Copy/paste the shape is one of the challenging functions to be applied in the system, so the developers team decided to perform the copy/paste for the specific shape based on user hand position on area targeted to apply command and this will help user drawings to be more customized.

Functionality 5: Choosing from colour scheme to colour shape drawn

Choosing the colour of a drawn shape is one of the available features from the system that will allow the user choose the colour needed from sequence of colours. These colours will be available on the screen where the user is drawing as in figure 1. Once the colour is chosen, it will be applied to the drawn shape and the user will have the freedom to choose another colour when needed to colour the drawn shape.

Functionality 6: Importing shapes to screen

This is one of the extra functionalities that the developer team were able to produce; it will help users to have more flexibility, more variety and more creativity to drawn shapes or designs. User perform pen down first and place his/her hand over desired area on screen canvas to import the circle and/or square shapes.

Functionality 7: Choosing a stroke size

This is one of the extra functionalities that the developer team were able to produce; it will help users to have more flexibility, more variety and more creativity to drawn shapes or designs. User can choose stroke size to be large or small before drawing or while drawing. While drawing that means user already is on pen down state if they clicked on stroke size with the other hand (one hand to draw, other to click) it will be instantly performed.

Functionality 8: Choosing stroke shape (brush stroke)

This is one of the extra functionalities that the developer team were able to produce; it will help users to have more flexibility, more variety and more creativity to drawn shapes or designs. User can choose stroke type dotted or flat and to be large or small in size before drawing or while drawing. While drawing that means user already is on pen down state if they clicked on stroke type with the other hand (one hand to draw, other to click) it will be instantly performed.

Functionality 9: Storing the shape drawn

This functionality will give the user the opportunity to save and retrieve the drawn shapes. Once the user finishes the drawing for the day he/she will have to click on saving button, then the file will be saved automatically. While drawing the user will be able to use wide range of editing features and once the user retrieve a previously drawn file then it will have same choices of editing features.

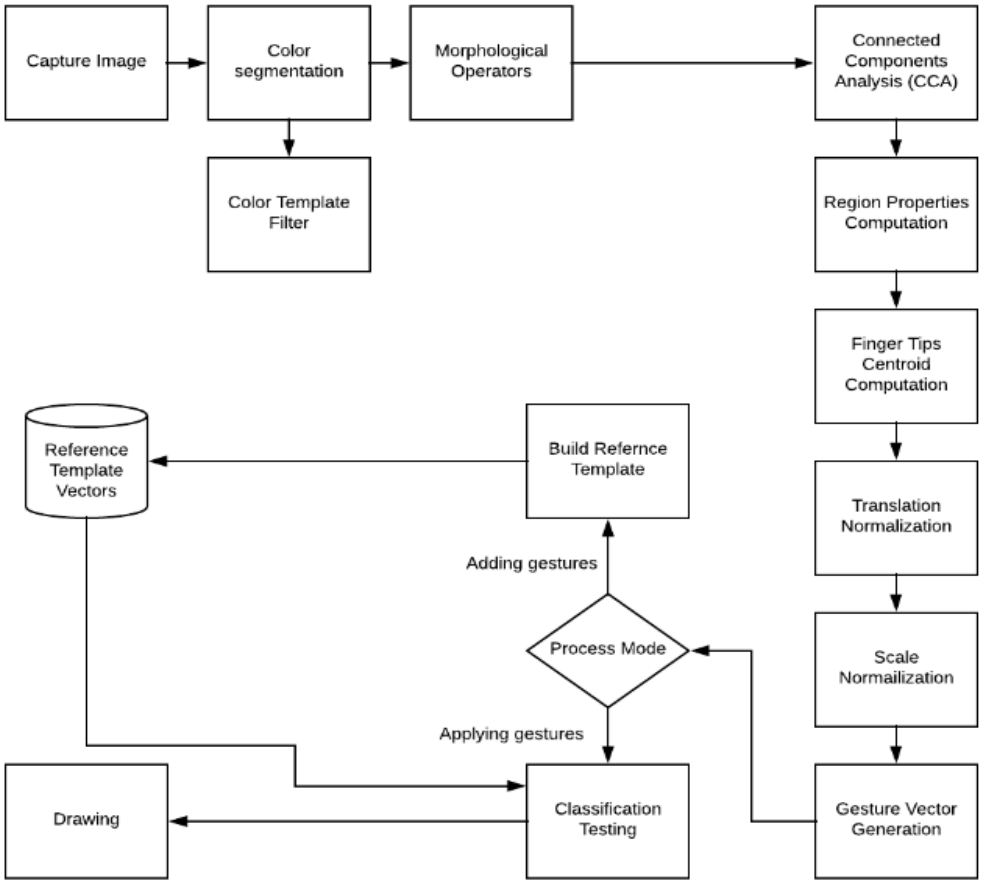

Figure 1: Process Model of the VIDS System.

In Figure 2, we present the Class Diagram of the VIDS system. As it can be seen, the software system consists of several classes which implements the functionalities described in the previous section. Further, in Figure 3, we present a Fishbone diagram for causes of software failures.

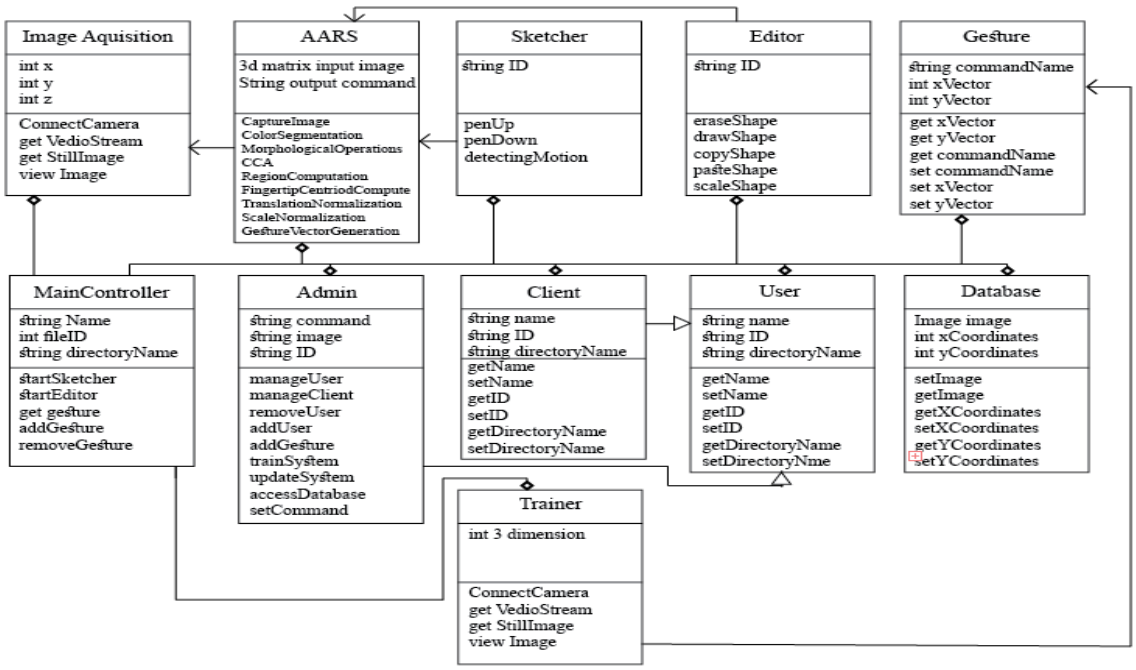

Figure 2: Class Diagram of the VIDS System

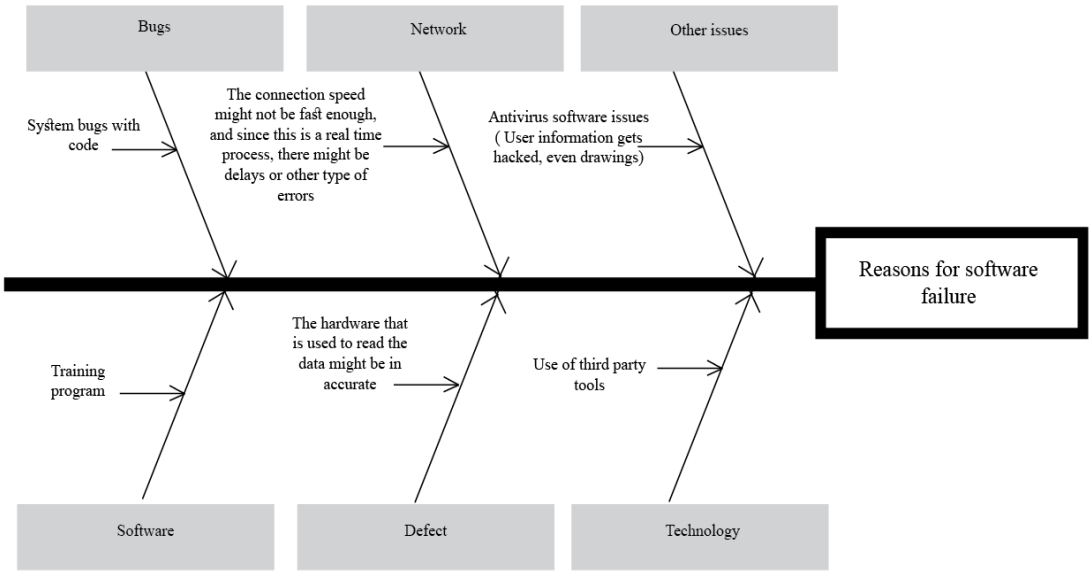

Figure 3: Fishbone Diagram of Software Failure Causes

RISK ANALYSIS

Risk analysis/assessment is an important part of deciding the methodology of the project as it defines all risks/dangers that are possible to occur throughout the lifetime of the project (application). As part of the risk analysis phase, a risk register defining all possible risks has been made to help in any decision-making process that solves future problems or issues. This risk register highlights important information regarding each listed risk (rank, probability, impact, etc.).

Table 1 presents a list of all possible risks and their corresponding evaluation: rank, category, response, probability and impact. Table 2 shows the probability matrix as the different risks are assessed and evaluated based on probability and impact to identify most critical risks.

Risks in Table 1 are all possible risks listed in a risk register – including the least-to-occur ones – to ensure a high-quality application delivery. Using Probability/Impact Matric to Calculate Risk Factors

Table 2 shows the probability and the impact of each risk so that the team members are aware of what could happen and how what its impact level on the project is. Risks are referred to as Risk 1, Risk 2…, Risk 5, as shown and identified in Table 1.

Table 1: Risk Analysis of the VIDS System

No. Rank	Risk	Description	Category	Root Cause	Potential Response	Probability	Impact
R1	Inaccuracy in translation	Providing inaccurate translation results	Performance Risk	Inaccurate gesture or Alogrithms	Make sure hand gestures are precise	Medium	High
R2	System Failure	System fails to launch	System Risks	System Crashes	Looking and correcting the bug	Low	High
R3	Failure in recognzing user's hand	System fails to recognize user's hand	Environmental Risk	Enivormental Requirements missing or glove missing	Provide correct background and glove color	Low	High
R4	Insecure system methods	User saved images leaked	Security Risks	Unreliable servers	Improve security methods	Low	High
R5	Recognzing more than one hand	System detects more than one hand	System Risks	Requirements to identify one hand is inaccurate	Provide correct code to recognize one hand	Low	Medium
R6	Slow response time	Reponse time is more than 4 seconds	Hardware Risks	Better hardware requirements	Provide quality hardware	Low	Medium

We conducted a survey to investigate the interest of our potential audience and the results are reported in Table 3. Some main observations of the survey include: high interest in drawing in general, high interest in virtual drawing, very high interest in learning gesture-based drawing. Finally, in Figure 4, we presented the components of the graphical user interface (GUI) of the developed system.

Table 2: Risk Evaluation based on Probability and Impact

		Impact		
bability	High			
	Medium			R1
	Low		R5 R6	R2 R3 R4
		Low	Medium	High

Table 3: Survey Results of Interest in VIDS System

Survey Questions	Results
Age	18-25:80% 25-45:20%
Gender	Female: 50% Male: 50%
Are you interested in Drawing? Any form?	Yes: 70% No: 30%
What is your profession?	Student: 70% Employee: 10% Both: 20%
Do you have any experience in drawing before?	Yes: 100% No:
Which drawing tool you used before?	Paint: 90% Applications on tablet: 5% Professional programs like Photoshop: 5% None: 0%
Are you interested to draw by free hand or by your own hand (like using kinect to draw)?	Hand: 30% Free Hand: 60% Other: 10%
What kind of features you would like to find to draw freely?	Copy/paste shape drawn: 20% Saving shape drawn: 30% Resize shape drawn: 10% Color shape drawn:30% Other: 10%
Are you willing to learn the gestures used to	Yes: 90%

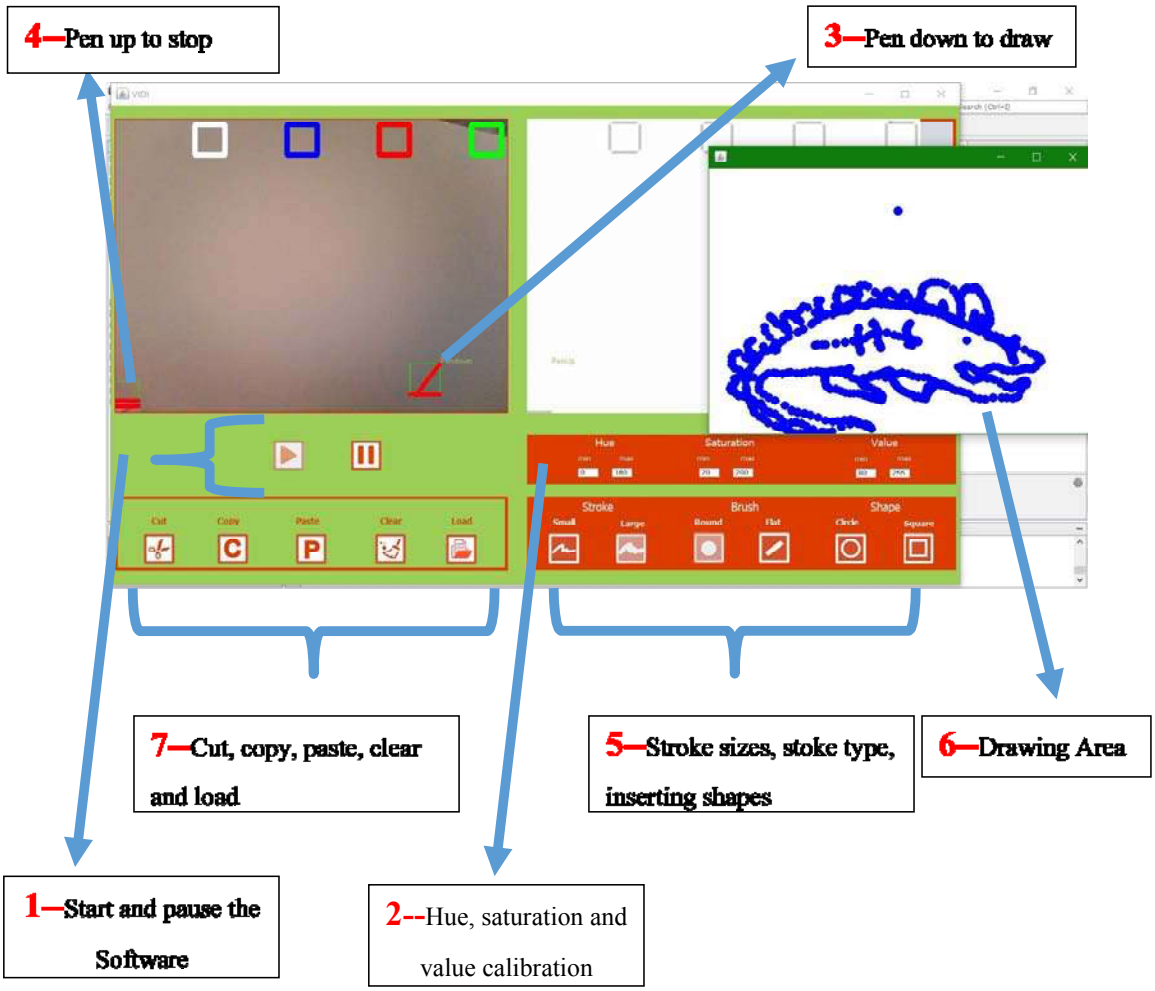

Figure 4: GUI of the developed VIDS system

CONCLUSION

We presented a proposed system for visual gesture recognition that can be used in computer drawing applications where none-traditional interaction methods are needed. We explored different algorithms from machine vision and image processing such as noise reduction, color space conversion, image segmentation, binary image operations, morphological operations, object detection and tracking, region of interest properties, etc. After significant testing, we identified an algorithm that works effectively. Keeping in mind the need for real-time response system, we considered OpenCV framework since it is based on a spectrum of optimized algorithms and widely used and tested. We faced some challenges including: detection algorithm, managing events, designing an effective and efficient GUI and real-time response. Our system was tested using real-world environment settings and showed reliable results. We hope that we can test our system with a wider audience specially with kids and evaluate their experience to see the potential of our system to enhance their experience in their learning environment.

REFERENCES

Rababaah Aaron (2013). Adaptive Decision Support Technique for Machine Vision-Based Security Systems to Detect Anomalous Events. *Precision Instrument and Mechanology* 2(2), 94-103.

Rababaah Aaron, Anggie Ramirez, Clifton White and Derik Robinson (2013). Design And Development Of Vision-Based American Sign Language Alphabets Auto Recognition System.*The Annual International Conference MTMI*, Virginia Beach, USA.

Ghotkar, A. S., R. Khatal, S. Khupase, S. Asati and M. Hadap (2012). Hand gesture recognition for Indian Sign Language. *International Conference on Computer Communication and Informatics*, 1, 1-4.

Shukla, J. and A. Dwivedi (2014). A Method for Hand Gesture Recognition. *Fourth International Conference on Communication Systems and Network Technologies*, Bhopal, 919-923.

Jeong, S., J. Jin, T. Song, K. Kwon and J. W. Jeon (2012), Single-camera dedicated television control system using gesture drawing. *IEEE Transactions on Consumer Electronics*, 58(4),1129-1137.

Chang, Y. S., B. Nuernberger, B. Luan, T. Höllerer and J. O'Donovan (2017). Gesture-based augmented reality annotation. *IEEE Virtual Reality (VR), Los Angeles, CA*, 469-470.

(A complete list of references is available upon request from Aaron Rababaah at: haroun01@gmail.com)

COMPUTER –AIDED DIAGNOSTIC SYSTEM FOR CHRONIC KIDNEY DISEASE USING HYBRID MACHINE LEARNING TECHNIQUE.

Rashmi Gupta, Dr.C.V. Raman University, Bilaspur, India (rashmigupta.1402@gmail.com)
Ragini Shukla, Dr.C.V. Raman University, Bilaspur, India (Raginishukla008@gmail.com)

ABSTRACT

Chronic kidney disease prediction is one of the most central problems in medical decision making. Therefore, there is a need to develop the automated tool for early prediction of this disease will be useful to cure. The main objective of this study is to develop the machine learning approaches to automate the detection. We were developed a computer-aided diagnosis system which uses classification and clustering K-means algorithm. By using K-means the incorrectly classified clusters instances (samples) are removed and finally models are constructed using various supervised classified algorithm. The 10-fold cross validation performance estimation of MLP, RBF, NB and J48, after K-means clustering with reducing 297 samples are 99.6, 98.98, 96.29, and 98.98 respectively. This result gives better result than other classification algorithm.

Keywords: Chronic Kidney disease, Radial Basis function, Naïve Bayes, J48, Multilayer Perceptron, K-means, clustering.

INTRODUCTION

Computer –Aided Diagnosis (CAD) holds the promise of improving the diagnosis accuracy and reducing cost. Machine learning, obvious provides a good result for diagnosis and prediction of disease but the main effort is for improving the accuracy and result. It plays a crucial role in the design of expert systems in medical diagnosis. In India, most of the people suffering from some sort of disease like cancer, diabetics, asthma and many more. We consider the disease chronic kidney disease for diagnosis.

Chronic Kidney disease is a term used for the heterogeneous disorders affecting the structure and function of the kidney. It is based on the presence of kidney damage (i.e. albuminuria) or decreased kidney for 3 month or more (Stevens, 2009). Kidney failure is traditionally regarded as the most serious outcome of chronic kidney disease and symptoms are usually caused by complications of reduced kidney function. Other outcomes include complication of reduced GFR (glomerular filtrations rate). Such as increased risk of cardiovascular disease, acute kidney injury, infection, cognitive impairment, and impaired physical function (James. et al. 2010). In this paper, we developed a new computer aided Hybrid diagnosis system for early prediction of kidney disease using classification and clustering algorithm and improving the accuracy and execution time.

Many researchers worked for improving the results and used different technology for diagnosis of different diseases. S.Ramya and Dr. N. Radha (2016) worked on diagnosis time and improvement of diagnosis accuracy using different classification algorithms of machine learning. The Proposed work deals with classification of different stages of CKD according to gravity. By analyzing different algorithms like BPN, NN, RBF and RF, the analysis results indicates that RBF algorithm gives better results than the other classifier and produces 85.3% accuracy.

Sahil Sharma, vinod Sharma (2016) has assessed 12 different classification algorithm on dataset which having 400 records and 24 attributes. They had compared their calculated results with actual results for calculating the accuracy of prediction results. They used assessment metrics like accuracy, sensitivity, precision and specificity. They find the decision tree technique gives accuracy up to 98.6 %, sensitivity of 0.97 and precision of 1 and specificity of 1.

Abhinandan Dubey et al. (2015) discussed the automated detection of disease using Machine Learning Techniques. The K-Means Clustering Algorithm with a single mean vector of centroid, to classify and make clusters of varying probability of likeliness of suspect being prone to CKD. The results are obtained from a Real Case Dataset from UCI Machine Learning Repository.

MATERIALS AND METHODS

Dataset:

The dataset for diagnosis of chronic kidney disease is taken from UCI repository datasets. There are 400 instances with 24 attributes, most of which are clinical in nature and rest is physiological. In the preprocessing of the data the missing values were dealt with by replacing numeric and discrete integer values by attribute mean of all the instances and nominal values were replaced using attribute mode.

In this study, we prose a hybrid method consisting of medical datasets followed by clustering for identification of incorrectly assigned cluster data point. Finally models are constructed with RBF, MLP, J48, NB. Figure 1 shows the block diagram of the proposed system.

Methods

Classification: Classification is referred to as supervised learning techniques because the classes are determined before investigative the data. Classification is mapping the data into predefined classes. In classification predict the objective class by diagnosis training dataset. This should be completed by searching appropriate borders for each objective class. It includes:

i. J48 Decision Tree:

J48 decision tree is an open source Java implementation of C4.5 decision tree algorithm in weak platform. It is the extension of the earlier ID3 algorithm, which is developed by Ross Quinlan. In J48 decision tree classification algorithm the branch of data distribution easily understandable and every leaf is pure gain evidence from the branch. J48 classification algorithm uses top-down greedy search methods for producing tree. J48 decision tree produces sorting tree whose; the leaf denotes the ending class and the internal attributes represent a possible number of outputs of the branch features.

ii. Naïve Bayes

Naïve Bayes is used for diagnosis and prediction of the problem. The Naïve Bayes algorithm requires a small amount of training data during classification to predict to evaluate the parameter. The Naïve Bayes classification method used to predict, an associate of each class. The class which has maximum probability is expected the most likelihood class. Below is Bayes theorem

$$P(Y/X) = P(X/Y) \cdot P(Y) / P(X)$$

P (X) is similar to whole classes.

P (Y) = relative frequency of class Y

iii. Radial Basis function:

Radial basis functions (RBF) networks are also feed forward, but have only one hidden layer. A RBF network has any number of inputs and has only one hidden layer with any number of units. It uses radial combination functions in the hidden layer, based on the squared Euclidean distance between the input vector and weight vector and uses exponential activation functions in the hidden layer. It has any number of outputs with any

activation function and has connection between the input layer and the hidden layer, and between the hidden layer and the output layer.

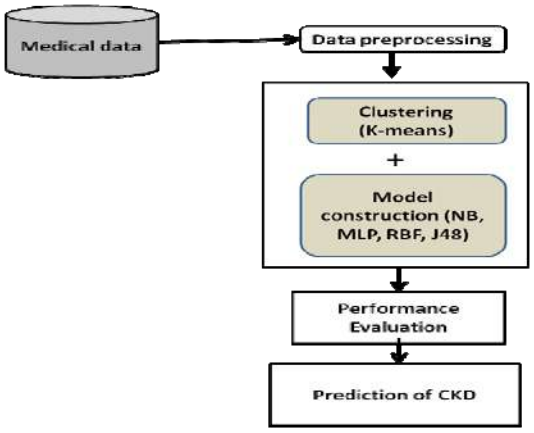

Figure. 1- Block diagram of proposed Hybrid Computer-aided Diagnosis System

iv. Multilayer Perceptron (MLP);

ANN is a mathematical model for information processing consisting of a number of highly interconnected elements organized into layers inspired by the human brain. It is trained with a part of the data to explore the association between inputs and outputs, and tested on the rest of the data. MLP is a popular ANN architecture used to model complex relationship between inputs and outputs.

Clustering: Clustering is similar to classification except that groups of data are not predefined and clustering is an unsupervised learning technique. In clustering, the evidence concerning the class label is not known. In clustering, similar data are put in the same place and dissimilar data placed in the other place. Values that fall outside the set or clusters may be considered as outliers. We use clustering for Simplifications, Pattern detection, Useful in data concept construction, unsupervised learning process.

i. K-Means Clustering

This is a centroid-based approach that takes the number of partitions k as input and creates k clusters of the input database consist of n objects or records by optimizing the rule of clustering. The resulting clusters with high similarity kept as intra-cluster and low similarity as inter-cluster. In this algorithm primarily a random set of k objects are chosen as the cluster centers to compute the mean value representing the cluster mean or centre. Each object can be involved in the cluster to exhibits high similarity. Updating of new cluster is made till there is no change found in the cluster structure.

IMPLEMENTATION AND RESULTS

For the implementation of hybrid computer aided diagnosis system, we have used MATLAB 2019b tool and Waikato Environment for Knowledge Analysis (WEKA) toolkit to analyze the performance gain that can be obtained by using various classifiers (Witten & Frank, 2005).

Hybrid Diagnostic system (HDS) for chronic Kidney disease in which the following iterative phases for the work have been identified which is shown in Figure1.

In the first phase, the medical datasets was obtained from the UCI repository of machine learning using chronic kidney disease dataset, in these data there were 400 data records and 24 instances.

In the second step of data preprocessing, missing data are replaced with suitable values of class label.

In the next step, K-means clustering algorithm is applied and incorrectly classified instances with wrong clusters are removed before the predictive model construction step. Table 1 shows the incorrectly classified cluster instances. Out of 400 instances 103 were incorrectly classified or indicated as wrong clusters. Therefore for better results and accuracy, we removed 103 samples from our datasets. Table 2 shows the confusion matrix for K-means cluster.

Table1. Error Percentage of incorrectly classified cluster instance

Cluster attribute (clusters)	Samples	Incorrectly classified	Error (%)
Cluster 0 <---- nckd Cluster 1 <-----ckd	400	103	25.75%

Table 2: Confusion matrix for K-means clusters

Assigned class		
Nckd	**Ckd**	**Actual class**
103	147	Ckd
150	0	nckd

In the next phase after removing incorrect assigned samples, the models were constructed with remaining data (297 samples). Supervised machine learning algorithm such as RBF, NB, MLP, and J48 are then used by using 10 fold cross validation method. Table 3 Shows that the performance estimation of selected classifier with 297 samples. The performance is evaluated by using the accuracy, sensitivity, specificity, precision, Kappa and RMSE is measured. Using these model we achieved prediction accuracy for MLP is 99.66%, RBF is 98.98%, NB is 96.295% and for J48 is 98.98%. The last phase is too compared with existing model and validated how the proposed model is better than existing model.Table3.

Table 3: The cross Validation 10-fold performance estimate of selected classier after K-means clustering with reducing 297 samples.

Proposed Method (K-means+ Classifier)	Acc	Se	Spec	Prec	Kappa	RMSE
MLP	99.66	0.99	1	0.99	0.99	0.052
RBF	98.98	0.99	1	0.9	0.97	0.09
NB	96.29	0.96	1	0.96	0.92	0.18
J48	98.98	0.9	1	0.9	0.97	0.09

Se:

sensitivity, spec: specificity, Acc: accuracy, RBF: Radial Basis function, LR: Logistic Regression
RMSE: Root mean square error, Prec: precision, MLP: Multilayer Perceptron

COMPARATIVE ANALYSIS

We compared accuracy achieved by the predicted model constructed with original (400 samples) in table 4 with the predicted model achieved with reducing datasets (297 samples) of hybridized model. When we used these classifiers with 297 samples, the accuracy of the classification model was improved by 1.91% in case of MLP, 0.98% in case of RBF, 1.79% in case of NB and 2.23% in case of J48. Results are comparable, promising and therefore the proposed hybrid model will be helpful in disease diagnosis

Table 4: The cross validation 10-fold performance estimate of selected classifier without K-means clustering (Original 400 instances)

Method	Acc	Se	Spec	Prec	Kappa	RMSE
MLP	97.75	0.97	0.99	0.97	0.95	0.12
RBF	98	.98	0.9	0.98	0.95	0.12
Naïve bayes	94.5	0.94	0.99	0.95	0.88	0.21
J48	96.75	0.96	0.95	0.96	0.93	0.17

CONCLUSION AND FUTURE PERSPECTIVE

In this study, we propose a hybrid method based on K-means clustering and classification algorithm and above experimental results shows hybridized classifier algorithm with K-means clustering gives better accuracy results as compared to single method. The prediction accuracy with reduced number of instances is better than original (400) number of instances. The classification algorithms that have been considered for prediction chronic Kidney disease are MLP, RBF, NB and J48.

The future perspective of this research is to use the same technique on optimized classifier to move irrelevant features and insignificant attributes for more accuracy and better results

REFERENCES

Alby, S., & Shivakumar, B.L. (2016, Feb 16). A prediction model for type 2 diabetes using adaptive neuro-fuzzy interface system, Biomedical Research, 63, 1-6.

Alba, E., & Chicano, J.F. (2004). Training neural networks with GA hybrid algorithms, Genetic and Evolutionary Computation GECCO, Springer, Berlin, Heidelberg, 3102, 852–863.

Bunke, H., Kandel, A. (2002, May). "Hybrid Methods in Pattern Recognition", World Scientific (Ed). doi.org /1142 /4871, 30,336.

Carvalho, M., & Ludermir, T. B. (2006). Hybrid training of feed-forward neural networks with Particle swarm optimization, Neural Information Processing, Springer, Berlin, Heidelberg, 4233, 1061–1070.

Dubey, A. (2015). A Classification of CKD Cases Using Multivariate K-Means Clustering. International Journal of Scientific and Research Publications, 5(8), 5.

Humar, K., & Novruz, A. (2008). Design of a hybrid system for the diabetes and heart diseases. Expert Systems with Applications, 35(1-2), 82-89.

Hussein, A.S., & Omar, W.M. (2012). Efficient chronic disease diagnosis prediction and recommendation system. In: Biomedical engineering and sciences (IECBES), IEEE EMBS conference on IEEE. 9–14.Https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease.

Kala, R., Shukla, A., & Tiwari, R. (2009). Comparative analysis of intelligent hybrid systems for detection of Pima Indian diabetes,World Congress on Nature & Biologically Inspired Computing, 947-952.

Mohebian, R., & Hamid, R.M. (2017). A Hybrid Computer –aided diagnosis system for prediction of Breast Cancer Recurrence (HPBCR) Using optimized Ensemble Learning, Computational and structural Biotechnology Journal, 15,75-85.

Nilashi, M., & shahmoradi, L. (2017). An Analytical Method for diseases Prediction Using Machine Learning Techniques, and Computers & Chemical Engineering, 04(10), 212-223.

National Kidney Foundation (NKF). (2002). Clinical Practice Guidelines for Chronic Kidney disease: Evaluation, Classification and Stratification. American Journal of Kidneydisease, 39, S1-S266.

Melin, p., & Castillo, O. (2005). “Hybrid Intelligent Systems for Pattern Recognition Using Soft computing”, springer. 172, 272.

Padmanaban, A., K.R., & Parthiban, G. (2016). Applying Machine Learning Techniques for Predicting the Risk of Chronic Kidney Disease. Indian Journal of Science and technology, 9(29), 1-5.

Patil, B.M., Joshi, R.C, Toshniwal, D. (2010). Hybrid prediction model for Type-2 diabetic patients. Expert System Appl; 37(12), 8102–8108.

Pujari, A.K. (2001). Data mining techniques. Universities press. Radhakrishnan, J., and Sumit, M. (2017). "KI Reports and World Kidney Day." Kidney international reports 2, 125-126.

Ramya, S., & Radha, N. (2016). "Diagnosis of chronic kidney disease using machine learning algorithms." International Journal of Innovative Research in Computer and Communication Engineering 4, 812- 820.

Sharma, S., Sharma, V., & Sharma, A. (2016). “Performance Based Various Machine Learning Classification Techniques for chronic kidney disease diagnosis, 4(3).

Stevens, L.A., & Levey, A.S. (2009). Current status and future perspectives for CKD testing”, American journal of kidney diseases, 53(3).

Temurtas, H., Yumusak, N., & Temurtas, F. (2009). A comparative study on diabetes disease diagnosis using neural networks, Expert Systems with Applications, 36(4), 8610–8615.

Verma, L., Srivastava, S., & Negi, P.C. (July 2016). A hybrid data mining model to predict coronary Artery disease cases using non-invasive clinical data. J Med System, 40(7), 1-7.

Vijayarani, S., & Dhayanand, S. (2015). "Kidney disease prediction using SVM and ANN algorithms." International Journal of Computing and Business Research (IJCBR), 6(2),1-12

Witten, H. I., & Frank, E. (2011). Data mining: Practical machine learning tools and techniques (2nd Ed.). Morgan Kaufmann Publisher, 655.

COMPARATIVE ANALYSIS OF ANN TECHNIQUES FOR STOCK MARKET PREDICTION

Richa Handa, Dr. C.V. Raman University, Bilaspur (C.G.), India (richihanda@yahoo.com,)
A.K. Shrivas, Dr. C.V. Raman University, Bilaspur (C.G.), India (akhilesh.mca29@gmail.com)

ABSTRACT

Prediction is required to make the strategy and planning of any stock market and also play an important role for financial experts. In this paper intelligent techniques are purposed to enhance the ability for accurate prediction of future value of stock data. This paper suggests three Artificial Neural Network (ANN) techniques: Error Back Propagation Network (EBPN), Radial Basis Function Network (RBFN) and Support Vector Regression (SVR) for stock market prediction. EBPN is a feedback neural network, RBFN uses radial basis functions as activation functions for prediction and SVR follows the generalization rule from large data set to balance the complexity of model. The comparative analysis of these three ANN techniques has been done and found that RBFN is outperform from other two ANN techniques with MAPE=0.039.

Keywords: Artificial Neural Network (ANN), Error Back Propagation Network (EBPN), Radial Basis Function Network (RBFN) and Support Vector Regression (SVR).

INTRODUCTION

Prediction of stock market creates problems in developing robust forecasting model for financial time series forecasting because of sudden fluctuating behavior of stock market and makes its nature volatile. Intelligent techniques are widely used for stock market prediction which always performs better than statistical techniques. This paper explores three intelligent techniques: Error Back Propagation Network (EBPN)(Mo, Wang, & Niu, 2016), Radial Basis Function Network (RBFN)(Mohammadi, Fatemi Ghomi, & Zeinali, 2014) and Support Vector Regression (SVR)(Kao, Chiu, Lu, & Yang, 2013). In this research work National Stock Exchange (NSE50) is used for prediction. We have collected one financial year data from April 01, 2018 to March 30, 2019. The data is normalized and then divided into training and testing data set using k-fold cross validation, where 90% data is used to train the model and 10% data used to test the model. After partitioning of data it is given to EBPN, RBFN and SVR models separately for next day ahead prediction of NSE 50 stock data. The comparative study has been done between all these three techniques after prediction and performance of model is compared by using error measure Mean Absolute Percentage Error (MAPE).

The rest of the study is organized as follows. Section 2 gives a brief overview of the literature. Section 3 describes dataset; section 4 briefly describes the methodology as we have used EBPN, RBFN and SVR. Section 5 explores the experimental results. Finally, Section 6 concludes the findings of the research work.

LITERATURE REVIEW

During last one decade researchers are focusing to develop prediction model based on neural network techniques. In literature, many research papers have been published in the field of time series data to develop a predictive model using different intelligent techniques. Authors (Handa, Hota, & Tandan, 2015) have developed models based on Back Propagation Network (BPN) and Radial Basis Function Network (RBFN). Authors(Mo et al., 2016) also used Exponent Back Propagation Neural Network (EBPN) to develop forecasting model and compared the model with EBPN model. Authors (Gan, Peng, & Dong, 2012) suggested a novel hybrid algorithm for automatic selection of input variables, number of hidden nodes of RBF network and optimizing network parameters.

Author (Niu & Wang, 2014) is performed in testing the forecasting effect of the EBPNN model and a comparison to back propagation neural network (BPNN). The outcome showed the advantages of EBPNN over BPNN. Authors (Sharma, Sharma, & Hota, 2016) have developed many models based on Back

Propagation Network (BPN) and Radial Basis Function Network (RBFN). Author (Wang, Wang, Zhang, & Guo, 2011) propose a new approach to forecasting the stock prices via the wavelet De-noising-based Back Propagation (WDBP) neural network. An effective algorithm for predicting the stock prices is developed. The To show the advantage of this new approach for stock index forecast, the WDBP neural network is compared with the single Back Propagation (BP) neural network using the real data set. Authors(de Oliveira & Ludermir, 2016) employs both linear and non linear patterns of a time series based on its volatility. This is explored using a hybrid evolutionary system composed by a simple exponential smoothing filter, ARIMA and autoregressive (AR) linear models and a SVR model.

DATASET USED

In this paper we have used one financial year daily NSE50 data from April 01, 2018 to March 30, 2019. Total 260 observations have been collected from site www.nseindia.com with five feature space: date, open, high, low and close. The dataset is partitioned into training and testing dataset using k-fold cross validation in which 90% data is used for training and 10% data is used for testing purpose. Data normalization is carried out by preparing the data in range between 0 and 1. The summary of dataset is shown in Table1.

Table 1: Summary of daily NSE50 stock data

Particular	**Detail**
Index Data	NSE50 stock data
Period	01-April-2018 to 30-March-2019 (01 year)
Total # of Samples	260
Downloaded From	www.nseindia.com
Data Partition	10 Fold cross validation
Total Observation	260 (Training-234. Testing-26)

METHODOLOGY

In this research work comparative analysis of three ANN techniques called EBPN, RBFN and SVR is used for next day ahead prediction of NSE50 stock data and k-fold cross validation (dynamic partitioning of data) is explained below:

1. Error Back Propagation Network (EBPN)

EBPN is popularly used forecasting technique (Mo et al., 2016) which can be trained using popular Error back propagation algorithm and it is a generalization of delta rule. The delta value for a given input vector compares the output vector to the correct answer. This technique is also known as the backward propagation of errors (Lahmiri, 2014) because after calculation of errors at the output it goes back through the network layers and repeats until the error comes to the desired output.

2. Radial Basis Function Network (RBFN)

Radial basis function network is an ANN used by many authors (Niu & Wang, 2014) which uses radial basis functions as activation functions. Radial basis function networks have many uses, including function approximation, time series forecasting, classification, and system control. RBFN is a three-layer architecture (Usman & Alaba, 2014), the first layer is called Input layer where source node is given, the second layer is

called hidden layer in which each neuron computes its output using radial basis function and this output is sent to the third layer called output layer. The MATLAB generated view of RBFN is shown in Figure 1 where 4 inputs are given to model.

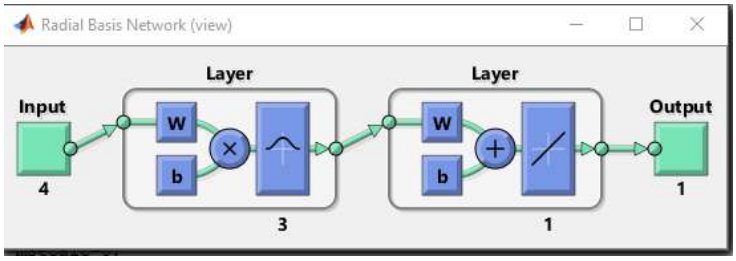

Figure 1: MATLAB generated view of RBFN.

3. Support Vector Regression

SVR method is proposed by (Vapnik & Chervonenkis, 2015) and it is based on the principal of Structure Risk Minimization (SRM), where data generalization has been done(Kao et al., 2013) from large data set to balance the complexity of model. SVR model is represented by equation 1.

$$f(x) = w.\emptyset(x) + b \tag{1}$$

Where x is input to the model, w is weight vector, b is bias and Ǿ is a mapping function. Weight vector is estimated by regularized risk function as shown in equation 2.

$$\frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l} L(y_i, f(x_i)) \tag{2}$$

SVR uses a loss function called ε-insensitive loss function shown in equation 3.

$$L(y, f(x)) = \begin{cases} 0 & |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon Otherwise \end{cases} \tag{3}$$

With ε-insensitive function SVR is formulated as equation 4 and 5.

$$\text{Minimize} \quad \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{l}(\xi_i - \xi_i^*) \tag{4}$$

$$\text{Subject to} \begin{cases} q_i - \left(v.\emptyset(x_i)\right) - b \leq \varepsilon - \xi_i \\ (v.\emptyset(x_i + b - q_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, for i = 1 \ldots . n \end{cases} \tag{5}$$

Where ξ and ξ^* are slack variables that reuse for error measurements came from the values outside boundaries.

This problem can be solved by Lagrangian theory, now the SVR function can be expressed as equation 6.

$$f(x, y) = f(x, \propto, \propto^*) = \sum_{i=1}^{n}(\propto_i - \propto_i^*)k(x, x_i) + b \tag{6}$$

Where α and α^* are Lagrangian multipliers and $k(x, x_i)$ is a kernel function, where$k(x, x_i) = \exp((-\|x_i - x_j\|^2)/2\sigma^2$, where σ is a parameter of Gaussian kernel.

4. K-Fold cross validation

K-fold validation is a technique where dynamically partitioned of data is done in training and testing data sets (Jiang & Chen, 2016). In k-fold cross- validation the data set is divided into k subsets. At each fold, one subset is used as the test set, and remaining subsets are merged together and used as a training set. Then the average error of all k trials is calculated so that each fold takes part in training and testing both.

RESULT AND ANALYSIS

In this research work the models are evaluated based on one year NSE50 stock data. A self written MATLAB code is generated from above mentioned three techniques as well as for k-fold cross validation. Here various parameters of RBFN and EBPN were adjusted using trial and error methods. Four parameters are given to all three techniques called input variable and there is one more parameter work as a dependent variable called target value which is next day close. Comparative results of all three models are obtained through MATLAB code by measuring the performance of models by error measure called mean Absolute Percentage Error (MAPE) as given in Table 2.

Table 2: Comparative MAPE of EBPN, RBFN and SVR for NSE50 stock data.

Techniques	**Data Partition**	
	Training	**Testing**
EBPN	0.600	0.660
RBFN	0.035	0.038
SVR	0.627	0.637

The empirical result depicted in table 2 clearly reflects that RBFN is performing better than other ANN techniques called EBPN and SVR with MAPE=0.035 for training and MAPE=0.038 for testing data. The comparative MAPE graph of actual and predicted values of NSE30 stock data using EBPN, RBFN and SVR is shown from figure 2 to 7 respectively.

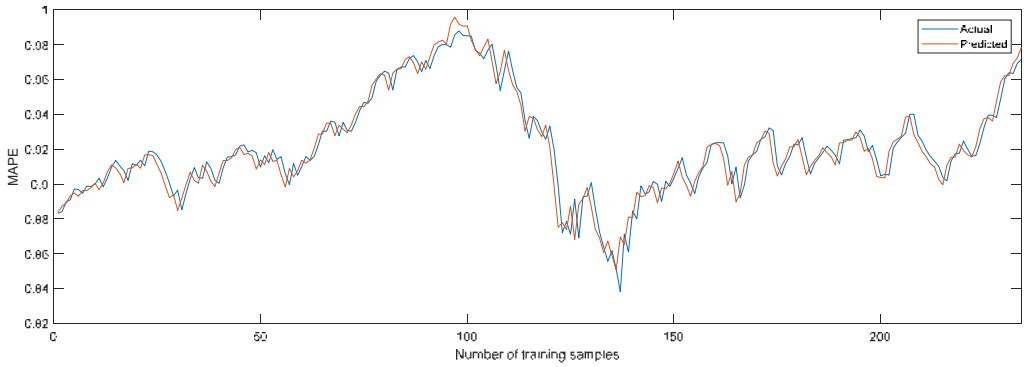

Figure 2. Comparative graph of predicted and actual value of training dataset of NSE50 using EBPN.

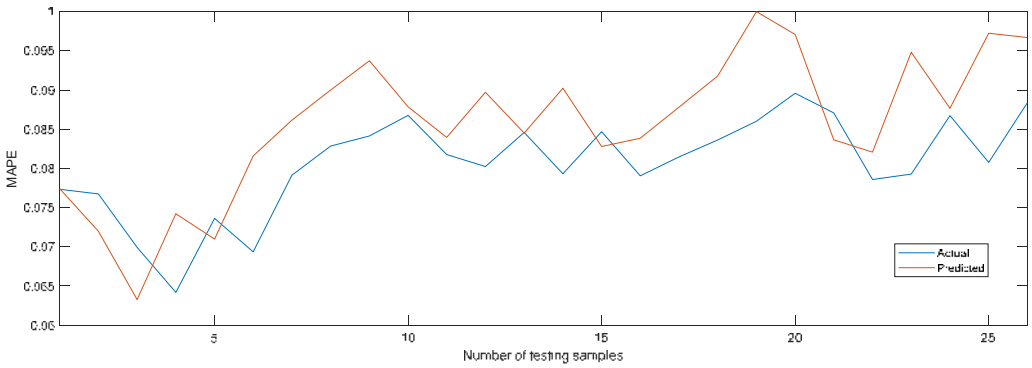

Figure 3. Comparative graph of predicted and actual value of testing dataset of NSE50 using EBPN.

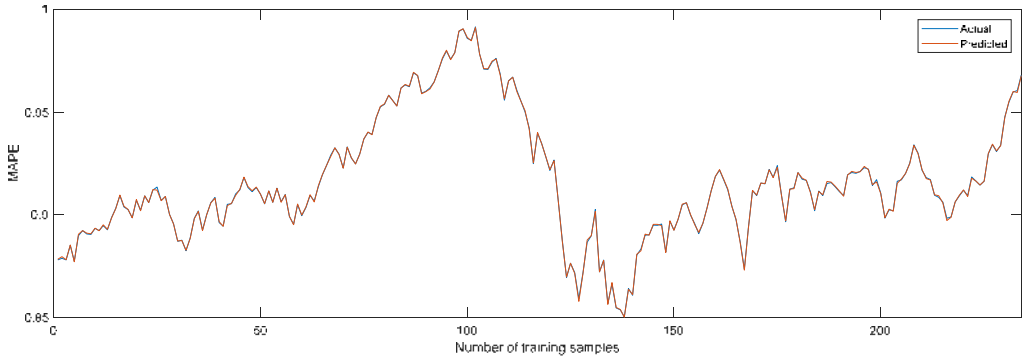

Figure 4. Comparative graph of predicted and actual value of training dataset of NSE50 using RBFN.

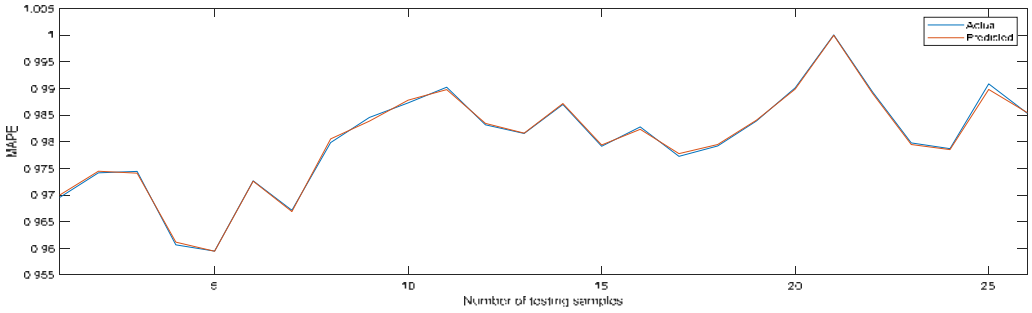

Figure 5. Comparative graph of predicted and actual value of testing dataset of NSE50 using RBFN.

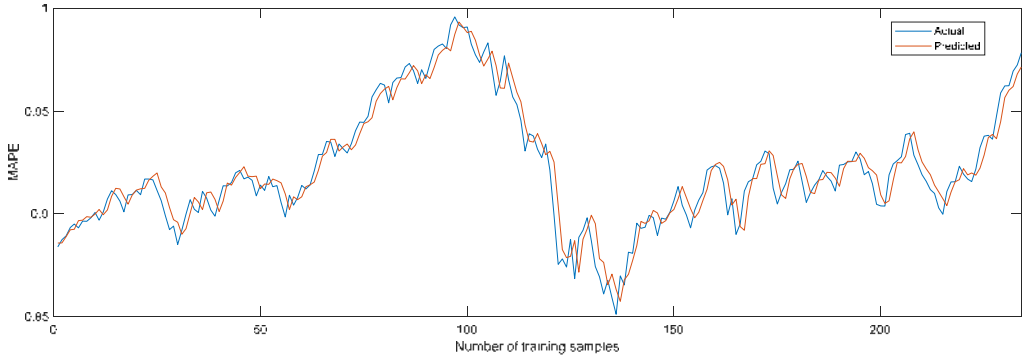

Figure 6. Comparative graph of predicted and actual value of training dataset of NSE50 using SVR.

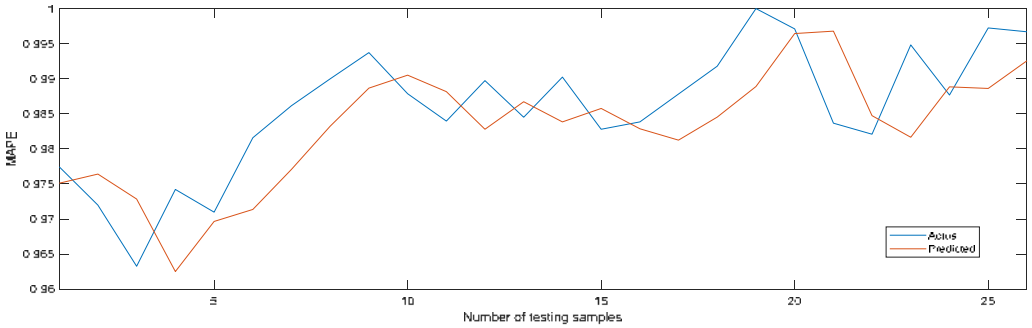

Figure 7. Comparative graph of predicted and actual value of testing dataset of NSE50 using SVR.

CONCLUSION

Stock market prediction is very tedious task due to its fluctuating behavior. Three ANN techniques EBPN, RBFN and SVR are used in this research paper for development of a robust forecasting model for NSE50 stock market prediction. 10-fold cross validation is used to train the models for more accurate prediction. The model using RBFN is performing better that EBPN and SVR by producing more accurate predictive result with MAPE=0.035 for training and MAPE=0.038 for testing.

REFERENCES

de Oliveira, J. F. L., & Ludermir, T. B. (2016). A hybrid evolutionary decomposition system for time series forecasting. *Neurocomputing*, *180*, 27–34. https://doi.org/10.1016/j.neucom.2015.07.113

Gan, M., Peng, H., & Dong, X. P. (2012). A hybrid algorithm to optimize RBF network architecture and parameters for nonlinear time series prediction. *Applied Mathematical Modelling*, *36*(7), 2911–2919. https://doi.org/10.1016/j.apm.2011.09.066

Handa, R., Hota, H. S., & Tandan, S. R. (2015). Stock Market Prediction with Various Technical Indicators using Neural Network Techniques. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, *3*(1), 604–608.

Jiang, P., & Chen, J. (2016). Displacement prediction of landslide based on generalized regression neural networks with K-fold cross-validation. *Neurocomputing*, *198*, 40–47. https://doi.org/10.1016/j.neucom.2015.08.118

Kao, L. J., Chiu, C. C., Lu, C. J., & Yang, J. L. (2013). Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. *Neurocomputing*, *99*, 534–542. https://doi.org/10.1016/j.neucom.2012.06.037

Lahmiri, S. (2014). Wavelet low- and high-frequency components as features for predicting stock prices with backpropagation neural networks. *Journal of King Saud University - Computer and Information Sciences*, *26*(2), 218–227. https://doi.org/10.1016/j.jksuci.2013.12.001

Mo, H., Wang, J., & Niu, H. (2016). Exponent back propagation neural network forecasting for financial cross-correlation relationship. *Expert Systems with Applications*, *53*, 106–116. https://doi.org/10.1016/j.eswa.2015.12.045

Mohammadi, R., Fatemi Ghomi, S. M. T., & Zeinali, F. (2014). A new hybrid evolutionary based RBF networks method for forecasting time series: A case study of forecasting emergency supply demand time series. *Engineering Applications of Artificial Intelligence*, *36*, 204–214. https://doi.org/10.1016/j.engappai.2014.07.022

Niu, H., & Wang, J. (2014). Financial time series prediction by a random data-time effective RBF neural network. *Soft Computing*, *18*(3), 497–508. https://doi.org/10.1007/s00500-013-1070-2

Sharma, H., Sharma, D. K., & Hota, H. S. (2016). A hybrid neuro-fuzzy model for foreign exchange rate prediction. *Academy of Accounting and Financial Studies Journal*, *20*(3), 1–13.

Usman, O. ., & Alaba, O. . (2014). Predicting Electricity Consumption Using Radial Basis Function (RBF) Network. *International Journal of Computer Science and Artificial Intelligence*, *4*(2), 54–63. https://doi.org/10.5963/ijcsai0402004

Vapnik, V. N., & Chervonenkis, A. Y. (2015). On the uniform convergence of relative frequencies of events to their probabilities. *Measures of Complexity: Festschrift for Alexey Chervonenkis*, *XVI*(2), 11–30. https://doi.org/10.1007/978-3-319-21852-6_3

Wang, J. Z., Wang, J. J., Zhang, Z. G., & Guo, S. P. (2011). Forecasting stock indices with back propagation neural network. *Expert Systems with Applications*, *38*(11), 14346–14355. https://doi.org/10.1016/j.eswa.2011.04.222

CITIZEN ACCEPTANCE AND USE OF MOBILE GOVERNMENT APPLICATIONS (MG-APPS) IN OMAN

Ashish Rastogi, IT Department, Higher College of Technology, Muscat, Sultanate of Oman
(profrastogi@gmail.com)

ABSTRACT

Governments all over the world are facing opportunities and challengesto provide high quality government services to their citizens in an efficient and effective way by using the modern technologies. Mobile Applications are considered as a very useful application of E-Governance to connect Government with citizens and use the government applications. The purpose of this study is to investigate the acceptability of four important mobile Government Apps (mG-Apps) Baladiyeti, HEAC, Khedmah and eSehaty belongs to different categories. The questionnaire has been created using Google forms and electronically distributed send to the 260 users selected randomly for data collection. The questionnaire consists of 22 survey questions, 5 demographic statements, and a question related with the use of mG-Apps. Data were collected from 217 respondent and all are valid responses. The response rate is around 83% and structural equation modeling (SEM) technique was used to evaluate the causal model and to examine the reliability and validity of the measurement model. The variables used in this study are perceived ease of use, perceived usefulness, Information Quality, Information System Quality, Attitude, Behavior Intention. The seven hypotheses have been tested. The findings of this research have provided theoretical contributions to the existing research on mG-Apps and practical implications to decision-makers involved in the development and implementation of mG-Apps in Oman.

Keywords: mobile Government Apps (mG-Apps), Baladiyeti, HEAC, Khedmah and eSehaty, E-Government, Structural Equation Modeling (SEM), Google Forms.

INTRODUCTION

E-government is simply the use of Government applications by using the applications information and communications technology on for example, The Internet, to improve the processes of government. E-government might seem to be little more than an effort to expand the market of e-commerce from business to government. Starting from using E-government applications from cloud computing, now we reach to M-Government(mG-Apps)((Rastogi, 2010). Surely there is some truth in this. E-commerce is marketing and sales via the Internet. Since governmental institutions take part in marketing and sales activities, both as buyers and sellers, it is not inconsistent to speak of e-government applications of e-commerce. Governments do after all conduct business.(Chadwick & May, 2003). Successful implementation of the e-government will be more often in the agenda to reduce the cost of the local and national governments and improve the services and user's satisfaction (Warkentin, Gefen, Pavlou, & Rose, 2002). Therefore, it is important to understand the needs of the entrepreneurs as a group of potential mG-Apps users and also help to reduce the administrative burdens of the enterprises. Citizen adoption of mG-Apps is an important issue for the success of e-government initiatives (Ozkan&Kanat, 2011). The center undertaking of government is administration, the activity of managing society, not showcasing and deals. In present day vote based systems, duty and power for control is partitioned up and shared around the authoritative, official and legal branches of government. Streamlining to some degree, the lawmaking body is in charge of making strategy as laws, the official for executing the approach and law authorization, and the legal for settling legitimate clashes. E-government is tied in with enhancing crafted by these branches of government. The Sultanate of Oman government has taken good initiative in using ICT in order to meet its goals and objectives and to provide structured services to all its stakeholders. Most of the large and international organizations in Oman have effective computer systems to efficiently conduct their business. A number of large organizations have spent

huge amounts of money on installing computer systems to support their business processes (Parker &Castleman, 2007).With the advancement of the Internet and supporting information and communication technologies, e-government has emerged as an effective means of delivering government services to citizens. In the recent past, e-government has become popular in many developing countries. Most notably are the Middle Eastern countries that have continued to invest significantly into e-government initiatives in the last five years. The purpose of this article is to highlight e-government initiatives inOman.

LITERATURE REVIEW

All over the world peoples are moving towards the use of smart phones from the traditional phones. Most of the country is going digital and by providing location based services to the citizens through smart mobile devices to enhance the quality of E-government services.(Albesher& Stone, 2016; Faisal &Talib, 2016) The important factors of the acceptance of the M-Government technologies are: a) wider acceptance of these technologies by the public sector b) M-Government services are cheaper than computer-based services c) Increase of penetration of mobile devices d) Ease of use for citizens e) Easier interoperability (ITU, OECD; 2011).The M-Government may be considered as an alternative to the existing laptop or desktop based E-Government services.Due to the high penetration of mobile technologies and mobile devices the use of M-Government has been increasing rapidly (Albesher& Stone, 2016). In developing countries, where the penetration of mobile growth is high, they could significantly increase the use of E-Government (Abanumy, Al-Badi, & Mayhew, 2005). (Susanto&Aljoza, 2015)discussed the dimensions of Perceived Ease of Use (PEOU), Perceived Usefulness (PU), and greater mattered aspects of one's decision of using e-government services. . Time, money and energy could be saved simply by making an individual see how useful it is to use an E-government service. Advance Software and hardware facilities encourage public and private institutions to offer a large number of services to mobile users (Payne, 2013). The digital co-operation between the government entities, government sector and the residents of the country is refereed as the E-government. The main purpose of the E-government is to provide fast, efficient and low cost services to the stakeholders with in the government and outside the government (Al-Azizi, Al-Badi, Al-Zrafi, & Sharma, 2017; Asongu & Nwachukwu, 2016).The researchers (Naqvi & Al-Shihi, 2014)discussed in his research report that e-Oman was established in 2013 to administer E-Government projects pertaining to infrastructure and latest technology to provide services to residents. An individual's perception on web navigation and whether he/she is able to use it anytime and anywhere depends on the dimensions of Perceived Ease of Use (PEU) of an online public service. Also, the intentions of an individual on whether to use an e-government service are affected by trust and social influence. Factors in designing, developing, managing a d promoting the e- government services should be welcomed and adopted by the government and e-government professionals. (Alshehri et al., 2012) altered UTAUT model on how accepting the user can be and use of e-government in KSA. (Colesca, 2009) states that all the government around the world are using technology or information system to achievetheir goal or governmentactivity.

TECHNOLOGY ACCEPTANCE MODEL (TAM)

Many researches use TAM as a framework to predict and explain a variety of human behaviors in the IT adoption context (Ajzen&Fishbein, 1980; Hu, Chau, Sheng, & Tam, 1999). TAM theorizes that causal linkages flow in a sequence of beliefs, attitudes, intentions, and behaviors. To examine an individual's actual system use, most studies focus on factors affecting the individual's intentions of system acceptance (Gefen, Karahanna, & Straub,2003; (Sharma, Gaur, Saddikuti, & Rastogi, 2017)).

A general model of TAM is shown in Fig. 1. Prior research suggests that perceived usefulness (PU) and perceived ease of use (PEOU) are two major influential emotional beliefs that determine a user's IT acceptance. (Davis, 1989)defined perceived usefulness as "the degree to which a person believes that using a particular system would enhance his or her job performance" and perceived ease of use as "the degree to which a person believes that using a particular system would be free of effort" (Venkatesh& Davis, 2000).

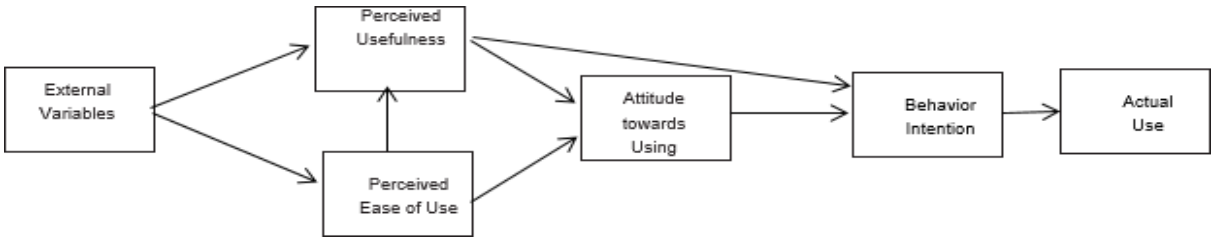

Figure 1: Technology Acceptance Model (Davis, 1989)

This study examines the validity of TAM in the e-Government setting in Oman and focuses on how Citizens behave differently, and exhibit different levels of acceptance, than other Mobile App user users. Information and communications technologies (ICTs) are playing an increasingly vital role in the daily life.(Taylor & Todd, 1995).(Larsen &Rainie, 2002)suggested that e-Government services include information for research, government forms and services, public policy information, employment and business opportunities, voting information, tax filing, license registration, or renewal, payment of fines, and submission of comments to government officials. However, the successful operation of e-Government does not depend on the technology, but rather on the people (Wang & Liao,2008).

RESEARCH HYPOTHESES AND MODEL

TAM asserts that intentions to perform behavior determine actual behavior. Intention itself represents an individual's attitude toward the behavior. The TAM indicates that both perceived usefulness (PU) and perceived ease of use (PEOU) are key, independent variables that can determine or influence potential user' attitudes (ATT) toward behavioral intention (BI).(Davis, 1989) called for further research to consider the role of additional external variables that influence PU and PEOU. Two important external variables – information systems quality (ISQ) and information quality (IQ) – have been consistently found to be influential factors that affect the perceived usefulness and ease of use of IT. According to (Ajzen&Fishbein, 1975), attitude and the subjective norms are important factors on the behavioral intention formation, a proposition that is supported by TAM. Users with a more positive attitude toward IT are likely to be more satisfied with system and view it as more useful (Ajzen&Fishbein, 1975) (Heeks, 2006). Therefore, user attitude is hypothesized to positively affect perceived usefulness andbehavioral intention.(DeLone& McLean, 1992) defined information system quality as quality manifested in a system's overall performance and measured by individuals' perceptions. Because citizens are faceless in e-Government interactions, the information system's quality becomes the “online storefront” upon which first impressions are formed. If a citizen perceives an e-Government system to be of high quality, that citizen will be more likely to use internet systems to submit applications or access other e-Government services online (Wang & Liao, 2008). Information quality (IQ), as assessed by citizens, usually influences their satisfaction and perceived usefulness (Moon & Kim, 2001; Aggelidid&Chatzoglou,2008).

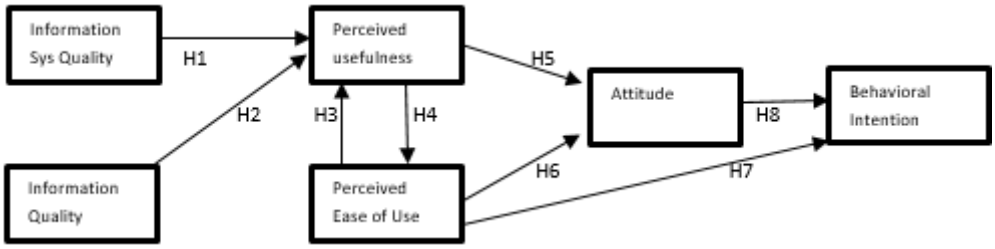

Figure 2: The Research Model

The mG-Apps promises to enhance the delivery of public services and information by redefining the traditional concept. Based on the technology acceptance model (TAM) theory (Lederer et al., 2000; Lin & Lu, 2000, Rastogi,2017) the current study presents the following hypotheses:

H1. The Information System Quality (ISQ) of mG-Apps positively affects the Perceived Usefulness (PU) of using the application in submitting application forms online.

H2. The Information Quality (IQ) of mG-Appsfiling systems positively affects the Perceived Usefulness (PU) of using the mG-Apps.

H3.The Perceived Ease of Use (PEOU) ofmG-Apps positively affects the Perceived Usefulness (PU) of using the mG-Apps to fill out applications.

H4.The Perceived Usefulness (PU) ofmG-Apps positively affects the Perceived Ease of Use (PEOU) of using the mG-Apps to fill out applications.

H5. The Perceived Usefulness (PU) of using mG-Appshas a positive effect on Attitudes of the user regarding use of the mG-Apps.

H6. The Perceived Ease of Use (PEOU) ofmG-Appshas a positive effect on user Attitudes toward the use of mG-Apps.

H7. The Perceived Ease of Use (PEOU) of the mG-Apps services has a positive effect on user Behavior Intentions.

H8. User attitude on using the mG-Appspositively affects user Behavior intentions.

SAMPLE DEMOGRAPHICS

The goal of this study was to apply and evaluate the proposed hypothetical model in using mGov applications. In this study, e-survey was used to collect the data. In many studies where e-surveys are used, the number of internet users is taken into account when determining the size of the sample(Couper, 2000). In our primary data it is observed that 45% respondents are diploma holders, 31.7% bachelor and 14.4% master degree holders. The gender ratio of the respondents is also 50:50 (approximately). The maximum number of respondents is from the age group of between 20 and 30 years.

DATA ANALYSIS AND RESULTS

Descriptive Statistics of the seven (7) constructs are shown in Table 1.The standard deviation of all the constructs are ranging from .802 to 1.02, and all the means above3.0.

Table 1. Descriptive Statistics

Items	N	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
ATT1	217	1.0	5.0	3.664	.0657	.9681	.937
ATT2	217	1.0	5.0	3.387	.0580	.8539	.729
ATT3	217	1.0	5.0	3.470	.0688	1.0139	1.028
PU1	217	1.0	5.0	3.078	.0580	.8544	.730
PU2	217	1.0	5.0	3.212	.0562	.8285	.686
PU3	217	1.0	5.0	3.378	.0625	.9205	.847
PEOU1	217	1.0	5.0	3.364	.0620	.9135	.834
PEOU2	217	1.0	5.0	3.452	.0688	1.0132	1.027
PEOU3	217	1.0	5.0	3.230	.0607	.8936	.799
PEOU4	217	1.0	5.0	3.410	.0661	.9730	.947
ISQ1	217	1.0	5.0	3.341	.0618	.9097	.828
ISQ2	217	1.0	5.0	3.415	.0610	.8993	.809
ISQ3	217	1.0	5.0	3.590	.0740	1.0897	1.187
ISQ4	217	1.0	5.0	3.415	.0617	.9095	.827
IQ1	217	1.0	5.0	3.313	.0621	.9146	.837
IQ2	217	1.0	5.0	3.263	.0676	.9954	.991
IQ3	217	1.0	5.0	3.207	.0546	.8041	.647
IQ4	217	1.0	5.0	3.300	.0561	.8264	.683
IQ5	217	1.0	5.0	3.304	.0606	.8922	.796
BI1	217	1.0	5.0	3.300	.0608	.8963	.803
BI2	217	1.0	5.0	3.244	.0595	.8768	.769
BI3	217	1.0	5.0	3.456	.0694	1.0225	1.046
Valid N (listwise)	217						

Construct validity and reliability have been tested to ensure the consistency and reliability of the results. Table 2 shows the reliability statistics in which the Cronbach's alpha is .938 which shows that our constructs are highly reliable.

Table 2: Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.938	.938	22

The following table 3 shows the item total statistics.

Table 3. Item-Total Statistics

Items	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
ATT1	70.129	166.354	.477	.415	.937
ATT2	70.406	165.001	.615	.596	.935
ATT3	70.323	164.016	.546	.518	.936
PU1	70.714	166.177	.559	.438	.936
PU2	70.581	167.569	.511	.477	.937
PU3	70.415	164.503	.587	.466	.936
PEOU1	70.429	162.681	.674	.559	.934
PEOU2	70.341	161.402	.652	.535	.934
PEOU3	70.562	163.877	.635	.521	.935
PEOU4	70.382	162.543	.634	.549	.935
ISQ1	70.452	164.258	.606	.528	.935
ISQ2	70.378	164.431	.606	.509	.935
ISQ3	70.203	159.986	.654	.554	.935
ISQ4	70.378	164.107	.613	.495	.935
IQ1	70.479	162.232	.693	.656	.934
IQ2	70.530	163.158	.592	.606	.935
IQ3	70.585	165.179	.648	.665	.935

IQ4	70.493	163.797	.696	.691	.934
IQ5	70.488	163.668	.646	.548	.935
BI1	70.493	163.677	.642	.648	.935
BI2	70.548	163.397	.671	.637	.934
BI3	70.336	161.002	.661	.570	.934

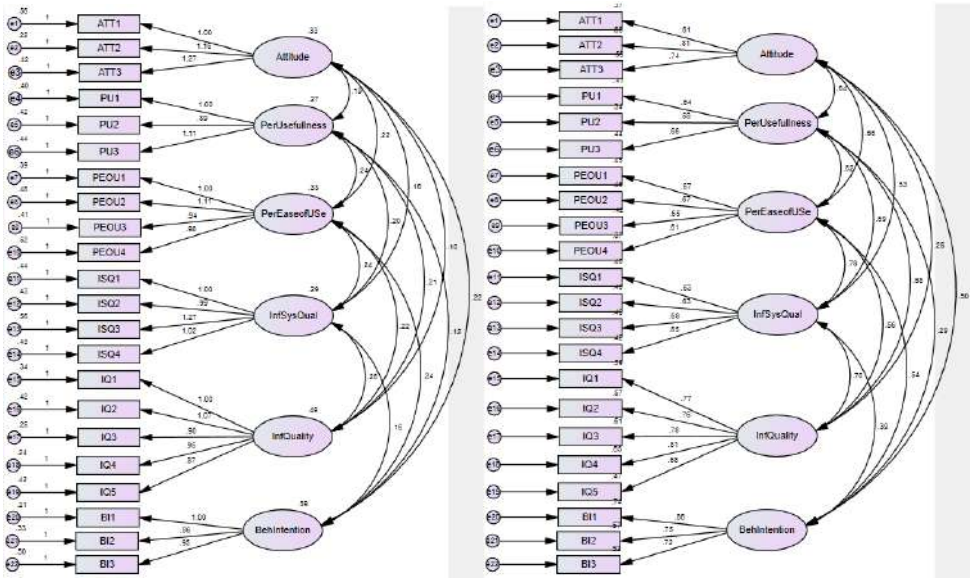

Figure 3: The Path Analysis

REGRESSION RESULTS OF HYPOTHESIS TESTING

By using the regression analysis the hypotheses aims to determine which predictors (independent variables) provide a meaningful contribution to the explanation of the dependent variables (Hair et al.).

H1. The Information System Quality (ISQ) of mG-Apps positively affects the Perceived Usefulness (PU) of using the application in submitting application forms online.

H2. The Information Quality (IQ) of mG-Apps filing systems positively affects the Perceived Usefulness (PU) of using the mG-Apps.

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.606[a]	.368	.362	.54837	2.177

a. Predictors: (Constant), IQ, ISQ

b. Dependent Variable: PU

ANOVA[a]						
Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	37.438	2	18.719	62.248	.000[b]
	Residual	64.353	214	.301		
	Total	101.790	216			

a. Dependent Variable: PU

b. Predictors: (Constant), IQ, ISQ

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.145	.190		6.032	.000
	ISQ	.305	.069	.332	4.428	.000
	IQ	.314	.072	.328	4.382	.000

a. Dependent Variable: PU

H3. The Perceived Ease of Use (PEOU) ofmG-Apps positively affects the Perceived Usefulness (PU) of using the mG-Apps to fill out applications.

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.639[a]	.409	.406	.52899	2.053
a. Predictors: (Constant), PEOU					
b. Dependent Variable: PU					

ANOVA[a]						
Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	41.627	1	41.627	148.756	.000[b]
	Residual	60.164	215	.280		
	Total	101.790	216			

a. Dependent Variable: PU

b. Predictors: (Constant), PEOU

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.245	.166		7.500	.000
	PEOU	.588	.048	.639	12.197	.000

a. Dependent Variable: PU

H4. The Perceived Usefulness (PU) of mG-Apps positively affects the Perceived Ease of Use (PEOU) of using the mG-Apps to fill out applications.

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.639[a]	.409	.406	.57553	1.934

a. Predictors: (Constant), PU

b. Dependent Variable: PEOU

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	49.274	1	49.274	148.756	.000[b]
	Residual	71.216	215	.331		
	Total	120.490	216			

a. Dependent Variable: PEOU

b. Predictors: (Constant), PU

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.122	.188		5.969	.000
	PU	.696	.057	.639	12.197	.000

a. Dependent Variable: PEOU

H5. The Perceived Usefulness (PU) of using mG-Apps has a positive effect on Attitudes of the user regarding use of the mG-Apps.

H6. The Perceived Ease of Use (PEOU) ofmG-Apps has a positive effect on user Attitudes toward the use of mG-Apps.

Model Summary[b]					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.622[a]	.387	.381	.62078	2.084

a. Predictors: (Constant), PEOU, PU

b. Dependent Variable: Attitude

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	51.992	2	25.996	67.458	.000[b]
	Residual	82.469	214	.385		
	Total	134.462	216			

a. Dependent Variable: Attitude

b. Predictors: (Constant), PEOU, PU

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	1.039	.219		4.745	.000
	PU	.307	.080	.267	3.838	.000
	PEOU	.439	.074	.416	5.974	.000

a. Dependent Variable: Attitude

H7. The Perceived Ease of Usefulness (PEOU) of the mG-Apps services has a positive effect on user Behavior Intentions.

H8. User attitude on using the mG-Apps positively affects behavior intentions

Model Summary					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.687[a]	.472	.467	.58103	1.764

a. Predictors: (Constant), Attitude, PEOU

b. Dependent Variable: BI

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	64.644	2	32.322	95.743	.000[b]
	Residual	72.245	214	.338		
	Total	136.889	216			

a. Dependent Variable: BI

b. Predictors: (Constant), Attitude, PEOU

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.687	.202		3.395	.001
	PEOU	.622	.065	.584	9.519	.000
	Attitude	.157	.062	.156	2.545	.012

a. Dependent Variable: BI

SUMMARY OF HYPOTHESIS TESTING

Hypothesis	Specification	Result
H1	The Information System Quality (ISQ) of mG-Apps positively affects the Perceived Usefulness (PU) of using the application in submitting application forms online.	Supported (R=.606, R^2=.368 $\beta = .305$, $p < .001$)
H2	The Information Quality (IQ) of mG-Apps filing systems positively affects the Perceived Usefulness (PU) of using the mG-Apps.	Supported (R=.612, R^2=.368 $\beta = .314$, $p < .001$)
H3	The Perceived Ease of Use (PEOU) of mG-Apps positively affects the Perceived Usefulness (PU) of using the mG-Apps to fill out applications.	Supported (R=.639, R^2=.409 $\beta = .588$, $p < .001$
H4	The Perceived Usefulness (PU) of mG-Apps positively affects the Perceived Ease of Use (PEOU) of using the mG-Apps to fill out applications.	Supported (R=.639 R^2=.409 $\beta = .696$, $p < .001$
H5	The Perceived Usefulness (PU) of using mG-Apps has a positive effect on Attitudes of the user regarding use of the mG-Apps.	Supported (R=.622 R^2=.387 $\beta = .307$, $p < .001$
H6	The Perceived Ease of Use (PEOU) of mG-Apps has a positive effect on user Attitudes toward the use of mG-Apps.	Supported (R=.622, R^2=.387 $\beta = .439$, $p < .001$)
H7	The Perceived Ease of Use (PEOU) of the mG-Apps services has a positive effect on user Behavior Intentions.	Supported (R=..687, R^2=..472 $\beta = .622$, $p < .001$)
H8	User attitude on using the mG-Apps positively affects behavior intentions.	Not Supported (R=.687, R^2=.472 $\beta = .157$, $p > .001$)

CONCLUSION

The data analysis of this study shows that the all the constructs of the proposed model have strong influence on the use of the various mg-Apps by the residents of Oman. Our data analysis results shows that the Information System Quality and Information Quality positively influence on the Perceived Usefulness (PU), this isconsistent with the results of (Lin, Fofanah, & Liang, 2011). Furthermore, Perceived Ease of Use (PEOU) of mG-Apps positively affects the Perceived Usefulness (PU) and Perceived Usefulness (PU) of mG-Apps positively affects the Perceived Ease of Use (PEOU) of using mg-Apps by the residents of Oman. This may be due to the User interface of the various mg-Apps are very simple and easy to understand. Even, the person without prior basic knowledge of the mg-Apps can use the Applications very easily. The mG-Apps discussed in the study are available in English and Arabic languages. It is also observed that usefulness and ease of use of mG-Apps has a strong linkage with the user attitudes. Some (Lai, 2017)respondent's attitude is very positive for using the applications in future whenever they require facilities from the government. This may due to the usability of the mG-Apps. Furthermore, The Perceived Ease of Use (PEOU) of the mG-Apps services has a very strong linkage on user Behavior Intentions. Although, the user attitude on using mG-Apps shows the weak linkage with behavior intention, and it is rejected. The Mobile Apps in Oman is significantly very useful and in near future government will add more mG-Apps. This study has five limitations, first the data is collected from capital of Sultanate of Oman, Muscat where the infrastructure is very good. If the data is collected from other cities the results may not be same. Secondly, most of the respondents are well educated and young, if the demographic data variates then it will affect the results. Third, the sample size in this study is moderate, if we increase the sample size then it may affect the results. Fourth, since all the mG-Apps are developed by the government organization and it may not be updated on regularly basis. In addition, the Omani government may develop public-private partnership (PPP) for intelligent and effective use of MG-Apps(Dwivedi et al., 2017; Sharma, Al-Badi, Rana, & Al-Azizi, 2018).Lastly, this study is to investigate the acceptability of four important mobile Government Apps (mG-Apps) Baladiyeti, HEAC, Khedmah and eSehaty. If the other mG-Apps are included in the study the results may be different.

REFERENCES

Al-Azizi, L., Al-Badi, A. H., Al-Zrafi, T., & Sharma, S. K. (2017). Exploring the adoption of mobile applications: Case studies in government agencies in Oman. In *29th International Business Information Management Association Conference - Education Excellence and Innovation Management through Vision 2020: From Regional Development Sustainability to Global Economic Growth*.

Aggelidid, V. P., &Chatzoglou, P. D. (2008). Using a modified technology acceptance model in hospitals, *Journal of Medical Informatics*, 78, 115−126.

Alshehri, M., Drew, S., Alhussain, T., &Alghamdi, R. (2012). The Effects of Website Quality on Adoption of E-Government Service: AnEmpirical Study Applying UTAUT Model Using SEM. *arXiv preprint arXiv: 1211.2410*.

Abanumy, A., Al-Badi, A., & Mayhew, P. (2005). E-Government Website accessibility: in-depth evaluation of Saudi Arabia and Oman, *The Electronic Journal of e-Government*, 3(3), 99-106.

Ajzen, I., &Fishbein, M. (1975). Belief, attitude, intention and behavior: An introduction to theory and research. In: Reading, *MA: Addison-Wesley*.

Ajzen, I., &Fishbein, M. (1980). Understanding attitudes and predicting social behavior.

Albesher, A. S., & Stone, R. T. (2016). Current state of m-government research: identifying future research opportunities, *International Journal of Electronic Governance*, 8(2), 119-139.

Asongu, S. A., & Nwachukwu, J. C. (2016). The role of governance in mobile phones for inclusive human development in Sub-Saharan Africa. *Technovation*., https://doi.org/10.1016/j.technovation.2016.04.002.

Colesca, S. E. (2009). Increasing e-trust: A solution to minimize risk in e-government adoption, *Journal of applied quantitative methods*, 4(1), 31-44.

Couper, M. P. (2000). Web surveys: A review of issues and approaches. *The Public Opinion Quarterly, 64*(4), 464-494.

Chadwick, A., & May, C. (2003). Interaction between States and Citizens in the Age of the Internet: "e□Government" in the United States, Britain, and the European Union, *Governance*, 16(2), 271-300.

Davis, F. D. (1989). MIS Quarterly, 13(null), 319.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable, *Information systems research*, 3(1), 60-95.

Dwivedi, Y. K., Rana, N. P., Janssen, M., Lal, B., Williams, M. D., & Clement, M. (2017). An empirical validation of a unified model of electronic government adoption (UMEGA). *Government Information Quarterly*. https://doi.org/10.1016/j.giq.2017.03.001.

Faisal, M. N., &Talib, F. (2016). E-government to m-government: a study in a developing economy, *International Journal of Mobile Communications*, 14(6), 568-592.

Gefen, D., Warkentin, M., Pavlou, P., & Rose, G. (2002). E-government adoption. AMCIS 2002 Proceedings, 83.

Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: an integrated model, *MIS quarterly*, 27(1), 51-90.

Hair, J., Blake, W., Babin, B., and Tatham, R. 2006. Multivariate Data Analysis. New Jersey: Prentice Hall.

Heeks, R. (2006). Health information systems: Failure, success and improvisation, *International journal of medical informatics*, 75(2), 125-137.

Hu, P. J., Chau, P. Y., Sheng, O. R. L., & Tam, K. Y. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. Journal of management information systems, 16(2), 91-112.

Hendrickson, A. R., Massey, P. D., &Cronan, T. P. (1993). On the test-retest reliability of perceived usefulness and perceived ease of use scales. *MIS quarterly*, 227-230.

Klopping, I. M., & McKinney, E. (2004). Extending the technology acceptance model and the task-technology fit model to consumer e-commerce, *Information Technology, Learning, and Performance Journal*, 22(1), 35.

Kaplowitz, M. D., Hadlock, T. D., & Levine, R. (2004). A comparison of web and mail survey response rates, *Public opinion quarterly*, 68(1),94-101.

Lin, J. C. C., & Lu, H. (2000). Towards an understanding of the behavioral intention to use a web site, *International journal of information management*, 20(3),197-208.

Lai, P. C. (2017). The literature review of technology adoption models and theories for the novelty technology. JISTEM-Journal of Information Systems and Technology Management, 14(1), 21-38.

Larsen, E., &Rainie, H. (2002). Digital Town Hall: How local officials use the Internet and the civic benefits they cite from dealing with constituents online: Pew Internet & American Life Project Washington, DC.

Lin, F., Fofanah, S. S., & Liang, D. (2011). Assessing citizen adoption of e-Government initiatives in Gambia: A validation of the technology acceptance model in information systems success, *Government Information Quarterly*, 28(2), 271-279.

Lederer, A. L., Maupin, D. J., Sena, M. P., & Zhuang, Y. (2000). The technology acceptance model and the World Wide Web, *Decision support systems*, 29(3), 269-282.

Lille, M., &Prause, G. (2009). E-governmental services in the Baltic Sea region (No. 10/2009), Hochschule Wismar, Wismar BusinessSchool.

Moon, J. W., & Kim, Y. G. (2001). Extending the TAM for a World-Wide-Web context Information Management, 38(4), 217−230.

Naqvi, S. J., & Al-Shihi, H. (2014). Factors Affecting M-commerce Adoption in Oman using Technology Acceptance Modeling Approach. *TEM Journal*.

Ozkan, S., &Kanat, I. E. (2011). e-Government adoption model based on theory of planned behavior: Empirical validation, *Government Information Quarterly*, 28(4), 503-513.

Parker, C., &Castleman, T. (2007). New directions for research on SME-eBusiness: insights from an analysis of journal articles from 2003-2006, *Journal of information systems and small business*, 1(1), 21-40.

Payne, J. (2013). Secure Mobile Application Development. *IT Professional*. , https://doi.org/10.1109/MITP.2013.46

Prause, G., &Reidolf, M. (2011). E-governmental Services for Entrepreneurs in the Baltic Sea Region-An Empirical Study, *China-USA Business Review*, 10(11).

Rastogi, A. (2017). An Empirical Investigation of Technology Acceptance Model of using C2C Mobile Business Application in Oman. In 2017 International Conference on Innovations in Control, Communication and Information Systems (ICICCI) (pp. 1-8). IEEE.

Rastogi, A. (2010). A Model based Approach to Implement Cloud Computing in E-Governance. *International Journal of Computer Applications*. https://doi.org/10.5120/1399-1888

Schuppan, Tino. (2010). E-Government Competencies: Looking Beyond Technology, *Handbook of Public Information Systems*, 3rd ed. Boca Raton: Taylor & Francis, 353-370.

Sharma, S. K., Al-Badi, A., Rana, N. P., & Al-Azizi, L. (2018). Mobile applications in government services (mG-App) from user's perspectives: A predictive modelling approach. *Government Information Quarterly*.https://doi.org/10.1016/j.giq.2018.07.002

Sharma, S. K., Gaur, A., Saddikuti, V., & Rastogi, A. (2017). Structural equation model (SEM)-neural network (NN) model for predicting quality determinants of e-learning management systems. *Behaviour and Information Technology*. https://doi.org/10.1080/0144929X.2017.1340973.

Susanto, T. D., &Aljoza, M. (2015). Individual acceptance of e-Government services in a developing country: Dimensions of perceived usefulness and perceived ease of use and the importance of trust and social influence, *Procedia Computer Science*, 72, 622-629.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models, *Information systems research*, 6(2), 144-176.

Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies, *Management science*, 46(2),186-204.

Warkentin, M., Gefen, D., Pavlou, P. A., & Rose, G. M. (2002). Encouraging citizen adoption of e-government by building trust, *Electronic markets*, 12(3), 157-162.

Wang, Y. S., & Liao, Y. W. (2008). Assessing eGovernment systems success: A validation of the DeLone and McLean model of information systems success, *Government Information Quarterly*, 25(4), 717-733.

Wu, L., & Chen, J. L. (2005). An extension of trust and TAM model with TPB in the initial adoption of on-line tax: an empirical study, *International Journal of Human-Computer Studies*, 62(6), 784-808.

CALL FOR PAPER

International Journal of Decision Science and Information Technology (IJDSIT)

International Journal of Decision Science and Information Technology (IJDSIT) is having ISSN No.1973-9013 (Print) and published yearly by the Department of Computer Science and Application, Atal Bihari Vajpayee University, Bilaspur (C.G.), India in association with Modern Technology and Management Institute, Inc. (MTMI) USA,invites submissions of high quality research papers in all areas of information technology and Decision Science for its next volume. IJDSIT is an academic and peer reviewed international journal and aims to publish original, theoretical and practical advances in the field of decision science and Information technology. Submitted papers should not be under review with any other journal and should follow the manuscript format requirements listed for papers at www.mtmi.us and www.bilaspuruniversity.ac.in/ijdsitjournal.

IJDSIT covers all topics of Decision Science and Information Technology. The main topic includes (But not limited to) are as below:

- Internet of Things (IoT).
- Big data and Data analytics.
- Cloud Computing.
- Software Engineering.
- Soft computing.
- Mobile Computing.
- Data Mining
- Blockchain Technology
- Bioinformatics
- Management Information System
- Nano Technology
- Robotics
- Artificial Intelligence
- Cryptography and Network Security
- Pattern recognition
- Machine Learning (ML) and Knowledge Mining (KM).
- Advanced Computer networking
- Computational Intelligence.
- Computer Networks and Inventive Communication Technologies.
- Decision Science
- Business Intelligence
- Cyber Security
- Mobile Adhoc Network
- Web Mining
- Signal and Image processing
- Information Security
- Human Computer Interaction
- Algorithms and Complexity.
- Software Engineering and Knowledge Engineering.
- Management and Information science.
- Smart Computing and Information Processing.
- Computational Intelligence and Analytics.
- Information Communication Technology (ICT)
- E-Commerce and M-Commerce
- Smart computing
- Wireless Sensor Network (WSN)
- Bio Medical Computing
- E-Governance
- Natural Language processing

Paper submission open for: Vol. : 09, January, 2020
Paper Submission Deadline: 15th November , 2019

Website: www.bilaspuruniversity.ac.in /ijdsitjournal and www.mtmi.us
E-mail for paper submission : ijdsitjournal@gmail.com