CSI Journal of COMPUTING

e-ISSN 2277-7091
Vol. No. 3, No. 2 | April 2020 - June 2020

CSI Executive Committee
(2019-20/21)

President
Mr. R K Vyas
(E) rkvyas05@gmail.com

Vice President cum President Elect
Dr. Anil Kumar Saini
(E) aksaini1960@gmail.com

Hon. Secretary
Dr. Vipin Tyagi
(E) dr.vipin.tyagi@gmail.com

Hon. Treasurer
Dr. Durgesh Mishra
(E) drdurgeshmishra@gmail.com

Immd. Past President
Prof. A K Nayak
(E) aknayak@iibm.in

REGIONAL VICE-PRESIDENTS

Region-I (2019-21)
Mr. Arvind Sharma
(E) arvindsha@hotmail.com

Region-II (2020-22)
Md. Shams Raza
(E) s_raza2000@yahoo.com

Region-III (2019-21)
Mr. Jayant Bhide
(E) jayantbhide2000@gmail.com

Region-IV (2020-22)
Mr. Mukesh Kumar
(E) Mukesh.kumar@sailcet.co.in

Region-V (2019-21)
Mr. M. S. Prasadbabu
(E) profmspbabu@gmail.com

Region-VI (2020-22)
Mr. Pradeep Rathi
(E) pjrathi61@gmail.com

Region-VII (2019-21)
Prof. Dr. M Sundaresan
(E) bu.sundaresan@gmail.com

DIVISION CHAIRPERSONS

Vacant

Division-I (2019-21)

Division-II (2020-22)
Dr. P. Sakthivel
(E) sakthi_velp@yahoo.com

Division-III (2019-21)
Prof. Suresh Chand Tyagi
(E) sct_35_2000@yahoo.com
sctyagi1963@gmail.com

Division-IV (2020-22)
Dr. B. Chidhambararajan
(E) dbcrajan@gmail.com

Division-V (2019-21)
Dr. Subhash Chandra Yadav
(E) scy123@gmail.com

NOMINATION COMMITTEE (2020-2021)

Mr. Anand Rao
(E) nanandrao@yahoo.com

Mr. Anil Ji Garg
(E) aniljigarg@gmail.com

Dr. Sharvari C Tamane
(E) hodit@jnec.ac.in,
sharvaree73@yahoo.com

Editorial Board

Journal Management Committee

Editorial Board Members

CONTENTS

ARTICLE	TITLE	AUTHOR	Page No.
	Preface		3
R1	Design of A Modern Museum Space	*Iftikhar B. Abbasov and Olga M. Kozak*	4
R2	A New Approach for Vehicle Authentication using Smart Card System	*M Y V Nagesh*	8
R3	Estimation of Soil Attributes using Spectroscopy	*Chitra M. Gaikwad, Sangeeta N. Kakarwal, Shubhangi Sapkal and Nilesh Uke*	12
R4	A Far down Illustration on Virtualization and its Blend with Cloud Computing Technology	*M. Senthil Kumar, B. Chidambara Rajan, P. Kiruthika and M. Rajakumar*	17
R5	A Novel Approach for Meat Adulteration Detection using the concepts of Image Processing and Machine Learning	*Prabha. B., Anjana M. P. and Pradeepa K.*	25
R6	Automatic Speech Recognition system in Marathi for Cerebral Palsy Disabled	*Maheshwari A. Ambewadikar and Mrs. Manasi R. Baheti*	32
R7	Multi-Dimensional Encryption Type 11 method for Video Files Security	*Boopathy D., M Sundaresan*	37
R8	Distinguishing Healthy and Infected Vegetable Crops using Hyperspectral Leaf Reflectance	*Anjana Ghule, Prof. Dr. R. R. Deshmukh and Chitra Gaikwad*	42
R9	Ensemble Learning approach for Named Entity Recognition from Hindi-English Tweets	*Suparna Arya, Amit Majumder, Sulabh Majumder, Aparajita kundu, Nuzhat Shamim, Pranati Rakshit*	47

PREFACE

This is the second issue of Volume 3. Earlier first issue of Volume 3 was published in January 2020.

The first paper of this issue entitled Design of A Modern Museum Space, is an invited paper by Professor Iftikhar B. Abbasov, Southern Federal University, Russia, highlights the work to develop a new design concept for museum space using modern exhibition multimedia technologies. Second paper entitled A New Approach for Vehicle Authentication using Smart Card System highlights use of smart card system for vehicle authentication to avoid real identity in the process of authentication can easily result in a leak of the privacy information of the vehicles. Third paper entitled Estimation of Soil Attributes using Spectroscopy highlights to find a technique to predict various nutrients in the soil, forth paper entitled A Far down Illustration on Virtualization and its Blend with Cloud Computing Technology highlights Virtualization is an additional software layer in a system that transforms the traditional system into virtual system, fifth paper entitled A Novel Approach for Meat Adulteration Detection using the concepts of Image Processing and Machine Learning highlights application whether the meat is adulterated by mixing with other red meat or not by using machine learning and image processing techniques, Sixth paper entitled Automatic Speech Recognition system in Marathi for Cerebral Palsy Disabled highlights techniques for speech recognition of Marathi digits and words of Cerebral Palsy (CP) Disabled person, Seventh paper deals with Multi-dimensional Encryption type 11 Method for Video Files Security, eighth paper Distinguishing Healthy and Infected Vegetable Crops using Hyperspectral Leaf Reflectance highlights to reduce the dimensionality and distinguish healthy and infected vegetable plants by selecting significant spectral region and subset of wavelengths and the last paper presents Ensemble Learning approach for Named Entity Recognition from Hindi-English Tweets.

My sincere gratitude to the members of CSI ExecCom for providing me the opportunity to serve as Editor-in-Chief for the CSI Journals of Computing. I received enormous support from immediate past and present office Bearers of CSI India to process this volume. Thanks to the President Prof. R. K. Vyas, Immediate Past President Prof. A. K. Nayak, Hony. Secretary Prof. Vipin Tyagi, Dr. D. D. Sharma, Chairman Publication Committee and other members.

This is the Second issue, I am editing this Journal. I hereby place record my sincere gratitude to the members of the Editorial Board from the country and abroad and the reviewers who have spend their valuable time for reviewing the papers submitted to the Journal and also provide constructive feedback to the authors.

The success of any Journal depends contribution from authors. I thank the authors and wish the Journal to attract more research papers and establish its impact in the computing field of research. I am thankful to my scholars of the MSL and HCI research laboratories, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) for their support during entire processing.

I hope this issue will be helpful to the researchers of various emerging areas.

Professor (Dr.) Ratnadeep R. Deshmukh

Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS)

Editor-in-Chief, CSI Journal of Computing

June 2020

Design of A Modern Museum Space

Iftikhar B. Abbasov[1] and Olga M. Kozak[2]

1 Prof. of Dept. of Engg. Graphics and Computer Design of Taganrog Technological Inst. of Southern Federal University
2 Master, Southern Federal University, Engineering-Technological Academy, Dept. of Engineering Graphics and Computer Design, Russia

The aim of this work is to develop a new design concept for museum space using modern exhibition technologies. The new design concept of the museum space will increase the interest of visitors to museums. For the development of criteria were compiled, a classification of modern museum spaces. The expositional premises of the museum are modeled, a concept using modern multimedia technologies is proposed. The proposed project will increase the tourist attractiveness of regional museums.

Keywords: Design concept, museum, exposition, multimedia technologies, museum space, renovation, zoning of premises, interior modeling.

1. Introduction

In recent years, the field of exhibition activity of museums and exhibition spaces has been actively developing. Modern society is experiencing an era of computerization, so there is a special need for communication, social communication for each member of the society [1]. As a result of technological progress, the exposition space with the use of multimedia technologies is becoming more familiar to museums and their visitors, many museums go into virtual space. This entails the need for a constant search for new methods of attracting the audience to the museum exposition.

This paper discusses the development of a design concept for a museum exhibition using multimedia technologies to enhance the educational function of museums and increase the flow of visitors. To achieve this goal, it is necessary to analyze modern museum expositions and propose a new concept of museum space using multimedia technologies. Therefore, we will consider a brief analysis of the state of museum expositions, existing analogues, and move on to the issues of creating a new concept.

2. Modern museum spaces

In the modern museum classification system, there are a huge variety of different options related to both traditional museology and modern technology [2,3]. Three main types of museum are distinguished, requiring different methods for selecting and displaying material [4,5]:

- Scientific - for specialists, represent a laboratory for research, or an archive of studies already conducted;
- Educational - for students, according to the requirements of the educational process, they must contain information for the assimilation of educational material;
- Public - for the mass visitor, everything in the museum should be aimed at satisfying his cultural needs.

Museum expositions should carry out the educational functions necessary for society [6,7,8]. By profile, museums can be classified into three levels:

- historical museums;
- natural science museums;
- art museums;
- museums of a complex profile.

In exposition museum practice there are three types:

- thematic exhibitions, which are based on a certain plot;
- stock exhibitions that introduce visitors to little-known and inaccessible collections;
- reporting exhibitions are created according to the results of restoration work, following the results of manning funds, new arrivals.

3. Museum space prototype analysis

When designing a museum space using modern technologies, it is worth considering the already existing museums that have successfully applied this technique in their practice. Today there are not so many, but exploring examples of domestic and foreign experience, it is worth highlighting such a museum as [9,10]:

- The archaeological museum of MARQ in the city of Alicante, Spain;
- Miraikan Museum of Science and Innovation in Tokyo;
- Jewish Museum and Tolerance Center, Moscow;
- historical multimedia park "Russia - My History", Ekaterinburg, Russia.

Raising the issue of studying expositions in museum space, one can rely on one's own experience and analysis of museums visited, as this gives a more complete picture of the material being studied. The Archaeological Museum in the city of Alicante was opened in 1932 in the Palace of Deputies, and later the museum was transferred to the building of the former San Juan Hospital. Along with historical exhibits, the updated museum also uses modern multimedia technologies. They allowed to make a visit to the museum interesting not only for people interested in archeology, but also attracted a new wider audience, including children. Getting historical information has become easier, more accessible and more interesting; the museum's fund has more than eighty thousand exhibits.

The Jewish Museum and the Center for Tolerance is located in an important architectural monument - in the building of the Bakhmetyevsky garage, which was designed by the first representatives of the Russian avant-garde K. Melnikov and V. Shukhov. At the entrance to the museum there is a feeling that it does not pass through the doors of the building, but through the portal to the future. With the help of modern computer and audiovisual technologies, information becomes voluminous, ubiquitous, affecting literally all senses. Architectural structures complement the established exposure, enhancing the visitor's involvement in the atmosphere with unexpected technological methods, static elements come to life when touched. The museum's exposition is not based on artifacts, but on information presented in an interactive mode, this allows the visitor to immerse themselves in the research process. Twelve thematic pavilions are equipped with panoramic cinemas and interactive screens. Audiovisual services are based on rare photographs, videos, interviews and documents.

4. The choice of the concept of forming a museum space

Approaching the design of the exposition space using multimedia technologies, several conceptual options for its creation were developed [11,12,13]. A more elaborated and fundamental concept was chosen, it can be chosen both for renovation within the existing building, as well as for the new museum building. In our case, within the framework of existing museum spaces, it is necessary to create and place a multimedia exhibition complex in them. The main task will be to ensure that multimedia elements are aimed at revealing the contents of the exposition and fit into the existing exposition space.

When designing the design concept, the thematic method of organizing the museum exposition was chosen, since it involves the disclosure of the concept by using additional methods of presenting information [14,15]. This, in fact, is the main task of applying modern technology in the exhibition space of the museum. Interactive is understood to be an exposition that involves the visitor in a dialogue with the exhibition [16.17]. Particularly effective in this regard are multimedia computer systems with touch screens.

The concept involves the introduction of a historical profile in museums, since a significant part of the information stored in these museums cannot be fully displayed in the hall (archives, letters, maps, etc.). Any spatial solution begins with zoning and planning [18,19]. Choosing a conditional room of a rectangular shape, we offer a symmetric division of space into several main zones: entrance, permanent exhibition, temporary exhibition, buffet, library, exit.

In Fig. 1 presents a plan of the museum premises, the numbers indicate the names of zones and expositions:

1. administrator;
2. virtual room;
3. temporary exposure;
4. permanent exposure;
5. cinema;
6. temporary exposure;
7. library;
8. cafeteria.

Often, when a visitor comes to a museum, he gets lost and already at the entrance to the exposition space, the sequential way of moving and exploring the exhibitions becomes obscure. The numbering of zones presented on the plan of the museum is also present in real space. It is expressed by large luminous numbers, and plays the role of hints that say in which direction to move. This technique helps visitors to the museum to consistently absorb information.

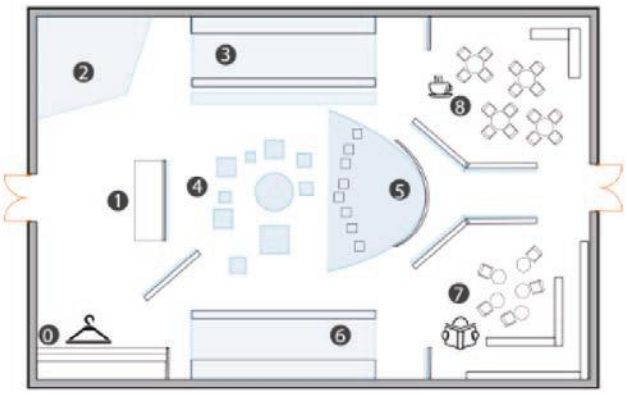


Fig. 1: Floor plan of the museum with sequential numbering of zones

In general, the entire museum space is formed around a permanent exhibition located in the center. This location helps the visitor understand the main informational content of the museum's activities by its largest exposition. When designing this space, we suggest using the following technologies [2,17]:

- interactive showcases;

- video mapping;
- touch table;
- touch kiosks;
- touch screens;
- It is supposed to use augmented reality glasses AR-guides.

In the center of the permanent exhibition is a large touch table. Its use can be diverse: from the image of geographical maps to the study of history. The main task of using this device is to disclose information about the history and exhibits presented in the museum or its funds. Also, the permanent exhibition is complemented by interactive showcases, touch kiosks and touch screens. Using this equipment gives the visitor the opportunity to choose the study of information of interest to him. However, do not forget that the main semantic load is still in the objects exhibited. The introduction of multimedia technologies only complement it and increase the information potential.

Further in the permanent exhibition hall is a kind of cinema with a large projection screen and places for viewing video sequences. Using video mapping makes it possible to adjust the screen size to a wall of any size, which allows you to conduct training lessons for an audience that is able to accommodate the space of the museum. The size of this zone depends on the initial premises of the museum, in which it is planned to carry out technological renovation of the exhibition spaces. It is worth noting that 3D-mapping allows you to demonstrate exhibits that for some reason cannot be demonstrated realistically, or to consider them in detail (structure of the bone, weapons, tools, etc.).

To remind visitors of the main museum objects - exhibits, it was decided to dilute the digital load of the museum with a real object - a sculpture of a bull [6,12]. The image of the bull is present in the history of many cultures, from rock paintings in the Paleolithic era and bull temples in Mesopotamia, to modern ones - bullfighting museums in Spain. Thanks to these animals, agriculture was mastered, which is of great importance for man and his history in general. Figure 2 shows the visualization of the entrance area of the developed museum concept. The concept was modeled in three-dimensional form based on the created floor plan [20].

Fig. 2. Museum entrance rendering

Our concept involves the use of a virtual room with projection display cases and audio-visual technologies. This exhibition space contains information about the museum itself, its building, creators, etc. The content of the exhibition is completely dependent on the museum information already available. The concept of museum space is completed by an exit zone with a library and a buffet adjacent to it.

Since the use of modern technologies in the museum's exposition is aimed at increasing the educational function of the museum by increasing the information content of the expositions, in the concept we offer, everything is aimed at solving this problem. The museum mostly lacks bright daylight. On the contrary, the concept of the museum is such that multimedia and other modern exhibition technologies are placed in a dark space. All technologies with their own glow from screens or projectors are better perceived in a dark room. The guides are LED spotlights located along temporary exposures. On the floor and on the ceiling, they create a kind of corridor leading visitors to the museum along the exhibitions. This technique is not only a natural guide, but also a design decision, reflecting the information load of the museum of historical profile.

In the center of the museum space, above the permanent exhibition is a large chandelier. She acts as a central emphasis and collects the entire composition of the design solution together (Fig. 3). Speaking about the color scheme of the museum space, it was decided to use mainly dark colors in the decoration of the room in order to emphasize the light flux emanating from the multimedia equipment of the expositions. Working on the contrast of light and dark planes, it creates the opportunity to immerse the visitor in the mysterious environment of the museum, which he has not yet explored. Small bench cubes (Fig. 3), designed to relax or view the footage, have a reflective surface, which visually dissolves them in the space of the entire room [21].

Fig. 3. Permanent exhibition hall

Wood trim is also visible in temporary exhibition areas. This solution allows you to dilute the dark environment, and

the shade of the tree supports the warm lighting coming from the LED illumination of the museum's travel routes, which in turn form a memorable associative image.

5. Conclusion

Currently, the search continues for new forms and methods of demonstrating museum objects, techniques and methods of attracting visitors. As a result, an analysis of museum exhibits using modern technologies was carried out, a design concept for a museum space with multimedia equipment was developed. Thus, the created conceptual model of museum space formation will increase the attractiveness of modern museums and increase the audience of visitors.

References

1] Bogolyubova N.M., Nikolaev Yu.V. Intercultural communication and international cultural exchange.-SPb.: SPbKO, 2009.-415 p.

2] Smirnova T.A. Digital technologies in the space of a museum exposition: technical, conceptual and communication capabilities // Culture: management, economics, law. - 2010. - No. 1. - P.16-19; URL:http://rosexpertpravo.ru/literature/tsifrovye-tekhnologii-v-prostranstve-muzeynoy-ekspozitsii-tekhnicheskie-kontseptua (accessed February 25, 2019).

3] Boroneva T.A. Art Museum and Modern Information Technologies // Digital Libraries: Prospects for Development: Materials from the Electronic Libraries Section of Baikal Inform. Forum, June 24-26, 2009, Ulan-Ude. - Ulan-Ude, 2009. - P.61-65.

4] Schmit F.I. The collection "Museums: an international study on the reform of public galleries" - M: 1931, 309 p.

5] Krasnokutskaya L.I. Museology. Teaching aid. Pyatigorsk, 2011.54 p.

6] Sotnikova S. I. Museology: a textbook for universities. M.: Bustard, 2010.381 p.

7] Mikhailovskaya A. I. Museum exposition. M., 1964.292 p.

8] Bonami Z. A. Literary Museum and Society // Museology. Museums of the world: Sat scientific Proceedings / Research Institute of Culture. M., 1991.64p.

9] Smirnova T.A. Museum in the 21st century: the integration of digital technology in the exposition space [Electronic resource] // Inform. about: magazine site. - 1999-2014.- URL: http://emag.iis.ru/arc/infosoc/emag.nsf/BPA/fb59b993f07873d144257a79004c0818 (accessed February 19, 2019).

10] Soshenina T.V. Interactive space in modern exhibition complexes // Architecture and Modern Information Technologies. - 2008. - No. 2 (3). - P.1-7.

11] Panchenko A.S. The exhibition as a form of interactive artistic communication ("Inter-time", loft project Floors) // Socio-humanitarian. Vestn. South of Russia. - 2010. - No. 8. - P.27-31.

12] Makhlina S. T. The semiotic aspect of the museum exposition // Museum in modern culture: collection of books. scientific Proceedings / St. Petersburg. state Academy of Culture. SPb., 1997. P.45-62.

13] Litvinov V.V. The practice of modern exposure. - M.: Poster, 1989.274 p.

14] Bezrukova E.A. The role of the designer of the exposition in the modern practice of museum activity // Tomsk State University Journal. Kemerovo state. University of Culture and Arts. - 2017. - No. 40. - P.75-79.

15] Veselitsky OV Features of the artistic design of museum displays in Leningrad in the 1960-1980s. Combine of painting and decorating art [Electronic resource] // Society. Wednesday. Development (Terra Humana). - 2010. - No. 3. - URL: https://cyberleninka.ru/article/n/osobennosti-hudozhestvennogo-proektirovaniya-muzeynyh-ekspozitsiy-v-leningrade-v-196 (accessed 12/20/2018).

16] Makarov D.V. The introduction of technologies and new techniques in the culture of the modern museum / D.V. Makarov, O.V. Shutova // Modern. problems of science and education. - 2015. - No. 2. - P.783-783

17] Glytina A.A. The introduction of multimedia technologies in the environment of museum and exhibition spaces. Conflicts, decision concepts, development prospects // New ideas of the new century: materials of the Intern. scientific conf. FAD TOU. - Khabarovsk, 2017 .- V. 3 .- P.20-26

18] Efimov A.V., Minervin G.B., Ermolaev A.P. Design of Architectural Environment. - M.: Architecture - S., 2004.- 304 p.

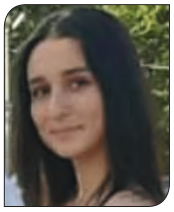
19] Shimko V.T. Basics of design and environmental design. Tutorial. M .: Publishing House: Architecture-S, 2005 .-- 270 p.

20] Abbasov I.B. The basics of three-dimensional modeling in the graphics system 3 ds Max 2018. Tutorial. Approved by UMO in the field of design, monumental and decorative arts. M .: DMK Press. 2017.186s. (third edition) ISBN: 978-5-97060-516-5 https://dmkpress.com/catalog/computer/graphics/978-5-97060-516-5/

21] Kozak O.M. The formation of cultural space in the context of the museum. Collection of articles of the VII International Scientific and Practical Conference "Priority Directions for the Development of Science and Education": - Penza: ICSN "Science and Education" - 2019. at 2 parts. Part 1, 248 p., P.187-189 https://elibrary.ru/item.asp?id=38180241

About the Authors

Iftikhar B. Abbasov in 1988 he received the diploma of engineer in the Taganrog radioengineering institute (Russia). In 1997 has protected scientific degree of the candidate of physic-mathematical sciences (degree PhD), in 2012 Doctor of Science (Prof). He is Professor of department of engineering graphics and computer design of Taganrog Technological Institute of Southern Federal University. His teaching and research interests are in the areas of computer-aided design and modelling.



Olga M. Kozak Master, Southern Federal University, Engineering-Technological Academy, Department of Engineering Graphics and Computer Design received a bachelor's degree in design from the Southern Federal University Academy of Architecture and Arts in 2017, in 2019, she also received a master's degree in design.

A New Approach for Vehicle Authentication using Smart Card System

M Y V Nagesh

CSE Dept., Raghu Engineering College, Dakamarri, Visakhapatnam. A.P. Email: nagesh.myv@raghuenggcollege.in

Vehicle checking is done manually by examining the associated documents to verify the authenticity. The paper documents are however prone to physical damage and probably may loss the print due to folding or other issues. There are a set of documents for insurance, registration, pollution etc., and checking frequency will become less. To address the above issue the existing paper document details are attached to smart card technology which is already been used for other daily routines. The use of smart card technology enables the user to have the flexibility of carrying it and to complain it online, if lost to have duplicated making things easier for the stakeholders of verification system.

The proposed smart card system will update the vehicle transactions with in no time making it readily available for the authentication of the vehicle. This paper proposes a framework for the smart vehicle authentication system and the performance of it can be anticipated to be efficient as the smart card usage is not a new concept in this present digital era.

Keywords: Smart Card, Vehicle Authentication, Membership Functions, Non linear System

1. Introduction

Now-a-days as per increased dependency on automobile industry leads to need for providing more security services simultaneously without any time lapse. The need for checking is increased with the increase in vehicle thefts making it a tedious job for both the owners and verifiers. This sometimes may lead to manual error in overlooking the documents and also leaving the vehicles without inspecting them. The users even though legitimate may have damaged documents due to various physical damages unnecessarily creating suspicion for the inspection team.

By implementing the Vehicle Authentication – Be it a bike, car or any other heavy duty vehicle, we need not have to carry all papers and instead can have only one card which replaces all the following documents:

- Driving License (optional)
- Vehicle registration certificate (RC book)
- Vehicle Insurance
- Pollution certificate
- Transfer Certificate (In case of purchasing vehicle from another person)

In case we fail to carry any one of the above documents we will be imposed with a fine or will be strictly punished basing on the type of document. This smart card system is like any other card has a chip embedded and is portable to carry and is mostly resistant to physical damages. The smart card can also have magnetic stripe but it is less tolerant to damages in comparison with the embedded chip. The card processing devices and servers are shown in below Fig. 1.

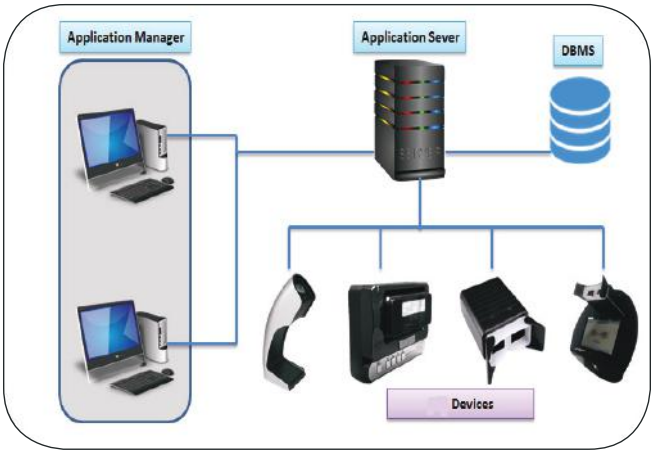


Fig. 1 : Card processing through devices

The card should be supported with the pin number to physical identity the credentials of the user otherwise the entire concept of the system fails. The vehicle if it is shared with other person then the pin details should be delivered to

prove this delegation. The card can have multiple copies if the vehicle is shared by multiple persons but should have only one pin number for identifying its credibility.

This paper organized as follows, Section 1 presents Smart Card history and developments, and Section 2 describes the design of smart card. Section 3 describes the steps involved in process of making results are shown in section 5 and is concludes it with section 6.

II. History

Smart Card History:

The origin of smart card technology have been initiated at the late seventies Michel Ugon (Guillou, 1992) [3] which was then tested for its applicability by the French group of bank cards, in 1985 the CB (Carte Bancaire) was created and has diffused 24 million devices in to the market. The first draft proposal was registered in 1983 for the physical characteristics and resulted in the standardization of the contact location after many discussions. The ISO/IEC 7816/1-4 standards have proposed the signals and protocols along with the cryptographic capabilities nevertheless of the insufficient RAM and processing power[2].

The usage of smart cards have been increased in recent years due to the wide spread of smart devices and online transactions. Almost everywhere the card technology is ranging from banking transactions [3] to children play arena. The proposed system implementation can be witnessed in nearby future.

III. Card Construction

The card have a secret information embedded on the chip or the magnetic stripe basing on the card type which can hold up to 32,000 bytes of information [7] also known as "e-wallet" which contains the valuables in electronic form.

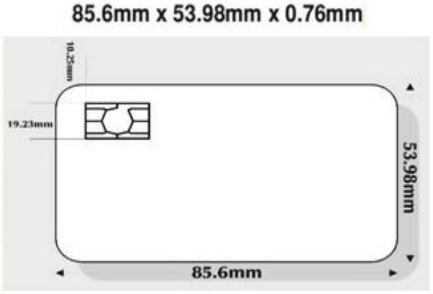


Fig 2 Card Constrution

Card Elements

Fig 3 Card Elements

III. Smart card authentication

Smart card authentication equipment is the smart card and the corresponding pin for the card. The smart card shape is like a playing card and if any one provides it with the correct pin then the authentication is accepted and then the output is checked to take the decision of either leaving the vehicle or seizing it, if the card misses some necessary details like registration or presents the details of the other vehicle instead of the vehicle being verified.

The card readers are based on the card types which may have the reader plate, which makes the contact of the card by couple of electrical connectors to read the contents and accepts the pin once card is verified to authenticate the user. After successful authentication the details are fetched from the database and are processed as per the operational guidelines. If the card is contact less then process of verification and authentication differs where the wireless communication is made to serve the purpose. The readers can be of two types one which is having the contact pad and the other which does not have a contact pad but have a communication like RF[1]. A card which is having both contact reader options will use reader option for huge data transaction and a contact less option for the mutual authentication purpose.

IV. Contact Smart cards

The implementation of the proposed smart card system is done by inserting the smart card into the reader device connected to the internet.

When the card is inserted the chip on the card is read and the corresponding details are fetched from the remote server for the verification process.

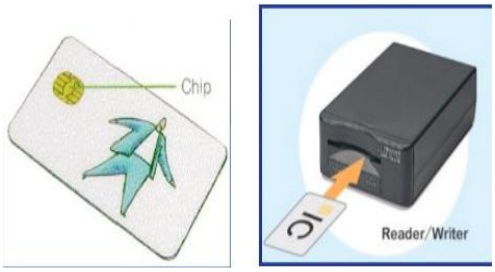


Fig. 4 : Contact smart card

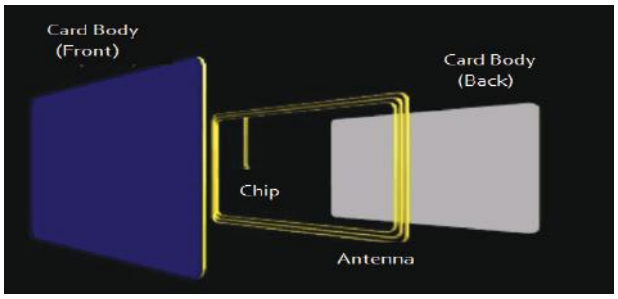


Fig. 5 : Contact less smart card

Smart cards are classified based on their components, card interfaces and card Operating systems. The classification of cards is done basing on their components on their interface like the magnetic strip, embedded chips. The cost of the card chip is more and thus the flexibility of it to use also more convenient, the magnetic strip may wore out if it is inserted more no of times requiring more replacements. The chip card on the other hand has an inbuilt processor, ROM and memory embedded on it [1]. The card technology is also been updating with the innovation of hardware devices and evolvement of new software.

As we can see from the above diagram, we start off with smart card sensitive user applications which understand smart cards. Once that is done, we move to the subsystem. Here, we either shown a user-interface or make a connection with smart card service providers. After this is done, the control moves to the smart card resource manager who ultimately analyzes and decides which resource is deemed fit to be used.

Next, we see drivers for different kinds of authentication like insurance, license etc. Finally, the processed data is passed into each different reader based on its purpose to be read and made into a proper card.

Verification Criteria:

The verification is divided into number of stages, which is being carried with the help of colored lights like traffic light system. The verification criteria can be explained with the following table.

Criteria	Colour	Document not verified
1	GREEN	All documents verified
2	YELLOW	'C' Book
3	RED	Insurance

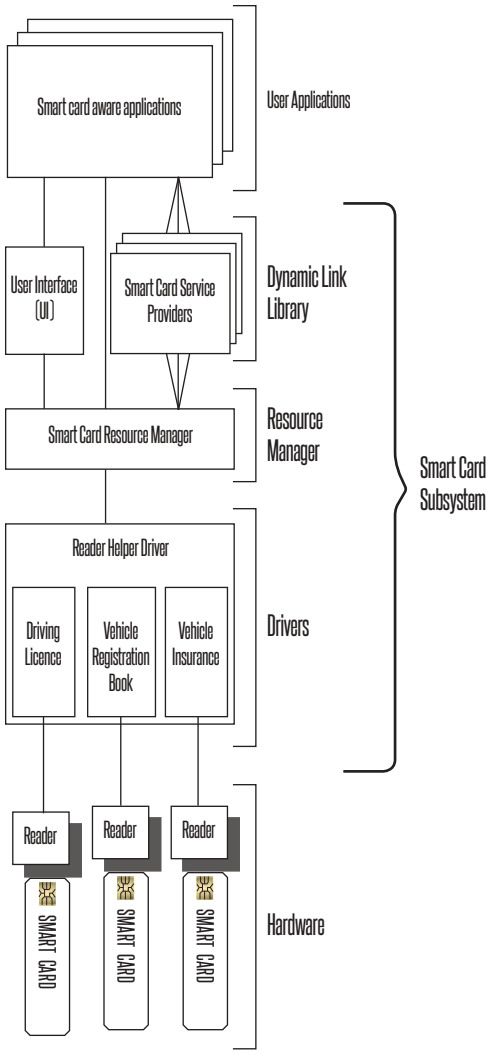


Fig. 6: Vehicle checking processing system

Criteria 1:

It will highlight the GREEN colored LED if the vehicle owner meets all the eligibility criteria.

Criteria 2:

It will highlight the YELLOW colored LED if 'C' book is not there.

Criteria 3:

It will highlight the RED colored LED if at all the vehicle is not insured.

The number of criteria increases with number of documents updated and thus the number of lights.

Conclusion

The inspection of the vehicle details done manually by verifying the paper documents is time consuming and has the disadvantage of damaging them if frequently been verified.

This paper has proposed a smart card solution to the above said problem which will make the checking faster and is resistant to the physical damage to certain extent. A single smart card replaces multiple paper documents and the technological advancement in smart card generates remainders for the user about renewal of insurance, pollution and so on. This system reduces manpower and reduces the verification time and increase the number of vehicle verifications.

The main drawback in this system is that the card reader should be always connected to online and poor connection to internet or any damage to the card reader may halt the entire process. But, the availability of smart devices and portable internet devices will make this to work efficient and a backup reader will never interrupt the system. The driving license attachment is left as optional if the vehicle is shared by many. The card can also have a bar code scanner to facilitate a faster verification but should be assisted with physical identity card.

References

1] A Survey on RFID Based Vehicle Authentication Using A Smart Card, Litty Rajan, Alpana Gopi, Divya P R, Surya Rajan International Journal of Computer Engineering In Research Trends International Journal of Computer Engineering In Research Trends, Volume 4, Issue 3, March-2017, pp. 106-110

2] Security Authentication System for In-Vehicle Network, Hiroshi UEDA*, Ryo Kurachi*, Hiroaki TAKADA, Tomohiro Mizutani, Masayuki INOUE and Satoshi Horihata, SEI Technical Review · Number 81 · OCTOBER 2015.

3] Vehicular Identification and Authentication System using Zigbee, Mr. Somnath A. Karmude, Prof. G. R. Gidveer, International Journal of Engineering Research & Technology), International Journal of Engineering Research & Technology (IJERT) Vol. 3 Issue 11, November-201

4] E-RTO Management System and Vehicle Authentication Using RFID, Alpana Gopi, Litty Rajan, Divya P R, Surya Rajan, International Research Journal of Engineering and Technology.

5] Automobile Driver Authentication system using Electronic Driving License, Makwana Foram R.1, Rahul R. Menon2, Patel Harsh K.3, International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 2, February 2016

6] Security in Automotive Networks: Lightweight Authentication and Authorization Philipp Mundhenk, Andrew Paverd, Artur Mrowca, arXiv:1703.03652v2 [cs.CR] 15 may 2017

About the Authors

M Y V Nagesh [CSI-F8003362] is presently working as Assistant Professor in Department of Computer Science & Engineering of Raghu Engineering College, Visakhapatnam, Andhra Pradesh.

Mail: nagesh.myv@raghuenggcollege.in

Estimation of Soil Attributes using Spectroscopy

Chitra M. Gaikwad[1], Sangeeta N. Kakarwal[2], Shubhangi Sapkal[3] and Nilesh Uke[4]

[1] Government College of Engineering, Aurangabad, India. Email: gaikwadchitra@gmail.com
[2] PES Engineering College, Aurangabad, India. Email: s_kakarwal@yahoo.com
[3] Government College of Engineering, Aurangabad, India. Email: shubhangisapkal24@gmail.com
[4] Trinity Academy of Engineering, Pune, India. Email: nilesh.uke@gmail.com

Soil health is very important for the growth of plants. If a farmer gets the information about the health of a soil, he can correctly identify which nutrients should be added to the soil for a good yield. This paper is an attempt to find a technique to predict various nutrients in the soil In this paper we are identifying the optimal wavelength ranges in electromagnetic spectrum sensitive to soil organic matter(SOM), pH and soil nitrogen(N). The soil reflectance (R) values were measured between 350-2500 nm wavelength in the electromagnetic range. The analysis is performed for soil contents such as SOM, N and pH. Three different pre processing techniques like first derivative R, first derivative log R and first derivative 1/log R were used. Comparison of partial least squares regression (PLSR) on preprocessed transformations of spectral data was carried out to predict the SOM, pH and N. PLSR model was applied to visible-near-infra-red (VisNIR) (400-1300 nm), near-infra red (NIR) (700-1300 nm) and mid-infra-red (midIR) (1300-2500 nm) region in electromagnetic spectrum. Partial least square regression analysis was carried out to find the sensitive wavelength range for SOM, N and pH. Comparison shows that the first derivative of reflectance, first derivative of log R and first derivative of reciprocal of log R has coefficient of determination as 0.99. The prediction accuracy with first derivative R, first derivative log R and first derivative 1/ log R is increased.

The predicted results demonstrate that soil nutrients can be identified by using spectroscopy in the VisNIR, NIR and midIR. SOM is effectively predicted in VisNIR, N in NIR and pH in Vis NIR and NIR.

Key Words: Soil reflectance, ASD Field spec4 spectroradiometer, Soil Organic Matter, Partial Least Square Regression (PLSR)

1. Introduction

Precision agriculture is the need of this age. Prediction of good yield in the farms requires the knowledge of soil quality. Traditional methods to test soil take more time and they are costly. Modern technique such as spectroscopy helps in soil quality assessment and provides environment friendly, cost effective and timely results. To determine soil contents using spectroscopy visible near infrared spectrum 350 -2500 nm is utilized. Soil organic matter (SOM), soil pH(pH) and nitrogen(N) are important properties of fertile soil.

The spectral reflectance of soil data consists of the information about various contents present in soil. The reflectance from soil surface relates to various soil contents. Thus it is used to study the contents of soil. The soil constituents are unclear due to the overlapping nature of reflectance. Thus we have to extract the optimal wavelengths from the spectral signatures by correlating reflectance with actual soil contents [1]. Thus there arises a need to analyse the spectral signatures to determine different amount of soil components. As the reflectance spectra consist of redundant information it needs to select suitable wavelengths to improve prediction accuracy. Therefore a model has to be developed predicting the relationship between spectral variables and the response variable at optimal wavelengths. The correlation and partial least square regression technique can serve the purpose [2]. The original soil reflectance spectra needs to be pre processed with spectral pre-processing techniques such as derivative analysis of soil reflectance and soil absorbance. Derivative analysis improves the performance of prediction and positive effect on soil analysis [3]. According to the study the correlation results show that first derivative, first derivative of log(R), 1/log(R) first derivative give maximum correlation coefficient between measured values of soil and reflectance [4]. Regression analysis is used in most of the studies as a basis for linear relationship between variables. Amongst the various regression techniques partial least square regression can be one of the techniques to extract the required data from soil reflectance [5]. In this study the raw spectra was pre-processed using 6 methods and after that PLSR was applied. Correlation analysis was applied to find the considerable wavelengths for soil organic matter. After modelling it with

the help of PLSR, R^2 was 0.986 [6]. The results predicted in this study used for soil organic matter R^2 is 0.98, for pH R^2 is 0.95, and for Total Nitrogen R^2 is 0.98 concentration were predicted using partial least square regression [7].

Our aim in this study is to analyze VisNIR, MidIR spectroscopic reflectance, correlation and PLSR on soil spectral signatures for predicting various attributes of soil like pH, Soil organic matter and Nitrogen from agricultural farms of the state of Maharashtra, India.

2. Data Preparation and Techniques

2.1 Collection and Analysis of Soil Data

Soil was collected from the upper layer of soil from eight different locations Bhivandi, Chikalthana, Kasara, Igatpuri, Kumbhephal, Naregaon, Shendra, University area Aurangabad of Maharashtra state in India. These samples were of yellow-brown, brown-red, and paddy types.

The chemical analysis of soil was carried out at a Rallis India Limited laboratory at Akola, Maharashtra, India. Table 1 below shows the color and texture of soils collected from different locations. The chemical properties of soil which are pH, Soil organic matter and nitrogen content in each sample were measured in the laboratory.

Table 1: Soil Color and Texture

Location	Color/ Texture
University	Black colour coarse material
Chikalthana	Faint Brownish Black colour coarse material
Shendra	Black colour coarse material
Kumbhephal	Faint Brownish Black colour coarse material
Naregaon	Faint Brownish Black colour coarse material
Kasara	Yellowish Brown colour coarse material
Igatpuri	Brown colour coarse material
Bhivandi	Brown colour coarse material

2.2 Soil Spectra Collection

The collected soils were first dried in air. Using the mesh of 2mm size the soil samples were sieved and the unwanted particles were removed. The soil reflectance from each sample was recorded with the aid of ASD Field Spec 4 spectroradiometer in the dark laboratory. The wavelength range over which the readings were taken is 350-2500 nm. Before actually reading the soil reflectance ASD FieldSpec spectroradiometer was calibrated using the white Spectralon panel to get the absolute reflectance values. A fiber optic probe was used to record the spectral signatures with an angle of 10 degrees. Each soil sample was placed in a Petri dish of 2 cm thickness and 20 cm diameter. To get more accurate results each sample was scanned 20 times. After recording the reflectance values for a sample 20 times its mean was calculated. The .ASD file of these spectra was saved with the help of RS3 software. Using the Viewspec pro software these values were exported to spreadsheet for further processing.

3. Preprocessing Methods

The soil spectra were obtained for soil samples and they were transformed using first derivative R, first derivative log(R), first derivative 1/log(R). The preprocessing transformations are shown in table 2.

Table 2: Transformations on Reflectance

Sr. No.	Transform	Formula
1	First derivative R	R'
2	First derivative Logarithm R	log(R)
3	First derivative 1/Logarithm R	1/log(R)'

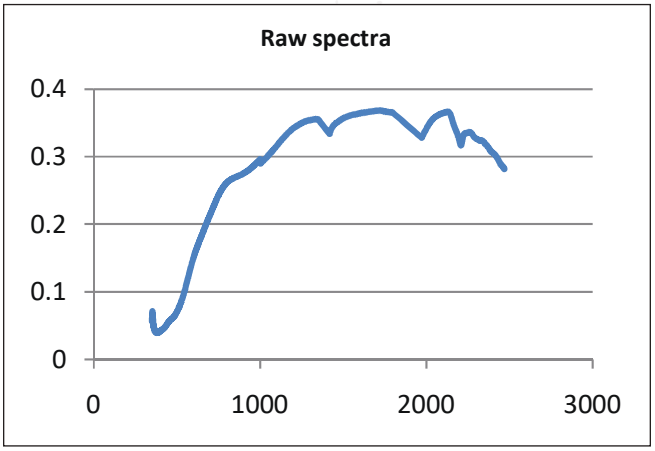


Fig. 1(a): Raw Spectra (Wavelength vs Reflectance)

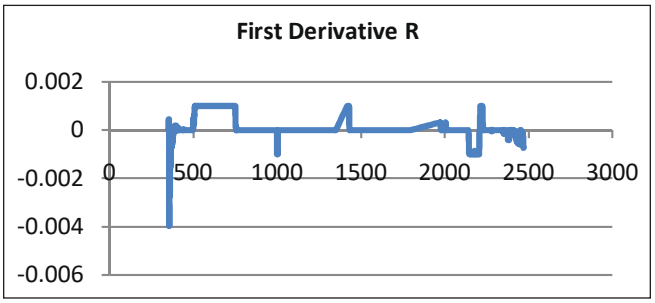


Fig. 1(b): First Derivative R (Wavelength vs Reflectance)

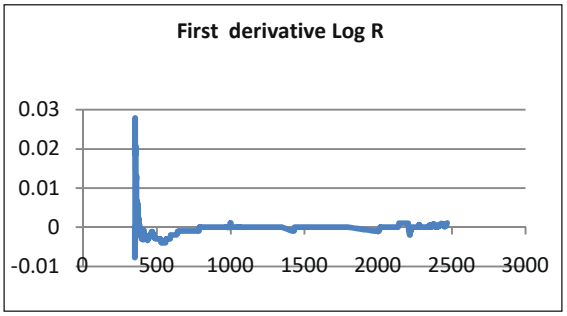


Fig. 1(c): First derivative Log R (Wavelength vs Reflectance)

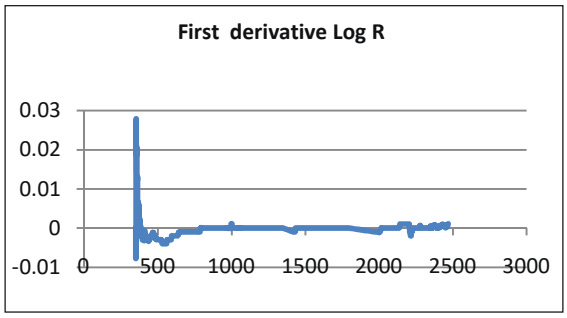


Fig. 1(d): First derivative 1/log(R) (Wavelength vs Reflectance)

Fig. 1 a, b, c, d shows the raw spectra, its first derivative spectra, first derivative of logarithm of spectra and first derivative of reciprocal of logarithm of spectra. In all the below figures X axis depicts the wavelength ranging from 350 – 2500 nm. Y axis depicts the spectral reflectance values recorded for the corresponding wavelengths.

4. Partial Least Square Regression Methods (PLSR)

PLSR is a statistic based multivariate analysis method. It combines Multiple Linear Regression Technique (MLR) and Principal Component Analysis (PCA) method. MLR can predict relation between the dependent and independent attributes and results in multicollinearity. The condition for MLR is that it needs less number of independent attributes than the number of samples. Principal Component Regression (PCR) generates predictors which contains eigenvectors and scores to solve the collinearity problem. This results in optimal number of independent variables into lesser number of principal components. It then regress these principal components against the responses. In Partial least squares regression dependent and independent variables are decomposed together and projected into mutually independent factors smaller in number. In this method decomposition of data with regression occurs in a one step with less number principal components unlike principal component regression. It overcomes the issue of collinearity and reduces the number of spectral bands without loss of spectral data [8].

5 Results

Correlation analysis was carried on measured and laboratory reflectance values from which five wavelengths were identified for soil organic matter. Multiple regression analysis was then carried out on these five wavelengths and the R2 values obtained were 0.92.

In the present study, partial least square regression was performed on VisNIR, NIR and MidIR range in electromagnetic spectrum for SOM, Nitrogen and pH values of soil. The R^2 values were recorded for reflectance values, the first derivatives of reflectance, first derivative of log of reflectance, first derivative of reciprocal of log of reflectance.

Table 3 shows the results of the measurement of SOM, Nitrogen, pH content using VisNIR, NIR, and midIR range spectral data with the respective preprocessing technique. Soil organic matter is predicted in the Vis NIR with first derivative of reflectance values with R^2 as 0.99 and first derivative of reciprocal of log R with R^2 as 0.99. Soil nitrogen is predicted in NIR region with first derivative of reciprocal of log R with R^2 0.99. Soil PH is predicted in Vis NIR and NIR region for first derivative of reflectance and first derivative of reciprocal of log R with R^2 0.99.

Table 3 shows soil organic matter can be predicted in VisNIR, NIR and MidIR with reflectance values, its first derivative, first derivative of log R and first derivative reciprocal of log R effectively. Soil nitrogen shows weak response in VisNIR, NIR for reflectance and first derivative of reciprocal of log R in mid IR. Also PH values can be predicted in VisNIR, NIR and midIR with reflectance, its first derivative, first derivative of log R and first derivative reciprocal of log R effectively.

Table 3: SOM, N, PH quantification using VisNIR, NIR and MidIR spectral ranges

Sr. No.	Pre processing method	VisNIR/ NIR/ Mid IR	Wavelength	R^2 (SOM)	R^2 (N)	R^2 (PH)
1	Reflectance (R) PLSR	VisNIR	400 - 1300	0.97	0.63	0.91
2	Reflectance PLSR	NIR	700 - 1300	0.95	0.69	0.93
3	1st Derivative of R PLSR	VisNIR	400 - 1300	0.99	0.97	0.99
4	1st Derivative of Log R PLSR	VisNIR	400 - 1300	0.97	0.98	0.98
5	1st Derivative of Log R PLSR	NIR	700 - 1300	0.93	0.93	0.92
6	1st Derivative of Log R PLSR	MidIR	1300 - 2500	0.89	0.71	0.96
7	1st Derivative of 1/Log R PLSR	VisNIR	400 - 1300	0.94	0.93	0.96
8	1st Derivative of 1/Log R PLSR	NIR	700 - 1300	0.99	0.99	0.99
9	1st Derivative of 1/Log R PLSR	MidIR	1300 - 2500	0.94	0.59	0.97

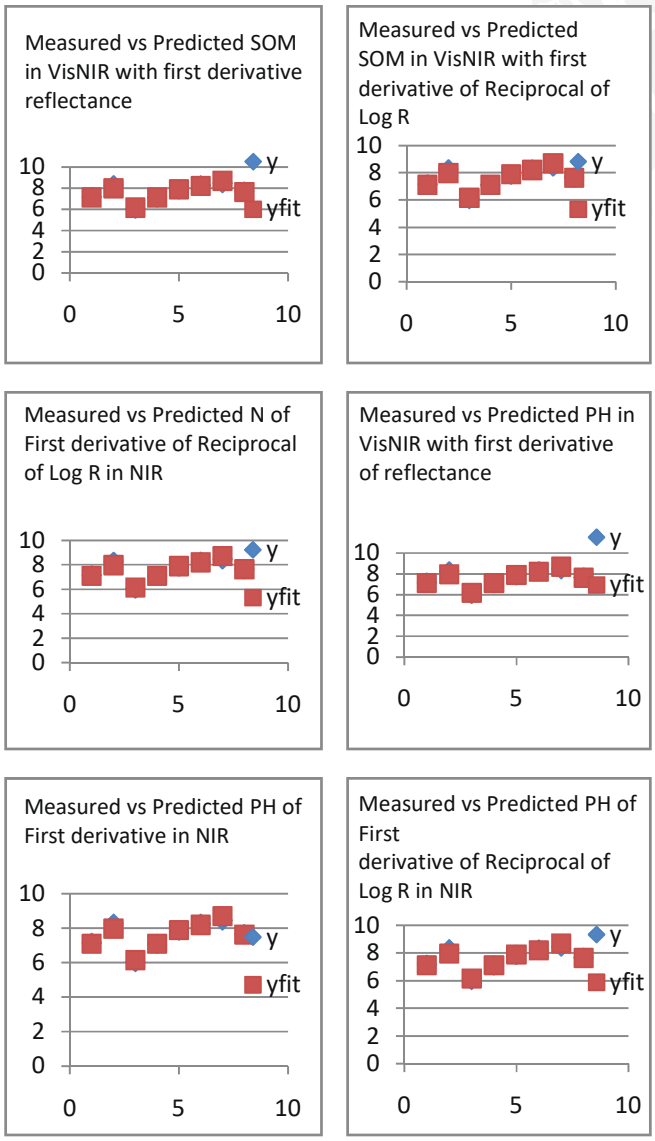


Fig. 2 Measured Vs predicted response of soil organic matter, Soil Nitrogen and Soil PH in Visible near infrared (Vis NIR), and near-infrared (NIR)

6 Conclusion

The three preprocessing techniques that is first derivative of reflectance, first derivative log R and first derivative of 1/log R were applied on the spectra of soils. The results of partial least square regression method were recorded on different preprocessed spectra to measure the occurrence of various soil contents such as soil organic matter, nitrogen and pH in different regions of electromagnetic spectrum. It was observed that the results of reciprocal of Log R are effective in VisNIR, NIR and MidIR for soil organic matter and PH except nitrogen. Soil organic matter has a effective response in the VisNIR with first derivative of reflectance values for R2 as 0.99 and first derivative of reciprocal of log R with R2 as 0.99. Soil nitrogen is predicted in NIR region with first derivative of reciprocal of log R with R2 0.99. Soil PH is predicted in VisNIR and NIR region for first derivative of reflectance and first derivative of reciprocal of log R with R2 0.99. This study can be utilized to some extent for nondestructive, less expensive testing of soil attributes. Further methods can be explored to improve and enhance the performance of soil testing.

7 Acknowledgement

Department of Science and Technology - Funds for Infrastructure under Science and Technology (DST-FIST) has supported the Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India, under section No. SR/ FST/ ET1340. The research is carried out under the support of this scheme. Authors are thankful to the Department and University authorities for providing the resources and support.

References

1] Antonio P. Leone, Raphael A. Viscarra-Rossel, Pietro Amenta and Andrea Buondonno, Prediction of Soil Properties with PLSR and Vis-NIR Spectroscopy: Application to Mediterranean Soils from Southern Italy, Current Analytical Chemistry, Bentham Science Publishers 2012, 283-299

2] M.C. Sarathjith, Bhabani Sankar Das, Suhas P. Wani, Kanwar L. Sahrawat, Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils, Geoderma, Elsevier, http://dx.doi.org/10.1016/j.geoderma.2015.12.031

3] Xing-Xing Qiao, Chao Wang, Mei-Chen Feng, Wu-De Yang, Guang-Wei Ding, Hui Sun, Zhuo-Ya Liang & Chao-Chao Shi, Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques, An International Journal for Rapid Communication, Taylor and Francis, ISSN: 0038-7010, Spectroscopy Letters, 2017, VOL. 50, NO. 3, 156–163

4] Mrs. Chitra Gaikwad, Prof. Dr. S. N. Kakarwal, Mrs. A. N.Ghule, Prof. Dr. R. R. Deshmukh, Dr. D. V. Kurmude, Use of spectral reflectance for sensitive waveband determination for soil organic matter, International Journal of Scientific & Technology Research, Vol 1, Issue 10, Oct 2019

5] Asa Gholizadeh , Lubos Boruvka , Mohammadmehdi Saberioon , Radim Vasat, Visible, Near-Infrared, and Mid-Infrared Spectroscopy Applications for Soil Assessment with Emphasis on Soil Organic Matter Content and Quality: State-of-the-Art and Key Issues, December 2013, Applied Spectroscopy 67(12):1349-62,

6] Xing-Xing Qiao, Chao Wang, Mei-Chen Feng, Wu-De Yang, Guang-Wei Ding, Hui Sun, Zhuo-Ya Liang & Chao-Chao Shi (2017), Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques, Spectroscopy Letters, 50:3, 156-163

7] Liu, J., Xie, J., Han, J. et al. Visible and near-infrared spectroscopy with chemometrics are able to predict soil physical and chemical properties. J Soils Sediments (2020). https://doi.org/10.1007/s11368-020-02623-1

8] Fekerte Arega Yitagesu, Freek van der Meer, Harald van der Werff, Wolter Zigterman, A Multivariate Regression Analysis For Deriving Engineering Parameters Of Expansive Soils From Spectral Reflectance, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B7. Beijing 2008

About the Authors

Mrs. Chitra M. Gaikwad [CSI-1191698] -Assistant Professor, Department of Information Technology, Government College of Engineering, Aurangabad, Maharashtra, India. She has 8 paper publications in International Journals, 3 in International Conference and 1 in National Conference. She can be reached at cmgaikwad@geca.ac.in

Dr. S. N. Kakarwal (Kurmude) [CSI-F8000602] - Professor in Department of Computer Science & Engineering, P.E.S. College of Engineering, Aurangabad, MS-India. Her research interests include Image Processing, Pattern Recognition. She has published 17 papers in International Journals, 28 papers in International Conference and 12 papers in National Conferences. She has published 12 articles in CSI magazine. She can be reached at s_kakarwal@yahoo.com

Dr. Shubhangi Sapkal Assistant Professor, Government College of Engineering since 20 years. She published 50 + papers in national and international journals and conferences. Her main research area is machine learning, information security and ANN.

Dr. Nilesh J. Uke [CSI-00127331] Ph.D. degrees in Computer Science from SRTM University India, in 2014. He is currently a Principal and Professor at Trinity Academy of Engineering, Pune, Maharashtra, Pune. His current research interest includes Visual Computing, Pattern Recognition, Artificial Intelligence, Human Computer Interface and Multimedia.

He is member of IEEE, ACM and Life Member of the Indian Society for Technical Education (ISTE) and Computer Society of India (CSI). He can be reached at nilesh.uke@gmail.com

A Far down Illustration on virtualization and its Blend with Cloud Computing Technology

M. Senthil Kumar[1], B. Chidambara Rajan[2], P. Kiruthika[3] and M. Rajakumar[4]

[1] Associate Professor in CS and Engg. Dept. at SRM Valliammai Engineering College of Tamil Nadu
[2] Professor/Principal at SRM Valliammai Engineering College, Affiliated to Anna University, Chennai
[3,4] student of SRM Valliammai Engineering College

Cloud computing is a top efficient imminent technologies in the world. Cloud computing is a convenient and user pay- per-use, methods. Cloud computing approaches and reaches the end users through their service and models. Cloud models include private, public and hybrid models. These models are access of data from a shared pool of resources. Cloud services include service of infrastructure, platform and software. These services serve the user's needs at reduced costs.

Virtualization is an additional software layer in a system that transforms the traditional system into virtual system. Virtualization is the imitation of physical layer. The core benefits of using virtualization techniques in cloud computing is that, its security purposes, hassle-free data transfer, avoidance of system failures and cost-effective technique. This paper explains trios, such as cloud computing, virtualization and also about virtualization in cloud computing.

Keywords: Cloud Computing Virtualization, Virtualization in Cloud.

1. Introduction

Virtualization is creation of a virtual version of the actual physical hardware [7]. Cloud service providers utilize virtualization techniques to adopt many benefits such as reducing cost, ease sharing of resources, ease management and migration of machines virtually [10]. The virtualization technique ensures separation of application from the operating system level and hardware level. Virtualization permits the cloud users to collectively use all its features like storage, processor and network with reduced cost [5]. As known, cloud computing is an emerging technology and has wide spread usage such as storing, manipulating and retrieval of data. This paper explains the elemental part of cloud computing with specifications such as cloud architecture, components, service modes and deployment models. Also, this paper suggests a detailed study on virtualization with its basics, components, levels, security challenges, vulnerabilities and threats to it.

II. Literature Review

Nadiah M. Aimutary and Khalil H. A. Al. Shqeerat describes the security challenges of virtualization technology in cloud computing. They also tend to identify the challenges and risks of virtualization and cloud computing environment. IssacOdun- Ayo, Blessing Idoko, Temidayo Abayomi – Zammu discussed about the identification of bandwith management mechanism in the cloud environment. Ms. Kalburgi Tayyata, Md. Ibrahim, Ms. Sajjan R. S., shared their ideas about the maintenance of virtualized environment with optimized resources. Manoj Mathew describes the virtualization and scheduling in cloud computing environment. And also, he discusses about the advantages and disadvantages of scheduling algorithms. Sonam Srivastava, S. P. Singh, are described about the Hypervisor-based technology in cloud computing. Durairaj M., Kannan P., are the authors are explained about the challenges and virtualization techniques in cloud computing. They also detailed an open source virtualization techniques and future research-direction. Upendra Singh, Prashant kumarBaheti are the authors describes the importance as well as applications of cloud computing and virtualization. Mani K. and Mohan R. are the authors explained about various work which was related with virtualization techniques. Anita H. M. and Jayarekha P. presented various algorithms and also described security challenges of virtualization in Cloud Environment.

III. Materials and Methods

1. Cloud Computing

1.1. Frame Work of Cloud

Cloud computing is the mix of outcomes of various technologies such as grid computing, utility computing, automatic computing and others. The services needed by the customers reach them through the internet and it is a technology driven by virtualized, dynamically scalable, computing functions and a large- scale distributed paradigm. The cost is as per the usage i.e. on demand pay-as-use. Cloud

architecture has 3 main parts

- Development model – the lowest layer
- Service model – the middle layer
- Mobile resources – the upper layer

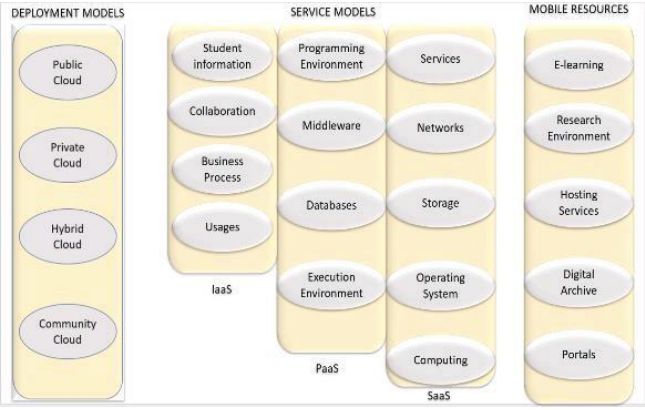


Fig. 1: Cloud Framework

Fig.1 represents cloud framework [11]. Mobile resources are the layer that is used by the client which includes E-learning research environment, Hosting services, Digital achieves and portals.

1.2. Constituents of Cloud Computing

The three constituents of cloud are Client, Data centres and distributed servers.

- Client – This refers to the Client's access which is through gadgets such as computers, laptops, tablets and so on.
- Data centres – Data centre is an area where all the data are being stored in a bulk and huge manner.
- Distributed server – The requested server is geographically far.

1.3. Service Models present in Cloud Framework

Service models are used to provide service to the client through internet [12]. Cloud services are:

- Platform as a Service (PaaS)
- Information as a Service (IaaS)
- Software as a Service (SaaS)

1.3.1. Platform as a Service (PaaS)

In PaaS the clients are provided a service such that it provides a base ground for running client web application. Clients can create their own application using tools such as libraries available, programming languages, services and patterns. The main advantage of PaaS is no separate environment for developing and testing are needed. In this service, clients can use private cloud to build their applications. Fig. 2 represents Platform as a Service (PaaS).

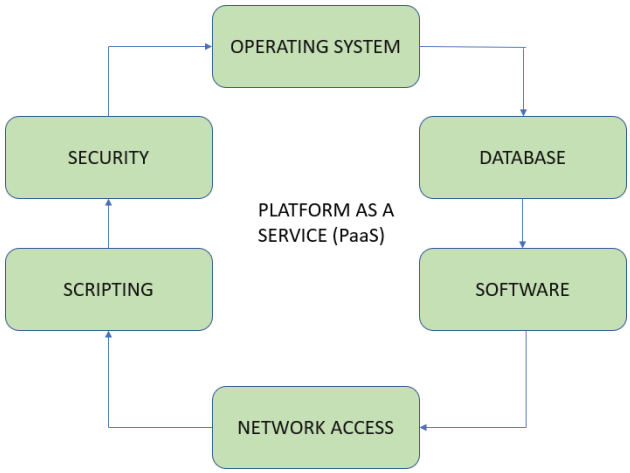


Fig. 2 : Platform as a Service (PaaS)

1.3.2. Software as a Service (SaaS)

In this, Software is provided as a Service. i.e. the required software can be purchased & installed in the computer by the client. This service is hosted by a cloud service provider and client can access to it by web browser. It heavily reduces the cost of client in developing of the software. Fig. 3 represents Software as a Service (SaaS).

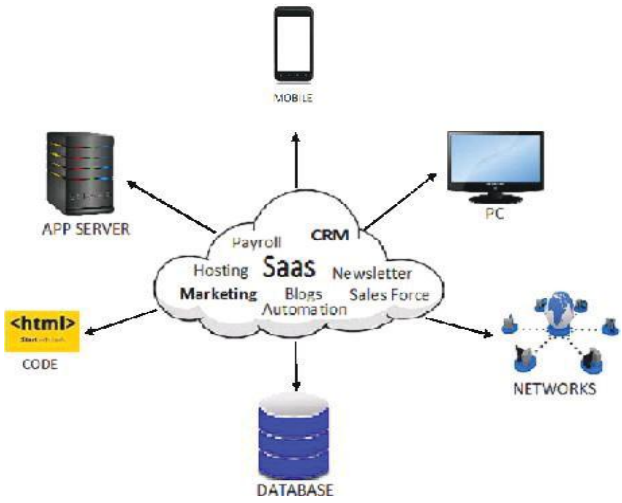


Fig. 3: Software as a Service (SaaS)

1.3.3. Information as a Service (IaaS)

IaaS is a service provider to the client. Without installing the operating system, itself, the clients can run the software, deploy cloud models and access fundamental resources. IaaS supports almost all operating system and it is an open source. Fig. 4 represents Information as a Service (IaaS).

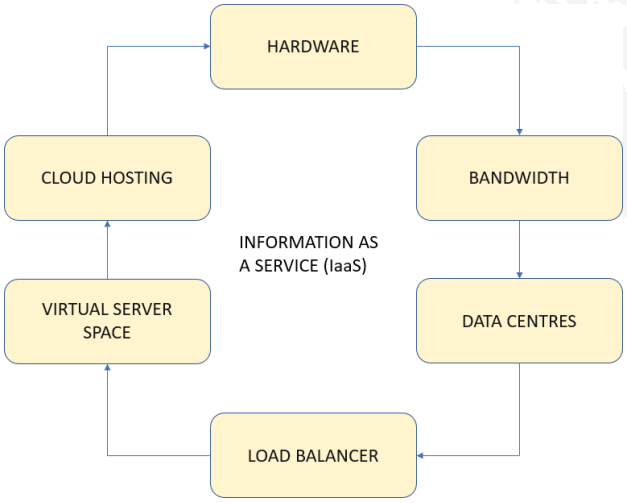


Fig. 4 : Information as a Service (IaaS)

1.4. Deployment Models in Cloud Framework

Depending on user's need for Privacy, Data security, Scalability and other features required by the client, the deployment models are of three types [12]

- Public cloud-deployment
- Private cloud-deployment
- Hybrid cloud-deployment

1.4.1. Public Cloud-Deployment

The services requested by the customer are provided via third party service provider through internet. When the requested service is available in the network publicly its public cloud deployment. Public cloud is available free to the clients and allows to develop their own services by the model. Though it is free, it requires some low cost based on the usage. Fig. 5 represents Public Cloud.

Fig. 5 : Public Cloud

1.4.2. Private Cloud-Deployment

Private cloud is chosen by a client when high security is needed. Most probably it is owned by a single company. Private cloud ensures authorized access to the information stored on the pool. The cost of private cloud is comparatively higher than that of Public cloud. Some of special services of private cloud are multi- tenancy, virtualization and access control.

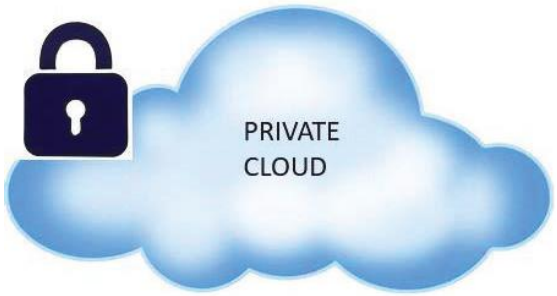


Fig. 6 : Private Cloud

Fig. 6 represents Private Cloud.

1.4.3. Hybrid Cloud-Deployment

Mix of both private cloud deployment and public cloud deployment is known as Hybrid cloud. It is comparatively more flexible and has better backup facilities. Fig. 7 represents Hybrid Cloud.

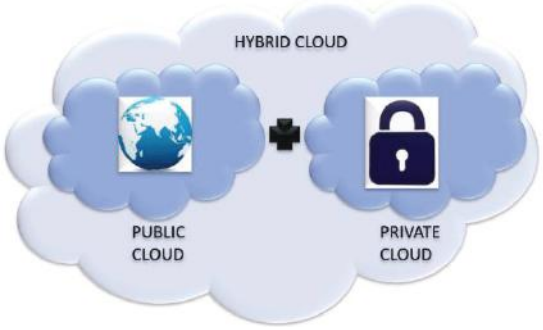


Fig. 7 : Hybrid Cloud

2. Virtualization

2.1 Components of Virtualization

There are five main components required for virtualization [8]. The bottommost layer is called hardware layer, then proceeds OS, hypervisor and Virtual Machine Monitor (VMM) then the virtual machines and the topmost layer is the users of cloud. Fig. 8 represents Components of Virtualization.

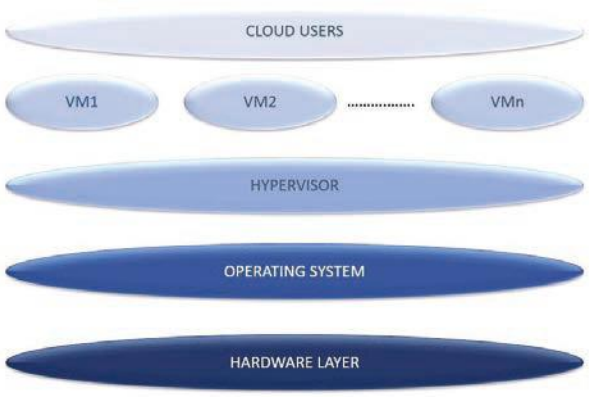


Fig. 8: Components of Virtualization

- **Virtual Machine**

Virtual machine can be seen as a emulation of a computer system. Virtual machine works as same as that of a hardware be imitating a hardware to run an application.

- **Hypervisor**

Hypervisor is a software program used for the operations of the virtual machines like launching when the task is allocated, migrating when the task is to be executed and shut down the virtual machine when the task is completed.

- **Operating System**

The Operating System of the owner machine is called Host Operating System and the Operating System of the virtual machine is called guest Operating System.

- **Hardware**

The resources including hardware is shared among the virtual machines

2.2. Traditional System

In traditional system, there are three layers namely, Hardware layer which is the lowest layer [8] [13]. The next is Operating system, the upper layer is Application that is being accessed by the client. In this traditional system, the hardware is exclusively allocated for a single operating system for some intended purpose requested by the application layer. Fig. 9 represents Traditional System.

Fig. 9 – Traditional System

2.3. Virtualization System

In addition to the layers of the traditional system, virtual system consists of virtualization layer [8] [13]. Virtualization layer is especially designed in virtual system to enfold the system hardware from the system software. Fig. 10 represents Virtualization system.

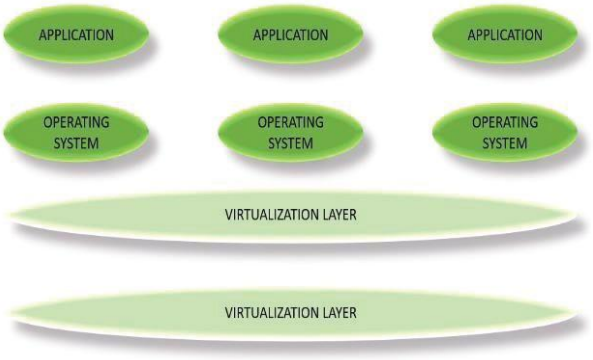


Fig. 10: Virtualization

Table – 1 : Represents Similarities between Traditional System and Virtual System.

Si. No.	Traditional System	Virtual System
1	Sharing of resources is not possible.	Sharing of resources is possible.
2	The allocated resources if not used cannot be allocated to other users.	The unused resources can be allocated to other users when required.
3.	Due to physical hardware and a single operating system migration of software is tedious.	Virtually software is portable from machine-to- machine.
4	Due to the migration process cost of setup is also high.	This kind of systems are cost- effective.

2.4. Levels of Virtualization

The concept of virtualization is to virtually create an executing environment to distribute the workload equally. This is termed as execution virtualization [4]. Execution virtualization can be further divide into

- Process virtualization
- System virtualization

2.4.1. Process Virtualization

Process level virtualization is held on the Operating System with full support of hardware.

- Operating System level

The user application that run concurrently is provided with a separated execution environment.

- Programming Language level

This level of virtualization is to ensure equal execution in various platforms. Some of these features are portability, ease of application deployment and execution management.

- Application level

This level is virtually creating the applications in multiple desktops but it actually be present in one location only. This serves one or more clients.

2.4.2. System Virtualization

System virtualization is held on the hardware without the support of Operating System.

- Hardware assisted virtualization

In this technique, the hardware provides the architectural support to run a software on a virtual machine.

- Full virtualization

Here the application programs run on the top of a virtual machine by complete stimulation of hardware. There is no modification in this level.

- Para virtualization

Para virtualization is modification of guest OS before its installment into a virtual machine for ease sharing of resources and other purposes.

- Partial virtualization

Partial virtualization provides partial emulation of the hardware resources to avoid complete execution of guest OS in complete isolation.

2.5. Security Challenges

Some common security challenges in cloud computing are [2]:

- User awareness

The cloud users should of aware of security challenges in the cloud environment priory, another meaning of user awareness is, the lack of cloud service providers to verify the customer sometimes, like the customers the attackers use the cloud and try to seal the content without any authentications.

- Insecure APIs

The cloud user may require any kind of service such as platform, software or infrastructure. The cloud service provider designs the interface and publish via Application Programming Interfaces (API). Insecure APIs is weak API which even allows unauthorized users.

- Weak authentication Authentication is a kind of mechanism which secures the sensitive data from unwanted users. It protects the data from being read, write or modified. Improper authentication may cause some insecurities to sensitive data.
- Incorrect VM isolation

There may be multiple VMs present and it is the hypervisor which should differentiate between those. Thus, incorrect VM isolation may lead to make access to other's VM and their applications which is present in the same host.

- VM diversity

VM diversity is a challenge which has to cope up with the risk such as executing unpatched software or older machines on the network.

- VM image sharing

VM image paves for easy deployment and restoring in virtual systems. It is used as a quick way for creation of VM. Besides, its benefits it also has a main drawback such that it affects the security and privacy of the cloud system.

- Lock of reliability

Reliability issues affects the cloud system and its users ensures the performance of the system.

- Insecure VM migration/mobility

Migration in virtualization is movement of application from one host to another. During this migration, the attackers may take the information or change the information. Thus, it should be processed.

2.6. Virtualization Threats and Attacks

Some virtualization threats and attacks are [2]:

- Cross – VM attack

This occurs to attack the co-located VM which cross the VM. This mostly occur when the isolation methods are properly implemented.

- VM Rollback

During the execution of a VM, the hypervisor is able to pause it and snapshot all current statuses with which the attackers can rollback, execute the VM and clears the history. This causes security issues.

- Data loss and Data leakage Data leakage is unauthorized use of information which has been audited and transferred previously. Data loss is loss of information due to improper encryption.
- Foot Printing attack

Foot printing is similar as tracking of the system by the attackers and which paves way for the attackers to carry out malicious attacks on the system.

- VM Sprawl

VM Sprawl is improper increase of VM's on the same host. This causes problem to cloud service providers because multiple VM's should be provided with memory.

- VM Escape

Malware escapes from the hypervisor control and continuous to run the malicious software in the VM.

- VM Hopping

The hypervisor is exploited with vulnerabilities because the malicious software jumps from one VM to another. This process is called VM Hopping.

- Hyper jacking

In this, the virtualized environment is controlled by a VM – based rootkits. Hyper jacking inserts duplicated hypervisor underneath the original. Since it is the most privileged it is harder even for an OS to find it.

2.7. Security Approaches

Here comes a statistic of the various security approaches handled by the cloud users to secure their cloud data.

Nearly 8% of users are unaware of threats to their cloud data. About 32% of users move towards secure framework. 13% of users have chosen trusted Intrusion Detected System (IDS). VM Isolation and VM lifecycle management is also a good choice which was chosen by 12% and 10% of users respectively. 11% of the cloud users selected sandbox as a secure approach. Fig. 11 represents Security approaches.

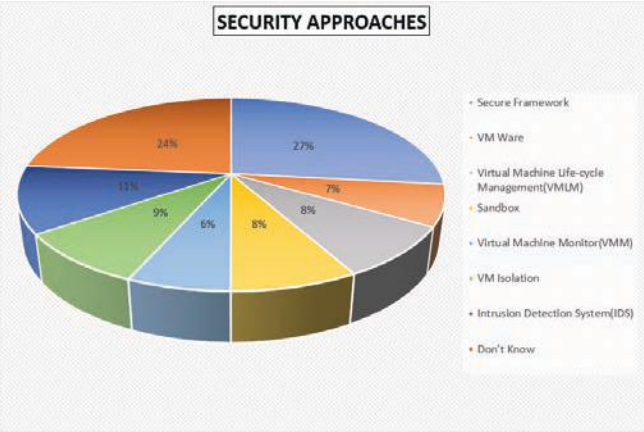


Fig. 11 – Security Approaches

2.8. Vulnerabilities

From the statistics, it is understood that more than 50% of vulnerabilities to cloud environment is because of the lack of user awareness. The user should be well explained about the security and vulnerabilities that can occur in the beginning itself. All the security policies must be explained clearly.

The next high contributions are from Incorrect VM isolation (above 20%), Insecure API's (above 30%), Insecure VM migration/mobility (above 20%) and less reliability (above 30%).

Incorrect VM isolation, Insecure API's, Incorrect VM migration/mobility and less reliability can lead to loss of data.

Weak authentication (above 25%) and VM image sharing (above 25%) these vulnerabilities mean that due to the weak authentication capabilities, the attackers can cause any kind of malicious attack to VM images.

20% of users consider VM diversity as a vulnerability due to which management is difficult.

The lowest vulnerability is VM transience which is less than 10%, i.e. being unaware of security problems. Fig. 12 represents Virtualization Vulnerabilities and Risks.

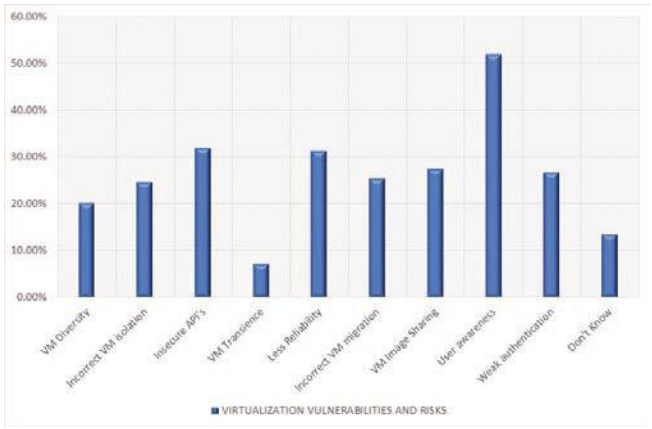


Fig. 12 : Virtualization Vulnerabilities and Risks

2.9. Virtualization Threats and Attacks

Virtualization techniques consists of some serious threats and issues. This statistic explains some virtualization threats and attacks.

Data leakage and data loss occupies the maximum which is 31%. VM side channels, Hyper jacking, VM Escape,

Table : 2 represents Security approaches

Techniques of Security Approaches	Number of Samples taken	Numbers of Samples passed	Percentage
Security Framework	1000	270	27
VM ware	1000	70	7
Virtual Machine Management	1000	80	8
Sandbox	1000	80	8
Virtual Machine Monitor	1000	60	6
VM Isolation	1000	90	9
Intrusion Detected Systems (IDS)	1000	110	11
Others	1000	240	24

Foot printing attack and VM Roll Back all occupies nearly same rank. VM Hopping and VM sprawl alone occupies less percentages in the overall chart. Fig. 13 represents Virtualization Threats and Attacks.

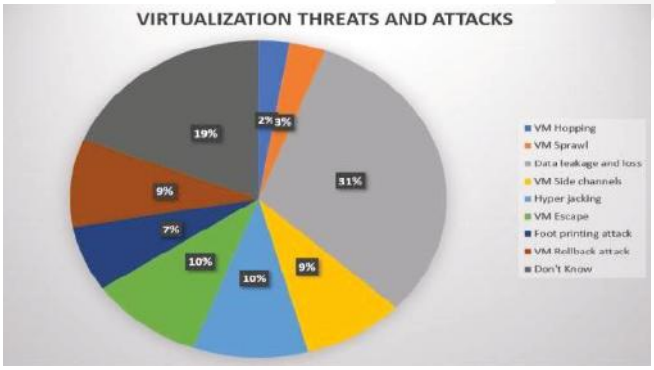


Fig. 13 : Virtualization Threats and attacks

Table 3 : represents Virtualization of Threats and attacks

Techniques on Virtualization and Attacks	Number of Samples taken	Numbers of Samples passed	Percentage
VM Hoping	1000	20	2
VM sprawl	1000	30	3
Data leakage and loss	1000	310	31
VM Side Channels	1000	90	9
Hyper jacking	1000	10	10
VM Escape	1000	100	10
Foot Printing Attack	1000	70	7
VM Rollback attack	1000	90	9
Others	1000	190	19

2.10. Virtualization in Cloud Technology

The terms so called "Virtualization" and "Cloud Computing", both can exist independently. But their combination is more worthful. Cloud computing without virtualization is somewhat inefficient. Nowadays, the tags for cloud computing such as "pay-per-use" and "Infinity" refers to virtualization techniques [11].

Virtualization is virtual which means unreal creation of either hardware, software, platform, infrastructure, operating system, network or memory. The management of virtual environment is somewhat tough than the physical environment because the virtual machine undergoes changes more quickly, scalability and agility properties has more improved in cloud because of virtualization. Fig. 14 represents Virtualization in Cloud Computing.

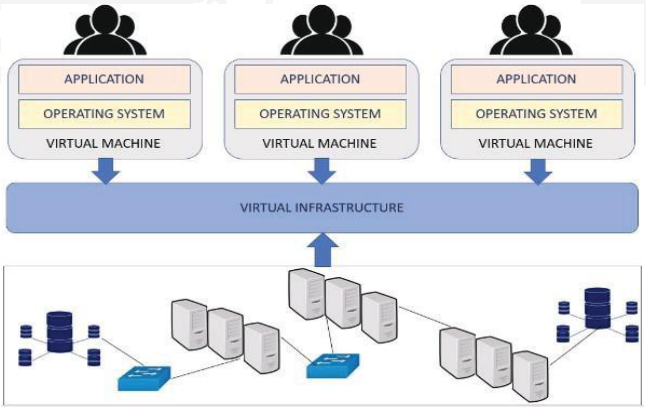


Fig. 14 – Virtualization in Cloud technology

Virtualization in cloud computing has the list of benefits as follows:

- Virtualization provides abstract virtual machine for running our applications which increases speed of completion of particular task.
- Virtualization isolates the hardware, memory and network. Each client is provided with separate VMs which improves privacy.
- An integrity of guest VM and Cloud infrastructure which improves security.
- Virtualization allows for sharing of resources and makes almost full virtualization of resources virtualization also improves reliability.
- Huge data centre's is being consolidated with virtualization techniques which reduces cost.
- Besides its benefits it also has few drawbacks:
  - Managing and migrating VM is difficult.
  - Monitoring the virtual resources is a critical task.
  - Utilization of resources for execution is also tough to deal.

IV. Discussion

The topics discussed in this paper basics of Cloud Computing and about the deployment models and service models were explained. Brief contents of virtualization its levels, security attacks and virtualization threats are provided. Present years security approaches, virtualization threats and vulnerabilities in virtualization were presented statistically. And finally, the essence of this paper which is the amalgam of cloud technology with virtualization was explained in brittle.

V. Conclusion

As there are huge data, problems due to data are also high. To solve this problem, Cloud is a transformative technology which provides a potential platform for huge data problems [9]. Due to the rapid growth of Cloud Computing and Virtualization technologies, its role in technical education is essential [6]. Virtualized lab environment and

Cloud resources will provide more benefits to students. Virtualization layer acts as a container and separates from physical machine [1]. And these technologies when combined can result better improvement in the technological world. The mix of these technologies would provide the world to greater extents.

VI. References

1] Ms. K. T. Md. Ibrahim, Ms. Sajjan R. S., "Cloud Resource Virtualization", Int. Journal of Computer Science and Engineering (IJCSE), Vol. 04, Issue 04, PP 21-27, 2016.

2] Nadiah M. Almutuiry and Khalil H. A. Al. Shqeerat, "A Survey on Security Challenges of Virtualization Technology in Cloud Computing", Int. Journal of Computer Science & Information technology (IJCSIT), Vol. 11, Issue 03, PP 95-105, 2019.

3] M. Mathew, "Virtualization and Scheduling in Cloud Computing Environment – A Study", IOSR Journal of Computer Engineering (IOSR - JCE), Vol. 20, Issue 04, PP 23-32, 2018

4] S. Srivastava, S. P. Singh, "A Survey on Virtualization and Hypervisor-based Technology in Cloud computing Environment", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 05, Issue 02, PP 430-435, 2016.

5] Durairaj H., Kanna P., "A Study on Virtualization Technique and Challenges in Cloud Computing", International Journal of Scientific & Technology Research (IJSTR), Vol. 03, Issue 11, PP 147-151, 2014.

6] Upendra Singh, Prashant Kumar Baheti, "Implication of Virtualization and Cloud Computing in Technical Education", International Journal of Engineering and Techniques (IJET), Vol. 03, Issue 06, PP 107- 110, 2017.

7] K. Mani, R. Mohan Krishnan, "A Survey on Cloud Computing Virtualization", International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE), Vol. 04, Issue 05, PP 9618-9623, 2016.

8] Anitha H. M., Jayarekha P., "Security Challenges of Virtualization in Cloud Environment", Int. Journal of Scientific Research in Computer Science and Engineering (IJSRCSE), Vol. 06, Issue 01, PP 37-43, 2018.

9] Shaikh Abdul Azeem, Satyentra Kumar Sharma, "Role of Network Virtualization in Cloud Computing and Network Convergence", International Journal of Advanced Conceptational Engineering and Networking (IJACEN), Vol. 05, Issue 09, PP 136-140, 2017.

10] Bohr Singh, Gursewak Singh, "A Study on Virtualization and Hypervisor in Cloud Computing", Int. Journal of Computer Science and Mobile Application (IJCSMA), Vol. 06, Issue 01, PP 17 – 22, 2018.

11] T. Swathi, S. R. Reddy, K. Srikanth, "Virtualization in Cloud Computing", International Journal of Computer Science and Mobile Computing (IJCSMC), Vol. 03, Issue 05, PP 540-546, 2014.

12] A. M. Chandrashekar, Shasikumar, "Cloud Computing service and Deployment Models", Information Journal for Research in Applied Science & Engineering Technology (IJRASET), Vol. 05, Issue 06, PP 1083 – 1088, 2017.

13] Malhotra L., Agarwal D., Jaiswal A., "Virtualization in Cloud Computing", Information Technology & Software Engineering (ITSE), Vol. 04, Issue 02, PP 1-3, 2014.

About the Authors

Dr. M. Senthil Kumar (LM- I1504760) is currently working as an Associate Professor in Computer Science and Engineering department at SRM Valliammai Engineering College of Tamil Nadu. He is a CSI-Student Branch Counsellor of the College. His research interests are in IOT, Big Data, Software Engineering and development of new tools for effort estimation.

Dr. B. Chidambara Rajan (LM-00063930) working as a professor/Principal at SRM Valliammai Engineering College, Affiliated to Anna University, Chennai. He has 25 years of teaching experience in government and reputed private institutions. He is a member of professional societies like CSI, IEEE, IETE, IEI, ISTE, ISOI, etc. He has published several technical papers in national and international journals and conferences His research interests include IOT, Big Data, Software Engineering and Networking.

Ms. P. Kiruthika [01491613] is a UG student of SRM Valliammai Engineering College Pursuing degree in B. TECH INFORMATION TECHNOLOGY. Her area of interests includes Cyber security, Blockchain, IoT and Bigdata.

Mr. M. Rajakumar (01491535) is a UG student of SRM Valliammai Engineering College in Computer Science and Engineering department. His areas of interests include IoT, Bigdata, Web designing and Cyber security.

A Novel Approach for Meat Adulteration Detection using the concepts of Image Processing and Machine Learning

Prabha. B[1], Anjana M. P.[2] and Pradeepa K.[3]

[1] Assistant Professor, Department of Information Technology, Loyola ICAM college of Engineering and Technology, Chennai, India. Email: prabha.b@licet.ac.in

[2,3] Student, 3rd year, Department of Information Technology, Loyola ICAM college of Engineering and Technology Chennai, India. Email: anjana.21it@licet.ac.in, pradeepa.21it@licet.ac.in

Red meat is an excellent source of essential nutrients such as protein, iron, niacin, zinc and vitamin B12. But in certain regions at butcher shop they mix one type of red meat with other in order to make quick money. For instance in Vijayawada, the butchers allegedly mixed beef with mutton as a measure to earn quick money as beef is much cheaper than mutton. The beef is available at an affordable price of ₹ 250 to ₹ 300 per kg, in comparison to the mutton being sold at approximately ₹ 500 to ₹ 600 per kg in Vijayawada. Meat Adulteration Detector helps to solve the problem of substitution of red meat. In a particular season and region sometimes one particular red meat can be expensive. In that case the meat vendors may substitute the required red meat with some other red meat like pork, buffalo meat, etc. This can be harmful to the consumers who are allergic to certain red meat. Using the meat adulteration detector application all that the consumer will have to do is to take a picture of the raw red meat that he/she is going to purchase and the application rightly tells whether the meat is adulterated by mixing with other red meat or not. The application is based on the concepts of machine learning and computer vision. The image captured by the consumer is compared with the repository of images in the dataset containing images of unadulterated red meat. Using Tensorflow – keras libraries the captured image is converted to gray scale and partitioned into layers, based on the layering done by Keras the image is categorized as adulterated or unadulterated. The features of the source (dataset images) and the target (image captured by the consumer) are extracted and compared using openCV. The feature matching lines are plotted using matplotlib between source and target. Based on the number of feature points matched in the unadulterated labeled image the percentage of freshness of the meat is calculated. This way the customers can ensure whether they are eating the red meat they asked for.

Keywords: Meat, Adulteration, Computer Vision, Machine Learning, Tensorflow, Keras and OpenCV

I. Introduction

The shifting costs and accessibility of food items from various birthplace give chances to the off-base statement of food segments, both from a quality and amount perspective. For instance, horsemeat might be named as hamburger. Furthermore, the enthusiasm for creature speciation in meat items depends on strict prerequisites (halal and fit). Strictly persuaded reviews, especially of pork in meat items or prepared nourishments, do exclude any specialized edge esteems. This implies absolutely subjective and especially touchy tests are required particularly for distinguishing the nearness of pork.

The significant tainting originates from outside source during dying, taking care of and handling. During dying, cleaning, and cutting, the primary wellsprings of microorganisms are the body portions of the creature and the intestinal tract. [1]

The sullying microbes on the blade before long will be found in meat in different pieces of the body, conveyed by blood and lymph. The outside of the creature harbors huge quantities of microorganisms from soil, water, feed and fertilizer, just as its regular surface greenery and the intestinal substance. Blades, fabrics, air and hands, apparel of the laborers can fill in as a middle wellspring of contaminants.

Hence, there is an increasing demand in the market for a meat adulteration detector so that the consumers can authenticate the meat they are purchasing. Consumers can

be confident that the meat they are purchasing is free from both adulteration and contamination.

II. Need:

"Food Fraud" is the burning question all over the world. Every food item that we consume is adulterated in one way or the other. It was found that 29% of horse meat is mixed with beef and is processed as patties for burgers in Britain. In Vijayawada, India three meat shops were sealed as they were found to adulterate mutton by mixing it with beef [5]. Hence, it is very much essential for the consumer to authenticate the meat that he/she is purchasing. Expending defile meat can prompt a path of medical issues, for example, Tapeworms can develop to somewhere in the range of 4.5 and 9 meters long and live in your digestive organs, food contamination, queasiness, spewing and looseness of the bowels, Heart and liver sickness, pneumonia, septicemia, peritonitis, tumors and abscesses.

III. Previous Methods:

Starting at now there is no portable application in the market to recognize meat contaminated of any kind. Meat legitimacy in food testing research facilities has been generally performed utilizing polymerase chain response (PCR) and catalyst connected immunosorbent examine (ELISA). These techniques require serious customization to accomplish the necessary affectability and exactness. What's more, the atomic data got is inadequate and information mining can't be performed post-examination. These are not kidding impediments for general wellbeing and sanitation examinations. A fluid chromatography-pair mass spectrometry (HPLC-MS/MS) strategy is utilized for the recognizable proof of meat marker peptides and quantitative recognition of normal exogenous meat in sheep defilement [5] [7].

IV. Field Work:

Shoppers give increasingly more consideration to the food security, with respect to the inception and substance of the food they purchase, for their best nutritive sources, taste, and flavor [6].

These are FSSAI standards for Raw Meat:

Hazards in meat safety includes the presence of:

- microbial pathogens
- various chemical and physical contaminants, cross-contamination
- food additives
- chemical and antibiotic residues

Because of the interest of value and the expanding in value, food corruption is still very basic in some food items that incorporate milk, wine, table oil, tea, espresso, and meats. For example, it has been asserted that 20% of the checked animal's meat were sure of manufactured in Guangdong, China, during 2014-2015. European criminal police association (EUPOL) and global criminal police association (INTERPOL) even have been cooperating in observing food security, prompting 2500-ton illicit and manufactured food. As a rule, corruption of meat is finished by replacement of low-valued or restricted meat types for that expensive one, similar to pork tainted inside the hamburger and lamb or other meat[2]. Fake of regular domesticated animals has cause extreme cultural issues, which not just overheads the estimations of strict concerned or hypersensitive customers yet in addition decays the credit of the ventures. For instance, numerous Hindus don't eat hamburger, while Islamic and Jewish laws outlaw the eating of pork meat.

There are some features that can be found in meat which are adulterated,

Feature inputs

Features	Beef	Pork	Carabeef	Mutton
Color of Fat	Yellow White	Whitish Yellow	Milky White	Fat is not visible
Color of Meat	Dark Red	Pale Red	Very Dark Red	Deep Red
Texture	Solid Structure	Flaky Structure	Solid Structure	Solid Structure

Fig 1: Feature inputs for Adulteration Detection in meat

Here we focus on Mutton for further research and implementation of the proposed solution

V. Proposed Solution:

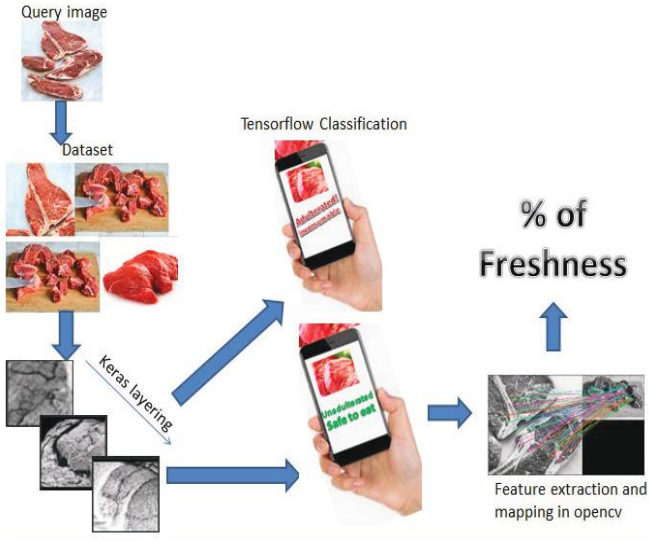


Fig. 2 – Workflow of the proposed solution

We propose a mobile application to resolve this issue of adulteration in meat we consume, that is to detect the meat adulteration using a mobile application where the buyer can easily take a picture of the raw meat which is to be purchased and can confirm the quality of the meat with its freshness percentage. The basic workflow is when a customer goes to

a butcher shop to buy mutton, he / she needs to capture the raw meat image, from our application this photo is sent to our model for prediction. This captured image is compared with a series of trained images and then categorized as adulterated or not using keras and tensorflow. After the categorization, if the image is labelled as unadulterated then its percentage of freshness is calculated using opencv. This application is modelled into mobile apk file using kivy python to provide credentials for users and grant camera access for the application to capture pictures[8].

VI. Process And Concepts Involved:

Understanding the Meat digital image structure with its intricate features and comparing them with voluminous similar dataset of images for predicting the adulterated meat is impossible for human interpretation, hence here machine learning, image processing and computer vision comes into play

The process of adulteration detection uses the machine learning techniques to compare the captured image with its existing set of images for its accurate results prediction.

a. Importing Libraries:

Keras is a python based neural network library used in Tensorflow. It is an open source library enabling its users to do image processing in an precise way. Keras is a human understandable Package which makes it easy to learn and use. A thick layer speaks to a framework vector augmentation. The qualities in the lattice are the trainable boundaries which get refreshed during iterative preparing. So you get a m-dimensional vector as yield. A thick layer along these lines is utilized to change the elements of your vector.

We mainly use Keras library for Layering the image and Tensorflow for image classification as adulterated or unadulterated and opencv to extract and compare the feature points of the source and the target image to calculate the freshness percentage of the meat[3].

b. Image conversion

The captured image needs to be converted to Grey scale for layering in keras and then it needs to be converted to TIFF (Tagged Image File Format) as it maintains high quality versatile lossless compression of raster graphics for feature extraction and plotting this is done with the use of PIL (pillow) library in python

c. Training and testing:

We train the model with 200 images of adulterated and unadulterated image datasets [4] we import the keras library and proceed the prediction using sequential keras layering. In consecutive layering we utilize thick technique for layering which is a customary layer of neurons in a neural system. Every neuron gets contribution from all the neurons in the past layer, in this manner thickly associated. We fix the model layer with three specific values, 128, 16, 2 and trained the 100 images with these layers after converting them into gray scale, we fixed the number of epochs to be 100, that is the count of times the model is trained and fitted back and forth in a iterative manner to improvise the accuracy here the accuracy varies with each epoch we can see that accuracy getting increased for each epoch on model testing and finally reached to be 1 that is the maximum accuracy, average of this epoch accuracies is the final test accuracy. We achieved the accuracy of 88% testing the model with the remaining 100 image datasets. We plot the images in a graph with labelled test results using matplotlib.

d. Freshness calculation:

Once the image is labelled to be unadulterated then the model calculates the meat's freshness percentage [9]. The imported library capacities for this model are numpy, opencv2 and pyplot from matplotlib. The general capacity of numpy is to convey a superior multidimensional exhibit and furthermore it contains fundamental devices to register with and control these clusters. We use Numpy to store the distance between the source and destination of matching feature points as set of array elements in ascending order later and while displaying the output, the least distance matched line is displayed first. The next library is opencv, opencv is an open source library utilized in computer vision for picture handling this library gives a few elevated level programming dialects like python. In particular we here import opencv2 which has a library of Python bindings designed to solve computer vision-based problems. All the OpenCV array structures are systematically changed to and from Numpy arrays,also the images which are to be read by the model are converted into grayscale with this package. This makes the model at ease to incorporate with other libraries that use Numpy such as SciPy and Matplotlib. Matplotlib is a tool used for plotting purpose it can produce 2 dimensional graphs in various milieus. The main use of matplotlib is to explore through the data and come to well defined conclusions through précised graph models. Matplotlib can do different types of plots like line plots, scatter plots, bar plots, histogram and multiple plots. We use pyplot in matplotlib it is a collection of facility style functions. Each pyplot function makes certain transformation to an image or the figure like the instances of creating a figure, and plotting an area in a figure, plots some lines in a plotting area, embellishes the plot with markers, etc. here we use pyplot to draw match line plots for the feature matching between images. Using BFMatcher the feature points from the source and target images are extracted and matched if they are same and based on the number of feature matching lines drawn we calculate the percentage of freshness of the meat.

We shall see the coding implementation done at the backend to detect the adulteration and to find the freshness percentage

VII.Code Implementation and Outcomes:

In [132]:

```
# TensorFlow and tf.keras
import tensorflow as tf
from tensorflow import keras

In [133]:
# Helper libraries
import numpy as np
import matplotlib.pyplot as plt
import glob, os
import re
In [134]:
# Pillow
import PIL
from PIL import Image
# Use Pillow library to convert an input
jpeg to a 8 bit grey scale image array for
processing.

In [136]:
 def jpeg_to_8_bit_greyscale(path, maxsize):
        img = Image.open(path).convert('L')
# convert image to 8-bit grayscale
        # Make aspect ratio as 1:1, by
applying image crop.
    # Please note, croping works for this
data set, but in general one
    # needs to locate the subject and then
crop or scale accordingly.
        WIDTH, HEIGHT = img.size
        if WIDTH != HEIGHT:
                m_min_d = min(WIDTH, HEIGHT)
                img = img.crop((0, 0, m_
min_d, m_min_d))
        # Scale the image to the requested
maxsize by Anti-alias sampling.
        img.thumbnail(maxsize, PIL.Image.
ANTIALIAS)
        return np.asarray(img)
```

In [137]:

```
def load_image_dataset(path_dir, maxsize):
       labels = []
        os.chdir(path_dir)
        for file in glob.glob("*.jpg"):
                img = jpeg_to_8_bit_
greyscale(file, maxsize)
               if re.match('ad.*', file):
                        images.append(img)
                        labels.append(0)
               elif re.match('unad.*', file):
                       images.append(img)
                                labels.append(1)
        return (np.asarray(images),
np.asarray(labels))
```

In [138]:

```
maxsize = 100, 100
```

In [139]:

```
(train_images, train_labels) = load_image_
dataset('C:/Users/Admin/Desktop/train',
maxsize)
(test_images, test_labels) = load_image_
dataset('C:/Users/Admin/Desktop/test',
maxsize)
```

In [140]:

```
class_names = ['ad', 'unad']
train_images.shape
```

In [140]:

```
(100, 100, 100)
```

In [142]:

```
def display_images(images, labels):
        plt.figure(figsize=(10,10))
        grid_size = min(25, len(images))
        for i in range(grid_size):
                plt.subplot(5, 5, i+1)
                plt.xticks([])
                plt.yticks([])
                plt.grid(False)
```

In [143]:

```
plt.imshow(images[i], cmap=plt.cm.binary)
```

In [144]:

```
plt.xlabel(class_names[labels[i]])
display_images(train_images, train_labels)
plt.show()
```

In [156]:

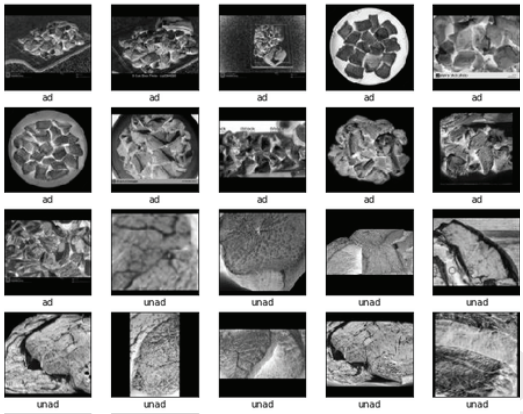


Fig 3 – Train image classification

In [145]:

```
test_images.shape
```

Out[145]:

```
(100, 100, 100)
```

In [148]:

```
train_images = train_images / 255.0
test_images = test_images / 255.0
```

In [149]:

```
# Setting up the layers.
model = keras.Sequential([
    keras.layers.Flatten(input_shape=(100,
100)),
        keras.layers.Dense(128,
activation=tf.nn.sigmoid),
        keras.layers.Dense(16,
activation=tf.nn.sigmoid),
   keras.layers.Dense(2, activation=tf.
nn.softmax),
    keras.layers.Flatten()])
```

In [150]:

```
sgd = keras.optimizers.SGD(lr=0.01,
decay=1e-5, momentum=0.7, nesterov=True)
model.compile(optimizer=sgd,
              loss='sparse_categorical_
crossentropy',
              metrics=['accuracy'])
```

In [151]:

```
Train on 100 samples
Epoch 1/100
100/100 [==============================] - 3s
145ms/sample - loss: 0.6986 - accuracy: 0.4091
Epoch 2/100
100/100 [==============================] - 0s
2ms/sample - loss: 0.6963 - accuracy: 0.5000
Epoch 3/100
...
Epoch 27/100
100/100 [==============================] - 0s
822us/sample - loss: 0.6226 - accuracy: 0.7727
Epoch 28/100
100/100 [==============================] - 0s
695us/sample - loss: 0.6200 - accuracy: 0.8182
Epoch 29/100
100/100 [==============================] - 0s
756us/sample - loss: 0.6174 - accuracy: 0.9091
...
Epoch 100/100
100/100 [==============================] - 0s
861us/sample - loss: 0.4348 - accuracy: 1.0000
<tensorflow.python.keras.callbacks.History at
0x1d704d3ccc0>
```

In [156]:

```
test_loss, test_acc = model.evaluate(test_
images, test_labels)

print('Test accuracy:', test_acc)
```

In [156]:

```
17/1 [==========] - 0s 13ms/sample
- loss: 0.6241 - accuracy: 0.8824
Test accuracy: 0.88235295
```

In [157]:

```
predictions = model.predict(test_images)
display_images(test_images,
np.argmax(predictions, axis = 1))
plt.show()
```

In [157]:

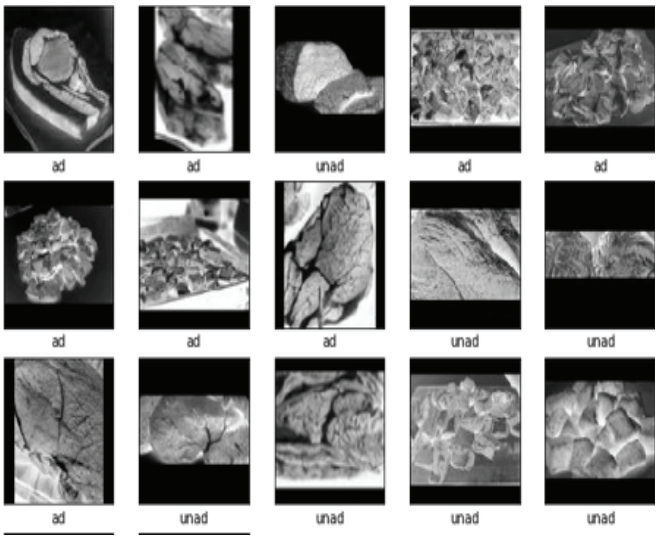


Fig 4 – Test image classification

In [15]:

```
#Individual image evaluation
predictions = model.predict(test_image)
display_images(test_image,
np.argmax(predictions, axis = 1))
plt.show()
imageNo1 = cv2.imread(fn, 0)
cv2.ocl.setUseOpenCL(False)
imageNo2 = cv2.imread('unad.15.tif',0)
imageNo1 = imageNo1[200:600, 200:1000]
keypoint1, descriptor1 = orb.
detectAndCompute(imageNo1,None)
```

```
keypoint2, descriptor2 = orb.
detectAndCompute(imageNo2,None)
bfMatcher = cv2.BFMatcher(cv2.NORM_HAMMING,
crossCheck=True)
matchResults = bfMatcher.
match(descriptor1,descriptor2)
matchResults = sorted(matchResults, key =
lambda x:x.distance)
resultImage = cv2.drawMatches(imageNo1,keypo
int1,imageNo2,keypoint2,matchResults[:100],
None, flags=2)
plt.imshow(resultImage),plt.show()
if (matchResults[170:250]):
       print ("90% ]")
elif (matchResults[160:169]):
         print ("80% ")
elif (matchResults[140:159]):
          print ("70% ")
elif (matchResults[1:69]):
          print ("50% ")
```

Out [157]:

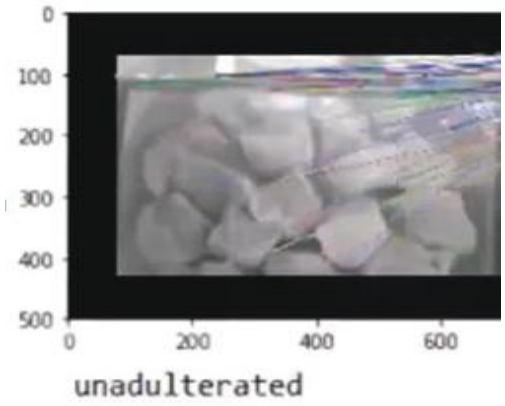


Fig. 5 : Results Prediction (adulterated or unadulterated)

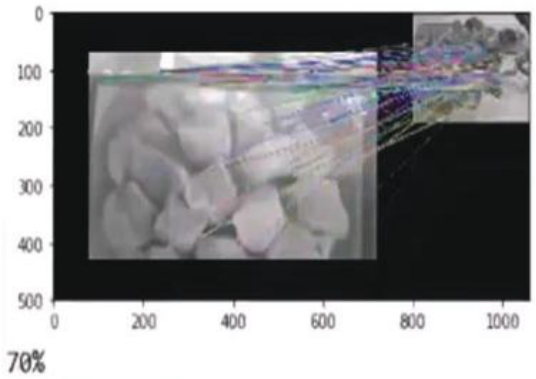


Fig 6 – Freshness calculation

VIII. Future Enhancements:

As of now we have only one module which detects the presence of other types of red meat mixed with mutton. We are planning to implement modules for every other red meat. In future a module to detect the presence of contamination also could be included. The meat adulteration detector is able to identify adulteration only in raw meat this could be extended for processed meat as well with the help of sensors.

IX. Conclusion:

This paper is based on the technology of digital image processing and novel machine learning techniques, can definitely anticipate a permanent solution for the burning question of adulteration in meat with the help of enhanced researches, required data sets and experiments and relieve the consumers from the intake of adulterated meat. The precision of the results rendered by this model would be more accurate as it is developed with the latest technology 'Tensorflow - Keras'. This application modelled using kivy would serve every consumer and the dealers in their day to day routine checking of the meat they buy and enable the meat eaters to eat the right meat.

References

1] N. Ballin, F. Vogensen and A. Karlsson, "Species determination – Can we detect and quantify meat adulteration?", 2020.

2] S. Soares, J. Amaral, I. Mafra and M. Oliveira, "Quantitative detection of poultry meat adulteration with pork by a duplex PCR assay", 2020.

3] Q. Hua, S. Azeem and S. Ahmed, *Machine learning with TensorFlow 1.x.*

4] Github : Dataset [Online]. Available: https://github.com/dc1234ap/Meat-adulteration-detection-through-digital-image-analysis

5] H. Velioglu, B. Sezer, G. Bilge, S. Baytur and I. Boyaci, "Identification of offal adulteration in beef by laser induced breakdown spectroscopy (LIBS)", Meat Science, vol. 138, pp. 28-33, 2018. Available: 10.1016/j.meatsci.2017.12.003.

6] "Adulteration and the Inspection of Meat: Recent Belgian Legislation.", *The Lancet*, vol. 140, no. 3604, pp. 750-751, 1892. Available: 10.1016/s0140-6736(01)87852-1.

7] R. Wissiack, B. de la Calle, G. Bordin and A. Rodriguez, "Screening test to detect meat adulteration through the determination of hemoglobin by cation exchange chromatography with diode array detection", *Meat Science*, vol. 64, no. 4, pp. 427-432, 2003. Available: 10.1016/s0309-1740(02)00211-5.

8] L. Coelho and W. Richert, *Building machine learning systems with Python.*

9] Anjana. M.P, Pradeepa. K, "Detecting food adulterants using the concepts of machine learning", *Ripublication.com*, 2020. Available: https://www.ripublication.com/ijaerspl2019/ijaerv14n5spl_05.pdf. [Accessed: 23- Feb- 2020].

About the Authors

Ms. Prabha B. is working as Assistant Professor in the Department of Information Technology from Loyola-ICAM College of Engineering and Technology, Chennai since 2012. Her area of specialization includes Wireless sensor Networks, Cloud computing, IoT and Machine learning. Published nearly 10 papers in leading International conferences and Scopus Indexed Journals

Anjana M. P. is pursuing B.Tech IT in Loyola ICAM College of Engineering and Technology currently living in Chennai. She has published a paper in the International Journal of Computer Sciences and Engineering. She is highly interested in the field of Artificial Intelligence and Machine Learning. She is highly energetic and has the ability to work under the stipulated time efficiently. She is a self- motivated person who does things in a very constructive and innovative manner. She aims to contribute her knowledge and findings for the welfare of the society. She is keen to find technical solutions for the challenges faced in the medical fields.

Pradeepa K. is pursuing B.Tech IT in Loyola ICAM College of Engineering and Technology currently living in Chennai who has published two papers in the International Journal of Computer Sciences and Engineering. She is interested in the domain of Machine Learning and Computer Vision. She is highly proactive and looks forward in contributing her knowledge for technological developments, looks everything in a different and productive way.

Automatic Speech Recognition system in Marathi for Cerebral Palsy Disabled

Maheshwari A. Ambewadikar[1], Manasi R.Baheti[2]

Department of Computer Science And Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Maharastra, Aurangabad. Email: mahiambewadikar@gmail.com, manasibaheti@gmail.com

Speech has been the catalyst in the human development, being the natural mode to communicate, to interact. When it comes to the interaction between human and computer, it is accomplished via Automatic Speech Recognition (ASR). Many times, Speech is disordered, due to many reasons. The speech disorder in Cerebral Palsy disabled person motives speech impairment and fails to communicate. This paper presents the work for the speech recognition of Marathi digits and words of Cerebral Palsy (CP) Disabled person. The design and improvement (development) of the Automatic Speech Recognition (ASR) system in Marathi language, that will recognize the digits and some selected words used in daily communication, is the main task handled here.[2] The proposed system uses some techniques of the speech recognition, such as ZCR for endpoint detection, MFCC for the feature extraction of the pronounced word/digit, and DTW for the pattern matching. The database of the utterances of Cerebral Palsy (CP) persons is created and Automatic Speech recognition technique is applied for recognition. Such system can bridge the gap between Cerebral Palsy Disabled person and society as the system can be embedded in many applications like phone dialer, hands-free operations etc….

Keywords : Cerebral Palsy (CP), Automatic Speech Recognition (ASR), ZCR, MFCC, DTW.

I. Introduction

Speech Recognition, also called as Automatic Speech Recognition (ASR) or

Computer Speech Recognition, implies understanding speech by the computer for performing any required task [1].With the rapid development of the communication technologies, a reliable speech enabled technique is the necessity of the time [1].

A person can suffer from any form of the disability. Speech disability can come because of many reasons; this work considers the speech disorder due to Cerebral Palsy. Speech Recognition system for Cerebral palsy is this System that compares the utterance of the cerebral palsy person as input ,with reference utterance stored in database [2]. It is studied as of now that, currently no such system is available in the Marathi language. This paper presents the work that will develop the interface in Marathi language, to aid the CP disabled persons, to interact with the society, so that at least few, daily transactions in communication they can do with the help of the technology. Speech is variable in nature.[12] It is the ambitious task to recognize the speech. Basically it means talking to the computer/ or automated system, and recognize what is being said. It is very hard to recognize (identify) the speech of person with the disorder, like CP. Cerebral palsy is a disability, Where person having a Speech problem to produce Speech appropriately (properly) .Therefore such persons have to rehash their words for communication.[8] It is aimed to develop an interface that will take input speech of speech disordered person and get it correctly recognized. It is a system where the stored speech references in the data base can be perceived (recognized) according to speech of CP disabled[2].The system development includes creation of the database, in which digits in Marathi language and some selected words that are spoken in daily communication[2]. Additional, algorithm are, Zero Crossing Rate (ZCR) utilized for the endpoint detection, Mel-Frequency Cepstral Coefficients (MFCC) using for the feature extraction, and Dynamic Time Warping (DTW) utilized for the pattern matching [2].

Objective:

The Requirement for quick and broad application of Speech enabled systems is turning out to be evident Marathi is one of the native (regional) Indian languages. There are 12 vowels and 36 consonants present in Marathi languages. The objective of the research is to create the speech recognition system for Cerebral Palsy person to recognize digits in Marathi language.

Motivation:

Cerebral palsy can affect a person's Speech ability to finely coordinate the muscles around the mouth and tongue

that are required (need) for speech. Few people with cerebral palsy will most likely be unable to produce any sounds, others might have the option to produce sounds but have difficulty controlling their movement enough to produce speech that is clear and comprehended by others.

There is no any Speech recognition system Developed in Marathi language for cerebral palsy people. A system existing in Marathi language so to overcome this language barrier want to develop this system.

To utilize of technology for disable people to bridge the gap between society & disable person.

Literature Review:

The Objective of this survey is to report the right research work and development done in speech recognition and their applications to HCI .The special emphasis is given to studies done with respect to application in Indian language.

In this the design of speech recognition system, relies upon the different issues: such as definition of various type of speech classes, speech representation, feature extraction techniques, speech classifiers, database, language models and performance evolution. The problems that exist in Automatic Speech Recognition (ASR) and the different techniques developed by various research workers. while literature review observe mostly used like Mel-frequency Cepstral Coefficient (Mfcc) for feature extraction using for Pattern Matching, Dynamic Time Warping (DTW) to solve these problems using various technique /Method and algorithms.

II. Methodology and Experimental Work

Methodology

Speech recognition system performs two principal activities: signal modeling and pattern matching .signal modeling represents process of changing (converting) speech signal into a set of parameters. Speech when recorded and saved to the system, becomes digitized. Speech wave part (components) are called as features but before the feature extraction we have a need of data collection speech samples from speaker and then extracted feature for further analysis. The technique followed by us for the proposed work is shown in figure.

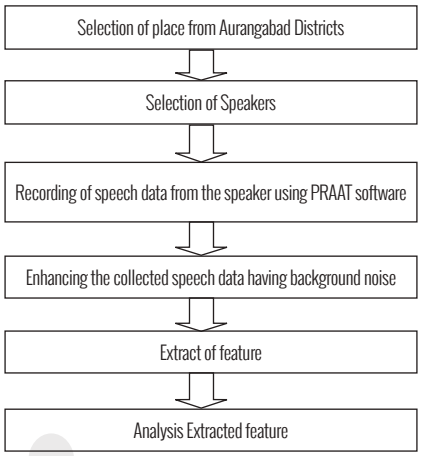


Fig. Methodology adopted for the proposed work

Implementation

1. Data collection Procedure

1.1 Speaker Selection:

The speech data will be created from the essence (Native) speakers of Marathi language. The favored speakers will be from region of Aurangabad district. they would be comfortable with reading & speaking the Marathi language the speaker are classified on the basic gender.

1.2 Speech Data Choice

I needed to have a restricted vocabulary to ensure the capture process was light weight, yet at the same time have enough verity for models trained on the data to potentially be utilize full for some applications I also needed the dataset to be usable in comparable ways to common proprietary collections like Digits this drove me to stop 10 common digits and Words as the core of our vocabulary.

1.3 Speech Data Collection:

Data Collection

To access our strategy,a list of 10 digit short phrases in Marathi was compiled consisting of numbers & Words , commands to a typical information–access applications each entry was selected because it is either a single digit or a short phrase given our goal of high accuracy, small vocabulary speech recognition the vocabulary size was kept to a maximum limit of ten digit one target language were chosen .[13]

1.4 Recording Environment:

Speech Data Recovery is first step to words working of Speech Recognition system. The proficiency of recognition develop on the Speech Data used to prepare the system. At this moment the measures taken for gathering Speech information, to developed a ground-breaking Speech Recognition System as a grammatical feature. To achieve a high sound viewpoint, the account was done in the normal (ordinary) room without uproarious sound and impact of reverberation. The look at recurrence for all Recording was set to be 16000 Hz at the room temperature. The speakers were ask to sit in front from the amplifier with the separation of around 12-15 cm. The Speech data was assemble with the help of receiver 'praat' programming utilizing the single channel. The preprocessing was finished with the assistance of computerized speech laboratory facility (CSL). The computerized speech laboratory (CSL) is one of the examination framework for discourse and voice. Computerized speech laboratory (CSL) is one of the Input/output recording gadget for a PC, which has exceptional element for positive acoustic estimation.

III. Experimental Work

III. I Creating Database-

Creating database is the most significant task. As speech is variant in nature, task oriented database is very basic

requirement in order to develop the system. Database was created for the proposed system.[5] Speech data acquisition is the initial step to words building of speech recognition system .The accuracy of recognition relies upon the speech data used to train the system[6].

The data base is creates consist of selected digit in Marathi language used in daily life .digits from १ to १० (१,२,३,४,५,६,७,८,९,१०.) And word selected are Here used the cerebral palsy person's database which has scattered by age wise. The basic need of this database is to develop an automatic speech recognition system. The accuracy of Speech recognition depends upon the nature of the discourse information gathered and the nature of training set data. The database is also created for this study. It is aim to developed speech recognition system that will recognize Marathi Digits.

Database of audio recordings of the persons with Cerebral Palsy and persons without cerebral palsy were taken[13]. First of all, Marathi digits 0 to 9 are recorded in the closed environment in the lab, with the help of the software Praat and by using the headphone with mike. This is very difficult task, as to have recordings at required frequency, at required amplitude, pitch and duration. Various utterances were taken before making it final for the recordings of the CP person.

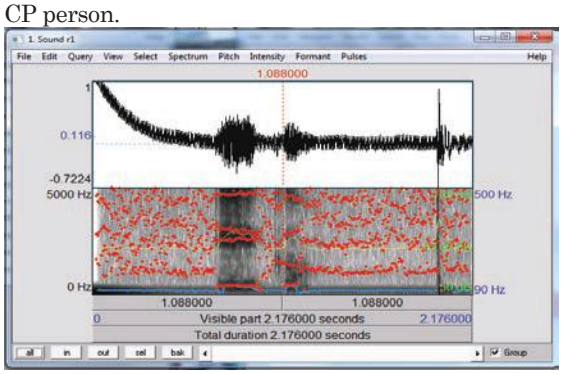


Fig. 1. Speech recorded in Praat for the word

The proposed ASR system will recognize Marathi digits and words, spoken by Cerebral Palsy disabled person that is the system would be speaker independent .So, the ASR system needs to be trained with speech of lot of people so that it can handle variation in accent and speaking style. Recordings of five CP persons of different age group, both male and female were taken, each digit and word were uttered 10 times. Similarly recordings of same digits and words, with five persons with no disorder in speech were taken. In this recording, all speakers were local (native) speakers of Marathi language[14].

In this recording, all speakers were local (native) speakers of Marathi language[14]. This group of speaker of is a combination of different gender, education level, and age. Secondly, the speakers with CP disability also belong to Marathi language as their mother tongue. Following are the words and digits in Marathi taken in to the work. This Database having two type of people one is normal and another one Cerebral Palsy which is having Speech dieses but both are native person which is familiar with Marathi Language and both Speaker having same Digit to speak.

Table 1. Marathi Digits.

Sr. No.	Digit	Gender	Utterances
1	१	M and F	10
2	२	M and F	10
3	३	M and F	10
4	४	M and F	10
5	५	M and F	10
6	६	M and F	10
7	७	M and F	10
8	८	M and F	10
9	९	M and F	10
10	१०	M and F	10

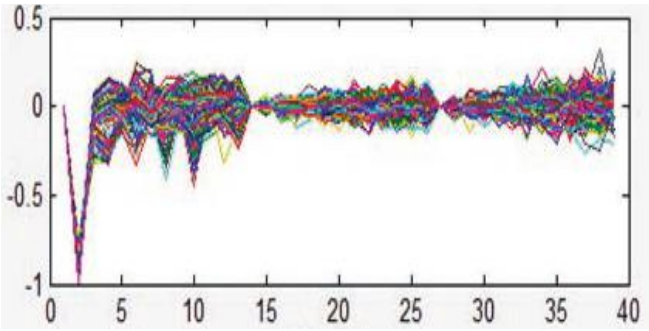
Automatic Speech Recognition can be achieved by a extensive variety of techniques. Basically, speech recognition starts with the digital sampling of speech. Most of the techniques include spectral analysis; e.g. LPC analysis (Linear Predictive Coding), MFCC (Mel Frequency Cepstral Coefficients), and many more. MFCC (Mel Frequency Cepstral Coefficient) is the next stage is recognition of phonemes, groups of phonemes and words [5]. After the MFCC (Mel Frequency Cepstral Coefficients) stage can be achieved by many processes such as DTW (Dynamic Time Warping). Automatic Speech Recognition system performs two fundamental operations: signal modeling and pattern matching [10]. Signal modeling represents process of converting speech signal into a set of parameters [9]. In this, features are extracted from the speech. Next step is Pattern matching is the task of finding parameter set from memory which closely matches the parameter set obtained from the input speech signal. The signal modeling involves with four basic operations: like spectral shaping, feature extraction, parametric transformation, and statistical modeling [9]. After database creation, next task was to apply feature extraction algorithm for looking into the details of the digitized speech signal. For this, most widely used technique of MFCC is applied and features are extracted.

III. II Features Extraction

The first step in any automatic speech recognition system is to extract features that is Recognize (identify) the component of the audio signal [15] This will identify the syntactic (linguistic) content [7]. The major purpose of the MFCC process is to mimic the behavior of the human ears [8]. The method of filters the spaced linear at low frequencies and logarithmically at high frequencies have been used in order to capture the phonetically important characteristics of the voice [7]. MFCC extracts 13 (12parameters +1energy) features, these features can be applied with delta (26 features)

and double delta (39 features), for more recognition rate. [7] Following figure shows the window for the speech signal recorded using Praat. These audio recordings were called in Matlab to extract the MFCC features, following figure shows the plot of the extracted features [12].

MFCC (Mel Frequency Cepstral Coefficients) as the feature extraction technique The different values initialized during the computation of MFCC ((Mel Frequency Cepstral Coefficients) following.



MFCC features

After Feature Extraction Sampling frequency using 16000 and Hamming window is used with the window length of 25 millisecond per window given step time 10 millisecond and the minimum frequency is 0 (Zero) (lowest band edge mel filters(HZ) and maximum frequency 4000(the highest band edge of mel filters(HZ)set).

III. IV Fast Fourier Transformation (FFT):

Fast Fourier Transformation converted each frame N samples from time domain in to the frequency domain (DFT), which is defined on the set of N samples $\{X_n\}$, as follows.

$$X_k = \sum_{n=0}^{n-1} Xn\, e^{-j2\pi kn/N}, \quad K=0,1,2\ldots N-1$$

In general X_k are complex numbers and we only consider their absolute values (frequency magnitude). The resulting sequence $\{X_k\}$ is interpreted as follow positive frequencies $0 \le f < F_s 2$, correspond to values $0 \le n < N/2-1$, while negative frequencies $-F_s/2 < f < 0$ corresponds to as spectrum or periodogram. [15]

III. V. Pattern Matching

After features are extracted, these needs to be matched. There are various pattern matching techniques, here, Dynamic Time warping (DTW) is a well know dynamic programming technique for finding the best alignment between two times series pattern is used [9]. DTW calculates the minimum distance between two words [12].

A local distance matrix is form for all the segment in the sample word and template word .This warping between two time series can then be used to find matching region between the two time series or to determine the similarity between the two time series [6]. Creation of the database, feature extraction and pattern matching is the main process in this speech recognition [9].

As speech disorder is to be recognizing, it is very challenging task. The speech of the CP person is matched with the speech of the person without speech disorder. Further, the actual recognition rate is to be found out. But, as MFCC features are taken double delta time, the recognition rate will be increased. This type of system provides a prototype of an interface that will recognize the speech of speech disordered person with CP.

The significance of this work will be that, this can serve as the input for further linguistic study, also this can be the foundation to develop the various interfaces in Marathi language, such as phone dialer system, hands free operation system etc.

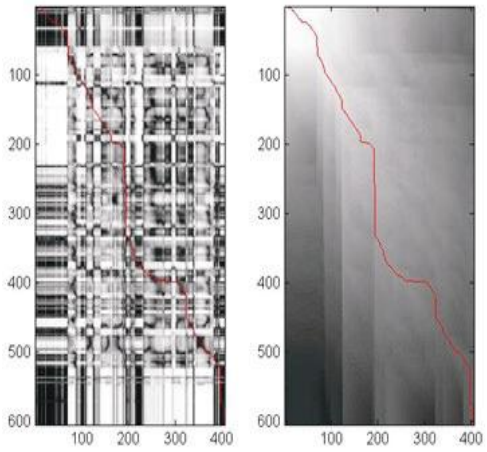


Fig. 3. DTW of two patterns of the same word.

Future Scope

Once Digit Recognition is done we can develop different (various) applications for Cerebral Palsy (CP) people. Like digit recognition for mobile dialing, playing simple information. Using various combination of feature extraction techniques we can develop our database more accurately and efficient.

Conclusion

The objective of this work is mainly to develop ASR System In Marathi Language for speech disorder of the CP person. In order to meet this objective, the first necessary step was to develop the database required. This system is intended at improving on the current Human-computer interface by developing a voice interface, which will prove to have so many advantages to the applications for the speech disorder. Today, even though much work in done regarding HCI in Speech Recognition; it is limited up to English language only. The concept of making these advanced systems reach up to speech disabled person and solve the problem of communication gap is the main problem handled over here.

References

1] Ingyin Khine, Chaw Su "speech recognition system using MFCC and DTW" International Journal of Electrical, Electronics and Data Communication, ISSN(p): 2320-2084, ISSN(e): 2321-2950 Volume-6, Issue-7, Jul.-2018, http://iraj.in.

2] Mohd Hafidz Mohamad Jamil, S.A.R. Al-Haddad, CheeKyun, "A Flexible Speech Recognition System for Cerebral PalsyDisabled", https://link.springer.com/ chapter/10.1007/978-3-642-25327-0 5.

3] Mr. Kashyap Patel Dr. R. K. Prasad "Speech Recognition and Verification Using MFCC & VQ" International Journal of Advanced Research in Computer Science and Software Engineering www.ijarcsse.com Volume 3, Issue 5, May 2013 ISSN: 2277 128X.

4] Shweta Bansal, "Isolated Word Speech Recognition System for Object identification", International journal of Scientific and engineering research , Volume 4, ISSUE 1, January-2013.

5] Rahana fathema, Raseema P.E. "Gammatone-Cepstral-Coefficient-for Speaker", International journal of scientific and engineering research, volume 4, Issue 10, October-2013 ISSN 2229-5518

6] Lindasalwa muda, mumtaj begam and i. elamvazuthi "Voice Recognition Algorithms Using Mel-Frequency Cepsrtal Coefficient", Journal of computer, value 2, Issue 3, March 2010, ISSN 2151-9617 https://sites google.com/site/Journal of computing.

7] Prasanth P.S. "Speaker Recognition Using Vocal Tract Features", International Journal of Engineering Inventions e-ISSN:2278 7461, P-ISSN: 2319-6491 Volume 3, Issue 1 (August 2013) PP. 26-30.

8] Risha Mal, R. K. Sharma "Intoxicated speech Detection using Mfcc feature Extraction and Vector Quantization", International Journal of electronic and electrical engineering ISSN 0974-2174. Volume 7. Number 3(2014). PP. 269-280.

9] Manish P. Kesarkar "Feature extraction for speech Recognition", M.tech credit seminar Report, electronic systems group, EE. Dept, IIT Bombay , Submitted November 2003.

10] S.A. Mohd yusof, P.M.Raj "Speech Recognition Application Based on Malaysian Spoken Vowels Using Autoregressive Model of the vocal tract", Processing of the international conference on electrical engineering and informatics institute technology Bondug, Indonesia. June 17-19, 2007.

11] Irfan Ahmad, Nasir Ahmd et.al "The Development of isolated words pashto Automatic Speech Recognition System", http:// www.researchgate.net/publication/235726972 September 2012.

12] https://www.abidibo.net/blog/2013/10/02/how-create-your-own-speech-recognition application-tasker.

13] pete warden " Speech commands: A dataset for limited-vocabulary Speech Recognition", google brain, mountain view, coliformia petevarden@google.com April 2018, 1804,03209

14] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, Vishal B. Waghmare, "Indian Language Speech Database: A Review", Dept. of CS and IT Dr. B. A. M. University, Aurangabad-431004, India, International Journal of Computer Applications (0975 – 888) Volume 47– No.5, June 2012.

15] Pukhraj P. Shrishrimal, Ratnadeep R. Deshmukh, "Design and devlopment of Spoken Marathi Isolated words Database for Agriculture Purpose and its Analysis" Dept. of CS and IT Dr. B. A. M. University, Aurangabad-431004, April 2013.

16] Manasi Ram Bheti, Dr .S. C. Mehrotra "Speech Recognition System For Rural Applications Using Interactive Voice Response System", Dept. of CS and IT Dr. B. A. M. University, Aurangabad-431004, Feb-2016.

About the Authors

Maheshwari A. Ambewadikar Working as M.Phil. Research Student Department of Computer Science And Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra. She can be reached at mahiambewadikar@gmail.com

Dr. Manasi Ram Baheti [Working as Assistant Professor Department of Computer Science And Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra. She can be reached at manasibaheti@gmail.com

Multi-Dimensional Encryption Type 11 method for Video Files Security

Boopathy D.[1], M Sundaresan[2]

[1] Research Scholar, Department of Information Technology, Bharathiar University, Coimbatore.
Email: ndboopathy@gmail.com
[2] Professor and Head, Department of Information Technology, Bharathiar University, Coimbatore.
Email: bu.sundaresan@gmail.com

Video encryption is a course of action that is used to hide and maintain the information of a video file in a secure manner. The Multi-Dimensional Encryption Type 11 has been implemented and tested previously on different image file formats. This paper extends that implemented Multi-Dimensional Encryption Type 11 into the different video files, due to positive progressive results of this Multi-Dimensional Encryption Type 11 on the different color and grayscale image files. The features of the proposed video encryption have been compared with the existing encryption techniques that include Advanced Encryption Standard and Data Encryption Standard. The standard comparison parameters were tested in the image file itself, so in this paper the time taken to encrypt the file and time taken to decrypt the file only were considered. The features of the proposed encryption technique on the video files show that the Multi-Dimensional Encryption Type 11 is possible and it can provide better security than the existing video encryption techniques.

Keywords : Video Encryption, Multi-Dimensional Encryption Technique, Deep Fake, Encryption and Decryption, Data Security.

1. Introduction

The encryption and decryption is the basic concept which was used in the ancient period to protect the information from the intruders and illicit persons. The digital and online sectors are now made the word *"security"* as vital one. When the internet was introduced, the network concept was treated as a medium to exchange the information between two distinct nodes. The importance of *"security"* was not realized at that time. Later on when the network concept rolled out to commercial and public use, the word *"security"* became more important one. The encryption concepts were implemented on documents, audios, content base information, passwords, secret tokens and many other things. However, the encryption concepts are used rarely in the video files. Due to the *"Deep Learning's Deep Fake"* concept, the video files also need to be seriously taken care of during encryption implementation.

The Deep Fake concept can learn the user's facial expressions and their words pronunciation from their video. Once the deep fake concept learned that emotions and reactions of the users from their video files, then it can easily create fake video files. The video files need encryption technique [1, 2] support to avoid unwanted / fake users who would make the video files vulnerable.

1.1. Types of Encryption Algorithms

There are two types of encryption concepts available and they are: *Symmetric key Algorithms and Asymmetric key Algorithms.*

Symmetric-key algorithms are meant for cryptography, which use the same cryptographic keys for both encryption of plain text and decryption of cipher text. The keys may be identical or there may be a simple transformation to go between the two keys. The Data Encryption Standard and Advanced Encryption Standards are a few examples of Symmetric-key algorithms.

Asymmetric encryption uses a pair of keys (one public and one private). Anything encrypted with one of the keys, can only be decrypted with the other. The public key can also be used for sender authentication in certain scenarios. RSA Algorithm, Diffie Hellman Algorithm and Elliptic Curve Cryptography are some of the algorithm examples of Asymmetric encryption.

The objective of the research Work is to check if it is possible to validate, test and implement the multi-dimensional encryption type 11 on video files. *"Deep Fake from the deep learning"* concept and existing security threats for the video files have motivated the researchers to extend the multi-dimensional encryption type 11 research work on the video files.

The section two explains the review of the related literature, and section three illustrates the methodology

of the multi-dimensional encryption type 11 on the video files. Section four discusses the results, discussions and comparisons between the proposed and the existing techniques. Finally, the fifth section concludes the research with its conclusion and future enhancements.

2. Review of Related Literature

Babatunde et al. have done the survey on some of the existing video encryption techniques with an explanation on the concept of video compression. The review of the paper which also explores the performance metrics used in the evaluation and comparison of the performance of video encryption algorithms [3].

Madhvi Soni et al. explain that the encryption algorithms which were developed to secure text data are not suitable for video content due to the particular requirements and special properties of video data. The authors of this paper have proposed a classification scheme and the description of various video encryption methodologies with their performance evaluation [4].

Abomhara et al. illustrate the description and comparison between encryption methods and representative video algorithms which were already presented. With due respect not only for the encryption speed but also the security level and stream size, it must be pointed out that a trade-off between quality of video streaming and choice of encryption algorithm were shown in their paper. It denotes that achieving an efficiency, flexibility and security is a challenge for researchers in the video encryption [5].

Zao Mobile Application is one of the Deep Fake Applications which can be used to create a fake video file, based on the user's emotions and reactions from their available video files [6 - 8].

3. Multi-Dimensional Encryption Type 11 (MDET) for Video Files

The Multi-Dimensional Encryption Type 11 (MDET) was used in the Secured Cloud Data Storage Prototype Model (SCDSPM) and it has been proposed with four modules to overcome some of the data-security-related issues in cloud storage and it also tries to maintain the safety measures within the data ownership at the country level.

In the Encryption and Decryption part of the SCDSPM the new encryption and decryption scheme was proposed, coded, designed and implemented in testing files and the results obtained were compared with some of the existing encryption and decryption algorithms [9, 10]. Those comparisons are shown that the proposed new encryption and decryption scheme is better than existing algorithms. In the next level, the proposed new encryption and decryption scheme in the SCDSPM also can be extended and implemented into the Video files too. The pseudo code and procedure for the video file encryption are explained and illustrated in the upcoming sub sections under the third section. Figure 01 (a) and 01(b) shows the proposed model's flow.

3.1. Flow Chart of MDET

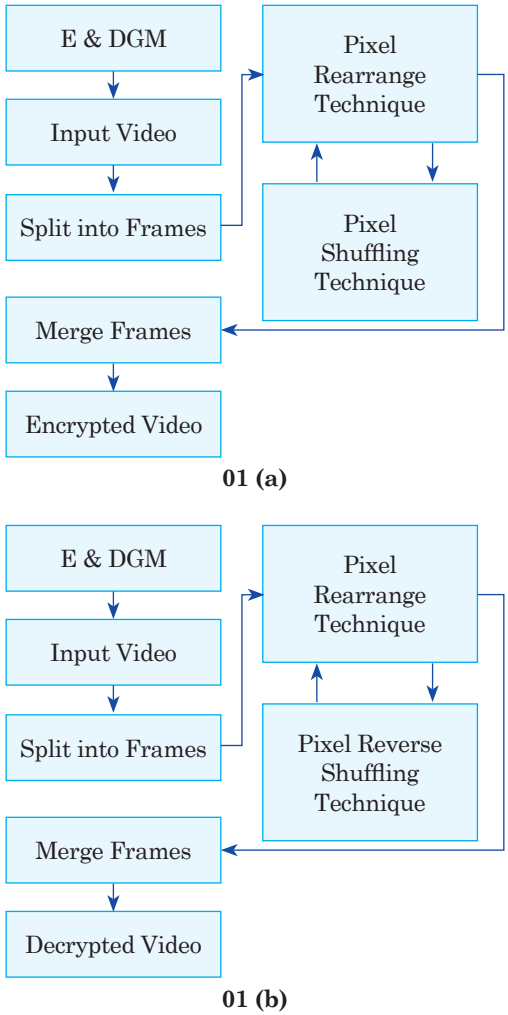


Fig. 01 (a) Encryption Flow Chart and Fig. 01 (b) Decryption Flow Chart

3.2. Pseudo Code

The pseudo code for the proposed video encryption is illustrated in Table 01 and the pseudo code for the decryption is illustrated in Table 02.

Table 01. Pseudo for Encryption

Step 01:	Start the process
Step 02:	Get the input file from the user
Step 03:	Verify the video file format
Step 04:	Analyze the number of the frames per second in that video file
Step 05:	Split the video file format into frames per second and store it as sequence of images

Step 06:	Select and split the images into tri band and store it as respective tri-band color images
Step 07:	Apply the MDET process on all the sequence of tri band images up to the last sequence of tri band image. For Decryption of a encrypted video file, the reverse MDET process need to apply on that encrypted video file
Step 08:	Merge the processed tri band color images to get a frame of encrypted image
Step 09:	Merge all the encrypted images into sequence of frames to get a video file
Step 10:	Prepare the decryption key automatically, based on the images constant pixel value
Step 11:	Verify the encrypted video file
Step 12:	Stop the process

Table 02. Pseudo for Decryption

Step 01:	Start the process
Step 02:	Get the Encrypted video as an input file from the user
Step 03:	Verify the video file format
Step 04:	Get the decryption key from the user to decrypt the video file
Step 05:	Verify the decryption key: If decryption key is correct, then continue the process, else send key error message to the user
Step 06:	Analyze the number of frames per second in that video file
Step 07:	Split the video file format into frames per second and store it as sequence of images
Step 08:	Select and split the images into tri-band and store it as respective tri band color images
Step 09:	Apply the Reverse MDET process on all the sequence of tri band images up-to the last sequence of tri band image. For Decryption of an encrypted video file, the reverse MDET process need to apply on that encrypted video file
Step 10:	Merge the processed tri band color images to get a frame of encrypted image
Step 11:	Merge all the decrypted images into sequence of frames to get a decrypted video file
Step 12:	Verify the decrypted video file
Step 13:	Stop the process

The above mentioned procedure is applied to the encryption process of a video file and the same process can be applied to the encrypted video file to get a decrypted video file. However, in decryption process, the decryption technique needs to be applied instead of applying encryption technique in the above mentioned procedure.

3.3. Encryption and Decryption Scheme

The user needs to upload the video file for the encryption purpose. Once the video file is received from the user, then it needs to be validated for the purpose of applying the encryption technique. Once the video file is validated, then the video file's frames are calculated per second and those frames will be split and arranged in a sequential order. After that, the MDET needs to be applied on all the sequentially arranged frames. When the MDET [10] finishes its work,

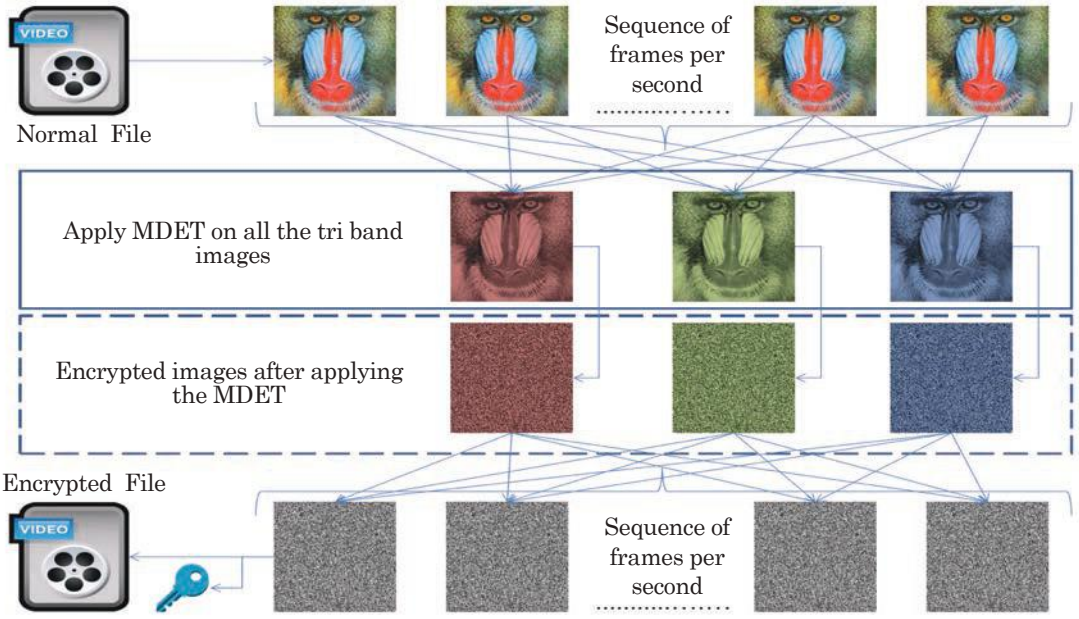


Fig. 02 : MDET-based framework for Video Encryption

then the sequentially arranged frames need to be merged and saved as a video file to get the encrypted video file. At the end of the process the decryption key for that video file will be automatically generated and provided to the user of that video file. All these things also will be taken care of by MDET. Figure 02 illustrates the MDET based encryption framework for Video file

For decryption, the same procedure needs to be followed, but at the beginning stage of the process the user needs to provide the decryption key. When the decryption key is validated, the reverse MDET will be applied on that encrypted video file to get the decrypted video file.

3.4. Test Files

The three files with different time were taken for the testing purpose with the proposed methodology. Those files are mentioned in the table 03.

Table 03. Test File Details

S. No.	File Name	File Length
01	First File	01 Second
02	Second File	02 Seconds
03	Third File	03 Seconds

The file mentioned in the table 03 which includes *"one second video file, two second video file and three second video file"* are taken for testing purposes for proposed and existing methods. The file format of the video files is Windows Media Video format.

4. Results and Discussion

There are many performances analyzing metrics available to compare the performance between the different encryption techniques. During the process, efficiency, flexibility and security are taken as important performance analyzing metrics. The above mentioned scheme is compared with the efficiency of the AES and DES techniques only.

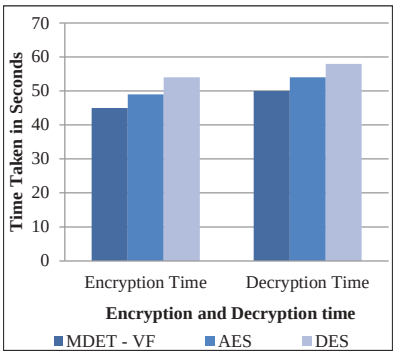
The image comparison with standard parameters that include the Mean Square Error, Peak Signal Noise Ratio, Information Entropy, Correlation Coefficient and isequal() are already tested on the image files [09, 10]. And also the Number of Changing Pixel Rate and the Unified Averaged Changed Intensity were also compared to verify the strength of the proposed model's efficiency with existing models. These comparisons are also already done and shown in the author's existing papers [09 - 11].

The proposed and the above mentioned two techniques were tested on the sample files. The efficiency of the time taken to encrypt and decrypt the video files are shown and illustrated in Table 04 and Figure 03 (a), 03 (b) and 03 (c). The results and illustrations show that the proposed MDET for the video file is better in efficiency than the existing AES and DES techniques.

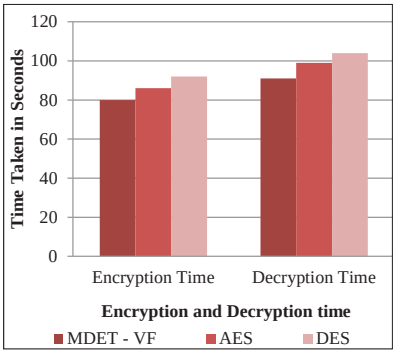
While comparing the proposed method with the existing method in the one second video file, the proposed method MDET took 7.4% and 17.3% lesser time for encryption than AES and DES methods and also the proposed method took 6.3% and 14.6% lesser time for decryption than AES and DES methods.

For the two second video file, the proposed method MDET took 7.6% and 17.8% lesser time for encryption than AES and DES methods and also the proposed method took 6.6% and 14.2% lesser time for decryption than AES and DES methods.

For the three second video file, the proposed method MDET took 7.9% and 13.5% lesser time for encryption than AES and DES methods. For decryption, the proposed method took 9.2% and 9.4% lesser time for decryption than AES and DES methods.



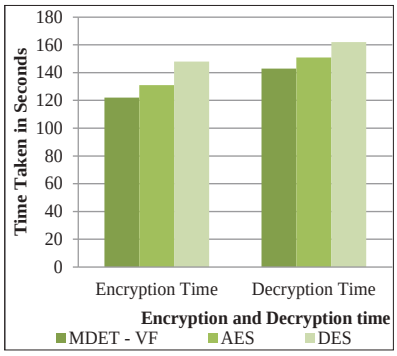
03 (a)



03 (b)

Table 04. Encryption and Decryption Time to Process a File

Time Taken for the process	First Test File			Second Test File			Third Test File		
	MDET - VF	AES	DES	MDET - VF	AES	DES	MDET - VF	AES	DES
Encryption Time	45	49	54	80	86	92	122	131	148
Decryption Time	50	54	58	91	99	104	143	151	162



03 (a)

Fig. 03 (a) ***Encryption and Decryption time for one second video,*** **03 (b)** ***Encryption and Decryption time for two second video,*** **03 (c)** ***Encryption and Decryption time for three second video***

5. Conclusion and Future Enhancement

Encryption algorithms and techniques need to be redefined with certain time intervals, and only then the security standards can be maintained in an effective manner. The proposed MDET has already performed well on the image files. Based on the performance, the MDET was proposed and tested on the one second sample video file, two second video file and three second video file. The testing scenario provides the positive result for the MDET. The MDET needs to be tested on different format video files with different video duration time.

The proposed method was presented tested on the sample files only, the obtained results were showing that this proposed MDET can be implemented on different kinds of file formats too. When coming to the application sector, the different application sectors who are seeking the secured manner to transfer their confidential video file in the unsecured network also can be done by using this proposed MDET. These things can be taken as an extension for a future work.

References:

1] https://en.wikipedia.org/wiki/Encryption. Accessed on 14/029/2019.

2] https://en.wikipedia.org/wiki/Video_Content_Protection_System. Accessed on 14/029/2019.

3] Babatunde A.N., Jimoh, R.G., Abikoye O.C., Isiaka B. Y., *"Survey of Video Encryption Algorithms"*, Covenant Journal of Informatics & Communication Technology. Vol.05 No.01, June 2017, pp. 65 – 80.

4] Madhvi Soni , Sapna Chaudhary, *"A Survey of Video Encryption Methodologies"*, International Journal of Innovative Research in Computer and Communication Engineering, Vol.03, No.10, October 2015, pp. 9858 – 9863.

5] M. Abomhara, Omar Zakaria, Othman O. Khalifa, *"An Overview of Video Encryption Techniques"*, International Journal of Computer Theory and Engineering, Vol.02, No.01 February 2010, pp. 103 – 110.

6] https://www.theverge.com/2019/9/2/20844338/zao-deepfake-app-movie-tv-show-face-replace-privacy-policy-concerns. Accessed on 14/09/2019.

7] https://www.news18.com/news/buzz/zao-a-new-chinese-ai-app-lets-you-swap-your-face-with-any-celebrity-in-8-seconds-2295115.html. Accessed on 14/09/2019.

8] https://www.engadget.com/2019/09/02/zao-face-swapping-app-privacy-issues/. Accessed on 14/09/2019.

9] Boopathy.D and Dr. M. Sundaresan, *"Multidimensional Approach for Securing Images on Cloud"*, ICGST Journal of Graphics, Vision and Image Processing Journal, Vol.17, No.02, December 2017, pp. 31 – 38.

10] Boopathy.D and Dr. M. Sundaresan, *"A Novel Multi-Dimensional Encryption Technique to Secure the Grayscale Images and Color Images in Public Cloud Storage"*, Journal of Innovations in Systems and Software Engineering (A NASA Journal), Vol.15, No.01, January 2019, pp. 43 - 64.

11] Boopathy.D and Dr.M.Sundaresan, *"Secured Cloud Data Storage – Prototype Trust Model for Public Cloud Storage"*. International Conference on ICT for Sustainable Development (ICT4SD-2015), Ahmadabad, India, July 3 – 4, 2015, pp 329 - 337

About the Authors

Mr. D. Boopathy (CSI Life Membership No. I1500567) is a Research Scholar doing his PhD in Computer Science in the Department of Information Technology at Bharathiar University. He is qualified with M.Sc. (IT) and MCA from Bharathiar University. He did his Master of Philosophy in Computer Science at Dr. G.R.D College of Science, Coimbatore. His areas of interests are Information Security, Data Privacy and Cloud Computing. He is a Life Member of Computer Society of India and Indian Science Congress Association. So far he has co-authored 4 book chapters for 4 edited books (2 for IGI Global USA).

Dr. M. Sundaresan (CSI Life Membership No. I0007384) is currently Professor and Head of the Department of Information Technology at Bharathiar University, Coimbatore, India. He holds PhD in computer science. He has contributed more than 75 research papers in different areas of Computer Science such as Image Processing, Data Compression, Natural Language Processing, Speech Processing and Cloud Computing in reputed journals. He is a Senior and Life Member of Professional Bodies such as Computer Society of India, Indian Science Congress Association, and Indian Society for Technical Education and IACSIT. He is also in editorial board of five journals. He is the Past Sectional President for Information and Communication Science & Technology (including Computer Sciences) section in 105th Indian Science Congress. He is the Regional Vice President for Regional VII in Computer Society of India. So far he has authored 4 book chapters for 4 edited books (2 for IGI Global USA).

Distinguishing Healthy and Infected Vegetable Crops using Hyperspectral Leaf Reflectance

Anjana Ghule[1], Prof. Dr. R. R. Deshmukh[2] and Chitra Gaikwad[3]

[1] Assistant Professor, IT Department, GEC Aurangabad. Email: anjanaghule@gmail.com
[2] Professor, CS and IT Department, Dr. B.A.M. University Aurangabad. Email: rrdeshmukh.csit@bamu.ac.in
[3] Assistant Professor, IT Department, GEC Aurangabad. Email: gaikwadchitra@gmail.com

Hyperspectral Remote Sensing provides data over a large number of contiguous wavebands. Primary objective of this research was to reduce the dimensionality and distinguish healthy and infected vegetable plants by selecting significant spectral region and subset of wavelengths. For the experimentation, spectral measurements of healthy and diseased leaves of Brinjal, Cluster Beans and Long Beans were divided into five regions of electromagnetic spectrum, i) Visible: 350nm-670nm ii) Red Edge: 671nm-780nm iii) Near Infrared: 781nm-1000nm iv) Shortwave I: 1301nm-1500nm and v) Shortwave II: 1701 nm-1900 nm. PCA and LDA methods were used as pre-filters for dimensionality reduction on each region, before applying Random Forest classifier. The results obtained revealed that classification performance of Visible and Red Edge regions was better than NIR, SW1 and SWII regions. The PCA method provided better accuracy and kappa values as compared to LDA.

Keywords: Hyperspectral Remote Sensing, Spectral subset selection, Vegetable crops, Dimensionality reduction, Classification

1. Introduction

India, the second largest populated country and majority of Indians prefer vegetarian food. Broadly we can say that the health of Indians depends on healthy crops. Various types of vegetables are grown in different regions of India. Considering the public health and economy of the country, it is essential to monitor the health of vegetables and crops.

Hyperspectral Remote Sensing has extensive applications in the vegetation and agriculture over conventional methods [1]. Crop species discrimination, monitoring of crop growth, crop identification, crop type classification, vegetation health monitoring is also very crucial in the field of agriculture[2]. Plant health or canopy health has direct impact on the ecosystem. Major part of the plant above the soil is covered with leaves. Plant health can be monitored by studying leaves and we can consider leaves as representatives of the canopy. For the study of crop health, it is important to characterize and quantify vegetation parameters. Nutrient deficiencies, water deficiencies, floods and disease are major factors of crop stress [3].

Crops have different biophysical and biochemical characteristics, and this has a direct impact on the spectral profile. Measurement and study of attributes and their inter-relationship can provide significant information about plant productivity, plant health, stress and availability of nutrients. The visible domain is in the range of 400nm -700nm., major photosynthetic pigments (chl a) chlorophyll a and (chl b) chlorophyll b shows absorption in this region. In 700-1300nm NIR region, absorption is very low as cellulose and leaf pigments are almost transparent. 1300-2500nm SWIR region is largely influenced by water contents. Other parameters such as protein, cellulose, lignin and starch also influence SWIR region [3][4].

Red Edge Position is extensively used for the assessment of crop stress. Sudden inflection is observed between 680 - 780 nm in the reflectance spectra of vegetation, this is called as Red Edge Region. Red Edge Position is the wavelength at which maximum magnitude of First Derivative curve (FD) is observed in the red-edge region [5].When single peak in the First Derivative curve is obtained Maximum First Derivative (MFD), Maximum First Derivative Spread –mean (MFDS-m) methods are used to find Red Edge Position for the study of crop health [6].

Hyperspectral reflectance provides measurements over large number of narrow and contiguous bands, in various regions of electromagnetic spectrum. However, high correlation and redundant neighboring wavebands makes further analysis challenging [7]. Therefore, to improve the classification speed and accuracy, selection of significant

wavelengths is used as a pre-requisite for eliminating curse of dimensionality in the applications of hyperspectral remote sensing. [8][9].

To discriminate between healthy and unhealthy crops, reflectance at specific wavelengths in Visible, Red Edge, NIR and SWIR regions are found useful. In ground based hyperspectral remote sensing, researchers have also used various non-REP methods to assess the crops health. Principal Component Analysis and Linear Discriminant analysis were used to find significant wavebands and hyperspectral vegetation indices to discriminate between the potato species and to assess disease, water and nutrient stress of potato crop [10]. To Discriminate Orchard Species, ANOVA and PCA and Random Forest classifier were applied one by one for reducing the number of wavelengths and obtained optimal discriminating wavelengths without losing significant wavelengths, Discriminant Analysis was performed to check spectral separability [11]. For the discrimination between healthy and infected oil palms, both spectral reflectance and first derivatives transform of the samples was recorded. Optimal spectral bands were selected using ANOVA and, classification was done using a maximum likelihood classifier [12].

Random Forest classifier is an ensemble machine learning method and has been widely used for Hyperspectral data. This tree based classifier has an ability of noise reduction and removing irrelevant features. Here we have also used two pre-filtering approaches PCA and LDA for reducing dimensions before the classifier is build. Performance of PCA and LDA was compared by classification accuracy measures.

2. Material and Methods

2.1 Study area and sample collection

Aurangabad city (Lat 19.846011 and Long 75.282556) is situated in Maharashtra, India. In-field hyperspectral spectral data was collected 27th August 2016 from a farm of 100 Acres, where different vegetables are grown. Healthy and infected plants of Brinjal, Cluster Beans and Long Beans were identified by visual judgment. Leaves without stalks were hand-picked and transferred in sealed polythene bags. Total 25 leaf samples of each were collected. Leaf spectra were measured within 3 hrs after collection.

2.2 Spectral Data Acquisition

2.2.1 Leaf Spectra Measurements

The spectral reflectance of leaf sample was acquired with an ASD FieldSpecPro spectroradiometer, in the Multispectral Lab of Computer Science and Information Technology Department of Dr. Babasaheb Ambedkar Marathwada University, Aurangbad. This instrument acquires reflectance in 350-2500 nm spectral range and sampling step is 1 nm.

Spectral measurements were recorded in the dark room using a fiber optic with a 8° field of view. The leaf blade was kept 15cm below the sensor. The target was illuminated by 50 W halogen lamp. White reference scans were done for optimization of the signal and calibration of accuracy. The spectral data was collected in the range of 350 to 2500 nm, containing reflectance at 2151 wavelengths. Reflectance spectra of the samples was recorded with 10 iterations each using RS3 software [13].

Fig:1 shows spectral reflectance curves of healthy and infected leaves of Brinjal, Long Beans and Cluster Beans in the range of 350nm-2500nm.

2.2.2 Spectral Data Preprocessing

The raw .asd files were further processed by ViewSpec Pro version 6.2. Splice Correction was done for removing noise and interpolating missing data. Reflectance curves were exported as text data files. These data files were transferred to Excel software and files were converted to .csv format. In spectral data processing, if we represent the dataset by matrix X, then each row of the matrix Xj is contains the intensity value corresponding each wavelength for ith sample. Each column Xi, is an intensity value of a wavelength for all samples.

2.3. Dimensionality Reduction and Classification

Hyperspetral data has a curse of dimensionality. A reflectance spectrum comprised 2151 wavelengths was available for experimentation, but not all were useful. Selection of significant wavelengths is a crucial task. This study aspires to obtain subset of significant wavelengths for discrimination between healthy and unhealthy vegetable leaves. For achieving this, dimensionality reduction process was carried out and most significant features were selected in the different regions of the electromagnetic spectrum. LDA (Linear Discriminant Analysis) and PCA (Principal Component Analysis) were applied to find discriminative wavelegths. Performance of these methods was evaluated using Random Forest supervised classifier on the basis of Accuracy and Kappa values.

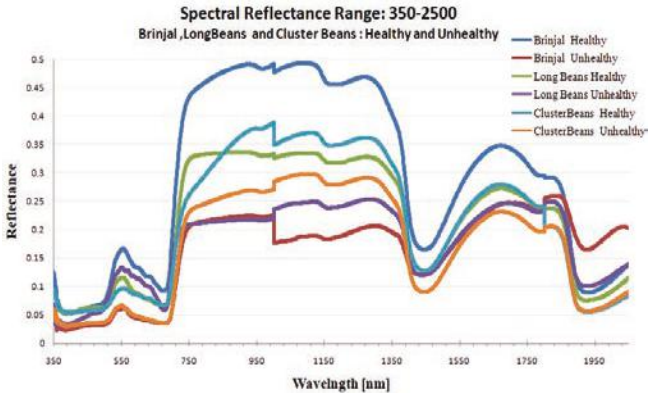


Fig.1: Spectral Reflectance of healthy and unhealthy leaves of Brinjal, Long Beans and Cluster Beans in the range of 350 nm -2500nm. (Wavelength vs Reflectance)

2.3.1. Linear Discriminant Analysis (LDA)

LDA is a supervised dimensionality reduction method which is based on Fisher Criterion. This method aims to find a linear transformation by projecting high dimensional space onto a low-dimensional space, in which, it maximizes the Between-Class scatter and minimizes the Within-Class scatter. Two measures are defined for all samples in both the classes i) Sb: between class scatter matrix and ii) Sw : within-class scatter matrix.

$$S_b = \sum_{k=1}^{c} n_k (\mu_k - \mu)(\mu_k - \mu)^T \quad (1)$$

$$S_w = \sum_{k=1}^{c} \sum_{i \in C_k} (x_i - \mu_k)(x_i - \mu_k)^T \quad (2)$$

where C_k is the index, μ_k is mean and n_k is number of samples, in the k^{th} class respectively.

$$\mu = \sum_{k=1}^{c} n_k \mu_k \quad (3)$$

is the overall mean of original data space[14][15].

2.3.2. Principal Component Analysis (PCA)

PCA transforms original data set X, containing *p* column vectors into another dataset Y is having *d* column vectors, where d< <p. First Principal Component has the highest variance with original data set. The second PC, is orthogonal to the first PC and also has highest variance to the first PC. Likewise each PC is orthogonal to the previous one, this property eliminates the problem of colinearity. This variance is captured in first few PCs' which represents the original data set. Higher numbered PCs are ignored, as they contain very less information [16].

If Y_1 is the first single largest variance among all linear combination, it is called as first PC and calculates as

$$Y_1 = e_1^T X \quad (4)$$

i^{th} PC is calculated as,

$$Y_i = e_i^T X \quad (5)$$

where e_i is called i^{th} loading vector and Y_i is i^{th} PC.

2.3.3. Random Forest Classifier (RF)

Random forest creates set of multiple decision trees. It selects samples randomly and generates separate tree for each sample. Prediction result of each tree is used to provide vote for selecting best classification. Prediction results having majority of votes is selected. RF does not suffer from overfitting because of randomly selected samples and number of trees participating in voting by prediction result. Given a training set $X = x_1, ..., x_i$ with response variable $Y = y_1, ..., y_i$, discrimination function is defines as

$$H(x) = argmax_Y \sum_{i=1}^{k} (I(hi(X, \theta_k) = Y) \quad (6)$$

Where I.() is the indicator function, hi is i^{th} single decision tree, Y is the class label and $argmax_Y$ is the Y value by maximizing

$$\sum_{i=1}^{k} (I(hi(X, \theta_k) = Y) \quad (7)$$

3. Experimentation and Results

As discussed in section 2.2 data set is pre-processed and following steps were carried out for further experimentation 1) Mean of multiple iterations of each sample was obtained 2) Reflectance spectra was divided into five distinct regions of electromagnetic spectrum and subset of wavelengths from different regions were selected viz. i) Visible : 350nm-670nm ii) Red Edge : 671nm-780nm iii) Near Infrared : 781nm-1000nm iv) Shortwave I: 1301nm-1500nm and v) Shortwave II: 1701 nm-1900 nm.

Further processing was done with Scikit-Learn library of python. In the first step, training and testing data was split. In the second step, constants and quasi-constant features were removed. In the third step, duplicate variables were removed. In the fourth step highly correlated features were detected and removed.

To reduce processing time and improve accuracy LDA and PCA were applied for dimensionality reduction, before applying Random Forest classifier. First five components of LDA and PCA are used for classification.

In case of Brinjal, when reflectance spectra of healthy leaves was visually compared with diseased spectra, decrease in the reflectance magnitude in the visible, Red Edge , NIR regions and increase in the reflectance of SWII regions was observed, where as not much difference was observed in the first water absorption SWI region. As shown in Table no 1, highest accuracy of .95 was obtained in the visible region in both cases. Here minimum classification accuracy was obtained in the NIR region.

Table 1: Comparison of classification accuracy of Brinjal

Regions	Filter	Accuracy	Kappa
Visible	LDA	0.95	0.89
	PCA	0.95	0.9
Red Edge	LDA	0.89	0.75
	PCA	0.9	0.76
N IR	LDA	0.61	0.11
	PCA	0.65	0.14
SW I	LDA	0.89	0.75
	PCA	0.88	0.75
SW II	LDA	0.89	0.75
	PCA	0.94	0.75

In Spectral reflectance of Cluster Beans, it is seen that reflectance magnitude of visible, RE and NIR region was decreased, whereas increase in reflectance of water absorption SW1 and SW2 region. But in some leaves difference in the

NIR region of some diseased leaves samples was observed. Table 2 displays, classification results. Accuracy of .97 was obtained in the visible region using both PCA and LDA, but in RE and NIR regions classification accuracy of PCA was greater than LDA components, which was .96 and .95 respectively. Average accuracy of classification of Cluster Beans, considering five regions was .90 for LDA and .92 for PCA.

Table 2: Comparison of classification accuracy of Cluster Beans

Regions	Filter	Accuracy	Kappa
Visible	LDA	0.97	0.92
	PCA	0.97	0.93
Red Edge	LDA	0.91	0.8
	PCA	0.96	0.92
N IR	LDA	0.94	0.88
	PCA	0.95	0.9
SW I	LDA	0.87	0.72
	PCA	0.87	0.72
SW II	LDA	0.84	0.64
	PCA	0.86	0.7

On visual judgment of healthy and diseased Long Beans leaf spectras, it is observed that there increase in reflectance magnitude of Visible, decrease in RE and NIR regions of diseased leaves, where was no difference in the SW1 and SW2 regions. As mentioned results in Table 3, maximum classification accuracy of 1 was obtained in RE region for PC components, average accuracy was .88 and .90 respectively for LDA and PCA.

Table 4 shows the mean classification accuracy obtained by RF classifier after PCA and LDA filter methods.

Table 3: Comparison of classification accuracy of Long Beans

Regions	Filter	Accuracy	Kappa
Visible	LDA	0.94	0.87
	PCA	0.94	0.87
Red Edge	LDA	0.95	0.89
	PCA	1	1
N IR	LDA	0.77	0.4
	PCA	0.8	0.49
SW I	LDA	0.89	0.75
	PCA	0.9	0.78
SW II	LDA	0.89	0.75
	PCA	0.89	0.75

Table 4: Comparison of Classification performance (mean values of accuracy and Kappa)

Regions	Filter	Accuracy	Kappa
Brinjal	LDA	0.846	0.65
	PCA	0.864	0.66
Cluster Beans	LDA	0.9	0.79
	PCA	0.92	0.83
Long Beans	LDA	0.89	0.73
	PCA	0.9	0.78

As shown in the Fig.2 , Mean accuracy and mean Kappa values of PCA filtering are better compared to LDA

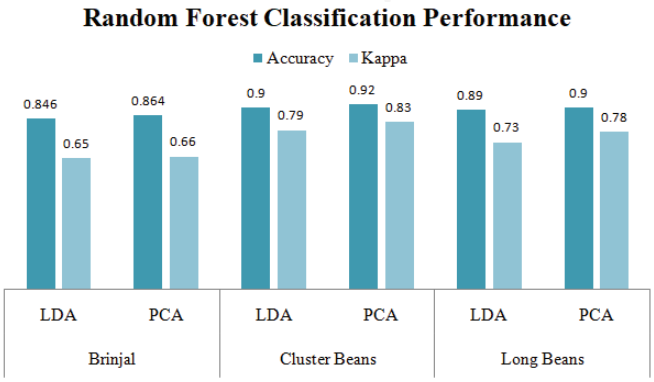


Fig. 2: Random Forest Classification Performance

4. Discussion and Conclusion

Hyperspectral Remote sensing provides large number of features in the form of wavelengths. Number of wavelengths was greater than sample size. Selection of significant wavelengths for a particular application is a critical job. This experiment was done for narrowing down to a small region and selecting only few features useful for further classification of healthy and unhealthy vegetable leaves. Two filter methods LDA and PCA were applied and Random Forest classification was performed on five different regions.

Experimental results have shown that, for distinguishing healthy and unhealthy vegetable leaves, we can achieve good results by focusing on Visible and Red Edge regions of the spectral signature. Performance of PCA and LDA was similar in some cases, but overall accuracy of Random Forest classification was better with PCA.

5. Acknowledgement

Department of Science and Technology - Funds for Infrastructure under Science and Technology (DST-FIST) has supported the Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India , under section No. SR/ FST/ ET1340. This research work has been done under the same support. Authors wish to extend sincere gratitude towards the Department and University authorities for providing required resources.

References:

1] Prasad S. Thenkabail, M. K. (2014). Hyperspectral Hyperion Images and Spectral Libraries of Agricultural Crops,Hyperspectral Remote Sensing of Vegetation and Agricultural Crops. *Photogrammetric Engineering & Remote Sensing , 80* (8), 694-723.12.

2] Prasad S. Thenkabail, M.I.(2013). Selection of hyperspectral narrowbands (HNBs) and composition of hyperspectral twoband vegetation indices (HVIs) for biophysical characterization and discrimination of crop types using field reflectance and Hyperion/ EO-1 data. *IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing,6* (2) , 427-439.

3] R. N. Sahoo, S. S. (2015). Hyperspectral remote sensing of agriculture. *CURRENT SCIENCE,Special Section Hyperspectral Remote Sensing , 108* (5), 848-859.

4] Christoph Römer, M. W. (2012). Early Drought Stress Detection in Cereals: Simplex Volume Maximisation for Hyperspectral Image Analysis. *Functional Plant Biology , 39* (11), 878–890.

5] Dawson, T. &. (1998). A new technique for interpolating the reflectance red edge position. *International Journal of Remote Sensing , 19* (11), 2133-2139.

6] Anjana Ghule, P. D. (2019). MFDS-m Red Edge Position Detection Algorithm for Discrimination between Healthy and Unhealthy Vegetable Plants. *International Conference on Recent Trends in Image Processing & Pattern Recognition*, (pp. 372-379).

7] Mozhgan Abbasi, ,. J. (2019). Optimal SpectralWavelengths for Discriminating Orchard Species Using Multivariate Statistical Techniques. *Remote Sensing , 12* (63), 1-16.

8] Prospere K., M. K. (2014). Plant Species Discrimination in a Tropical Wetland Using In Situ Hyperspectral Data. *Remote Sensing , 6* (9), 8494-8523.

9] Mirzaei M., M. S. (2019). Scenario-based discrimination of common grapevine varieties using in-field hyperspectral data in the western of Iran. *International Journal of Applied Earth Observation and Geoinformation , 80*, 26–37.

10] Shibendu Shankar Ray, J. P. (2010). Use of Hyperspectral Remote Sensing Data for Crop Stress Detection: Ground-Based Studies. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science , 38* (8), 562-570.

11] Junshi Xia, P. D. (2014). Hyperspectral Remote Sensing Image Classification Based on Rotation Forest. *IEEE Geoscience And Remote Sensing Letters , 11* (1), 239-243.

12] Helmi Z. M. Shafri, M. I. (2011). Spectral Discrimination of Healthy and Ganoderma-Infected Oil Palms from Hyperspectral Data. *International Journal of Remote Sensing , 32* (22), 7111–7129.

13] http://www.asdi.com/applications/remote-sensing/spectral-remote-sensing. (n.d.). Retrieved Janurary 21, 2016, from http://www.asdi.com/applications/remote-sensing/spectral- remote-sensing.

14] Aleix M. Martinez, A. C. (2001). PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence , 23* (2), 228-233.

15] PU, R. (2009). Broadleaf species recognition with in situ hyperspectral data. *International Journal of Remote Sensing , 30* (11), 2759 — 2779.

16] Tom Howley a, M. G.-L. (2005). The effect of principal component analysis on machine learning accuracy with high-dimensional spectral data. *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (pp. 363–370). Cambridge: Proceedings of AI. ■

About the Authors

Mrs. Anjana N Ghule (CSI LM. - 95255) M.E. Computer Science and Engineering from Dr. B. A. M University Aurangabad (MS) India. Currently working as an Assistant Professor, Department of Information Technology, Government College of Engineering, Aurangabad, (MS), India. Life member CSI, ISTE and IETE. EC member, IETE Aurangabad Centre. Published research papers in International conferences and reputed journals. Her area of research is Remote Sensing. She can be reached at anghule@geca.ac.in.

Prof. Dr. R. R. Deshmuk (CSI LM - 00100518) M.E., M.Sc. (CSE) Ph.D. FIETE, PEIN Fellow, Working as Professor, Department of CSIT, Dr. B.A.M. University, Aurangabad, (MS), India. Coordinator of DST-FIST program, University Coordinator, MHRD GIAN program, Chairman, IETE Aurangabad Centre 2014–2018, Organized Zonal ISF-2016, Sectional President, ICT Section ISCA-2019, Coordinator, महाराष्ट्र राज्य मराठी विश्वकोश निर्मिती मंडळ, महाराष्ट्र शासन, Life member ISCA, CSI, ISTE, IEEE, IAEng, CSTA, IDES, ACEE. Management Council, Senate, Academic Council Member at University, Edited Twelve books, Published more than 225 research papers in reputed Journals, Editor-In-Chief, CSI Journal of Computing. Visited Russia, USA, China, Spain, Philippines, Uzbekistan, Thailand for academic work. He can be reached at rrdeshmukh.csit@bamu.ac.in

Mrs. Chitra M. Gaikwad [CSI-1191698] -Assistant Professor, Department of Information Technology, Government College of Engineering, Aurangabad, Maharashtra, India. She has 8 paper publications in International Journals, 3 in International Conference and 1 in National Conference. Her research interests include Image Processing and Remote Sensing. She can be reached at cmgaikwad@geca.ac.in

Ensemble Learning approach for Named Entity Recognition from Hindi-English Tweets

Suparna Arya[1], Amit Majumder[2], Sulabh Majumder[3], Aparajita Kundu[4], Nuzhat Shamim[5], Pranati Rakshit[6]

JIS College of Engineering, West Bengal
[1] suparnaarya1997@gmail.com, [2] amit.majumder@jiscollege.ac.in, [3]sulabh21198@gmail.com,
[4] aparajitakundu1998@gmail.com, [5] shamimsaba96@gmail.com, [6] pranati.rakshit@jiscollege.ac.in

Named Entity Recognition (NER) is a machine learning paradigm that identifies named entities from content written in Natural Language. It accepts a string of tokens in the form of sentence or paragraph as input and recognizes the relevant nouns like names of people, places, organizations etc. that are resides inside the corresponding sentence accepted by the model. This model also belongs from the field of Information Extraction of Natural Language Processing (NLP). There is lots of research work performed on Named entities recognition (NER), but most of them have worked with resource rich languages and domains. It is very challenging engagement to work with informal tweets which make the process more complex with its unstructured and incomplete data. In this paper, we recommend a process of recognizing named entities form tweets mixed in hindi and English languages using various ensemble based machine learning algorithms with the help of content and contextual features that are extracted from hindi-english tweets.

1. Introduction

We all regularly through huge amount of online data generated by various online platforms, media houses, banks, machines etc. which are mostly unstructured and complex. This huge and complex unstructured data needs to be organized efficiently to make effective utilization of each datum. Named Entity Extraction is a machine learning model that automatically identifies all sort of useful details related to people, organizations, and places that are already mentioned in the content. Multiple language speaker sometimes switch between different languages and when they state or write something they involve multiple code mixing which indicates the usage of multiple linguistic units from various languages in utterance of single line of sentence. Without going into the depth of code mixing, it is important to first account all complications with hindi-english twitter tweets itself. First, comes the acronyms part and online slangs of various blogs and tweets that makes them more complex to understand. Accordingly, the uncertainty is a major issue as semantic notation process cannot easily interpret the usage of references and information. Second, the acronyms shows much more variation and that apt to be less grammatical than large posts, consists of untraditional capitalization, and make normal use of emojis, acronyms and hash-tags, which provides a large part of the meaning. With all these complications, we like to challenge ourselves with a supervised learning model, which combines the predictions of various machine learning algorithms to build an improvised learning algorithm know as ensemble algorithm. These sorts of algorithms improve the prediction process of the derived model.

The named entity recognition identifies proper nouns in texts and classifies them into a set of predefined types. Over the years there has been considerable work done for NER in English and other languages, but, on code-mixed language it is quite low. It motivated us to design one NER system using natural language processing technique to extract named entities automatically from Hindi-English code-mixed Tweets.

2. Background and Related work

Bali et al[1]. performed analysis of data from Face- book posts generated by English-Hindi bilingual users. Analysis depicted that significant amount of code- mixing was present in the posts. Vyas et al.[1,2] formalized the problem, created a POS tag annotated Hindi-English code- mixed corpus and reported the challenges and problems in the Hindi- English code- mixed text. They also performed experiments on language identification, transliteration, normalization and POS tagging of the Dataset. Sharma et al.[3] addressed the problem of shallow parsing of Hindi-English code-mixed social media text and developed a system for Hindi- English code-mixed text that can identify the language of the words, normalize them to their standard forms, assign them their POS tag and segment into chunks.

In Named Entity Recognition there has been significant research done so far in English and other resource rich languages Morwal et al.[4], but same cannot be said for

code-mixed text due to lack of structured resources in this domain. Significant work was carried out on bengali data and code-mixed data for named entity recognition by Ekbal et al[6,7]. Bhargava et al.[5] proposed a hybrid model for NER on Hindi-English and Tamil-English code-mixed Dataset. Bhat et al[8]. proposed a neural network architecture for NER on Hindi-English code- mixed Dataset. Named Entity Recognition for Hindi-English Code- Mixed Social Media Text was also addressed by Singh et al[9]. In the work of Named Entity Recognition in Tweets [10], it has been seen that the performance of the existing named extraction systems are not so good. There is a need of Named Entity Extraction system which is tuned on Tweets dataset. In the work of named entity extraction for four leading Indian languages, namely Bengali, Hindi, Telugu, and Oriya [18], weighted-voting based ensemble approach was used and the technique was effective.

3. Features

Since we are working with hindi-english tweets as dataset, deriving syntactical features is very complex. Therefore, we decide to work with content and contextual features just to carry out named entity recognition. The features that are described are used in our experiment.

Word: This feature is describes the current original word.

Word_lower: This describes lowercase form of current word. For example, if there is a word say "Cricket", then it will be calculate the feature value as "cricket".

Word_1stUpper: This feature provides Boolean value. It describes true value if first letter of the current word is written in upper-case. For example, if there is a word say "Cricket", then it will provide feature value as 'True' since the word initiate with capital letter.

Word_isAlpha: This feature provides Boolean value. It describes true value if current word is consists of alphabets.

Word_isdigit: This feature provides Boolean value. It describes true value if current word is consists of numbers.

Word_isupper: This feature provides Boolean value. It describes true value if current word consists of upper-case characters.

Word_startsWith#: This feature provides Boolean value. It describes true value if current word precedes with a '#' character.

Word_startsWith@: This feature provides Boolean value. It describes true value if the current word precedes with a '@' character.

Word-1: This feature describes the word before current word.

Word-1_1stUpper: This feature provides Boolean value. It describes true value if previous word of the current word consists of first upper- case character.

Word-1_isAlpha: This feature provides Boolean value. It describes true value if previous word of current word consists of alphabets.

Word-1_isdigit: This feature provides Boolean value. It describes true value if the previous word contains all numeric characters.

Word-1_isupper : This feature provides Boolean value. It describes true value if the previous word is consists of all upper-case characters.

Word-1_lower: This feature provides Boolean value. It describes true value if the previous word consists of all lower-case characters.

Word-1_startsWith#: This feature provides Boolean value. It describes true value if the previous word precedes with a '#' character.

Word-1_startsWith@: This feature provides Boolean value. It describes true value if the previous word precedes with a '@' character.

Word+1: This feature describes the word after current word.

Word+1_1stUpper: This feature provides Boolean value. It describes true value if next word of the current word consists of first upper-case character.

Word+1_isAlpha: This feature provides Boolean value. It describes true value if next word of the current word is consists of alphabets.

Word+1_isdigit: This feature provides Boolean value. It describes true value if the next word contains all numeric characters

Word+1_isupper: This feature provides Boolean value. It describes true value if the next word is consists of all upper-case characters..

Word+1_lower: This feature provides Boolean value. It describes true value if the next word of the current word consists of all lower-case characters.

Word+1_startsWith#: This feature provides Boolean value. It describes true value if the next word precedes with a '#' character. Word+1_startsWith@: This feature provides Boolean value. It describes true value if the next word precedes with a '@' character.

N-gram: This is the character level of n-gram features containing several sub-words of having length equals to n, where n is the number of characters. Here we used 1-gram, 2- gram and 3- gram features.

BOS: This feature provides Boolean value. It describes true value if the current word is the beginning of the sentence.

EOS: This feature provides Boolean value. It describes true value if current word is the end of the sentence.

4. Dataset and Experimental Results

We used dataset that consist of hindi-english tweets on several topics like movie reviews, politics, social events, sports, etc. originated from an Asian country India. After extracting features from the dataset it provides several tags that are labeled was 'Person', 'Organization' and 'Location' based on BIO standard. Statistics of the labeled dataset is described in Table-1.

Table 1: Statistical Dataset

Entity Tag	Count
B-Per	2138
Other	63499
B-Org	1432
I-Org	90
B-Loc	762
I-Loc	31
I-Per	554
Total	68506

At first we performed supervised learning approach by using Naïve Bayes, Random Forest and Decision Tree types algorithms for named entities extraction from hindi-english twitter tweets. Whose results after experiments have been described in Table-2, Table-3 and Table-4. We show our experimental results in terms of recall, precision and F1-score which are representation as:

$$recall = \frac{\text{Number of Events correctly identified by system}}{\text{Number of events in the gold standard test data}} \quad (1)$$

$$recision = \frac{\text{Number of events correctly identified by syste}}{\text{Number of events identified by the system}} \quad (2)$$

$$F1\_score = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}} \quad (3)$$

Table 2: Result using Naïve Bayes

Entity Tag	Precision	Recall	F1-score
I-Loc	0.00	0.05	0.00
B-Org	0.14	0.19	0.16
I-Per	0.05	0.20	0.08
Other	0.97	0.73	0.84
B-Per	0.28	0.43	0.34
I- Org	0.01	0.47	0.01
B-Loc	0.05	0.25	0.09
Accuracy	0.70	0.70	0.70
Macro avg	0.21	0.33	0.22
Weighted avg	0.91	0.70	0.79

Table 3: Result using Random Forest

Entity Tag	Precision	Recall	F1-score
I-Loc	0.00	0.00	0.00
B-Org	0.96	0.26	0.41
I-Per	0.88	0.02	0.03
Other	0.94	1.00	0.97
B-Per	0.82	0.32	0.46
I- Org	0.00	0.00	0.00
B-Loc	0.00	0.00	0.00
Accuracy	0.94	0.94	0.94
Macro avg	0.51	0.23	0.27
Weighted avg	0.93	0.94	0.92

Table 3: Result using Random Forest

Entity Tag	Precision	Recall	F1-score
I-Loc	0.03	0.05	0.04
B-Org	0.45	0.51	0.48
I-Per	0.26	0.22	0.24
Other	0.96	0.96	0.96
B-Per	0.53	0.51	0.52
I- Org	0.08	0.03	0.05
B-Loc	0.32	0.32	0.32
accuracy	0.92	0.92	0.92
Macro avg	0.38	0.37	0.37
Weighted avg	0.92	0.92	0.92

5. Ensemble learning approach

After getting prediction values from the above algorithms, we decide to work with another supervised model known as Ensemble Learning. This model combines the predictions of Random Forest, Naïve Bayes and Decision Tree for better performance in refers to any of this constituent algorithm. In this paper we use Averaging, Max voting, and Weighted average algorithms for Ensemble Learning approach and their prediction is described in Table-5, Table-6 and Table-7.

Table 3: Result using Random Forest

Entity Tag	Precision	Recall	F1-score
I-Loc	0.00	0.00	0.00
B-Org	0.69	0.11	0.10
I-Per	0.03	0.20	0.05
Other	0.97	0.79	0.87
B-Per	0.11	0.54	0.18
I- Org	0.01	0.02	0.01
B-Loc	0.91	0.10	0.17
Accuracy	0.76	0.76	0.76
Macro avg	0.39	0.25	0.21
Weighted avg	0.92	0.76	0.82

Table 6: Result using Max Voting

Entity Tag	Precision	Recall	F1-score
I-Loc	0.00	0.00	0.00
B-Org	0.02	0.02	0.02
I-Per	0.01	0.16	0.02
Other	0.93	0.81	0.86
B-Per	0.03	0.02	0.02
I- Org	0.00	0.00	0.00
B-Loc	0.03	0.01	0.01
Accuracy	0.75	0.75	0.75
Macro avg	0.15	0.15	0.13
Weighted avg	0.86	0.75	0.80

Table 7: Result using Weighted Voting

Entity Tag	Precision	Recall	F1-score
I-Loc	0.00	0.23	0.01
B-Org	0.80	0.36	0.50
I-Per	0.56	0.17	0.26
Other	0.96	0.97	0.97
B-Per	0.76	0.46	0.57
I- Org	0.18	0.06	0.09
B-Loc	0.58	0.18	0.27
accuracy	0.93	0.93	0.93
Macro avg	0.55	0.35	0.38
Weighted avg	0.94	0.93	0.93

6. Result Analysis and Comparison

At first we compare the prediction results of the three classification algorithms that is, Decision Tree, Naïve Bayes and Random Forest and realized that Random Forest is turns out as best performing algorithm. We compared our model with the existing named entity extraction system[9] which was implemented with same set of data using CRF (Conditional Random Field) model and we realized that our Random forest based model is comparatively better than the existing CRF-based model (depending upon F1- score value).

We also compared our model with existing Name Entity Recognition model of Vinay Singh[1], Deepanshu Vijay[2], Syed S. Akhtar[3], Manish Shrivastavathat[4] and identifies that they extract entities from same data set. Their model used CRF with 'c1=0.1' and 'c2=0.1' and 'L-BFGS' algorithm and CRF with 'Avg. Perceptron' algorithm both provides highest F1-score. Whereas, in our model Random Forest algorithm provides highest F1-score. Then, we approached for ensemble learning algorithms where we implements Averaging, Max Voting and Weighted Average. After implementing this three ensemble learning algorithm we got better predictive performance from Weighted Average among all three ensemble learning algorithm.

7. Conclusion

In this experiment we implemented three different classification based algorithms that are S, Decision Tree Naïve Bayes and Random Forest and extracted named entities based on content and contextual features. After extraction we realized that Random Forest provides better performance in comparison of other algorithms. Then we reached out for Ensemble learning algorithm for better predictive performance by combing the prediction of three classification algorithm, and build improvised model. Here, we build three different combinational models and realized Weighted Average provides better result than ensemble algorithm. In our future we will try our best to improve much more accuracy of this model using different classification algorithm, ensemble algorithm, different datasets and more efficient features for better performance check of the model. In this paper, we remain deprived from deep learning approach. In future work we will like to introduce deep learning approach for named entities extraction.

References

1] Hansen, L.K., Salamon, P.: Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(10) (1990) 993-1001

2] Schapire, R.E.: The strength of weak learnability. Machine Learning 5(2) (1990) 197-227

3] Krogh, A., Vedelsby, J.: Neural network ensembles, cross validation, and active learning. In Tesauro, G., Touretzky, D.S., Leen, T.K., eds.: Advances in Neural Information Processing Systems 7. MIT Press, Cambridge, MA (1995) 231-238

4] Kuncheva, L.I., Whitaker, C.J.: Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. Machine Learning 51(2) (2003) 181-207

5] Freund, Y., Schapire, R.E.: A decision- theoretic generalization of on-line learning and an application to Boosting. Journal of Computer and System Sciences 55(1) (1997) 119-139

6] Breiman, L.: Bagging predictors. Machine Learning 24(2) (1996) 123–140 7. Wolpert, D.H.: Stacked generalization. Neural Networks 5(2) (1992) 241–260 Breiman, L.: Random forests. Machine Learning 45(1) (2001) 5-32.

7] Kalika Bali, Jatin Sharma, Monojit Choudhury, and Yogarshi Vyas. 2014. " i am borrowing ya mix-ing?" an analysis of english-hindi code mixing in facebook. In Proceedings of the First Workshop on Computational Approaches to Code Switching, pages 116-126.

8] Yogarshi Vyas, Spandana Gella, Jatin Sharma, Ka-lika Bali, and Monojit Choudhury. 2014. Pos tagging of english- hindi code-mixed social media con-tent. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 974-979.

9] Arnav Sharma, Sakshi Gupta, Raveesh Motlani, Piyush Bansal, Manish Srivastava, Radhika Mamidi, and Dipti M Sharma. 2016. Shallow parsing pipeline for hindi-english code-mixed social media text. arXiv preprint arXiv:1604.03136.

10] Sudha Morwal, Nusrat Jahan, and Deepti Chopra. 2012.

Named entity recognition using hidden markov model (hmm). International Journal on Natural Language Computing (IJNLC), 1(4):15-23.

11] Rupal Bhargava, Yashvardhan Sharma, and Shubham Sharma. 2016a. Sentiment analysis for mixed script indic sentences. In Advances in Computing, Communications and Informatics (ICACCI), 2016 Inter-national Conference on, pages 524–529. IEEE.

12] Asif Ekbal and Sivaji Bandyopadhyay. 2008. Bengali named entity recognition using support vector machine. In Proceedings of the IJCNLP- 08 Workshop on Named Entity

13] [13] Recognition for South and South East Asian Languages.

14] Deepak Gupta, Shubham Tripathi, Asif Ekbal, and Pushpak Bhattacharyya. 2016. A hybrid approach for entity extraction in code-mixed social media data. MONEY, 25:66.

15] Irshad Ahmad Bhat, Manish Shrivastava, and Riyaz Ahmad Bhat. 2016. Code mixed entity extraction in indian languages using neural networks. In FIRE (Working Notes), pages 296-297.

16] Vinay Singh, Deepanshu Vijay, Syed S. Akhtar, Manish Shrivastava. Named Entity Recognition for Hindi- English Code-Mixed Social Media Text. In Proceedings of the Seventh Named Entities Workshop, pages 27-35, Melbourne, Australia, July 20, 2018, Association for Computational Linguistics

17] Alan Ritter, Sam Clark, Mausam, Oren Etzioni; Named Entity Recognition in Tweets: An Experimental Study; in Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, July, Year: 2011, Address:, Edinburgh, Scotland, UK..

18] Asif Ekbal, Sriparna Saha; "Weighted Vote-Based Classifier Ensemble for Named Entity Recognition: A Genetic Algorithm-Based Approach". ACM Trans. Asian Lang. Inf. Process. 10(2): 9:1-9:37 (2011) ■

About the Authors

Suparna Arya is currently doing her B.Tech from Computer science and Engineering department, JIS College of Engineering, West Bengal, India. She has interest in Machine Learning, Information Extraction, Natural Language Processing.

Amit Majumder has received B.Tech degree in Computer Science and Engineering from Kalyani Govt. Engineering College, Kalyani, West Bengal, India, and ME degree in Computer Science and Engineering from Jadavpur University, Kolkata, West Bengal, India. Currently he is working as Assistant Professor in CSE department at JIS College of Engineering, Kalyani, West Bengal, India. He has interest in areas of Artificial Intelleigence, Soft Computing, Natural Language Processing, Machine Learning, Information Extraction. Currently he is doing his work on extracting bio- molecular event using Bio-NLP technique.

Sulabh Majumder is currently doing his B.Tech from Computer science and Engineering department, JIS College of Engineering, West Bengal, India. She has interest in Machine Learning, Information Extraction, Natural Language Processing.

Aparajita Kundu is currently doing her B.Tech from Computer science and Engineering department, JIS College of Engineering, West Bengal, India. She has interest in Machine Learning, Information Extraction, Natural Language Processing.

Nuzhat Shamim is currently doing her B.Tech from Computer science and Engineering department, JIS College of Engineering, West Bengal, India. She has interest in Machine Learning, Information Extraction, Natural Language Processing.

Dr. Pranati Rakshit is an Assistant Professor in the Department of Computer Science and Engineering of JIS College of Engineering, India. She has more than 19 years of work experience associated with teaching and research. She has completed her Ph.D. degree from Jadavpur University, Kolkata, West Bengal, in the field of Pattern Recognition and Medical image analysis. She has completed her Master degree and B.E. Degree in Computer Science & Engineering. She has worked in the field of Data Mining, image Processing, IOT also. She has supervised more than 20 M.Tech Projects. She has a good number of research publications. She is a life member of Indian Society of Technical Education

Guidelines to Authors

General Submission Guidelines

Journal of Computing

Published by the Computer Society of India

General

- The manuscript submitted should be author's original work; citations to others work should have appropriate citations with due credit and must adhere to standard copyright rules. It is the responsibility of the authors to obtain the necessary disclosure permissions for the material used (be it Government, Corporations, Societies, Publishers or authors) and not the CSI (CSI has a zero tolerance policy on plagiarism; cf. http://www.csi-india.org/policies) The paper should not have been published earlier and must not be under consideration for publication elsewhere.
- Regular papers will be generally 15 to 20 pages in the double column of size 8.5" x 11" (or 21.6cm x 27.9cm) each column of 3.5" (width 6.45cm).

Manuscript Formats

a. Full papers generally consist of the title, author affiliation, footnote (including any financial support acknowledgment), abstract, introduction, body, conclusions, reference list, list of figures and table captions, and original figures and tables for reproduction.

b. A paper may also include appendixes, a glossary of symbols, and an acknowledgment of nonfinancial support. All manuscript pages, footnotes, equations, and references should be labelled in consecutive numerical order. Illustrations and tables should be cited in text in numerical order.

c. ***Abstract:*** The abstract should be limited to 50–200 words and should concisely state what was done, how it was done, principal results, and their significance. The abstract will appear later in various abstracts journals and should contain the most critical information of the paper.

d. ***References:*** A numbered list of references must be provided at the end of the paper. The list should be arranged in alphabetical order. List only one reference per reference number. In text, each reference number should be enclosed by square brackets. Footnotes or other words and phrases that are not part of the reference format do not belong on the reference list.

e. ***Section Headings:*** Primary section headings within papers are enumerated by Roman numerals and are centered above the text. For the purpose of typing the manuscript only, primary headings should be capital letters.

Peer Review

- After the Editor/Editor-in-Chief of a publication determines that a paper is suitable for his/her publication, it will be forwarded to a group of reviewers (generally at least three) selected for their expertise in a given field. During this process, an author is often asked to expand, rewrite, or explain further the content of his/her paper. It is not uncommon that an author is asked to provide another draft with the suggested changes for further review.
- Generally it may take up to six months from the time of submission.
- Authors have the right to withdraw the manuscript at any stage of the review process.

Final Acceptance:

- Once a manuscript has received the final approval of the reviewers and Editor-in-Chief, the author will be notified and sent an CSI Copyright Form. He/she will be asked to prepare the manuscript for final electronic publication along with a reprint/ colour reproduction invoice.
- ***Final Manuscript Submission:*** For accurate and efficient transferal of your manuscript, particularly those containing extensive mathematics, it is recommended that the authors may use TEX or LATEX programs. A CSI LATEX style file can be found at www.csijournal.org/guidelines.html. The .cls file can be downloaded here. Also, in your file use CSI header "CSI Journal of Computing" using markboth as indicated in the illustrating file. You may use the bibliography style as in IEEE (number the references) style format (TRANS-JOUR). Doc users can use the style guide provided at this address. For headers, use the CSI Journal of Computing. The copyright form can be downloaded here. Each author needs to submit a passport photograph in jpg format, with photo size less than 30 KB. Also, a short bio (about 6 lines) need to be provided for each author.

Page Charges

- No page charges.
- Colored reproduction of figures, photographs at no cost for the web version. For print version, actual print charges to be borne by the authors. If the cost is not borne, the same would be reproduced in black and white. At the time of proofs being sent, appropriate invoice will be sent.

Journal of Computing

BEST PAPER AWARD

The best paper in each volume of the

CSI Journal of Computing will be given an award in one of the

CSI sponsored conferences.

The details will follow.

Published by
Prof. A K Nayak on behalf of
COMPUTER SOCIETY OF INDIA™

Designed & Printed by
DSK Enteprises, Thane, Maharashtra
(M) 96196 60794 (E) dskenterprises42@gmail.com

Published from
Computer Society of India
Samruddhi Venture Park, Unit-3, 4th Floor,
Marol Industrial Area, Andheri (East), Mumbai 400 093.
Tel.: 022-29261700 Fax: 022-2830 2133
Email: csi.journal@csi-india.org
Editor-in-Chief : **Ratnadeep R. Deshmukh**