

# INTERNATIONAL JOURNAL OF RESEARCH IN COMMERCE, IT & MANAGEMENT

I  
J  
R  
C  
M



A Monthly Double-Blind Peer Reviewed (Refereed/Juried) Open Access International e-Journal - Included in the International Serial Directories

*Indexed & Listed at:*

Ulrich's Periodicals Directory ©, ProQuest, U.S.A., Cabell's Directories of Publishing Opportunities, U.S.A., Google Scholar,  
Indian Citation Index (ICI), J-Gate, India [link of the same is duly available at Infibnet of University Grants Commission (U.G.C.)],  
Index Copernicus Publishers Panel, Poland with IC Value of 5.09 (2012) & number of libraries all around the world.

Circulated all over the world & Google has verified that scholars of more than 7835 Cities in 197 countries/territories are visiting our journal on regular basis.

Ground Floor, Building No. 1041-C-1, Devi Bhawan Bazar, JAGADHRI – 135 003, Yamunanagar, Haryana, INDIA

<http://ijrcm.org.in/>

# CONTENTS

Sr. No.	TITLE & NAME OF THE AUTHOR (S)	Page No.
1.	<p style="text-align: center;">A GENETIC ALGORITHM BASED IMAGE AUTHENTICATION TECHNIQUE IN FREQUENCY DOMAIN USING HAAR WAVELET TRANSFORM (AGAIAFDHWTT)</p> <p style="text-align: center;"><i>Dr. AMRITA KHAMRUI</i></p>	1
2.	<p style="text-align: center;">QUANTIFYING THE EXPLAINABILITY OF MACHINE LEARNING MODELS: METRICS AND BENCHMARKS</p> <p style="text-align: center;"><i>K.ROOPA</i></p>	6
	<b>REQUEST FOR FEEDBACK &amp; DISCLAIMER</b>	<b>11</b>

**FOUNDER PATRON****Late Sh. RAM BHAJAN AGGARWAL**

Former State Minister for Home & Tourism, Government of Haryana  
Former Vice-President, Dadri Education Society, Charkhi Dadri  
Former President, Chinar Syntex Ltd. (Textile Mills), Bhiwani

**CO-ORDINATOR****Dr. BHAVET**

Former Faculty, Shree Ram Institute of Engineering & Technology, Urjani

**ADVISOR****Prof. S. L. MAHANDRU**

Principal (Retd.), Maharaja Agrasen College, Jagadhri

**EDITOR****Dr. G. BRINDHA**

Professor & Head, Dr.M.G.R. Educational & Research Institute (Deemed to be University), Chennai

**CO-EDITOR****Dr. A. SASI KUMAR**

Professor, Vels Institute of Science, Technology & Advanced Studies (Deemed to be University), Pallavaram

**EDITORIAL ADVISORY BOARD****Dr. S. P. TIWARI**

Head, Department of Economics & Rural Development, Dr. Ram Manohar Lohia Avadh University, Faizabad

**Dr. CHRISTIAN EHIUBUCHE**

Professor of Global Business/Management, Larry L Luing School of Business, Berkeley College, USA

**Dr. SIKANDER KUMAR**

Vice Chancellor, Himachal Pradesh University, Shimla, Himachal Pradesh

**Dr. JOSÉ G. VARGAS-HERNÁNDEZ**

Research Professor, University Center for Economic & Managerial Sciences, University of Guadalajara, Guadalajara, Mexico

**Dr. TEGUH WIDODO**

Dean, Faculty of Applied Science, Telkom University, Bandung Technoplex, Jl. Telekomunikasi, Indonesia

**Dr. M. S. SENAM RAJU**

Professor, School of Management Studies, I.G.N.O.U., New Delhi

**Dr. A SAJEEVAN RAO**

Professor & Director, Accurate Institute of Advanced Management, Greater Noida

**Dr. D. S. CHAUBEY**

Professor & Dean (Research & Studies), Uttaranchal University, Dehradun

**Dr. CLIFFORD OBIYO OFURUM**

Professor of Accounting & Finance, Faculty of Management Sciences, University of Port Harcourt, Nigeria

**Dr. KAUP MOHAMED**

Dean & Managing Director, London American City College/ICBEST, United Arab Emirates

**Dr. VIRENDRA KUMAR SHRIVASTAVA**

Director, Asia Pacific Institute of Information Technology, Panipat

**Dr. MIKE AMUHAYA IRAVO**

Principal, Jomo Kenyatta University of Agriculture & Tech., Westlands Campus, Nairobi-Kenya

**Dr. SYED TABASSUM SULTANA**

Principal, Matrusri Institute of Post Graduate Studies, Hyderabad

**Dr. BOYINA RUPINI**

Director, School of ITS, Indira Gandhi National Open University, New Delhi

**Dr. NEPOMUCENO TIU**

Chief Librarian & Professor, Lyceum of the Philippines University, Laguna, Philippines

**Dr. SANJIV MITTAL**

Professor & Dean, University School of Management Studies, GGS Indraprastha University, Delhi

**Dr. RAJENDER GUPTA**

Convener, Board of Studies in Economics, University of Jammu, Jammu

**Dr. SHIB SHANKAR ROY**

Professor, Department of Marketing, University of Rajshahi, Rajshahi, Bangladesh

**Dr. SRINIVAS MADISHETTI**

Professor, School of Business, Mzumbe University, Tanzania

**Dr. NAWAB ALI KHAN**

Professor & Dean, Faculty of Commerce, Aligarh Muslim University, Aligarh, U.P.

**MUDENDA COLLINS**

Head, Operations & Supply Chain, School of Business, The Copperbelt University, Zambia

**Dr. EGWAKHE A. JOHNSON**

Professor & Director, Babcock Centre for Executive Development, Babcock University, Nigeria

**Dr. A. SURYANARAYANA**

Professor, Department of Business Management, Osmania University, Hyderabad

**P. SARVAHARANA**

Asst. Registrar, Indian Institute of Technology (IIT), Madras

**Dr. MURAT DARÇIN**

Associate Dean, Gendarmerie and Coast Guard Academy, Ankara, Turkey

**Dr. ABHAY BANSAL**

Head, Department of Information Technology, Amity School of Engg. & Tech., Amity University, Noida

**Dr. YOUNOS VAKIL ALROAIA**

Head of International Center, DOS in Management, Semnan Branch, Islamic Azad University, Semnan, Iran

**WILLIAM NKOMO**

Asst. Head of the Department, Faculty of Computing, Botho University, Francistown, Botswana

**Dr. JAYASHREE SHANTARAM PATIL (DAKE)**

Faculty in Economics, KPB Hinduja College of Commerce, Mumbai

**SHASHI KHURANA**

Associate Professor, S. M. S. Khalsa Lubana Girls College, Barara, Ambala

**Dr. SEOW TA WEEA**

Associate Professor, Universiti Tun Hussein Onn Malaysia, Parit Raja, Malaysia

**Dr. OKAN VELI ŞAFAKLI**

Professor & Dean, European University of Lefke, Lefke, Cyprus

**Dr. MOHENDER KUMAR GUPTA**

Associate Professor, Government College, Hodal

**Dr. BORIS MILOVIC**

Associate Professor, Faculty of Sport, Union Nikola Tesla University, Belgrade, Serbia

**Dr. LALIT KUMAR**

Course Director, Faculty of Financial Management, Haryana Institute of Public Administration, Gurugram

**Dr. MOHAMMAD TALHA**

Associate Professor, Department of Accounting & MIS, College of Industrial Management, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

**Dr. V. SELVAM**

Associate Professor, SSL, VIT University, Vellore

**Dr. IQBAL THONSE HAWALDAR**

Associate Professor, College of Business Administration, Kingdom University, Bahrain

**Dr. PARDEEP AHLAWAT**

Associate Professor, Institute of Management Studies & Research, Maharshi Dayanand University, Rohtak

**Dr. ALEXANDER MOSESOV**

Associate Professor, Kazakh-British Technical University (KBTU), Almaty, Kazakhstan

**Dr. ASHOK KUMAR CHAUHAN**

Reader, Department of Economics, Kurukshetra University, Kurukshetra

**Dr. BHAVET**

Former Faculty, Shree Ram Institute of Engineering & Technology, Urjani

**YU-BING WANG**

Faculty, department of Marketing, Feng Chia University, Taichung, Taiwan

**SURJEET SINGH**

Faculty, Department of Computer Science, G. M. N. (P.G.) College, Ambala Cantt.

**Dr. TITUS AMODU UMORU**

Professor, Kwara State University, Kwara State, Nigeria

**Dr. RAJESH MODI**

Faculty, Yanbu Industrial College, Kingdom of Saudi Arabia

**Dr. SAMBHAVNA**

Faculty, I.I.T.M., Delhi

**Dr. THAMPOE MANAGALESWARAN**

Faculty, Vavuniya Campus, University of Jaffna, Sri Lanka

**Dr. SHIVAKUMAR DEENE**

Faculty, Dept. of Commerce, School of Business Studies, Central University of Karnataka, Gulbarga

**SURAJ GAUDEL**

BBA Program Coordinator, LA GRANDEE International College, Simalchaur - 8, Pokhara, Nepal

***FORMER TECHNICAL ADVISOR***

**AMITA**

***FINANCIAL ADVISOR***

**NEENA**

Investment Consultant, Chambaghat, Solan, Himachal Pradesh

***LEGAL ADVISORS***

**JITENDER S. CHAHAL**

Advocate, Punjab & Haryana High Court, Chandigarh U.T.

**CHANDER BHUSHAN SHARMA**

Advocate & Consultant, District Courts, Yamunanagar at Jagadhri

***SUPERINTENDENT***

**SURENDER KUMAR POONIA**

## **CALL FOR MANUSCRIPTS**

We invite unpublished novel, original, empirical and high quality research work pertaining to the recent developments & practices in the areas of Computer Science & Applications; Commerce; Business; Finance; Marketing; Human Resource Management; General Management; Banking; Economics; Tourism Administration & Management; Education; Law; Library & Information Science; Defence & Strategic Studies; Electronic Science; Corporate Governance; Industrial Relations; and emerging paradigms in allied subjects like Accounting; Accounting Information Systems; Accounting Theory & Practice; Auditing; Behavioral Accounting; Behavioral Economics; Corporate Finance; Cost Accounting; Econometrics; Economic Development; Economic History; Financial Institutions & Markets; Financial Services; Fiscal Policy; Government & Non Profit Accounting; Industrial Organization; International Economics & Trade; International Finance; Macro Economics; Micro Economics; Rural Economics; Co-operation; Demography; Development Planning; Development Studies; Applied Economics; Development Economics; Business Economics; Monetary Policy; Public Policy Economics; Real Estate; Regional Economics; Political Science; Continuing Education; Labour Welfare; Philosophy; Psychology; Sociology; Tax Accounting; Advertising & Promotion Management; Management Information Systems (MIS); Business Law; Public Responsibility & Ethics; Communication; Direct Marketing; E-Commerce; Global Business; Health Care Administration; Labour Relations & Human Resource Management; Marketing Research; Marketing Theory & Applications; Non-Profit Organizations; Office Administration/Management; Operations Research/Statistics; Organizational Behavior & Theory; Organizational Development; Production/Operations; International Relations; Human Rights & Duties; Public Administration; Population Studies; Purchasing/Materials Management; Retailing; Sales/Selling; Services; Small Business Entrepreneurship; Strategic Management Policy; Technology/Innovation; Tourism & Hospitality; Transportation Distribution; Algorithms; Artificial Intelligence; Compilers & Translation; Computer Aided Design (CAD); Computer Aided Manufacturing; Computer Graphics; Computer Organization & Architecture; Database Structures & Systems; Discrete Structures; Internet; Management Information Systems; Modeling & Simulation; Neural Systems/Neural Networks; Numerical Analysis/Scientific Computing; Object Oriented Programming; Operating Systems; Programming Languages; Robotics; Symbolic & Formal Logic; Web Design and emerging paradigms in allied subjects.

Anybody can submit the **soft copy** of unpublished novel; original; empirical and high quality **research work/manuscript** **anytime** in **M.S. Word format** after preparing the same as per our **GUIDELINES FOR SUBMISSION**; at our email address i.e. [infoijrcm@gmail.com](mailto:infoijrcm@gmail.com) or online by clicking the link **online submission** as given on our website ([FOR ONLINE SUBMISSION, CLICK HERE](#)).

## **GUIDELINES FOR SUBMISSION OF MANUSCRIPT**

### 1. **COVERING LETTER FOR SUBMISSION:**

DATED: \_\_\_\_\_

**THE EDITOR**

IJRCM

**Subject:** SUBMISSION OF MANUSCRIPT IN THE AREA OF \_\_\_\_\_.

**(e.g. Finance/Mkt./HRM/General Mgt./Engineering/Economics/Computer/IT/ Education/Psychology/Law/Math/other, please specify)**

**DEAR SIR/MADAM**

Please find my submission of manuscript titled ' \_\_\_\_\_ ' for likely publication in one of your journals.

I hereby affirm that the contents of this manuscript are original. Furthermore, it has neither been published anywhere in any language fully or partly, nor it is under review for publication elsewhere.

I affirm that all the co-authors of this manuscript have seen the submitted version of the manuscript and have agreed to inclusion of their names as co-authors.

Also, if my/our manuscript is accepted, I agree to comply with the formalities as given on the website of the journal. The Journal has discretion to publish our contribution in any of its journals.

<b>NAME OF CORRESPONDING AUTHOR</b>	:
Designation/Post*	:
Institution/College/University with full address & Pin Code	:
Residential address with Pin Code	:
Mobile Number (s) with country ISD code	:
Is WhatsApp or Viber active on your above noted Mobile Number (Yes/No)	:
Landline Number (s) with country ISD code	:
E-mail Address	:
Alternate E-mail Address	:
Nationality	:

\* i.e. Alumnus (Male Alumni), Alumna (Female Alumni), Student, Research Scholar (M. Phil), Research Scholar (Ph. D.), JRF, Research Assistant, Assistant Lecturer, Lecturer, Senior Lecturer, Junior Assistant Professor, Assistant Professor, Senior Assistant Professor, Co-ordinator, Reader, Associate Professor, Professor, Head, Vice-Principal, Dy. Director, Principal, Director, Dean, President, Vice Chancellor, Industry Designation **etc.** **The qualification of author is not acceptable for the purpose.**

**NOTES:**

- a) The whole manuscript has to be in **ONE MS WORD FILE** only, which will start from the covering letter, inside the manuscript. **pdf. version is liable to be rejected without any consideration.**
  - b) The sender is required to mention the following in the **SUBJECT COLUMN of the mail:**  
**New Manuscript for Review in the area of** (e.g. Finance/Marketing/HRM/General Mgt./Engineering/Economics/Computer/IT/ Education/Psychology/Law/Math/other, please specify)
  - c) There is no need to give any text in the body of the mail, except the cases where the author wishes to give any **specific message** w.r.t. to the manuscript.
  - d) The total size of the file containing the manuscript is expected to be below **1000 KB**.
  - e) Only the **Abstract will not be considered for review** and the author is required to submit the **complete manuscript** in the first instance.
  - f) **The journal gives acknowledgement w.r.t. the receipt of every email within twenty-four hours** and in case of non-receipt of acknowledgment from the journal, w.r.t. the submission of the manuscript, within two days of its submission, the corresponding author is required to demand for the same by sending a separate mail to the journal.
  - g) The author (s) name or details should not appear anywhere on the body of the manuscript, except on the covering letter and the cover page of the manuscript, in the manner as mentioned in the guidelines.
2. **MANUSCRIPT TITLE:** The title of the paper should be typed in **bold letters, centered and fully capitalised**.
  3. **AUTHOR NAME (S) & AFFILIATIONS:** Author (s) name, designation, affiliation (s), address, mobile/landline number (s), and email/alternate email address should be given underneath the title.
  4. **ACKNOWLEDGMENTS:** Acknowledgements can be given to reviewers, guides, funding institutions, etc., if any.
  5. **ABSTRACT:** Abstract should be in **fully Italic printing**, ranging between **150 to 300 words**. The abstract must be informative and elucidating the background, aims, methods, results & conclusion in a **SINGLE PARA. Abbreviations must be mentioned in full.**
  6. **KEYWORDS:** Abstract must be followed by a list of keywords, subject to the maximum of **five**. These should be arranged in alphabetic order separated by commas and full stop at the end. All words of the keywords, including the first one should be in small letters, except special words e.g. name of the Countries, abbreviations etc.
  7. **JEL CODE:** Provide the appropriate Journal of Economic Literature Classification System code (s). JEL codes are available at [www.aea-web.org/econlit/jelCodes.php](http://www.aea-web.org/econlit/jelCodes.php). However, mentioning of JEL Code is not mandatory.
  8. **MANUSCRIPT:** Manuscript must be in **BRITISH ENGLISH** prepared on a standard A4 size **PORTRAIT SETTING PAPER. It should be free from any errors i.e. grammatical, spelling or punctuation. It must be thoroughly edited at your end.**
  9. **HEADINGS:** All the headings must be bold-faced, aligned left and fully capitalised. Leave a blank line before each heading.
  10. **SUB-HEADINGS:** All the sub-headings must be bold-faced, aligned left and fully capitalised.
  11. **MAIN TEXT:**

**THE MAIN TEXT SHOULD FOLLOW THE FOLLOWING SEQUENCE:****INTRODUCTION****REVIEW OF LITERATURE****NEED/IMPORTANCE OF THE STUDY****STATEMENT OF THE PROBLEM****OBJECTIVES****HYPOTHESIS (ES)****RESEARCH METHODOLOGY****RESULTS & DISCUSSION****FINDINGS****RECOMMENDATIONS/SUGGESTIONS****CONCLUSIONS****LIMITATIONS****SCOPE FOR FURTHER RESEARCH****REFERENCES****APPENDIX/ANNEXURE****The manuscript should preferably be in 2000 to 5000 WORDS, But the limits can vary depending on the nature of the manuscript.**

12. **FIGURES & TABLES:** These should be simple, crystal **CLEAR, centered, separately numbered** & self-explained, and the **titles must be above the table/figure. Sources of data should be mentioned below the table/figure. It should be ensured that the tables/figures are referred to from the main text.**
13. **EQUATIONS/FORMULAE:** These should be consecutively numbered in parenthesis, left aligned with equation/formulae number placed at the right. The equation editor provided with standard versions of Microsoft Word may be utilised. If any other equation editor is utilised, author must confirm that these equations may be viewed and edited in versions of Microsoft Office that does not have the editor.
14. **ACRONYMS:** These should not be used in the abstract. The use of acronyms is elsewhere is acceptable. Acronyms should be defined on its first use in each section e.g. Reserve Bank of India (RBI). Acronyms should be redefined on first use in subsequent sections.
15. **REFERENCES:** The list of all references should be alphabetically arranged. **The author (s) should mention only the actually utilised references in the preparation of manuscript** and they may follow Harvard Style of Referencing. **Also check to ensure that everything that you are including in the reference section is duly cited in the paper.** The author (s) are supposed to follow the references as per the following:
- All works cited in the text (including sources for tables and figures) should be listed alphabetically.
  - Use (ed.) for one editor, and (ed.s) for multiple editors.
  - When listing two or more works by one author, use --- (20xx), such as after Kohl (1997), use --- (2001), etc., in chronologically ascending order.
  - Indicate (opening and closing) page numbers for articles in journals and for chapters in books.
  - The title of books and journals should be in italic printing. Double quotation marks are used for titles of journal articles, book chapters, dissertations, reports, working papers, unpublished material, etc.
  - For titles in a language other than English, provide an English translation in parenthesis.
  - **Headers, footers, endnotes and footnotes should not be used in the document.** However, **you can mention short notes to elucidate some specific point**, which may be placed in number orders before the references.

**PLEASE USE THE FOLLOWING FOR STYLE AND PUNCTUATION IN REFERENCES:**

**BOOKS**

- Bowersox, Donald J., Closs, David J., (1996), "Logistical Management." Tata McGraw, Hill, New Delhi.
- Hunker, H.L. and A.J. Wright (1963), "Factors of Industrial Location in Ohio" Ohio State University, Nigeria.

**CONTRIBUTIONS TO BOOKS**

- Sharma T., Kwatra, G. (2008) Effectiveness of Social Advertising: A Study of Selected Campaigns, Corporate Social Responsibility, Edited by David Crowther & Nicholas Capaldi, Ashgate Research Companion to Corporate Social Responsibility, Chapter 15, pp 287-303.

**JOURNAL AND OTHER ARTICLES**

- Schemenner, R.W., Huber, J.C. and Cook, R.L. (1987), "Geographic Differences and the Location of New Manufacturing Facilities," Journal of Urban Economics, Vol. 21, No. 1, pp. 83-104.

**CONFERENCE PAPERS**

- Garg, Sambhav (2011): "Business Ethics" Paper presented at the Annual International Conference for the All India Management Association, New Delhi, India, 19–23

**UNPUBLISHED DISSERTATIONS**

- Kumar S. (2011): "Customer Value: A Comparative Study of Rural and Urban Customers," Thesis, Kurukshetra University, Kurukshetra.

**ONLINE RESOURCES**

- Always indicate the date that the source was accessed, as online resources are frequently updated or removed.

**WEBSITES**

- Garg, Bhavet (2011): Towards a New Gas Policy, Political Weekly, Viewed on January 01, 2012 <http://epw.in/user/viewabstract.jsp>

# A GENETIC ALGORITHM BASED IMAGE AUTHENTICATION TECHNIQUE IN FREQUENCY DOMAIN USING HAAR WAVELET TRANSFORM (AGAIAFDHWT)

**Dr. AMRITA KHAMRUI**  
**ASST. PROFESSOR**  
**DEPARTMENT OF COMPUTER SCIENCE**  
**VIVEKANANDA COLLEGE**  
**MADHYAMGRAM**

## ABSTRACT

*In this study A Genetic Algorithm based image authentication technique in frequency domain using Haar Wavelet Transform has been proposed. This is frequency domain-based image authentication technique where Haar Wavelet Transform is applied to generate transformed domain coefficients. A  $2 \times 2$  mask is taken from the source image in row major order. Haar Wavelet Transform is applied on it. It generates four frequency coefficients. Three bits of the authenticating image is embedded onto each coefficient of the mask. Embedded mask is processed through reverse Haar Wavelet Transform to generate spatial domain coefficients. Stego sub intermediate image is obtained by applying this process for the whole cover image. Stego sub intermediate image is processed through Genetic Algorithm to generate optimized stego image. New Generation followed by Crossover and Mutation are performed as part of Genetic Algorithm. Genetic Algorithm is applied as optimization process to generate optimized stego image. The proposed technique has been compared with existing approaches. It has been shown that the proposed technique obtains better result as compared to the existing approaches.*

## KEYWORDS

haar wavelet transform (HWT), image fidelity (IF), mean square error (MSE), peak signal to noise ratio (PSNR), structural similarity index measure (SSIM).

## JEL CODE

O31

## 1. INTRODUCTION

In today's scenario image authentication is achieved through Steganography. Image Steganography is a way of sending messages through digital image such that apart from sender and receiver no one realize that there is some hidden information. Image authentication can be achieved in two ways, using spatial domain steganography [11], [12] and using frequency domain steganography [10]. To detect unauthorized access information security and image authentication are widely used. The secret information is embedded inside the source image by keeping its visibility intact. The widely used methods to achieved this is to use the least-significant bit (LSB) replacement [7] by masking [17], [8], filtering and transformations [9] on the source image. Usually, the cover image for steganography can be picture [11], video, sound file etc. A message may be hidden invisibly by using some steganographic algorithms [13] to ensure the security [14] which is main concern today for trafficking across the network. Security may be achieved by hiding information into images. Hiding information in the image without changing its visibility is an important task for image authentication. Ownership verification [15] is very crucial for military people, research institute and scientists. Security of information and authentication of images is very important for the digital image document to prevent them for unauthorized access. Data hiding refers to the nearly invisible embedding of information within a host data set as message, image or video. The data hiding is a process of creating hypermedia document or image, which is very less convenient to manipulate. Steganography is a process of hiding the message/image in the source image by some algorithm and cryptography is a process to hide the message content. The process is to hide a message inside an image by keeping the visibility intact. The most common method of steganography is least-significant bit (LSB) substitution developed through masking, filtering and transformations on the source image. In the proposed technique a frequency domain-based image authentication technique using Haar Wavelet Transform has been proposed.

Present proposal is an algorithm which would facilitate secure message transmission through block base data hiding. Most of the works used minimum bits of the hidden image for embedding in transform domain, but the proposed algorithm embeds large amount of information in transformed domain with a minimum distortion of visual property.

Rest of the paper is organized as follows: section 2 deals with review of literature. Importance of the study has been discussed in section 3. Section 4 deals with the statement of the problem. Section 5 discuss the objective of the proposed technique. Section 6 discuss the hypothesis. The research methodology has been discussed in section 7. Results and discussion are discussed in section 8. Some findings are mentioned in section 9. Recommendations are given in section 10. Conclusion, limitation and scope for future research are presented in section 11, 12 and 13. References are drawn at end.

## 2. REVIEW OF LITERATURE

Some existing methodologies have been discussed in this section. In 2022, Tevaramani Saleem S et al. [1] proposed image steganography performance analysis using discrete wavelet transform and alpha blending for secure communication. In this technique, alpha is called a scaling parameter. Source image and authenticating images are of different types and dimensions, images from a webcam, and other predefined images of different formats have been normalized and pre-processed. A Haar Discrete Wavelet Transformation (DWT) is applied to both the source image and authenticating images. Stego image is generated by encrypting the authenticating image and embedded in the source image. In 2021 Varuikhina Vladmir et al. [2] proposed continuous Wavelet Transform applications in steganography. This paper focused on successful hiding of information with the help of Wavelet Transformation without any visible changes in the source image. This technique used single level 2D Discrete Wavelet Transformation such as Daubeshi, Haar and coiflet transformation and compare the result generated by each transformation with one another. In 2020 Govindasamy V et al. [3] proposed coverless image steganography using Haar Integer Wavelet Transform. Coverless image steganography is recent technique of steganography. It does not modify the source image that is carrying the secret information. This approach used Haar Integer Wavelet Transform to hide more secret information using coverless image steganography. Here the image has been divided into sub matrices. Then Integer Wavelet Transformation is used to generate coefficients of the submatrix. Then the sub-matrices are reshaped to form arrays. The coefficients are converted into binary bits. At last, the secret message has been fragmented into 8 bits to match it with the block and starting location in the array. The location map has been used to obtain the secret information by the receiver. In 2016 Houssein Essam H. et al. [4] proposed an image steganography algorithm using Haar Discrete Wavelet transform with advanced encryption system. In this paper Advanced Encryption System (AES) is used to encrypt data and Haar Discrete Wavelet Transform (HDWT) is used for embedding. HDWT decreases the complexity in image steganography and ensures less image distortion and lesser detect-ability. In 2010 Hui-Yu Huang et al proposed a lossless data hiding technique which is based on [5] discrete Haar wavelet transform. In 2016 Tushara M et al [6] proposed a review of image Steganography using discrete wavelet transform which presents a review on steganography techniques that use discrete wavelet transform.

## 3. IMPORTANCE OF THE STUDY

The existing techniques used different steganographic techniques with less embedding capacity. This paper presents an algorithm that would facilitate secure message transmission using frequency domain based data hiding procedure that uses Haar Wavelet Transformation. This method embeds large amount of secret information as compared to the existing works [3],[5],[6],[8] by keeping visual property intact.

4. STATEMENT OF THE PROBLEM

The detailed study of the Review of Literature reveals the following facts:

The proposed technique is frequency domain based image authentication approach which uses Genetic Algorithm for optimization. Most of the previous techniques having less capacity of hiding secret information but the proposed technique can embed huge amount of secret information. The proposed technique uses Genetic Algorithm as optimization technique but very few existing techniques used optimization.

5. OBJECTIVE

The objective of this research is to develop a new Genetic Algorithm based frequency domain based image authentication technique and can hide large message by keeping quality stego image.

6. HYPOTHESIS

The proposed technique uses some benchmark images taken from the USC-SIPI Image Database: Version 5, Original release: October 1997, Signal and Image Processing Institute, University of Southern California, Department of Electrical Engineering [20]. Source and authenticating images are considered as color images.

7. RESEARCH METHODOLOGY

In this section the technique (AGAIAFDHW) is illustrated with an example in detail. A mask of size 2 x 2 from source color image is taken in row major order. The mask is processed through haar wavelet transformation to generate frequency domain coefficients. Three bits of the authenticating image is embedded onto each transformed image coefficient sub mask. The image sub mask is transformed from frequency domain to spatial domain using inverse haar wavelet transform. The dimension of the hidden image is embedded first. Resulting stego image is processed through Genetic Algorithm to generate optimized stego image. As part of GA, New Generation followed by Crossover and Mutation are applied on it. In New Generation eight random chromosomes are taken as initial population. Roulette wheel selection has been applied to generate two fittest chromosomes using hash function. Two fittest chromosomes are processed through crossover and two new off springs has been generated in this process. Off springs are then processed through Mutation. In elitism most fitted offspring is chosen for next iteration. This process is repeated until the optimum value has been received. The Haar functions can be defined as follows:

$$\begin{aligned} \text{har}(0, \theta) &= 1, 0 \leq \theta \leq 1, \\ \text{har}(1, \theta) &= \begin{cases} 1, & 0 \leq \theta < 1/2, \\ -1, & 1/2 \leq \theta < 1, \end{cases} \\ \text{har}(2, \theta) &= \begin{cases} \sqrt{2}, & 0 \leq \theta < 1/4, \\ -\sqrt{2}, & 1/4 \leq \theta < 1/2, \\ 0, & 1/2 \leq \theta \leq 1, \end{cases} \\ \text{har}(3, \theta) &= \begin{cases} 0, & 0 \leq \theta < 1/2, \\ \sqrt{2}, & 1/2 \leq \theta < 3/4, \\ -\sqrt{2}, & 3/4 \leq \theta \leq 1, \end{cases} \\ &\vdots \\ \text{har}(2^p + n, \theta) &= \begin{cases} \sqrt{2^p}, & n/2^p \leq \theta < (n+1/2)/2^p, \\ -\sqrt{2^p}, & (n+1/2)/2^p \leq \theta < (n+1)/2^p, \\ 0, & 0 < \theta < \frac{n}{2^p} \text{ and } \frac{(n+1)}{2^p} < \theta < 1, \end{cases} \\ p &= 1, \dots; \quad n = 0, \dots, 2^p - 1. \end{aligned}$$

The original Haar defined as follows

$$\text{har}(k, 0) = \lim_{\theta \rightarrow 0, \theta > 0} \text{har}(k, \theta),$$

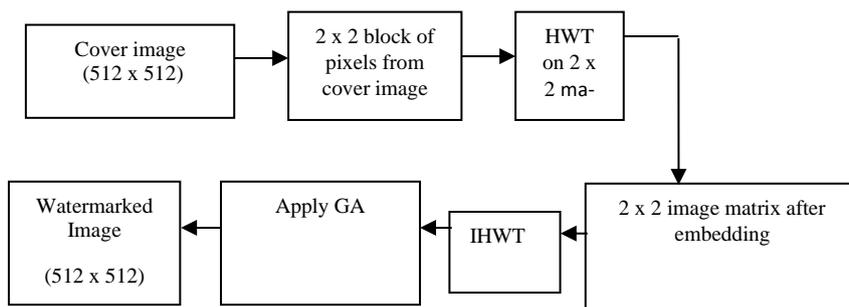
$$\text{har}(k, 1) = \lim_{\theta \rightarrow 1, \theta < 0} \text{har}(k, \theta),$$

A discrete Haar function is given in equation

$$r_n(t) = \sum_{k=0}^{2^n-1} \psi_{n,k}(t), \quad t \in [0, 1], \quad n \geq 0. \tag{1}$$

This is an orthogonal function. The process of AGAIAFDHW is shown in Figure 1. Process of embedding is explained in section 7.1 process of extraction is explained in section 7.2. An example of AGAIAFDHW is given in section 7.3.

FIGURE 1: SCHEMATIC DIAGRAM OF AGAIAFDHW



7.1 ALGORITHM FOR INSERTION

In this technique color image of size p x q is taken as source image. Color image of m x n is taken as authenticating image. Authenticating image bits are embedded in the transformed coefficients mask of source image.

Input: Source image of size p x q, authenticating image of size m x n.

Output: Optimized stego image of size p×q.

Method: Insertion of authenticating image bits into the source color image followed by GA

**Step 1:** Extract the dimension of the authenticating image from the header part of the authenticating image

**Step 2:** For each hidden message/image, read color source image mask of size 2×2 in row major order. Apply Discrete Haar wavelet transform to obtain transformed coefficients.

**Step 3:** The Dimension is embedded first followed by authenticating image pixel onto three LSBs of each coefficient. First sixteen embedding positions are reserved for width and next sixteen are reserved for height.

**Step 4:** Apply inverse haar Transform to back the mask into spatial domain.

**Step 5:** The process is repeated for the whole cover image to generate stego image.

**Step 6:** Stego image pixel is processed through GA. Eight random chromosomes are generated as initial population.

**Step 7:** The hidden information's are fabricated on last three LSB positions. Roulette Wheel selection is applied on it using a hash function to select two fittest chromosomes among eight initial populations.  $f(n) = 1/(\text{mod}(s(x,y) - c(x,y))+1)$

$f(n)$  is the fitness function,  $s(x,y)$  is intensity value of stego image pixel for the coordinate  $(x, y)$ .  $c(x, y)$  is intensity value of Host image/ source image for the same coordinate  $(x, y)$ . Best fittest value is one.

**Step 8:** Uniform crossover process is followed where first bit is from first parent and second bit is from second parent which will be started from fourth bit.

**Step 9:** Mutation is applied on the crossoverd chromosome by flipping a bit from one to zero or from zero to one which will be started from fourth bit.

**Step 10:** Elitism forwards the best chromosome in the next iteration by eliminating the weak chromosome. In elitism the average between the consecutive two pixels of the mutated image should be same as the average between the consecutive two pixels in the source image for the odd positions of the image and difference between the average and the odd position coefficient should be same as the previous one to ensure proper decoding.

**Step 11:** Step 7 to Step 10 is repeated for the whole stego image to generate optimized stego image.

**7.2 ALGORITHM FOR EXTRACTION**

The optimized stego image is received in spatial domain. The optimized stego image is taken as the input and the hidden message/ image size, content is extracted from it.

Input: optimized stego image of size p×q.

Output: Source image of size p×q, authenticating image of size m×n.

Method: Extract bits of authenticating image from embedded image

**Step 1:** Apply Discrete Haar wavelet transform onto the stego image mask to regenerate four frequency components

**Step 2:** Extract the authenticating image bits from three LSBs of each coefficient. Replace authenticating message/ image bit position in the block by '1'. For each eight extracted bits construct one image pixel of authenticating image. First sixteen bit extraction form the width of the authenticating image and next sixteen bit extraction form the height of the authenticating image.

**Step 3:** Repeat step 1 and 2 to regenerate authenticating image as per size of the authenticating image.

Step 4: Stop

**7.3 EXAMPLE**

**FIGURE 2**

226 137 91 209	165 7 15 51	161 3 11 51	220 124 96 204
2.i Source image mask	2.ii Transformed mask	2.iii Embedded mask	2.iv Embedded mask in spatial domain
230 140	230 140	246 156	220 124
2.v New Generated chromosome	2.vi Cross overed offspring	2.vii Mutated offspring	2.viii Resultant mask

Figure 2.i shows a 2 × 2 block of the source image. Haar wavelet transformation has been applied on it to generate four frequency components. Figure 2.ii Shows the transformed coefficients of the source image. Figure 2.iii shows the embedded image block. Three bits of the authenticating image is embedded onto each transformed coefficients of the source image, in three LSB positions. Figure 2.iv shows the embedded image block in spatial domain after applying reverse Haar Wavelet transformation. Embedded image is processed through Genetic Algorithm to generate optimized stego image. Each byte is processed through New Generation. Eight random chromosomes are generated for initial population as 3 LSB positions have been taken for embedding. Roulette Wheel selection has been applied to generate new generated image. Figure 2.v shows the new Generated chromosome. Two consecutive pixels are taken for Crossover. Uniform Crossover has been applied on two consecutive pixels after fourth bit position. Figure 2.vi shows the crossover offspring. Figure 2.vii shows the mutated offspring. Figure 2.viii shows the result of elitism.

**8. RESULTS AND DISCUSSION**

The proposed technique AGAIAFDHWT has been applied to some benchmark images taken from the USC-SIPI Image Database: Version 5, Original release: October 1997, Signal and Image Processing Institute, University of Southern California, Department of Electrical Engineering [16]. The machine specification is Intel Core 2 Duo CPU, 3.00 GB RAM and 2.00 GHz speed for experiment. The experiment had been performed in Linux platform with GCC compiler. Source images are of dimension 512 × 512, authenticating image is of size 256 × 256 and resultant image is of dimension 512 × 512. This section discusses the results of visual interpretation and performance of the algorithm based on some statistical parameters such as PSNR, MSE, IF and SSIM. Figure 3.i to 3.v shows the source images Lenna, Baboon, Pepper, Tahoe, Toucan. Figure 3.vi shows the authenticating image jet. Embedded images are shown from Figure 3.vii to 3.xi on embedding jet image using AGAIAFDHWT. It is clear from Figure 3 that there are no such changes visible in the embedded image. Table 1 show the PSNR, MSE, IF and SSIM [17] value for each of the embedded image against the source image. From Table 1, the maximum value of the PSNR shown is 39.954643 and that of minimum value of the PSNR is 36.987556. PSNR means peak signal to noise ratio and high value of PSNR in Table 1 indicates that the optimized stego image have better quality. MSE means mean square error. Low value of MSE indicates low error. IF is image fidelity and is used to compare between two images. IF is closer to 1 means two images are almost same. SSIM refers to structural similarity index measurement is obtained by comparing the pixel intensities. It is another metric for comparing the image degradation due to information hiding. If two images are same it will be 1. The formula for PSNR, MSE, IF and SSIM is given in equation (2), (3), (4) and (5) respectively, where the formula for  $\mu_x$  and  $\mu_y$  is given in equation (6) and  $\sigma_x^2$  and  $\sigma_y^2$  are given in equation (7) and (8). The formula for  $\sigma_{xy}$  is given in equation (9) and the constants C1, C2, K1, and K2 are given in equation (10). Table 2 compare with existing [2] and [5]. It has been shown that the proposed technique obtains better result compared to the existing [2] and [5]. The following formula is used to calculate PSNR, MSE and IF (image fidelity) and SSIM.

$$PSNR = 10 \log(\max(I_{m,n}^2)/MSE) \tag{2}$$

$$MSE = \frac{1}{MN} * \sum_{m,n} (I_1 m, n - I_2 m, n)^2 \tag{3}$$

$$IF = 1 - \sum_{m,n} (I_{1m,n} - I_{2m,n})^2 / \sum_{m,n} I_{2m,n}^2 \tag{4}$$

$$SSIM = 2(\mu_x\mu_y + C_1) (2\sigma_{xy} + C_2) / ((\mu_x^2 + \mu_y^2 + C_1) * (\sigma_x^2 + \sigma_y^2 + C_2)) \tag{5}$$

$$\mu_x = 1/N \sum_{i=1}^N X_i, \mu_y = 1/N \sum_{i=1}^N Y_i \tag{6}$$

$$\sigma_x^2 = 1/(N-1) \sum_{i=1}^N (X_i - \mu_x)^2 \tag{7}$$

$$\sigma_y^2 = 1/(N-1) \sum_{i=1}^N (Y_i - \mu_y)^2 \tag{8}$$

$$\sigma_{xy} = 1/(N-1) \sum_{i=1}^N (X_i - \mu_x) (Y_i - \mu_y) \tag{9}$$

The value of C<sub>1</sub>, C<sub>2</sub> and C<sub>3</sub> with K<sub>1</sub> and K<sub>2</sub> are from [17] given in equation (10)  
 C<sub>1</sub> = (K<sub>1</sub> L)<sup>2</sup>, C<sub>2</sub> = (K<sub>2</sub> L)<sup>2</sup>, K<sub>1</sub> = 0.01, K<sub>2</sub> = 0.03 (10)

FIGURE 3: VISUAL INTERPRETATION OF VARIOUS SOURCE AND OPTIMIZED STEGO IMAGES ALONG WITH AUTHENTICATING IMAGE

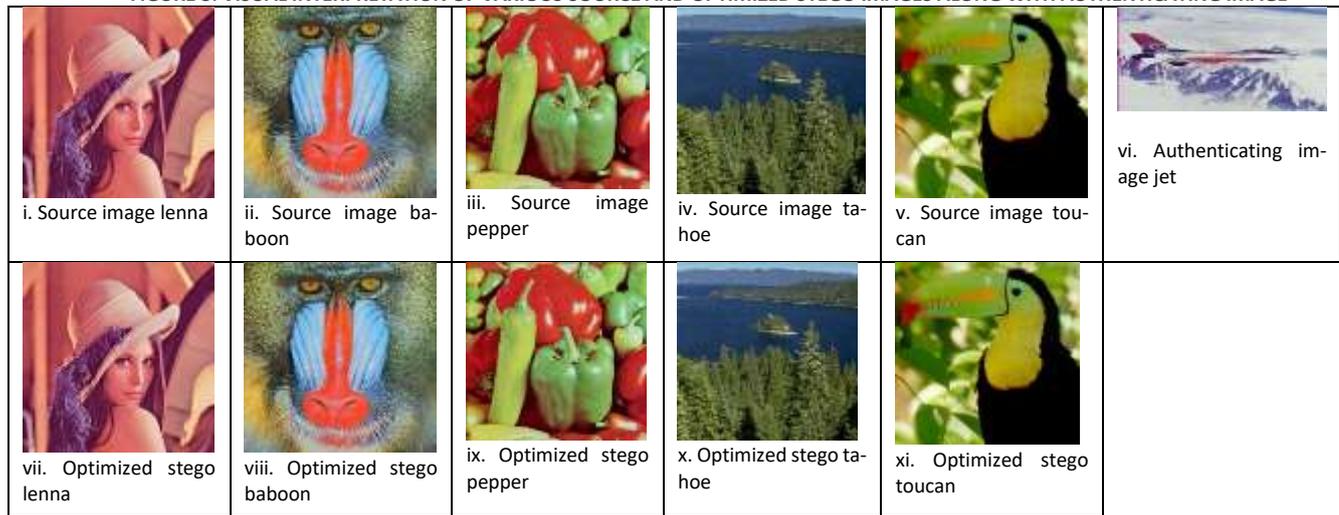


TABLE 1: PSNR, MSE, IF & SSIM VALUES OF VARIOUS OPTIMIZED STEGO IMAGES

Cover image	PSNR	MSE	IF	SSIM
JET	39.954643	6.570766	0.999812	0.999996
BABOON	38.825237	8.522297	0.999555	0.999999
TOUCAN	36.987556	13.011417	0.998222	0.999999
TAHOE	38.796883	8.578120	0.998945	0.999998
BOAT	38.875088	8.425030	0.999576	0.999998
HOUSE	39.052338	8.088099	0.999702	0.999999
LENA	38.588503	8.976978	0.999549	0.999768
SAFARI10	39.063881	8.066632	0.999106	0.993818
PEPPER	38,138615	9.982032	0.999399	0.999561
PEACEFUL	38.472198	9.244015	0.999022	0.995216

TABLE 2: COMPARISON OF PSNR VALUES OBTAINED FOR VARIOUS IMAGES USING AGAIAFDHWT AND EXISTING[2] AND [5]

Host Image	PSNR values of AGAIAFDHWT	Capacity of AGAIAFDHWT	PSNR values of existing[5]	Capacity of existing[5]	PSNR of existing[2]
Jet	39.954643	54000	28.49	39445	-
Lenna	38.588503	54000	29.68	39706	37.2
Boat	38.875088	54000	28.74	38948	-

9. FINDINGS

It has been observed that the proposed technique obtains consistent PSNR value compared to the capacity. IF value ensures that the source image and the stego images are identical.

10. RECOMMENDATION

The proposed technique can be finding applications in the following areas:

1) IMAGE AUTHENTICATION

This technique (AGAIAFDHWT) can be used for Image authentication. The image/ document that needs to be authenticated is considered as source image. Another small image/message can be taken as authenticating image. Authenticating image is embedded using this approach and send across the network by keeping the

visibility of the source image intact. At the receiver end it is extracted by the receiver using extraction algorithm. If the extracted image and the original authenticating images are same, then the image is authentic.

#### II) TELEMEDICINE

Telemedicine refers to the way of remote medical services via real time both way communication between patient and the doctor using audio or visual electronic means. If the image is communicated over the transmission media, then it should be authenticated. The process of authentication is stated in previous example. So AGAIAFDHWT can be used in image authentication.

#### III) DOCUMENT AUTHENTICATION

AGAIAFDHWT can be used in document authentication. Consider a legal image. It consists of an image part and a text part. If someone tampers the document it can be recognized through AGAIAFDHWT. From the text part, a digest can be generated using any message digest algorithm and it can be embedded in the image part of the legal document. When it needs to be authenticated the digest is extracted from the image part of the legal document and another digest is generated from the text part of the legal document. If both are same then the legal document is authenticated.

#### IV) BANK TRANSACTION

Bank can use secure e-payment method through steganography. Now a days there is huge demand of online shopping. This required online transaction. So there is an important task to protect the customers information. AGAIAFDHWT technique can be used in this purpose.

#### V) SMART CITY APPLICATION

AGAIAFDHWT can be used to hide the health record electronically for smart city application.

#### VI) SECURE MESSAGE TRANSMISSION

AGAIAFDHWT can be used for secure message transmission. The message that is to be transmitted in secure way should be hidden in the source image and can be transmitted over network. At the receiver end it is extracted from the transmitted image in reverse way.

### 11. CONCLUSION

The technique AGAIAFDHWT is a frequency domain based color image authentication technique uses Discrete Haar Wavelet Transformation. This technique hides large message with a little distortion by using Discrete Haar Wavelet Transformation and Genetic Algorithm is used to maintain the quality of the optimized stego image. This technique has been compared with some existing Wavelet Transformation-based approaches and the result shows that the technique works better than those approaches.

### 12. LIMITATION

The proposed technique is a frequency domain approach. The hiding capacity can be increased.

### 13. SCOPE FOR FUTURE RESEARCH

The proposed technique uses Haar Wavelet Transform. It can be applicable for other transformation also and hence future scope of the research.

### REFERENCES

1. Tevaramani Saleem S et al. "Image steganography performance analysis using discrete wavelet transform and alpha blending for secure communication ", Global Transitions Proceedings Volume 3, Issue 1, June 2022, Pages 208-214.
2. Varuikhina Vladimir et al. "Continuous Wavelet Transform Applications In Steganography", 14th International Symposium on Intelligent Systems, Procedia Computer Science 186 (2021) 580–587.
3. Govindasamy V et al. "Coverless Image Steganography using Haar Integer Wavelet Transform ", Proceedings of the Fourth International Conference on Computing Methodologies and Communication (ICCMC 2020) IEEE Xplore Part Number:CFP20K25-ART; ISBN:978-1-7281-4889-2
4. Houssein Essam H. et al. "An Image Steganography Algorithm using Haar Discrete Wavelet Transform with Advanced Encryption System", Proceedings of the Federated Conference on Computer Science and Information Systems pp. 641–644, DOI: 10.15439/2016F521, ACSIS, Vol. 8. ISSN 2300-5963, 978-83-60810-90-3/\$25.00 c 2016, IEEE.
5. Hui-Yu Huang et al. "A lossless data hiding based on discrete Haar wavelet transform", CIT 2010, 978-0-7695-4108-2/10 \$26.00 © 2010, IEEE DOI 10.1109/CIT.2010.276
6. Tushara M et al. "Image Steganography using discrete wavelet transform- A Review", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, ISSN: 2321-2004, Vol 3, Special Issue 1, February 2016
7. Lifang Yu et al "Improved Adaptive LSB steganography based on Chaos and Genetic Algorithm", EURASIP Journal on Advances in Signal Processing, volume 2010, Article ID 876946.
8. Khamrui, A et al., "A Genetic Algorithm based Steganography on Color Images (GASCI)" International journal of Signal and Imaging Systems Engineering (IJSISE), 2014 Vol.7 No.1, pp.59 – 63, Inderscience publishers.
9. Khamrui, A et al., "An Image Authentication Technique in Frequency Domain using Genetic Algorithm (IAFDGA)", International Journal of Software Engineering & Applications (IJSEA), ISSN: 0975-9018 (Online),0976-2221(Print), Indexed on DOAJ,EBSCO,INSPEC,ProQuest etc.,DOI:10.5121/ijsea.2012.3504, Vol. 3, No. 5, pp.39-46, Sept. 2012.(http://aircse.org/journal/ijsea/papers/3512ijsea04.pdf),2012.
10. Khamrui, A. et al. "A Genetic Algorithm based approach in image authentication using Z transform" (GAIAZT), AMSE Journals 2014 series: Advances B, Vol. 57,N<sup>o</sup>1, pp 20-30.
11. Khamrui A et al., 'A Data Embedding Technique for Gray Scale Image Using Genetic Algorithm (DEGGA)', IEEE Sponsored International Conference on Communication, Computing & Security (Proceedings by Excel India Publishers ISBN-978-93-80697-505), NIT Rourkela, pp. 19–22 7–9 January 2010.
12. Khamrui A. et al., "A Data- Hiding Scheme for Digital Image using Pixel Value Differencing (DHPVD)", International Symposium on Electronic System Design (ISED),Print ISBN:978-1-4577-1880-9, 19-21 December, Kochi, India, 2011.
13. Hashad A. I. et al "A Robust Steganography Technique using Discrete Cosine Transform Insertion" Information and communication Technology, 2005. Enabling Technologies for the New Knowledge Society: ITI 3<sup>rd</sup> International Conference.ISBN:0-7803-9270-1.
14. Khamrui A. et. al. "A Novel Genetic Algorithm Based Data Embedding Technique in Frequency Domain Using Z Transform (ANGAFDZT)", DPPR- 2012, Advances in Intelligent Systems and Computing, 177, ISSN No.: 2194-5357, Springer, pp. 885-893, July 13 -15, Chennai, India (2012).
15. Khamrui A., Mandal, J. K, "A spatial Domain Image Authentication Technique using Genetic algorithm", International Conference on Computational Intelligence, Communications, and Business Analytics, CICBA 2017, Computational Intelligence, Communications, and Business Analytics pp 577–584, Springer Link.
16. Allan G. Weber, The USC-SIPI Image Database: Version 5, Original release: October 1997, Signal and Image Processing Institute, University of Southern California, Department of Electrical Engineering. http://sipi.usc.edu/database/(Last accessed on 20<sup>th</sup> January, 2011).
17. Mandal Jyotsna Kumar, "Reversible Steganography and Authentication via Transform Encoding", Studies in Computational Intelligence 901, Springer Nature.

**QUANTIFYING THE EXPLAINABILITY OF MACHINE LEARNING MODELS: METRICS AND BENCHMARKS****K.ROOPA****ASST. PROFESSOR****MADANAPALLE INSTITUTE OF TECHNOLOGY & SCIENCE  
ANGALLU****ABSTRACT**

*With the increasing adoption of machine learning models in domains of high societal impact, ensuring their explainability has become paramount. This study delves into the intricate balance between model performance and interpretability, shedding light on the challenges in quantifying explainability. We highlight the multifaceted nature of interpretability, ranging from the subjectivity of explanations to the diverse needs across domains. Employing a range of datasets and model architectures, we evaluate various techniques aiming to enhance model transparency. Our findings underscore the pressing need for holistic explainability frameworks, domain-adapted solutions, and community-driven benchmarks. As we integrate AI deeper into decision-making processes, this research emphasizes that the path forward is not only about achieving high model accuracy but also about fostering trust and understanding. The goal is clear: a future where AI systems are both powerful and transparent, ensuring the benefits are accessible, comprehensible, and equitable for all.*

**KEYWORDS**

explainability, machine learning models, interpretability metrics, trustworthy ai, model benchmarks, transparency, feature importance, local and global explanations.

**JEL CODE**

O31

**INTRODUCTION**

In recent years, machine learning (ML) has witnessed unprecedented advancements, leading to transformative applications across myriad sectors—from healthcare diagnostics to financial forecasting. These powerful algorithms, capable of processing vast amounts of data and detecting intricate patterns, hold the potential to significantly impact decision-making processes. However, alongside their burgeoning capabilities, a fundamental concern has arisen: the "black-box" nature of many state-of-the-art models.

The term "black-box" alludes to complex models, like deep neural networks or certain ensemble methods, which, despite their high predictive accuracy, often lack transparent reasoning behind their predictions. This opacity can be particularly concerning in high-stakes environments, such as medical diagnoses or credit approvals, where the consequences of decisions are profound, and stakeholders demand clarity.

But, what does it mean for a model to be "explainable"? And how can we quantitatively measure such a qualitative property? This research dives deep into these questions, exploring the nuances of model explainability, the metrics to gauge it, and the benchmarks to test these metrics. As we stand at the crossroads of an AI-driven era, it's imperative to ensure that these tools are not just performant, but also interpretable, trustworthy, and accountable to their human users.

In the following sections, we will unpack the challenges of quantifying explainability, investigate various metrics developed to measure it, and evaluate these metrics against standardized benchmarks. Through this exploration, we aim to offer a comprehensive perspective on where the field currently stands and the path it needs to traverse to ensure a harmonious future where humans and machines collaboratively drive decisions.

**OBJECTIVES OF THE STUDY**

1. To devise robust quantitative metrics for evaluating the explainability of machine learning models. These metrics should encompass various aspects of explainability, such as fidelity, comprehensibility, stability, and consistency, to provide a comprehensive assessment.
2. To establish a standardized benchmarking framework that enables consistent evaluation of explainability techniques across different model architectures, datasets, and application domains.

**RESEARCH METHODOLOGY**

The study is based on both primary and secondary data collected through various journals, magazines and websites.

**LITERATURE REVIEW**

Early work by Doshi-Velez and Kim (2017) introduced the concept of interpretability and emphasized the need for a rigorous science of interpretable machine learning.

Ribeiro et al. (2016) proposed the LIME framework for explaining the predictions of any classifier, providing a model-agnostic approach to local interpretability.

Lundberg and Lee (2017) developed SHAP (SHapley Additive exPlanations), a unified approach to interpreting model predictions based on Shapley values, which has gained widespread adoption in the field.

**THE NEED FOR EXPLAINABILITY**

As machine learning (ML) models become more sophisticated, their internal mechanisms often become harder to interpret. These so-called "black-box" models, which include deep neural networks, ensemble methods, and others, can produce highly accurate predictions, but understanding how they arrive at these predictions remains challenging. The opacity of these models has given rise to a critical demand for explainability in the ML community.

**CRITICAL REAL-WORLD IMPLICATIONS****HEALTHCARE**

In healthcare, where ML models assist in diagnosis, treatment recommendations, and prognosis predictions, understanding the rationale behind a model's decision is paramount. Wrong predictions without explanations can be life-threatening.

**FINANCE**

In the financial sector, ML models are used for credit scoring, fraud detection, and investment strategies. Providing reasons for a credit denial or recognizing why a particular transaction was flagged as fraudulent is not just a regulatory requirement but also central to customer trust.

**LEGAL AND CRIMINAL JUSTICE**

ML models are increasingly being used for risk assessment in criminal justice settings. The decisions made here can affect individuals' liberties, making it essential to ensure transparency and avoid biases.

## ETHICAL CONSIDERATIONS

**Accountability:** If a model's decision leads to a negative outcome, there needs to be a clear understanding of how that decision was made to hold the relevant parties accountable.

**Bias and Fairness:** Without explainability, biases hidden within models—often arising from biased training data—can go unnoticed and uncorrected. This can perpetuate systemic injustices and inequalities.

**Trust:** For end-users, especially domain experts like doctors or financial analysts, to trust an ML model, they need to understand its decision-making process. Trust is crucial for the broader adoption of ML solutions.

## REGULATORY MANDATES

Many sectors have regulations that require decisions made by algorithms to be explainable:

**General Data Protection Regulation (GDPR):** The European Union's GDPR has provisions that can be interpreted as giving individuals the right to an explanation when subjected to automated decisions.

**Financial Services:** Regulations often require that customers be given reasons for decisions, such as loan denials.

## CHALLENGES IN ACHIEVING EXPLAINABILITY

- **Trade-off with Model Complexity:** Simpler models are often more interpretable but may not achieve the same accuracy as more complex models.
- **Subjectivity:** What's considered "explainable" can be subjective and vary among users. A technical explanation may suffice for a data scientist, while a domain expert might need contextual reasoning.

## EXPLAINABILITY Vs. ACCURACY TRADE-OFF

One of the long-standing tensions in machine learning revolves around the trade-off between explainability and accuracy. As models grow more complex, they often yield better accuracy, but their interpretability diminishes, turning them into "black-box" systems.

## THE RISE OF COMPLEX MODELS

With the advent of deep learning and ensemble methods, the machine learning community has achieved unprecedented accuracies in tasks ranging from image recognition to natural language processing. These models, however:

- **Incorporate Millions of Parameters:** Deep neural networks, especially, can have millions of tunable parameters, making it difficult to discern how any specific input influences the output.
- **Non-linearities:** Many modern models employ complex non-linear functions that challenge straightforward interpretation.

## THE ALLURE OF SIMPLICITY

On the flip side, simpler models like linear regression or decision trees, while more interpretable, might not capture intricate patterns in the data. They offer:

- **Clear Mechanisms:** The decision boundaries or logic are often transparent and can be visually represented.
- **Feature Importance:** These models can usually provide a clear ranking of feature importance.

## REAL-WORLD IMPLICATIONS OF THE TRADE-OFF

- **Healthcare:** A highly accurate model that predicts patient risk might be favored, but if doctors can't understand its predictions, they might be hesitant to rely on it.
- **Finance:** Investment models might yield great returns, but if they're not interpretable, they can't be audited or easily adjusted by human experts.

## STRIKING A BALANCE

- **Model-Agnostic Explanations:** Techniques like LIME or SHAP aim to offer explanations for any model by approximating its decisions using simpler, interpretable models.
- **Regularization for Simplicity:** Some methods introduce regularization to make models like neural networks sparser, and thus more interpretable, without significant accuracy losses.
- **Attention Mechanisms:** In deep learning, attention mechanisms can highlight parts of the input data (like words in a sentence) that were pivotal in making a decision.

## METRICS FOR QUANTIFYING EXPLAINABILITY

Explainability in machine learning refers to the degree to which a human can understand the cause of a decision made by a model. This is especially important in applications where understanding model behavior is crucial for trust, compliance, or debugging. Several metrics have been proposed to measure and benchmark the explainability of models.

### FEATURE IMPORTANCE METRICS

These metrics quantify the importance of input features in determining the predictions:

- **Permutation Importance:** Evaluates the change in model performance (e.g., accuracy) when the values of a specific feature are randomly shuffled.
- **SHAP (SHapley Additive exPlanations):** Based on cooperative game theory, it provides a unified measure of feature importance by averaging all possible combinations of features.

### MODEL FIDELITY METRICS

These metrics evaluate how well a simpler, interpretable model can approximate the predictions of a complex model:

- **LIME (Local Interpretable Model-agnostic Explanations):** Fits a simpler model (like a linear model) to the predictions of a complex model but only in the locality of a specific instance.

### VISUAL INTERPRETABILITY METRICS

Visualization tools and metrics that provide insights into model behavior:

- **Saliency Maps:** In deep learning, these maps highlight regions in input data (like an image) that were most influential in a model's decision.
- **Activation Maximization:** Visualizes the input that would maximally activate a particular neuron in a neural network.

### TEXT-BASED EXPLANATIONS

Metrics evaluating the quality and understandability of textual explanations associated with model predictions:

- **Counterfactual Explanations:** Describes an instance by highlighting what minimal changes would need to be made for the model to change its prediction.
- **Influence Functions:** Identify training instances that most influenced a particular prediction.

## USER STUDIES

While not a direct metric, gathering feedback from end-users or domain experts provides qualitative insights into a model's explainability:

- **Comprehension Metrics:** Measures based on users' ability to understand, predict, or trust model predictions after seeing explanations.
- **Satisfaction Ratings:** User feedback on the quality, usefulness, or trustworthiness of provided explanations.

## AGGREGATED INTERPRETABILITY METRICS

Some metrics aim to provide a holistic view of explainability by aggregating multiple dimensions:

- **Model Complexity vs. Performance Plots:** Graphs that plot a model's performance against its complexity, highlighting the trade-off between accuracy and interpretability.

## TECHNIQUES FOR ENHANCING EXPLAINABILITY

Understanding the logic and reasoning behind a machine learning model's prediction is crucial for user trust, regulatory compliance, and system debugging. Several techniques have been proposed to provide or improve this interpretability.

### MODEL-SPECIFIC TECHNIQUES

These are techniques that are tailored for specific types of models:

- **Linear Models Coefficients:** For linear regression and logistic regression, the coefficients of the model can be directly interpreted as the importance of each feature.
- **Decision Trees Visualization:** Decision trees are inherently interpretable as they make hierarchical decisions based on feature values. They can be visualized to understand the decision-making process.
- **Attention Mechanisms in Neural Networks:** In deep learning, especially in models like Transformers, attention weights can provide insights into which parts of the input (e.g., words in a sentence) were pivotal in making a decision.

### MODEL-AGNOSTIC TECHNIQUES

Techniques that can be applied regardless of the model's internal workings:

- **LIME (Local Interpretable Model-agnostic Explanations):** LIME approximates a complex model using a simpler, interpretable model (like linear regression) but only in the locality of a specific instance. It then provides explanations based on this simpler model.
- **SHAP (SHapley Additive exPlanations):** Based on game theory, SHAP values give each feature an importance score for a particular prediction.
- **Feature Importance via Permutation:** By shuffling one feature at a time and observing the deterioration in model performance, the importance of each feature can be gauged.
- **Counterfactual Explanations:** These provide insights by describing what minimal changes to the input are needed to change the model's prediction.

### VISUALIZATION TECHNIQUES

- **Saliency Maps:** For neural networks, especially in image tasks, saliency maps highlight regions in the input that were most influential in the model's decision.
- **Activation Maximization:** Visualizes the input that would maximize the activation of a particular neuron, helping in understanding what features that neuron captures.
- **Partial Dependence Plots:** These plots show the relationship between a target response and a set of features, detailing how predictions change as feature values change.

### TECHNIQUES FOR TEXT DATA

Word Embedding Projections: Techniques like t-SNE or PCA can be used to visualize high-dimensional word embeddings in 2D or 3D space, offering insights into semantic relationships.

- **Topic Modeling:** Algorithms like Latent Dirichlet Allocation (LDA) can extract topics from large volumes of text, offering a high-level view of the text's themes.

### POST-HOC TECHNIQUES

- **Model Distillation:** This involves training a simpler, interpretable model (e.g., a decision tree) to mimic a complex model. The simpler model serves as a proxy for understanding the complex model's decisions.

### BENCHMARKS FOR MODEL EXPLAINABILITY

As the need for explainability in machine learning grows, so does the demand for standardized benchmarks to evaluate and compare various explainability techniques. Benchmarks offer a consistent framework for assessment, facilitating the development of more effective and universally applicable explainability methods.

### DATA BENCHMARKS

Several datasets are popularly used to evaluate explainability techniques:

- **Tabular Datasets:** Datasets like UCI's Adult Income or Breast Cancer datasets have been used to evaluate explanations for classical machine learning models.
- **Image Datasets:** Datasets such as ImageNet or CIFAR-10/100 are often used to test the explainability of convolutional neural networks, especially with techniques like saliency maps.
- **Text Datasets:** For NLP models, datasets like IMDb reviews or newsgroups can be used in conjunction with attention mechanisms and other interpretability tools.

### EXPLAINABILITY METRICS

Benchmarks need metrics, and for explainability, some commonly proposed metrics include:

- **Fidelity:** Measures how well the explanation represents the model's behavior.
- **Consistency:** Assesses whether similar instances receive similar explanations.
- **Stability:** Determines if slight perturbations to the input lead to significant changes in explanations.
- **Comprehensibility:** Often evaluated using human studies to determine if the explanations are easily understood.

## COMPARATIVE FRAMEWORKS

Frameworks that facilitate the comparison of multiple explainability techniques on standard tasks:

- **Interpretable Baselines:** Comparing the performance and interpretability of simpler models (e.g., linear regression, decision trees) with more complex ones.
- **Model-Agnostic Explanations:** Comparing techniques like LIME, SHAP, or counterfactual explanations across different model architectures to determine their universality.

## REAL-WORLD SCENARIOS AND CASE STUDIES

Benchmarks can also comprise real-world tasks to evaluate the practical applicability and usefulness of explanations:

**Medical Imaging:** Understanding which regions in a medical image led to a particular diagnosis.

**Credit Decisions:** Explaining why a loan application was approved or denied.

## HUMAN STUDIES

Incorporating feedback from domain experts or end-users:

**User Surveys:** Assessing if the explanations provided by models align with human intuition or domain expertise.

**Interactive Feedback:** Platforms where users can interact with models, modify inputs, and see how explanations change.

## CHALLENGES IN BENCHMARKING EXPLAINABILITY

**Subjectivity:** What's considered a "good" explanation can be subjective and vary among individuals.

**Varying Needs:** An explanation sufficient for a data scientist might not be adequate for a domain expert or layperson.

## CHALLENGES IN QUANTIFYING EXPLAINABILITY

As machine learning models find applications in critical domains, ensuring that these models are interpretable and explainable becomes essential. However, truly quantifying explainability is a challenging endeavor, with various complexities to consider.

## SUBJECTIVITY OF INTERPRETATION

- **Varying Perspectives:** What one expert considers a clear explanation might be seen as convoluted or inadequate by another. There's no universal definition of what makes a "good" explanation.
- **Domain Dependency:** A satisfactory explanation in healthcare might differ significantly from one in finance or automotive safety. Domain expertise heavily influences interpretability expectations.

## EXPLAINABILITY VS. ACCURACY TRADE-OFF

- **Model Simplicity:** While simpler models are inherently more interpretable, they might not capture intricate patterns in data as effectively as complex models.
- **Complex Models:** Deep learning and ensemble methods, though powerful, often act as black boxes, making interpretations challenging.

## INCONSISTENCY ACROSS METHODS

- **Model-Agnostic vs. Model-Specific:** Different explainability techniques might produce varying explanations for the same model and data point.
- **Global vs. Local Explanations:** An explanation that holds for a specific instance might not generalize to the overall model behavior, and vice versa.

## SCALABILITY AND COMPUTATIONAL ISSUES

- **High-Dimensional Data:** With data that has thousands of features, like gene expression data or high-resolution images, generating concise and meaningful explanations becomes challenging.
- **Computational Overheads:** Some explainability techniques, especially model-agnostic ones, can be computationally expensive, making real-time explanations difficult.

## LACK OF GROUND TRUTH

- **Absence of Benchmarks:** Unlike accuracy or loss metrics, there's no definitive ground truth for explainability. This makes comparative evaluations challenging.
- **Human Studies Limitations:** While user studies can provide feedback, they are often subjective, and large-scale studies can be resource-intensive.

## POTENTIAL MISLEADING INTERPRETATIONS

- **Over-reliance on Explanations:** There's a risk that users might place undue trust in models if explanations are provided, even if the underlying model is flawed.
- **Simplistic Explanations:** An overly simplified explanation might not capture the nuances of model decisions, potentially leading to misinterpretations.

## MODEL AND DATA DIVERSITY

- **Evolving Models:** As ML research advances, newer model architectures emerge. Ensuring explainability techniques remain relevant and effective for these is challenging.
- **Diverse Data Types:** Text, images, time series, and structured data all have distinct characteristics, and a one-size-fits-all explainability solution is elusive.

## CONCLUSION & FUTURE DIRECTIONS

### CONCLUSION

Machine learning, with its promise of automating complex tasks and unveiling patterns hidden in vast amounts of data, has seen tremendous growth in both research and applications. However, as we integrate these models deeper into society's fabric, particularly in critical decision-making areas, the "black-box" nature of many sophisticated algorithms has raised valid concerns. The push towards explainable AI is not merely a technical challenge but an ethical imperative, ensuring that the advancements benefit all and cause no inadvertent harm.

Our exploration underscored the multifaceted challenges in quantifying explainability – from the inherent subjectivity of what constitutes a "good" explanation to the intricacies of diverse data and model architectures. However, these challenges also illuminate the path forward, highlighting areas that require focused research, interdisciplinary collaboration, and sustained dialogue with end-users.

### FUTURE DIRECTIONS

**Holistic Explainability Frameworks:** There's a growing need for frameworks that don't just offer piecemeal explanations but provide a comprehensive understanding of models, integrating global and local explanations, textual summaries, and visualizations.

- **Ethics of Explainability:** As the field evolves, ethical guidelines for creating and presenting explanations will become paramount, ensuring that they are not misleading and genuinely promote understanding.
- **Interactive Explainability Platforms:** Future solutions might not be static explanations but interactive platforms where users can query models, tweak inputs, and visualize changes in real-time.
- **Explainability in Emerging Model Architectures:** With continual advancements in machine learning, ensuring that novel model architectures are interpretable will remain an ongoing challenge.
- **Domain-specific Solutions:** Recognizing that explainability needs vary across sectors, there will be a push towards domain-adapted explainability methods, tailor-made for areas like healthcare, finance, or autonomous systems.
- **Community-driven Benchmarks:** The development of widely accepted, standardized benchmarks for evaluating explainability will be pivotal in driving research and ensuring consistent evaluation.
- **Educating Stakeholders:** Beyond just developing explainability techniques, there's a need to educate stakeholders, from developers to end-users, on the importance, nuances, and utilization of these methods.

## REFERENCES

1. Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
2. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.
3. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., & Elhadad, N. (2015, August). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1721-1730).
4. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE international conference on computer vision (pp. 618-626).
5. Holzinger, A., Biemann, C., Pattichis, C. S., & Kell, D. B. (2017). What do we need to build explainable AI systems for the medical domain?. arXiv preprint arXiv:1712.09923.
6. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). IEEE access, 6, 52138-52160.
7. Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018, October). Explaining explanations: An overview of interpretability of machine learning. In 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA) (pp. 80-89). IEEE.
8. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). "Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

## **REQUEST FOR FEEDBACK**

**Dear Readers**

At the very outset, International Journal of Research in Commerce, IT & Management (IJRCM) acknowledges & appreciates your efforts in showing interest in our present issue under your kind perusal.

I would like to request you to supply your critical comments and suggestions about the material published in this issue, as well as on the journal as a whole, on our e-mail [infoijrcm@gmail.com](mailto:infoijrcm@gmail.com) for further improvements in the interest of research.

If you have any queries, please feel free to contact us on our e-mail [infoijrcm@gmail.com](mailto:infoijrcm@gmail.com).

I am sure that your feedback and deliberations would make future issues better – a result of our joint effort.

Looking forward to an appropriate consideration.

With sincere regards

Thanking you profoundly

**Academically yours**

Sd/-

**Co-ordinator**

## **DISCLAIMER**

The information and opinions presented in the Journal reflect the views of the authors and not of the Journal or its Editorial Board or the Publishers/Editors. Publication does not constitute endorsement by the journal. Neither the Journal nor its publishers/Editors/Editorial Board nor anyone else involved in creating, producing or delivering the journal or the materials contained therein, assumes any liability or responsibility for the accuracy, completeness, or usefulness of any information provided in the journal, nor shall they be liable for any direct, indirect, incidental, special, consequential or punitive damages arising out of the use of information/material contained in the journal. The journal, neither its publishers/Editors/ Editorial Board, nor any other party involved in the preparation of material contained in the journal represents or warrants that the information contained herein is in every respect accurate or complete, and they are not responsible for any errors or omissions or for the results obtained from the use of such material. Readers are encouraged to confirm the information contained herein with other sources. The responsibility of the contents and the opinions expressed in this journal are exclusively of the author (s) concerned.

## ABOUT THE JOURNAL

In this age of Commerce, Economics, Computer, I.T. & Management and cut throat competition, a group of intellectuals felt the need to have some platform, where young and budding managers and academicians could express their views and discuss the problems among their peers. This journal was conceived with this noble intention in view. This journal has been introduced to give an opportunity for expressing refined and innovative ideas in this field. It is our humble endeavour to provide a springboard to the upcoming specialists and give a chance to know about the latest in the sphere of research and knowledge. We have taken a small step and we hope that with the active co-operation of like-minded scholars, we shall be able to serve the society with our humble efforts.

## *Our Other Journals*

