

# Genetically Modified Compression Approach for Multimedia Data on cloud storage

Amanjot Kaur Sandhu<sup>[1]</sup>, Er. Anupama Kaur<sup>[2]</sup>  
<sup>[1]</sup>M.tech Scholar, <sup>[2]</sup> Assistant Professor.

Department of Comp. Sc. and Engg, SUS College of Engineering & Technology, Tangori, Mohali, India.

<sup>[1]</sup>[jotsandhu24@gmail.com](mailto:jotsandhu24@gmail.com), <sup>[2]</sup>[saimbianupama@gmail.com](mailto:saimbianupama@gmail.com)

---

**Abstract:** Today, a large number of organizations are moving towards cloud for storing a large amount of digital data. Storage is the most important concern in cloud. There are many storage issues of cloud computing which are related to space and time, validation, access control etc. The high demand for data processing leads to high computational requirement which is usually not available at the user's end. Storage in Cloud computing forms a very important part as the need of virtual space to store our large data has grown over these years. Many compression techniques are available in cloud computing but some are efficient in terms of time and some are space efficient. But the need is to propose a technique that is the combination of both. The algorithm named Genetically Modified Compression Algorithm solves the problem of storage space and time taken to upload data in cloud. The Genetically Modified Compression technique is compared with exiting approaches using definite parameters to propose a improvised cloud storage model. The implementation is Cloud Sim Tool Kit integrated with Eclipse IDE and java language is used

**Keywords:** Storage, Cloud computing, Compression, Genetic Algorithm, Image Compression, Inverse Wavelet Transform.

---

## 1.Introduction

Cloud Computing is one of the most inspiring technology in IT world. It is the Internet based technology where users can share resources among different CSP and cloud vendors CV. The cloud in cloud computing is a set of hardware, software, network, storage that combines together and deliver and aspects of computing as a service. As cloud Computing offers infrastructure to the clients on pay as per use basis so called utility computing. Organizations can provide hardware for cloud internally or a third party can provide it externally.

There are many types of clouds- it might be restricted to a single organization or group (private cloud), available to the public over the Internet (public cloud) or shared by multiple organizations (hybrid cloud). Data Storage is an important aspect and there are various ways to store it on the systems. A special attention is given to modified compression which can be achieved by using hybridization of optimization and compression algorithm.

### 1.1 Motivation and Research Problem

Today, a large number of organizations are moving towards cloud for storing a large amount of data. Storage is the most important concern in cloud. There are many storage issues of cloud computing which are related to space, time, data validation, access control etc. The impact of data storage and the extent of loss that is suffered due to compression technique applied to cloud data motivates to take the problem as a challenge and come up with feasible solutions that can store the data in efficient manner. Many compression techniques are available in cloud computing but some are efficient in terms of time and some are space efficient. But the need is to propose a technique that is the combination of both. The algorithm named Genetically Modified Compression Algorithm solves the problem of security of data in cloud.

### 1.2 Cloud Computing Architecture

Cloud architecture refers to the components and sub components required for cloud computing. It is systems architecture of the software systems concerned in the delivery of cloud computing, usually involves multiple cloud components communicating with each other on application programming interfaces, usually web services and 3-tier architecture. The front end and back end are vital components of cloud computing architecture. The front end is the computer user. The client's network (or computer) and the applications used to access the cloud via a user interface such as a web browser. Cloud computing architecture's back end is the 'cloud' itself, comprising various resources required to provide cloud computing services. It comprises of virtual machines, services, deployment models, servers, etc. The architecture of cloud is shown in Figure 1.

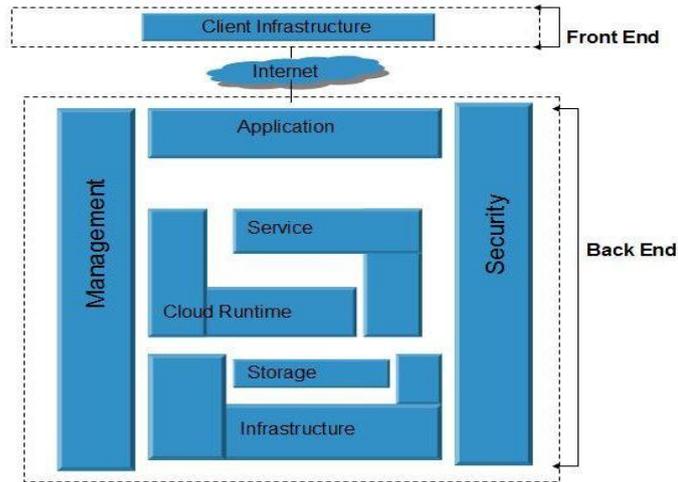


Figure 1: Cloud computing architecture

### 1.2.1 Cloud Deployment models

Cloud services can be deployed in different ways, depending upon the organization structure and provisioning locations. According to NIST definition, cloud deployment models are categorized into four types: Private, Public, Hybrid and Community cloud.

1) **Public Cloud:** In this model, the cloud infrastructure is available to general public. It is usually operated by a large organization, government and a combination of both. Customers access resources and pay for the operating resources. The cloud services and cloud resources are taken from large resource pools that are shared by all end users. If a cloud service provider runs several datacenters, resources can be assigned in such a way that load is uniformly distributed between all centers. The benefits of this type of cloud is its scalability and low capital cost. The best known examples of public cloud systems are Amazon Web Services (AWS) which contains Elastic Compute Cloud (EC2) and the Simple Storage Service (S3) which forms an IaaS cloud offering and the other is Google App Engine which provides PaaS to its customers.

2) **Private Cloud:** It provides a limited access to its resources to consumers that belong to the same organization that owns the cloud. It is managed and operated by the organization, third party or may be by the combination of them. It offers high degree of data security, making it a popular option for organizations uncomfortable with storing information on someone else's infrastructure. Private cloud systems make use of virtualization solutions. This type of cloud may be off premise and on premise.

3) **Hybrid Cloud:** A hybrid cloud is a combination of two or more cloud infrastructures i.e. private, public or community. It includes standardized technologies that enable the cloud portability. The mission critical applications of the organizations are placed on private cloud and remaining applications that are not mission critical stored in public cloud.

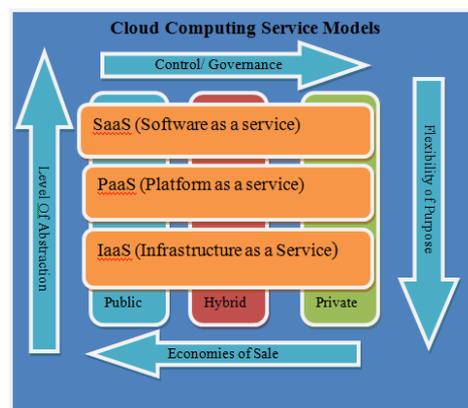


Figure 2: Cloud Service and delivery mode

4) **Community Cloud:** This type of cloud infrastructure is shared by several organizations of a specific community. It is operated and managed by organizations, third party or combination of them. It is costlier than the public cloud because cost is distributed over fewer customers.

### 1.2.2 Compression on cloud

The large sized data as multimedia data has drastically increased on cloud cause inefficient use of Cloud storage service provided by cloud service provider. Compressing of data is obligatory for managing large data on cloud storage. A number of compression techniques are offered for compressing data. Compression techniques compress the data for better data storage usage and transmission thus reducing bandwidth, transmission time. From study of numerous papers, highly efficient technique for compression is mandatory for management of cloud storage. There are two type of compression techniques lossy compression and lossless compression. In lossy compression some loss of information is tolerable such as in videoconferencing or in image where loss is imperceptible to human eye. But for critical financial data lossless compression is vital in which data loss does not occur after compression.

### 2. Literature Survey

During past few years, cloud computing has been growing very fast. It is one of the today's most emerging technologies. It has grown from being a promising business idea to one of the fastest expanding part of the IT industry. It is most inspiring because of it does reduce cost associated with it while increasing flexibility and scalability for processes and resources. Due to increasing demand for cloud, storage is the main concern for IT world. The literature survey presents the various issues of cloud storage the various techniques available for compressing data.

#### 2.1 Issues

1. **Chi Yang et al** [1] discussed various challenges over cloud such as memory bottlenecks and storage space due to raise of big data, multimedia data on cloud. The spatiotemporal compression based on spatial and temporal features of data is proposed and results obtained in terms of data size and fidelity loss.

2. **K. Govinda et al** [2] examined the problem with Simple storage service (S3) in July 2008 suffers from immense loss as it is failed to route user's request to appropriate physical storage and proposed LZW compression technique for optimization of cloud storage and achieves 50 percent of compression.

3. **Khobragade P. B. et al** [3] reviewed a number of compression techniques for compressing image such as Lempel-Ziv-Welch, Huffman coding, Run Length encoding, compression based on discrete cosine transform, discrete wavelet transform (DWT), integer wavelet transform (IWT) based compression and concluded that integer wavelet transform based compression techniques along with lifting scheme gives better compression ratio and retains quality of data.

4. **Mridul Kumar Mathur et al** [4] proposed lossless Huffman compression method for image that is type of statistical coding that attempts to reduce the amount of bits which are requisite to serve as string of symbols. This technique results 65% compression for gray level images.

5. **Mukherjee, Tilak et al** [6] proposed lossless compression approach using wavelets in which integer wavelet transforms are used and wavelet coefficients are converted to integer values and lifting scheme is proposed for obtaining better peak signal to noise ratio and less time execution for compression, hence resulted in good compression results.

#### 2.2 Compression techniques

1) LZ77, LZ78, LZW (Lempel-Ziv-Welch) Compression: These are dictionary based algorithms. LZW technique replaces the repeated occurrences to dictionary that is pre-initialized with all possible symbols. For image data, pixels of image are checked and dictionary is maintained in order to find correspondence.

2) Run Length Encoding: Run Length encoding compresses the image data as series of recurring pixels values are replaced by token. First part of token includes value of pixel and second includes the number of repetitive values. For example 1116669999 can be replaced with (1,3) (6,3) (9,4) and results in lossless compression.

3) Huffman Coding: Based on probability of occurrences of data symbols, table is build based on probabilities and variable length bit code is assigned to each data symbol and probabilities arranged in descending order and merging is done till two are left and high probability symbols have short length codes.[3]

4) Compression based on spatial redundancies: In these type of compression methods the spatial redundancies of image are removed based on spatial correlation. The temporal redundancies are being observed in videos also useful for compression to take place [9]. The spatial domain is converted into frequency domain using transformations such as Discrete Fourier Transformations (DFT) or Discrete Cosine Transformations (DCT). Quantization is done for actual reduction and Zigzag scan is done to obtain the DC component and for better result lossless entropy encoding [10] is done, hence results in compressed data. The compression using DCT is implemented for making comparative analysis between this existing technique and proposed one e.g. JPEG compression using DCT.

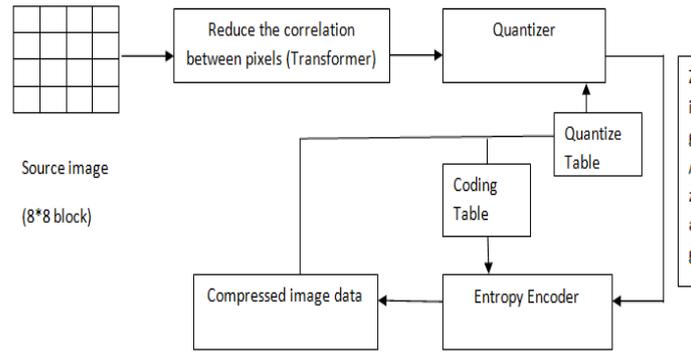


Figure 3: Image compression

4) Lossless Compression based on Integer Wavelet Transforms: Wavelet transforms that are applied to image and de-correlates the image into different size lines and producing sub bands, preserving much of its spatial correlation. Discrete wavelet transform converts the image into the high pass wavelet coefficients and low pass wavelet coefficient, thus results in enhanced execution time but thresholding is done to eliminate wavelet coefficients less than desired TH. Then quantization leads to loss of information .Other useful transformations such that IWT [5] in which wavelet coefficients are rounded off to integer value and Lifting are widely used. Wavelet based compression methods produce good results, so to optimize these results, optimization techniques have great attention.

### 3. Proposed Methodology

The genetically modified compression algorithm is proposed to compress the multimedia data say image. The proposed algorithm comprises of following steps:

1) **Genetic Algorithm:** Genetic Algorithms (GAs) are novel kind of searching algorithms which are in general based on the natural selection process. The basic principle is the survival of fittest in which strong lives and weak are eliminated. This heuristic is routinely used to generate valuable solutions to search problems such as in image compression GAs, which generate solutions to optimization problems using methods inspired by inheritance, mutation, selection, and crossover. In this algorithm the input file is been passed to generate R G B values after which Genetic Operator will be applied to the tree after initializing population then doing crossover and then mutation operation.

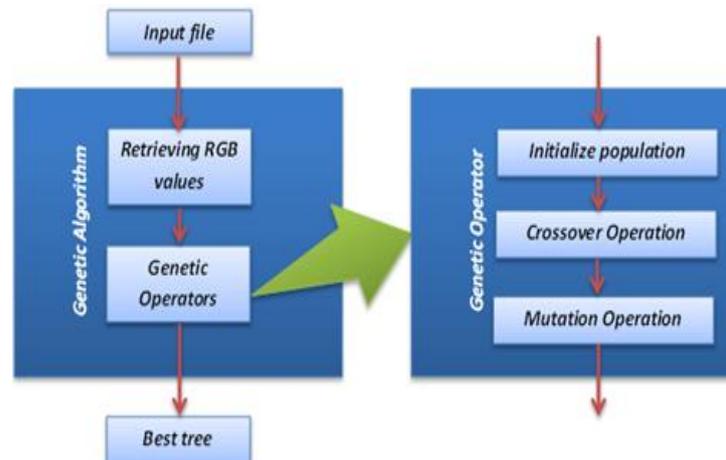


Figure 4: Flow of Genetic Algorithm

GA will be used to find valuable solutions to optimize problems such as compression. Genetic algorithm converts the input data image to  $T_1T_2T_3$  space i.e. best solution by applying Genetic operators.  $T_1$  represents luminance information and  $T_2, T_3$  are chrominance information.

a) Population Initialization: Creating the Individual & the Initial Population first, the image file is wanted to be compressed and find the best tree. It has used files which are read from the proposed system depending on the image; the initial population is created, where an individual is produced and from this individual (chromosome) the other individuals of the population are found randomly.

b) Crossover Operation: It is combining parent chromosomes to produce children chromosomes. It maintains a list of superior chromosomes that are passed to the children. In the proposed system, the cycle crossover (CX) has been chosen as one of the permutation crossover operators that mate the matching with the problem. This type of

crossover gives a variety of individuals; in addition it avoids the conflict in the genes constructing the chromosome (individual), which is the most important property that must be available in the proposed system.

c) Mutation Operation: From computing the probabilities, it is determined whether there is mutation or not. The mutation probability ( $P_m$ ) is 0.0009, if the gene's probability is less than or equal to the  $P_m$  then mutation occurs at that gene. On condition mutation occurs, the gene's location that happened at it the mutation is exchanged with the succeeding gene. If the mutation occurs at the last gene, in this condition this gene's location is exchanged with the first gene. These are basic operations that are carried to find best space  $T_1 T_2 T_3$ .

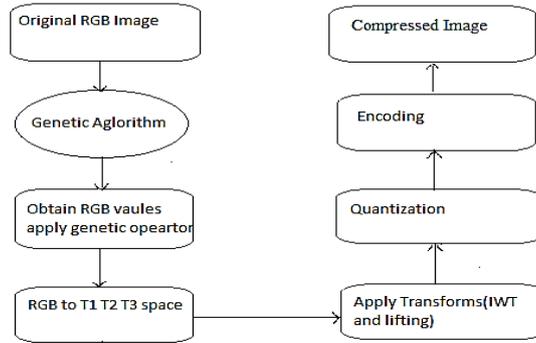


Figure 5: Flow of Genetically modified compression Algorithm

**2) Transform Functions:** After GA further steps of compression are done. The wavelet transforms are applied to image to evaluate the wavelet coefficients. In this implementation Integer Wavelet transforms are applied to the image that wavelet coefficients are rounded off to nearest integers as they need less storage as compared with floating point operations. In this both approximation and detailed content are decomposed. The main advantages of using wavelet coefficients than others because these can represent slow variations and also sharp edges of image efficiently.

The lifting scheme is used in which input signal is splitted into odd and even samples using filter functions and values stored in polyphase matrix such as  $(\text{odd}_{j-1}, \text{even}_{j-1}) = \text{Split}(s_j)$ . The samples are correlated, so it is easy to predict odd samples from even samples that in actual are even values. The difference between the actual odd samples and the prediction result in wavelet coefficients forms prediction error. And after that update operation updates the even values by adding to prediction error.  $d_{j-1} = \text{odd}_{j-1} - P(\text{even}_{j-1})$ ,  $s_{j-1} = \text{even}_{j-1} + U(d_{j-1})$ . The odd elements are being replaced by the difference and the even elements by the averages provides detail and approximation content. Similarly reverse operation is performed by Update, Predict and Merge operations and these transforms are performed in spatial domain.

**3) Quantization:** The quantization is done to obtain only necessary information used to represent.

**4) Encoding:** This process encodes the quantized data using lossless encoder such Huffman encoder [7].

**5) Compressed data:** The size of image is reduced hence compression is done and storage utilization increases.

### 3.1 Result and Performance Evaluation:

The whole implementation of the work proposed was done in JAVA language with eclipse as an Integrated Development Environment (IDE) and cloud sim for simulation of the task submitted over the cloud and based on the socket programming in java. The graphical user interface (GUI) is created, the client and server part which communicate with each other with encrypted messages. The MySQL database is installed on the server side to validate and register the user details. At the time of login it checks the validation of the user in the database. The GUI for proposed work is shown as follows in which after selecting GA based that is proposed approach the results are calculated and then compressed data is uploaded and uploading time is calculated using cloud simulations.

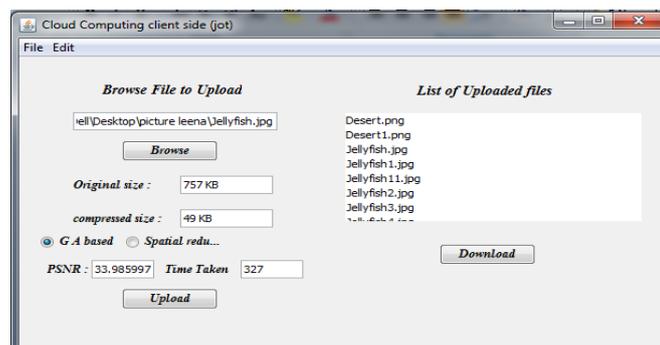


Figure 6: Graphical user interface for proposed work

The parameters for performance evaluation are:

1. **Compressed Size:** Compressed size is size of image or file after compression.
2. **Time Execution:** It is time taken for compressing the image. It is calculated using
3. **Peak Signal to noise Ratio:** It is the measurement of the peak error between the compressed image and original image. The higher the PSNR contains better quality of image. To compute the PSNR first of all MSE (mean square error) is computed. The mean-squared error (MSE) between two images  $I_1 (m,n)$  and  $I_2 (m,n)$  is  

$$MSE = 1/MN \sum [I_1 (m,n) - I_2 (m,n)]^2$$
 where M and N are the number of rows and columns in the input images respectively.  

$$PSNR = 10 \cdot \log_{10} (R^2 / MSE)$$
 and R is maximum fluctuation in input image.
4. **Upload Time:** It is time taken to upload the compressed file to storage and is calculated in terms of data size.

In the proposed work Genetic algorithm is combined with IWT to develop an algorithm of compression. Input image of different type format is compressed using compression based on removal of spatial redundancies such as lossless JPEG compression (existing method). The same tasks are performed for existing method to make comparative analysis. The following Tables show the results of corresponding parameters of compression for proposed method (Figure 5) and existing compression method (Figure 3).

Table1: Results for Genetically Modified Compression (Proposed)

Input Type	Original Size (in Kilo-bytes)	Compressed Size (in Kilo-bytes)	Time Execution (in milli-seconds)	Upload Time (in milli-seconds)	Peak Signal to Noise Ratio
jpg	757	49	327	203.05	33.98
png	462	24	156	101.63	32.73
jpg	47	33	160	135.86	33.47
png	311	84	343	345.42	30.26
jpg	606	63	409	261.16	34.22
png	326	96	297	396.16	31.95

The results for compression based (Figure 3) on removable of spatial redundancies using transformations (DCT) are evaluated and shown in following Table 2.

Table 2: Results for Existing Compression Method

Input Type	Original Size (in Kilo-bytes)	Compressed Size (in Kilo-bytes)	Time Execution (in milli-seconds)	Upload Time (in milli-seconds)	Peak Signal to Noise Ratio
jpg	757	58	593	239.63	18.14
png	462	30	265	123.89	12.70
jpg	47	39	253	162.39	19.36
png	311	100	437	412.72	14.95
jpg	606	79	491	327.09	14.57
png	326	105	306	433.4	25.67

The following graph shows the compression or storage space required by an image of different types before and after compression based on size reduced after compression

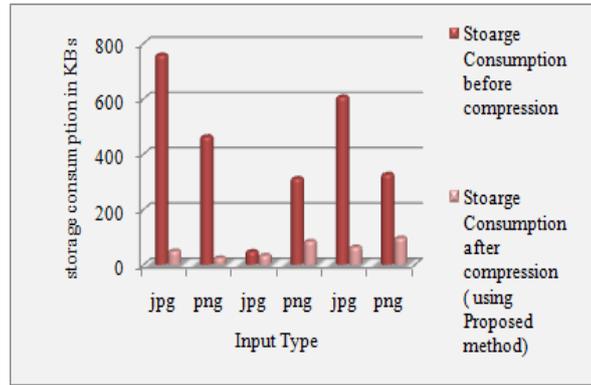


Figure 7: Storage Space Consumption

The following graphs depict the performance analysis of implemented methods and comparison of existing compression based on DCT transformations and Genetically modified compression with IWT , the proposed compression approach based on compressed size ,PSNR values and Time taken for compression. Proposed algorithm maintains the PSNR ration (Figure 9) of the uploaded image after compression and also takes less time for compression as in Figure 10.

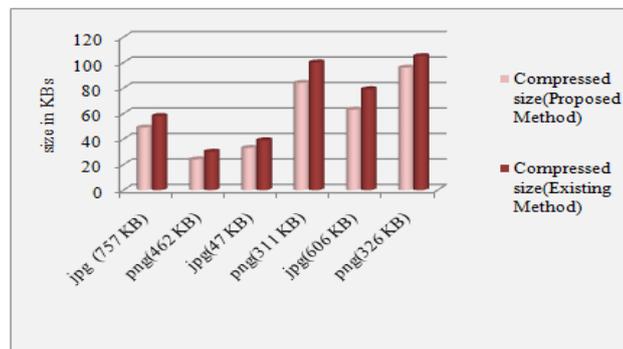


Figure 8: Compressed size after compression

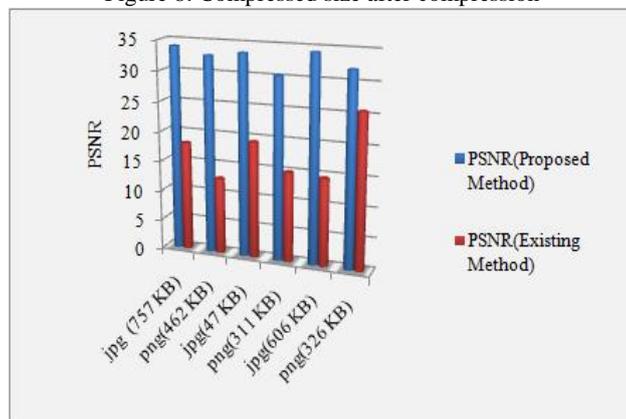


Figure 9: Peak to Signal noise ratio Comparison

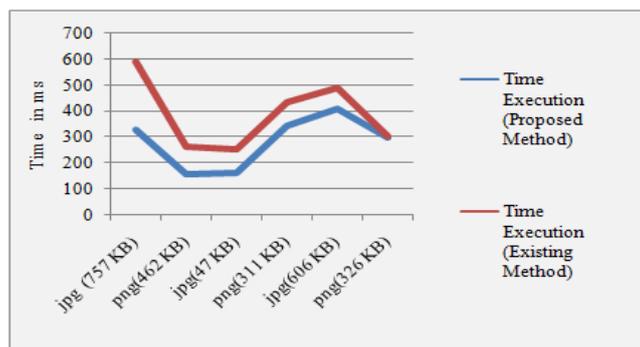


Figure 10: Execution time for compression

The compressed data is uploaded to storage and time taken to upload the data is calculated using cloud simulations. Since PSNR ration is been maintained keeping the compression size on the lower side, the proposed algorithm is showing better results than existing implemented compression method.

#### 4. Conclusion and Future Scope:

The proposed Genetically Modified Compression Algorithm showed the better results by keep both the PSNR and compression size in balance. The execution time of the algorithm is also manageable to run it over the cloud to compress images and hence saves the storage space. The quality of image is preserved after compression .The present research work is done on different type of image data so in future work can be extended to audio and video data. As the time advances there are several optimization algorithm on which research is going on, so a future reference such algorithm can also be applied on multimedia data so that compression can be done in better way over the cloud.

#### REFERENCES

- [1] Yang, Chi, et al. "A spatiotemporal compression based approach for efficient big data processing on Cloud." *Journal of Computer and System Sciences* 80.8 (2014): 1563-1583.
- [2] Govinda, K., and Yuvaraj Kumar. "Storage Optimization in Cloud Environment using Compression Algorithm."
- [3] Khobragade, P. B., and S. S. Thakare. "Image Compression Techniques-A Review." *International Journal of Computer Science and Information Technologies (IJCSIT)* 5.1 (2014): 272-275.
- [4] Mathur, Mridul Kumar, Seema Loonker, and Dheeraj Saxena. "LOSSLESS HUFFMAN CODING TECHNIQUE FOR IMAGE COMPRESSION AND RECONSTRUCTION USING BINARY TREES." *International Journal of Computer Technology & Applications* 3.1 (2012).
- [5] Raman, Vaibhav, and Richa Gupta. "JPEG Multi-Resolution Decomposition of Image Compression using Integer Wavelets." *International Journal of Computer Applications* 95.10 (2014).
- [6] Mukherjee, Tilak, and M. Koteswara Rao. "Efficient Performance of Lifting Scheme Along With Integer Wavelet Transform In Image Compression." *International Journal of Engineering Research and Applications (IJERA)* 3.4 (2013): 1950-1953.
- [7] Shahbahrani, Asadollah, et al. "Evaluation of Huffman and arithmetic algorithms for multimedia compression standards." *arXiv preprint arXiv:1109.0216* (2011).
- [8] Kapoor, Anjali, and Renu Dhir. "Image Compression Using Fast 2-D DCT."
- [9] Shaikh, Muhammad Aakif, and Sagar S. Badnerkar. "Video Compression Algorithm Using Motion Compensation Technique."
- [10] Raid, A. M., et al. "Jpeg Image Compression Using Discrete Cosine Transform-A Survey." *arXiv preprint arXiv:1405.6147* (2014).
- [11] Gupta, Maneesha, and Amit Kumar Garg. "Analysis of image compression algorithm using DCT." *International Journal of Engineering Research and Applications (IJERA) ISSN: 22489622 Vol 2* (2012): 515-521.
- [12] Jain, Chesta, et al. "Performance analysis of integer wavelet transform for image compression." *Electronics Computer Technology (ICECT), 2011 3rd International Conference on*. Vol. 3. IEEE, 2011.
- [13] Mitchell, Melanie. *An introduction to genetic algorithms*. MIT press, 1998.
- [14] Paulinas, Mantas, and Andrius Ušinskas. "A survey of genetic algorithms applications for image enhancement and segmentation." *Information Technology and control* 36.3 (2015).
- [15] Muneshwara, M. S., M. S. Swetha, and G. N. Anil. "A Smarter Way of Securing and Managing Data for Cloud Storage Applications Using High Throughput Compression in the Cloud Environment." (2014).
- [16] Calheiros, Rodrigo N., et al. "CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms." *Software: Practice and Experience* 41.1 (2011): 23-50.