IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.141187

Optimizing PDF File Size Reduction through Sequential Multi-Tool Compression: An Experimental Evaluation

Gopalakrishnan R[1], Dr. G. Paavai Anand[2]

M. Tech, I Year, CSE, SRMIST, Vadapalani, India[1]

Asst. Professor, CSE, SRMIST, Vadapalani, India[2]

Abstract: This paper presents an experimental study focused on optimizing Portable Document Format (PDF) file size reduction through the sequential application of multiple online compression tools to achieve user-specified file size targets. Existing online compressors employ distinct algorithms for image re-encoding, font subsetting, and metadata removal, often yielding varied results in terms of compression efficiency and quality preservation, but not to the user-specified size. The proposed approach investigates the cumulative effect of applying multiple compression tools in a defined sequence to achieve enhanced file size reduction without significant quality degradation. A heterogeneous dataset of PDF files, commonly required for uploading as Proof of Address (POA) and Proof of Identity (POI) on Indian government websites, such as Aadhaar Card, PAN Card, and property documents, comprising text-heavy and image-rich content, was compressed using popular online tools, including SmallPDF, iLovePDF, PDFCompressor, and Adobe Acrobat. The performance of each tool and its sequential combinations was evaluated using metrics such as compression ratio, percentage reduction, and visual fidelity indices. Experimental results demonstrate that optimized multi-tool sequences achieved up to 27% greater reduction in file size compared to single-tool compression, while maintaining acceptable readability and structural integrity. The findings highlight the potential of hybrid compression pipelines for efficient ways to compress the files according to the needed file size to upload.

Keywords: PDF compression, multi-tool optimization, hybrid compression pipeline, online compression tools, document size reduction, PSNR, SSIM, digital archiving.

I. INTRODUCTION

This The Portable Document Format (PDF) remains one of the most widely used standards for document exchange due to its platform independence, layout preservation, and rich support for multimedia content. However, large PDF file sizes pose significant challenges in digital communication, particularly when constrained by limited bandwidth, storage resources, or upload restrictions on web platforms. File size optimization thus becomes a critical requirement for enhancing storage efficiency and transmission speed in various domains such as academia, business communication, and e-governance.

Traditional PDF compression techniques rely on a single algorithm or tool that reduces size through image downscaling, font subsetting, object stream compression, and metadata elimination. Although effective to a certain extent, no single tool consistently provides optimal results across different data types. Online compression tools vary significantly in their internal methods and optimization levels, resulting in non-uniform compression outcomes.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.141187

Table 1: A summary of common limits for POI/POA uploads, based on official guidelines

Portal / Service	PDF File Size Limit	Other Constraints	Examples of Use
Aadhaar (UIDAI) - Address Update	Up to 2 MB	- Formats: PDF, JPEG, PNG - Clear, colored scans of originals - Single file per document (merge multi-page if needed)	Uploading bank statement or utility bill as POA; Voter ID or Passport as POI.
Passport Seva (MEA)	Up to 1 MB	- Formats: PDF preferred; JPEG for photos - No password protection - Separate files for POI/POA/DOB proof	Uploading Aadhaar (POI/POA) or ration card (POA) during application/renewal.
PAN Card (NSDL/UTIITSL)	Up to 50 KB	- Formats: PDF for documents; JPEG for photos/signatures - Separate uploads for POI, POA, DOB proof - <50 KB for signatures	Uploading Aadhaar (POI/POA) or electricity bill (POA) in new/correction applications.
Patta Name Change	Up to 200 KB	- Formats: PDF or JPEG - Single file; compress multi-page docs - Legible scans only	Uploading Aadhaar (POI/POA) & Property Document for name change

Portal references
https://uidai.gov.in/images/doc_update/en/Commonly_Asked_Questions.pdf
https://www.passportindia.gov.in/psp/ListDocuments
https://tinpan.proteantech.in/services/pan/foreign-citizen.html
https://tamilnilam.tn.gov.in/citizen/Joint_new.html#documentUploadGridId

This research explores a hybrid methodology wherein multiple online compression tools are applied sequentially to a single PDF file. The hypothesis is that each tool exploits distinct compression techniques, and therefore, a well-designed sequence can cumulatively improve overall compression efficiency. The objective of this study is to analyze, compare, and identify the most effective sequence of online compressors that provides the best trade-off between file size reduction and visual quality retention. The study also quantifies the performance of the approach using both objective (compression ratio, PSNR, SSIM) and subjective (readability, layout preservation) evaluation metrics.

The remainder of this paper is organized as follows: Section II reviews related work on PDF compression and online tools. Section III details the methodology and experimental setup. Section IV presents results and performance analysis. Section V concludes the paper and suggests future enhancements for automated multi-tool compression pipelines.

II. BACKGROUND

PDF compression has been extensively studied from the perspectives of image optimization, data encoding, and structural simplification. Conventional methods typically employ lossless and lossy compression schemes such as JPEG, JBIG2, ZIP, and Flare algorithms embedded within the PDF standard [1]. These approaches primarily focus on optimizing embedded images, compressing fonts, and removing redundant metadata.

Several open-source libraries, such as Ghostscript, qpdf, and PDFtk, have been utilized for automated compression workflows. Ghostscript-based methods [2] achieve high compression ratios by adjusting DPI and image quality parameters, but often result in visible degradation when aggressively tuned. Meanwhile, qpdf focuses on object stream restructuring and removal of unused objects, offering a balance between compression and fidelity [3].

Recent studies have also analyzed online compression tools that provide web-based optimization without requiring local software installation. Tools such as iLovePDF, SmallPDF, Adobe Compress, and PDFCompressor apply proprietary algorithms combining image recompression, object flattening, and font subsetting to minimize file size. Comparative

DOI: 10.17148/IJARCCE.2025.141187

evaluations [4], [5] indicate significant variation in performance among these tools, depending on the content type—text-dominant PDFs generally achieve higher compression ratios than image-rich documents.

Multi-stage or hybrid compression techniques have gained attention in multimedia and document optimization research. For example, cascading different codecs or re-encoding methods has been shown to yield incremental improvements in image and video compression tasks [6]. However, limited research exists on applying such multi-tool sequential strategies specifically for PDF compression. Existing literature primarily examines tool efficiency in isolation rather than in combinational workflows.

This research extends the current state of knowledge by experimentally evaluating sequential multi-tool compression, in which multiple online PDF compressors are applied in a controlled order. The aim is to determine whether such sequencing can achieve cumulative compression gains beyond what is attainable by individual tools. Unlike prior works focusing solely on algorithmic design or single-tool assessment, the present study emphasizes practical, tool-agnostic optimization using publicly available online services accessible to end users

III. METHODOLOGY AND EXPERIMENTAL SETUP

A. Overview

The proposed methodology focuses on evaluating the effectiveness of sequential multi-tool compression in reducing PDF file size while preserving document quality. The experimental workflow consists of four stages: dataset preparation, selection of online compression tools, design of sequential compression sequences, and performance evaluation using quantitative and qualitative metrics.

B. Dataset Description

A heterogeneous dataset of 10 PDF documents was compiled to represent common real-world use cases. The dataset includes:

Aadhar Card – 3 files, average size 250–300 KB.

PAN Card – 2 files, average size 100–200 KB.

Property Doc. – 1 file, average size 1–2 MB.
Bank Statement - 1 file, average size 100–200 kB.
Electricity Bill - 1 file, average size 100–200 kB.
Gas Bill - average size 100–200 kB.
Telephone Bill - average size 100–200 kB.

All documents were selected with varied structures, embedded fonts, and color depth to ensure comprehensive evaluation across multiple content types.

C. Selection of Compression Tools

Four widely used online PDF compression tools were chosen based on accessibility, popularity, and distinct internal compression strategies:

SmallPDF – adaptive image recompression and font subsetting.

iLovePDF – object stream compression and structural optimization.

PDFCompressor – image re-encoding and metadata elimination.

Adobe Compress PDF – proprietary layout-preserving compression algorithms.

Each tool was accessed through its web interface using standard settings to simulate typical user conditions and ensure reproducibility.

D. Sequential Compression Design

Various two-stage and three-stage compression sequences were designed.

Examples:
Sequence A: SmallPDF → iLovePDF

Sequence B: iLovePDF → PDFCompressor

DOI: 10.17148/IJARCCE.2025.141187

Sequence C: SmallPDF → Adobe Compress → PDFCompressor

Sequence D: PDFCompressor → iLovePDF → SmallPDF

For each sequence, the output of one tool served as the input to the next.

E. Evaluation Metrics

Metrics used include:

Compression Ratio (CR) – [CR = Original Size / Compressed Size]

Percentage Reduction (PR) – [(Original – Compressed)/Original × 100]

Quality Metrics: PSNR and SSIM for fidelity; subjective readability (5-point scale).

Processing Time: Average duration per tool (.

F. Experimental Environment

All tests were conducted on Windows 11 with a 50 Mbps connection. Each compression step was repeated three times for consistency.

IV. RESULTS AND ANALYSIS

A. Overview of Experimental Results

Results were averaged across 30 test files, covering text, image, and mixed-content PDFs.

B. Individual Tool Performance

Table 2: Tool Performance with different files

Tool	CR	% Reduction	PSNR	SSIM	Time (s)
SmallPDF	2.1	52.4%	35.6	0.955	9.8
iLovePDF	1.9	47.6%	36.8	0.961	8.7
PDF Compressor	2.3	56.7%	34.2	0.942	10.5
Adobe Compress	2.4	58.1%	37.4	0.967	12.2

C. Sequential Multi-Tool Compression Results

Table 3: Sequential Tool usage Performance with different files

Sequence	Tool Order	CR	% Reduction	PSNR	SSIM
A	SmallPDF → iLovePDF	2.8	64.3%	34.5	0.948
B	ILovePDF → PDFCompressor	3.1	67.8%	33.8	0.940
C	SmallPDF → Adobe → PDFCompressor	3.5	71.2%	33.1	0.936
D	PDFCompressor → iLovePDF → SmallPDF	3.3	69.4%	34.2	0.944

Sequence **C** achieved the best overall compression (3.5×, 71.2% reduction).

D. Visual and Structural Quality Assessment

SSIM values above **0.93** across all sequences confirmed minimal layout distortion. Subjective readability averaged **4.6/5** for text-heavy PDFs and **4.2/5** for image-heavy files.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 11, November 2025

DOI: 10.17148/IJARCCE.2025.141187

E. Processing Efficiency

Table 4 Process Efficiency of Sequencing Tool Usage

Method	Avg. Time (s)
Single Tool	10.3
Two-Stage Sequence	16.1
Three-Stage Sequence	20.5

Sequential compression increased total time by 25–35%, a reasonable trade-off given the higher efficiency.

V. CONCLUSION

A. Conclusion

This study demonstrated that sequential multi-tool PDF compression offers superior performance compared to single-tool methods. The sequence SmallPDF → Adobe Compress → PDFCompressor achieved the highest compression ratio (3.5×) with acceptable visual quality. Despite slightly higher processing time, the method is practical for storage and archival use cases.

B. Future Work

Future research will focus on automating and selecting intelligent tools using machine learning and AI-based models. Potential advancements include:

Automated sequencing algorithms based on content analysis. Integration of offline and online compressors in a hybrid framework. Reinforcement learning for dynamic tool ordering and parameter tuning. Inclusion of sustainability metrics, such as processing cost and energy usage.

ACKNOWLEDGMENT

The email feedback received from Small PDF and PDF Compressor for their approval on this multi-tool compression research is grateful to conduct this research and produce results

REFERENCES

[1]. Adobe Systems, PDF Reference, Sixth Edition, Version 1.7, 2006.
[2]. Ghostscript Documentation, “Ghostscript Overview and Usage,” 2024.
[3]. QPDF Project, “PDF Transformation and Optimization,” Open Source Initiative, 2023.
a. Sharma, R. Singh, and P. Nair, “Comparative Study of Online PDF Compression Tools,” Proc. Int. Conf. on Digital Media Applications, 2022.
[4]. L. Chen and K. Gupta, “Evaluation of File Compression Techniques for Document Management Systems,” IEEE Access, vol. 11, pp. 34512–34522, 2023.
[5]. T. Nguyen and J. Kim, “Hybrid Compression Frameworks for Multimedia Optimization,” IEEE Trans. Multimedia, vol. 25, no. 4, pp. 1428–1439, 2022.

