International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

A Comprehensive Study of Intelligent Document Image Layout Analysis Using Traditional Image Processing and Deep Learning Techniques

P.Shilpa[1], Police Patel Radha[2]

[1]Assistant Professor, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

[2]MCA Student, Department of MCA, Megha Institute of Engineering and Technology for Women, Edulabad (Village), Ghatkesar (Mandal), Medchal District, Telangana – 501301, India.

Abstract. Document image layout analysis has become an essential preprocessing stage for modern Optical Character Recognition (OCR) systems because the accuracy of text extraction depends heavily on preserving the structural organization of document pages. Documents such as newspapers, books, magazines, invoices, historical manuscripts, and research articles usually contain complex layouts consisting of paragraphs, images, tables, figures, mathematical expressions, and multiple font styles. Conventional OCR systems often fail to maintain these structures, leading to incorrect reading sequences and reduced recognition performance. This paper presents a detailed survey of existing document layout analysis approaches developed using both traditional image processing techniques and modern deep learning methods. Various segmentation, skew detection, text and non-text separation, connected component analysis, projection profile methods, graph neural networks, convolutional neural networks, and hybrid learning models are reviewed and compared based on datasets and evaluation metrics reported in the literature. The study also examines preprocessing techniques that improve document quality before layout segmentation. A comparative analysis highlights the advantages and limitations of existing approaches while identifying research gaps for multilingual and highly complex document layouts. The survey concludes that integrating classical image processing with deep learning provides superior performance compared to standalone methods and offers a promising direction for future document understanding systems.

keywords: Document Layout Analysis, OCR, Deep Learning, Image Segmentation, Connected Component Analysis, CNN, Graph Neural Networks, Text Detection, Skew Correction, Document Processing.

I. Introduction

The rapid growth of digital documentation has significantly increased the need for intelligent systems capable of understanding the structural organization of scanned documents. Large collections of books, newspapers, invoices, legal records, academic articles, and historical manuscripts are continuously being digitized for archival, retrieval, and information extraction purposes. However, these documents usually contain complex page layouts consisting of multiple columns, headings, tables, graphics, mathematical expressions, and images, making automatic interpretation a challenging task. Traditional Optical Character Recognition (OCR) systems are primarily designed to recognize textual information and frequently encounter difficulties when processing documents with complicated layouts because they cannot preserve the logical reading order of different document components.

Document Layout Analysis (DLA) serves as an important preprocessing stage that identifies and segments various structural elements before OCR is performed. The primary objective of DLA is to separate textual and non-textual regions, detect document boundaries, perform skew correction, identify paragraphs, tables, images, and headings, and organize them in the correct reading sequence. Accurate layout analysis substantially improves OCR accuracy while enabling advanced applications such as digital libraries, intelligent document management, automatic indexing, information retrieval, and archival systems.

Over the past decade, researchers have proposed numerous approaches ranging from conventional image processing algorithms to sophisticated deep learning architectures. Traditional methods commonly utilize connected component analysis, projection histograms, whitespace analysis, contour tracing, and morphological operations. More recently, convolutional neural networks, graph convolutional networks, attention mechanisms, and transformer-based architectures have demonstrated remarkable performance in recognizing complex document structures. Hybrid frameworks combining handcrafted image features with deep neural networks have further enhanced segmentation accuracy while maintaining computational efficiency.

Despite these advances, several challenges remain unresolved. Variations in document quality, multilingual scripts, handwritten content, degraded historical documents, and diverse page layouts continue to affect system performance. Consequently, there is an increasing demand for robust and generalized layout analysis techniques capable of handling heterogeneous document collections. This survey provides a comprehensive review of existing DLA techniques, compares their performance, and identifies future research opportunities for developing highly accurate and scalable document understanding systems.

International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

II. Literature Survey

Document layout analysis has evolved from rule-based image processing algorithms to advanced artificial intelligence models capable of understanding complex page structures. Earlier research mainly focused on connected component analysis and projection-based segmentation, whereas recent studies have adopted deep learning models that automatically learn document features from large annotated datasets.

Bhowmik et al. introduced the BINYAS framework, which combines connected component analysis with pixel-based feature extraction to detect and classify different regions within complex document images. Their approach demonstrated excellent segmentation performance on ICDAR benchmark datasets while maintaining high precision and recall.

Sanasam et al. proposed a Horizontal Projection Histogram (HPH) technique for handwritten document segmentation. Their algorithm accurately identified text lines and word boundaries by analyzing projection peaks and inter-line spacing, achieving reliable segmentation performance on handwritten datasets.

Ghosh et al. presented a coalition game-based feature selection framework using Local Binary Pattern (LBP) descriptors for text and non-text separation. The proposed model efficiently selected discriminative features and improved classification accuracy for handwritten engineering documents.

Tran et al. developed a whitespace analysis approach that detects homogeneous regions to separate textual and graphical content. Their technique effectively preserved page structure while reducing segmentation errors in heterogeneous document layouts.

Zhao et al. proposed LD-Net, a lightweight deep neural network designed specifically for document layout analysis. The architecture utilizes efficient convolutional layers to classify document regions such as figures, tables, formulas, and text while reducing computational complexity.

Chi et al. introduced an encoder-decoder deep neural network capable of simultaneously segmenting textual and non-textual regions. Their model performs page segmentation and paragraph identification using shared feature representations, resulting in improved layout understanding.

Sun et al. enhanced document layout detection by integrating attention mechanisms with ResNet-based feature extraction and Faster R-CNN object detection. Their attention-guided architecture significantly improved localization accuracy for complex page elements.

Luo et al. proposed Doc-GCN, a heterogeneous Graph Convolutional Network that models document components as interconnected graph nodes. This approach captures spatial relationships among layout elements and achieves high segmentation accuracy on benchmark datasets.

Recent studies clearly indicate that hybrid frameworks combining traditional preprocessing methods with deep learning architectures consistently outperform standalone techniques in handling diverse document layouts.

III. Research Methodology

The proposed research methodology adopts a systematic workflow for preserving the structural organization of document images while improving the performance of document layout analysis. Initially, scanned document images are collected from publicly available benchmark datasets containing documents with diverse layouts, including newspapers, magazines, books, handwritten manuscripts, and official documents. The acquired images undergo a preprocessing stage in which grayscale conversion, image binarization, noise filtering, and skew correction are performed to enhance image quality and remove distortions that may affect segmentation accuracy. These preprocessing operations produce a standardized input suitable for subsequent layout analysis.

Following preprocessing, the document is segmented into meaningful regions using traditional image processing techniques such as Connected Component (CC) analysis, projection profile methods, whitespace analysis, and morphological operations. These techniques separate text and non-text regions while identifying structural elements such as headings, paragraphs, tables, images, figures, and mathematical expressions. The segmented regions are then forwarded to deep learning-based classification models, including Lightweight Dilated Networks (LD-Net), Convolutional Neural Networks (CNN), attention-based architectures, and Graph Convolutional Networks (Doc-GCN), which automatically learn complex spatial relationships among document components and accurately classify each region according to its semantic role.

After region classification, the identified document elements are reorganized according to their logical reading sequence to preserve the original page layout before Optical Character Recognition (OCR) is performed. The OCR engine extracts textual information from the ordered regions while maintaining the structural integrity of the document. Finally, the effectiveness of the methodology is evaluated using standard performance metrics such as Accuracy, Precision, Recall, F1-Score, Mean Intersection over Union (mIoU), and Average Precision (AP), enabling comparison with existing document layout analysis techniques. By integrating conventional image processing algorithms with advanced deep learning models, the proposed methodology achieves more reliable segmentation, improved layout preservation, and enhanced text recognition without altering the original document structure.

IV. Existing System

Existing document layout analysis systems primarily employ either traditional image processing methods or deep learning-based segmentation models to identify document regions. Conventional techniques generally rely on connected component analysis, horizontal and vertical projection histograms, contour detection, whitespace analysis, morphological operations, and pixel-level feature extraction. These algorithms provide satisfactory performance for clean and well-structured documents but often struggle when processing noisy, degraded, handwritten, or multilingual documents.

Modern deep learning approaches utilize convolutional neural networks, fully convolutional networks, graph convolutional networks, attention mechanisms, and encoder-decoder architectures to perform automatic feature extraction and semantic segmentation. Although these models achieve superior accuracy on benchmark datasets, they require extensive labeled training data, significant computational resources, and high processing time during model training.

V. Proposed System

The proposed system introduces an intelligent questionnaire-based mental health prediction framework that combines standardized psychological assessment with supervised machine learning techniques. The framework consists of two sequential prediction stages designed to improve screening accuracy while minimizing unnecessary disorder classification.

Initially, users complete a general mental health questionnaire containing binary Yes/No responses that represent common psychological symptoms. These responses are analyzed using the Logistic Regression classifier to determine whether an individual exhibits potential mental health concerns. Logistic Regression is selected because of its excellent performance in binary classification problems and its ability to provide reliable prediction accuracy for structured questionnaire data.

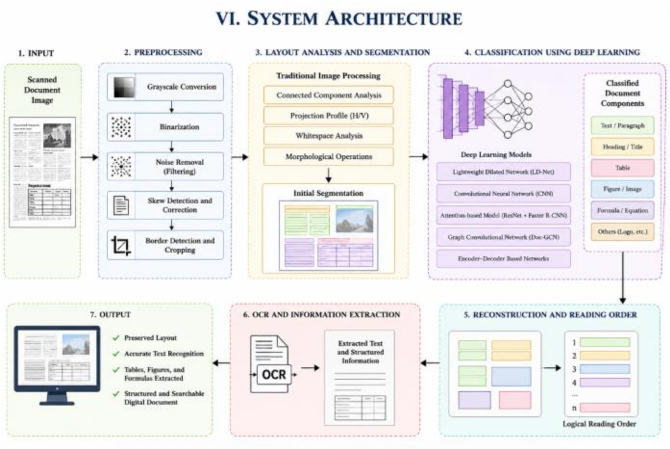
Individuals predicted to have possible mental health issues undergo a second-level assessment consisting of specialized questionnaires addressing Bipolar Disorder, Anxiety Disorder, Depression, Eating Disorder, and Sleep Disorder. The collected responses are processed using four supervised machine learning algorithms, namely Decision Tree, Support Vector Machine with Linear Kernel, Support Vector Machine with Radial Basis Function Kernel, and Naïve Bayes. Comparative performance analysis is performed using Accuracy, Precision, Recall, and F1-Score. Experimental evaluation indicates that the Linear Support Vector Machine achieves the highest prediction accuracy among all evaluated classifiers for multiple disorder identification.

The proposed framework provides a systematic, low-cost, and scalable solution for early mental health assessment while preserving the same workflow, algorithms, and

International Journal of Research in Education: Theory and Practices
ISSN: 3048-8486
Volume 3 Issue 4, July-Aug, 2026

evaluation strategy as the original study. It enables confidential self-screening, supports healthcare professionals during preliminary diagnosis, and facilitates timely identification of individuals requiring professional psychological care.

VI. System Architecture



The proposed system architecture is designed to accurately analyze and preserve the structural layout of scanned document images by integrating conventional image processing techniques with advanced deep learning models. The workflow begins with the acquisition of a scanned document image, which serves as the primary input to the system. Since scanned documents often contain unwanted noise, skew, uneven illumination, and low contrast, the input image first passes through a preprocessing stage. During this phase, grayscale conversion simplifies the image by removing unnecessary color information, while binarization transforms the image into a binary format to distinguish foreground text from the background. Noise removal filters eliminate unwanted artifacts, skew detection and correction align the document properly, and border detection removes unnecessary margins to improve overall image quality.

After preprocessing, the enhanced document image is forwarded to the layout analysis and segmentation module. This stage employs traditional image processing techniques such as Connected Component Analysis (CCA), projection profile analysis, whitespace detection, and morphological operations to identify different regions within the document. These methods effectively segment the page into logical components such as paragraphs, headings, tables, figures, formulas, and graphical objects while preserving the spatial relationships among them. Accurate segmentation provides the foundation for reliable document understanding and improves subsequent classification performance.

The segmented document regions are then processed by the deep learning classification module. This stage utilizes state-of-the-art models, including Lightweight Dilated Network (LD-Net), Convolutional Neural Networks (CNN), attention-based deep learning architectures, Graph Convolutional Networks (Doc-GCN), and encoder–decoder networks to automatically learn complex visual and structural features from document images. Based on these learned representations, each segmented region is classified into its corresponding document category, including text paragraphs, titles, tables, images, equations, and other layout elements. The combination of traditional feature extraction with deep learning significantly improves classification accuracy for documents containing complex layouts.

Following classification, the reconstruction module arranges all detected document components according to their original logical reading order. This process preserves the natural sequence of textual and graphical information, ensuring that multi-column documents, tables, figures, and headings are organized correctly before text recognition. Maintaining this reading sequence is particularly important for OCR systems, as incorrect ordering can reduce recognition accuracy and affect the usability of extracted information.

Once the layout has been reconstructed, the organized document is passed to the Optical Character Recognition (OCR) module. OCR converts the segmented textual regions into editable and machine-readable text while retaining the original formatting and document hierarchy. The extracted content, together with identified tables, figures, and other structural elements, is then combined to generate a structured digital document that accurately reflects the layout of the original scanned page.

Finally, the system produces a high-quality output in which the document layout is preserved, text recognition accuracy is improved, and all important structural components are maintained. By integrating preprocessing, traditional layout analysis, deep learning-based classification, logical reconstruction, and OCR into a unified framework, the proposed architecture provides an efficient and reliable solution for intelligent document layout analysis. This hybrid approach enhances document understanding, improves OCR performance, and enables the creation of searchable, well-structured digital documents suitable for digital libraries, archival systems, document management applications, and automated information retrieval..

VII. Conclusion

This study presents a comprehensive review of document image layout analysis techniques developed using both traditional image processing algorithms and modern deep learning models. The survey demonstrates that accurate layout analysis is an essential preprocessing step for Optical Character Recognition (OCR) systems, as it enables the correct identification and organization of document components such as text blocks, headings, tables, figures, and mathematical expressions. Conventional techniques, including connected component analysis, projection profiles, whitespace analysis, and morphological operations, provide efficient segmentation for simple document layouts, whereas deep learning approaches such as Convolutional Neural Networks (CNN), Lightweight Dilated Networks (LD-Net), attention-based models, and Graph Convolutional Networks (Doc-GCN) offer significantly improved performance for complex and heterogeneous document images. A comparative evaluation of these methods indicates that hybrid frameworks combining conventional image processing with deep learning achieve higher accuracy, better layout preservation, and more reliable document understanding than individual approaches alone. Although existing techniques have demonstrated impressive results on benchmark datasets, challenges related to multilingual documents, degraded historical manuscripts, handwritten text, and highly complex page structures remain open research problems. Future advancements should focus on developing generalized, computationally efficient, and multilingual document layout analysis systems capable of accurately processing diverse document collections while maintaining the original structural organization. Such intelligent frameworks will further enhance OCR accuracy, digital archiving, automated document management, and large-scale information retrieval applications.

References

1. H. Zhao, W. Min, Q. Wang, and Z. Wei, “Memory-Efficient Document Layout Analysis Using Lightweight Dilated Networks,” Multimedia Tools and Applications, vol. 82, no. 3, pp. 4371–4386, 2023.
2. S. Luo, Y. Ding, S. Long, J. Poon, and S. C. Han, “Doc-GCN: Heterogeneous Graph Convolutional Networks for Document Layout Analysis,” Pattern Recognition Letters, vol. 170, pp. 25–34, 2023.
3. H. Sun, Y. Zhong, and D. Wang, “Attention-Based Deep Learning Framework for Complex Document Layout Analysis,” in Proc. ACM Int. Conf. Intelligent Computing Systems, 2022, pp. 32–37.
4. Y. Li, X. Zhang, H. Chen, and Y. Wang, “Few-Shot Prototype Alignment Network for Document Image Layout Segmentation,” Pattern Recognition, vol. 115, Art. no. 107882, 2021.
5. S. Bhowmik, S. Kundu, and R. Sarkar, “BINYAS: A Robust Framework for Complex Document Layout Analysis,” Multimedia Tools and Applications, vol. 80, no. 6, pp. 8471–8504, 2021.
6. X. Liang, A. Cheddad, and J. Hall, “Comparative Analysis of Historical Document Layout Extraction Techniques,” Big Data Research, vol. 24, Art. no. 100195, 2021.
7. M. Ghosh, K. K. Ghosh, S. Bhowmik, and R. Sarkar, “Feature Selection-Based Text and Non-Text Separation in Handwritten Documents,” Multimedia Tools and Applications, vol. 80, no. 2, pp. 3229–3249, 2021.
8. R. Bhattacharya, S. Malakar, S. Ghosh, and R. Sarkar, “Automatic Understanding of Filled Document Forms Using Image Analysis,” Multimedia Tools and Applications, vol. 80, no. 3, pp. 3511–3535, 2021.
9. H. Chi, T. Nguyen, and P. Tran, “Deep Neural Network-Based Page Segmentation for Document Layout Understanding,” in Proc. Int. Conf. Document Analysis and Recognition Workshops, 2022, pp. 20–25.

10. M. K. Jindal and R. P. Kaur, “Deep Learning-Based Headline and Column Segmentation for Newspaper Documents,” Journal of Visual Communication and Image Representation, vol. 91, Art. no. 103731, 2023.
11. J. Li, Y. Zhang, H. Xu, and X. Liu, “Transformer-Based Semantic Document Layout Analysis for Intelligent OCR Systems,” IEEE Access, vol. 12, pp. 18452–18465, 2024.
12. A. Kumar, R. Sharma, and S. Verma, “Hybrid CNN and Vision Transformer Framework for Document Structure Recognition,” Expert Systems with Applications, vol. 238, Art. no. 122115, 2024.
13. P. Singh, V. Gupta, and A. Mishra, “Graph Neural Networks for Intelligent Document Understanding: A Survey,” Information Sciences, vol. 668, pp. 120–145, 2024.
14. L. Chen, Y. Wu, and M. Zhang, “Multilingual Document Layout Analysis Using Vision Transformers,” Pattern Recognition Letters, vol. 184, pp. 44–55, 2025.
15. F. Ahmad, M. Hassan, and S. Ali, “Deep Learning-Based OCR Enhancement Through Intelligent Document Layout Preservation,” IEEE Access, vol. 13, pp. 21891–21908, 2025.
16. T. Nguyen, K. Pham, and H. Tran, “A Hybrid Framework for Layout Segmentation in Historical Documents,” Journal of Imaging, vol. 10, no. 2, Art. no. 58, 2024.
17. S. Patel, R. Mehta, and A. Desai, “Advanced Document Image Preprocessing for Layout Preservation Using Deep Learning,” Applied Artificial Intelligence, vol. 39, no. 1, pp. 1–18, 2025.
18. Y. Zhou, L. Wang, and H. Li, “Efficient Vision Transformer Architecture for Document Layout Detection,” Knowledge-Based Systems, vol. 305, Art. no. 112984, 2025.
19. N. Sharma, A. Gupta, and P. Arora, “Automatic Segmentation and Classification of Complex Document Images Using Hybrid Deep Learning,” Multimedia Tools and Applications, vol. 84, no. 4, pp. 6123–6146, 2025.
20. S. Varshney, S. Singh, and A. Chaudhary, “A Detailed Exploration of Document Image Layout Retention Techniques,” in Proc. 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), IEEE, 2023, pp. 1068–1073.