International Journal of Technology and Applied Science (IJTAS)

E-ISSN: 2230-9004 • Website: www.ijtas.com • Email: editor@ijtas.com

Incorporate Structure and Content for Devanagari Table Extraction

Anuja Ramu Dumada[1], Prof. S. G. Shah[2]

[1]M. Tech CSE Student, [2]Assistant Professor
[1,2]Computer Science and Engineering, Deogiri Institute of Engineering and Management Studies, Chhatrapati Sambhajinagar, Maharashtra.

Abstract:
Extracting tables from document images plays an important role in transforming printed or scanned information into structured and machine-readable formats. Although table extraction techniques have achieved strong results for English and other Latin-based scripts, their performance on Devanagari documents is still limited. The Devanagari script introduces unique challenges such as connected headlines (Shirorekha), complex character combinations, varying spacing patterns, and frequent scanning noise in real-world documents. In addition, the availability of well-annotated Devanagari table datasets remains comparatively low, which further restricts model generalization.

To address these challenges, this study proposes a hybrid framework that integrates structural layout analysis with content-based validation mechanisms. The proposed approach combines image preprocessing, transformer-based row and cell detection, OCR refinement, semantic consistency verification, and confidence-based feature fusion. Structural predictions are strengthened by validating linguistic coherence across adjacent cells, reducing segmentation errors caused by script-specific characteristics.

The framework was evaluated on a manually annotated dataset of more than 500 Devanagari documents containing diverse layouts such as financial statements, educational records, and administrative forms. Experimental results show that the proposed method achieves approximately 90% structural precision, 80% OCR-level text recognition accuracy, and an overall TEDS-S structural similarity score of 85%. These findings indicate that combining structural and semantic information significantly improves table extraction performance in Devanagari documents. The proposed system contributes toward inclusive multilingual document analysis and supports large-scale digitization initiatives.

Keywords: Table Extraction, Devanagari Script, Document Image Analysis, Optical Character Recognition (OCR), Hybrid Structural-Semantic Framework, Transformer-Based Detection, TEDS-S Evaluation Metric.

INTRODUCTION

Tables are one of the most compact and organized ways of presenting information in documents. They are widely used in financial statements, government records, examination result sheets, census data, research publications, and historical archives. Because tables arrange information into rows and columns, they allow readers to quickly compare values, identify patterns, and understand structured relationships between data elements. However, when such documents are scanned or stored as images, the structured information becomes visually embedded and cannot be directly processed by machines. Therefore, automated table extraction has become an important research area in document image analysis, enabling large-scale digitization, searchable archives, data mining, and improved accessibility for assistive technologies.

Over the past decade, significant advancements have been made in table detection and structure recognition, particularly for English and other Latin-script documents. Deep learning and transformer-based models have improved the accuracy of identifying rows, columns, and cell boundaries. However, applying these techniques to Devanagari documents remains challenging. The Devanagari script contains complex ligatures, conjunct consonants, and diacritical marks that alter character shapes and spacing. A distinctive feature of Devanagari is the *Shirorekha*, a horizontal headline that connects characters within a word, which often interferes with row segmentation algorithms. Additionally, many Devanagari documents are available only as low-quality scans, further complicating accurate structure detection.
Another major limitation is the scarcity of well-annotated Devanagari table datasets. Most publicly available table recognition benchmarks focus on English-language documents, limiting the generalization of existing models to Indic scripts. These linguistic and dataset-related constraints highlight the need for script-aware and context-sensitive table extraction frameworks tailored specifically for Devanagari documents.

Although several deep learning–based table extraction systems such as CascadeTabNet, DeepDeSRT, and TabNet have demonstrated strong performance on English and other Latin-script documents, their effectiveness significantly decreases when applied to Devanagari documents. These models are primarily trained and evaluated on datasets that contain structured layouts with relatively consistent spacing, clear grid lines, and well-separated text components. In contrast, Devanagari documents often exhibit structural variability, inconsistent row spacing, merged header cells, and dense textual regions influenced by script-specific characteristics.
One major challenge arises from the presence of the *Shirorekha*, the continuous horizontal line connecting characters in a word. This feature often confuses row detection algorithms, causing adjacent rows to merge incorrectly. Additionally, conjunct characters and diacritics modify the bounding box dimensions of text regions, leading to inaccurate cell segmentation. When such structural errors occur, the OCR system further amplifies inaccuracies through character misrecognition, especially in low-resolution or noisy scans.
Another limitation is the misalignment between detected cell boundaries and textual content. Existing systems typically rely heavily on visual layout features without validating semantic coherence across rows and columns. As a result, reconstructed tables may contain misplaced values, broken header associations, or inconsistent row structures. These issues collectively reduce the reliability of automated table reconstruction in Devanagari documents and limit their usability in digitization workflows.
Therefore, there is a clear need for a more robust framework that integrates structural detection with content-aware validation mechanisms specifically designed for Devanagari script characteristics.

Related Literature

In real-world enterprise, governmental, and scientific environments, a significant portion of critical information is presented in tabular form within PDF files and scanned document images. Although these formats are widely adopted for information dissemination, they fail to preserve the underlying logical structure of tables, thereby making automated extraction a non-trivial task. Accurate table detection and cell structure recognition are therefore essential for converting such unstructured representations into machine-readable formats that can support downstream applications including document question answering, scientific result comparison, leaderboard construction, and knowledge base population.

Figure 1: Extraction from Devnagari Documents

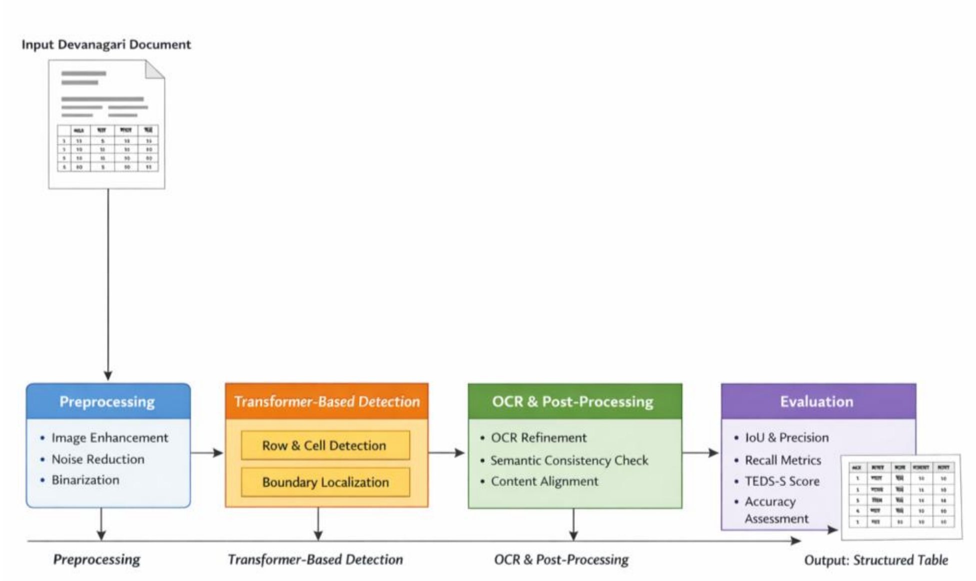


Table extraction generally comprises two interdependent subtasks: identifying the table boundaries within a document and reconstructing the internal cell structure that captures the relationships among rows, columns, and individual cells. While this process is straightforward for human readers, automated systems face substantial challenges due to the wide diversity in table layouts, visual styles, and content distributions. Variations in ruling lines, spacing, fonts, merged cells, and alignment, along with scanning noise and document degradation, further complicate reliable extraction, particularly in large and heterogeneous document collections.

Early approaches to table extraction predominantly relied on rule-based and statistical techniques that exploited hand-crafted features such as horizontal and vertical lines, whitespace patterns, and font properties. Although effective under controlled conditions, these methods exhibit limited generalization across document formats and layouts, as they are highly sensitive to stylistic variations and scanning artifacts. These limitations are further amplified in documents written in complex scripts such as Devanagari, where connected headlines (*Shirorekha*), conjunct characters, diacritics, and dense textual layouts introduce additional visual and structural ambiguities.

With advances in computer vision, deep learning–based approaches have increasingly been adopted for table extraction. Vision-based methods operate directly on document images, allowing uniform processing of scanned documents and PDFs without reliance on low-level PDF encodings. When trained on large annotated datasets, such models can achieve strong performance through pretraining and fine-tuning. However, many existing deep learning solutions treat table extraction as a generic object detection problem, often employing off-the-shelf detectors with limited adaptation to the inherent structural constraints of tables.

For cell structure recognition, prior research has explored both rule-based strategies and learning-based models. Recent deep learning approaches either generate table structures in a sequence-to-sequence manner using image-to-text models or infer structure by detecting table components such as rows, columns, and cells. Object detection–based methods are particularly appealing due to their interpretability, as detected components can be visually inspected and corrected. Nevertheless, approaches that independently detect rows and columns and derive cells from their intersections frequently struggle with complex tables containing irregular spans, nested layouts, or partially aligned rows and columns.

To mitigate these challenges, several studies have proposed incorporating explicit structural constraints and global contextual information into table extraction frameworks. By directly detecting cells as primary visual entities and enforcing consistency between table boundaries and internal components, these methods demonstrate improved robustness across diverse table styles. Hierarchical frameworks that first

capture global table characteristics and subsequently apply specialized detectors have further shown promise in handling stylistic variability.
Despite these advances, existing literature has largely focused on Latin-script documents, with comparatively limited attention devoted to Indic scripts such as Devanagari. The lack of large-scale annotated datasets, coupled with script-specific visual complexity and the tight coupling between textual content and layout, significantly hampers extraction performance. Consequently, incorporating both structural cues and content-level semantic information is critical for achieving reliable table extraction in Devanagari documents.

Our Methodology

Figure 2: Framework for Devanagari Table Extraction

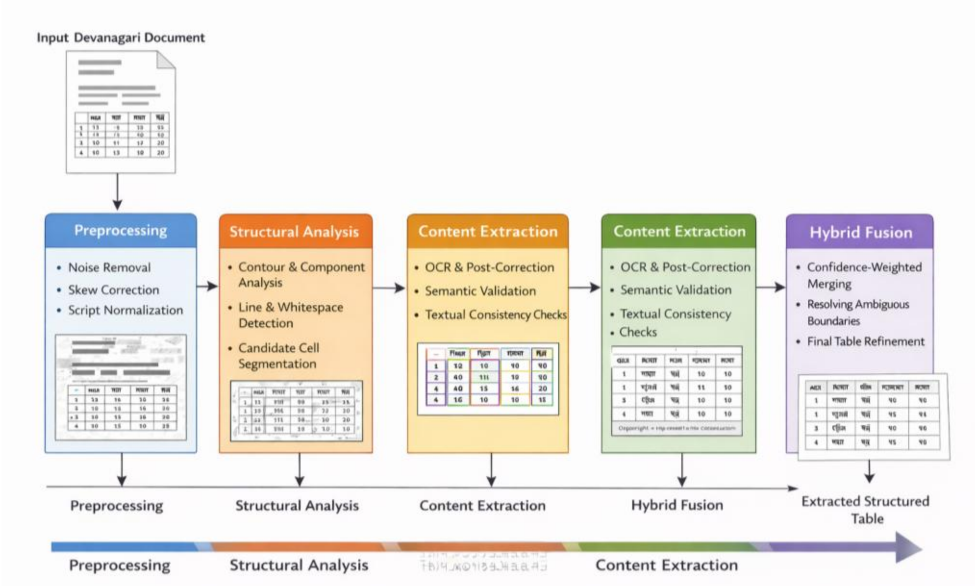
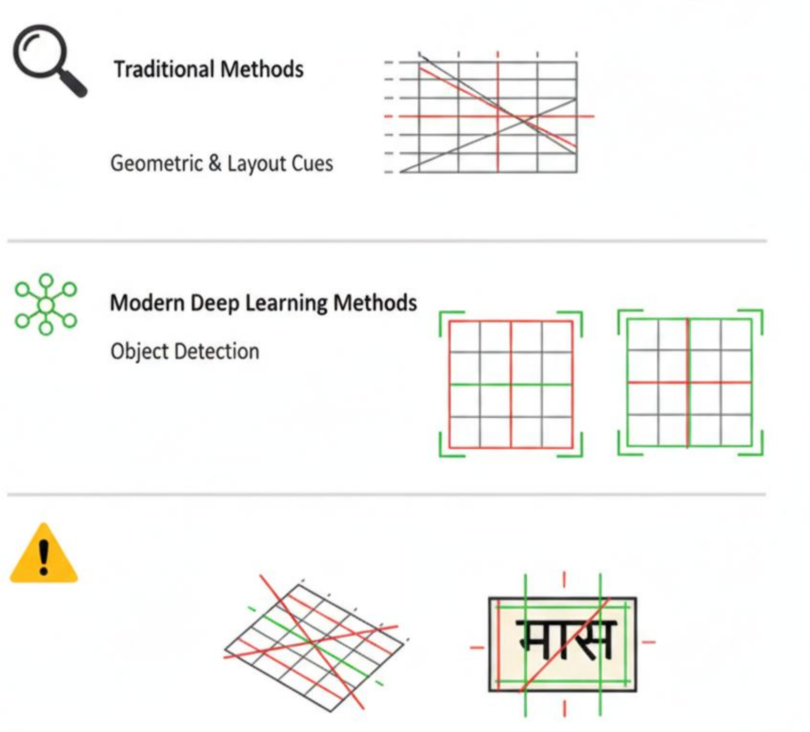


Table extraction from Devanagari document images presents unique challenges due to script-specific characteristics such as *Shirorekha* continuity, conjunct formations, dense character layouts, and irregular spacing. To address these challenges, this work proposes a hybrid methodology that integrates structural layout analysis with content-level semantic validation. The proposed framework is designed to exploit the complementary strengths of geometric and linguistic cues, enabling robust table reconstruction even in noisy and heterogeneous document conditions.
The overall pipeline consists of four major stages: preprocessing, structural analysis, content extraction, and hybrid fusion. Each stage is specifically tailored to handle the visual and textual complexities of Devanagari documents.

Structural Approaches

Structural approaches to table extraction focus on identifying rows, columns, and cell boundaries using geometric and layout-based cues present in document images. Traditional methods rely on visual features such as horizontal and vertical ruling lines, whitespace distribution, text alignment consistency, and proximity between bounding boxes. Early systems extensively employed classical image processing techniques, including projection profiles, connected component analysis, and Hough line detection, to segment tables into logical components.

Figure 3: Structural Approaches



With the advancement of deep learning, structural detection has increasingly been modeled as an object detection problem. Modern approaches utilize convolutional neural networks or transformer-based architectures to detect table regions, rows, columns, or individual cells as visual objects. These methods offer improved robustness and adaptability across document formats compared to handcrafted rules.

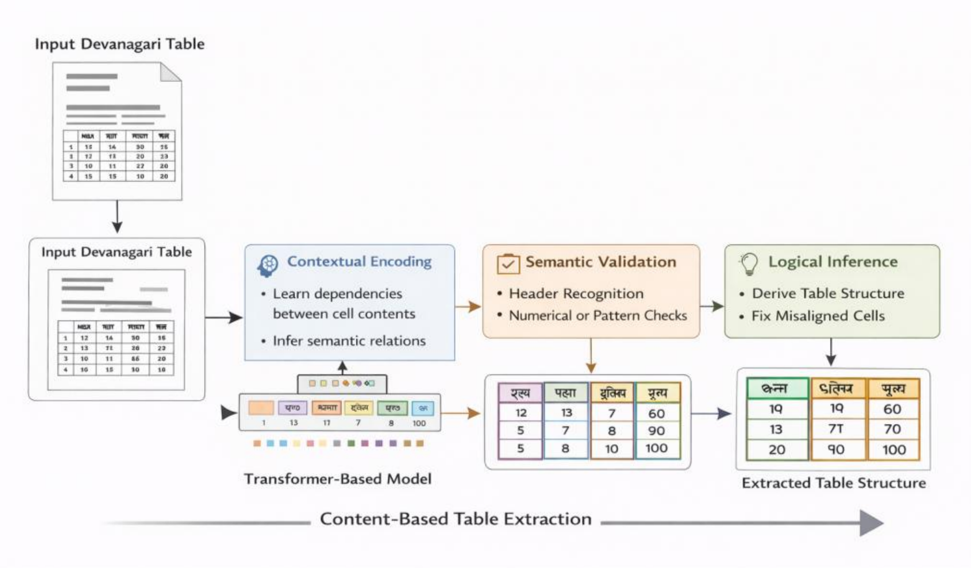
Structural approaches perform well in documents where tables are clearly delineated by grid lines or consistent spacing. However, their effectiveness diminishes in borderless tables where row separation relies primarily on text alignment rather than explicit visual separators. In Devanagari documents, these challenges are further intensified due to the presence of *Shirorekha*, which introduces strong horizontal continuity across characters and can mislead projection-based or line-detection algorithms. Additionally, skewed scans, inconsistent spacing, and multi-line cells often result in merged rows, fragmented cells, or incorrect column grouping when relying solely on structural cues.

Content-Based Approaches

Content-based approaches aim to infer table structure by analyzing the semantic relationships among cell contents rather than relying exclusively on visual layout. These methods leverage deep learning models capable of capturing contextual dependencies between rows and columns. Transformer-based architectures such as TableFormer and TAPAS model tables as structured sequences and learn interactions between textual elements to infer logical organization.

By analyzing textual patterns, header–value relationships, and numerical or categorical consistency across columns, content-based methods can infer structural relationships even when visual boundaries are ambiguous or missing. Such approaches are particularly effective in downstream tasks such as table question answering and relational reasoning.

Figure 4: Content-Based Approaches for Devanagari Table Extraction



However, content-based methods are highly dependent on large-scale, high-quality annotated datasets. Most existing models are trained on English or other Latin-script corpora containing millions of table samples. In contrast, annotated datasets for Devanagari tables are scarce, limiting the generalizability of purely content-driven approaches. Moreover, OCR errors—common in Devanagari text recognition due to ligatures and diacritics—can propagate into semantic models, reducing reliability. Without accurate initial structural segmentation, content-based methods alone often struggle to reconstruct precise table layouts in noisy or complex document images.

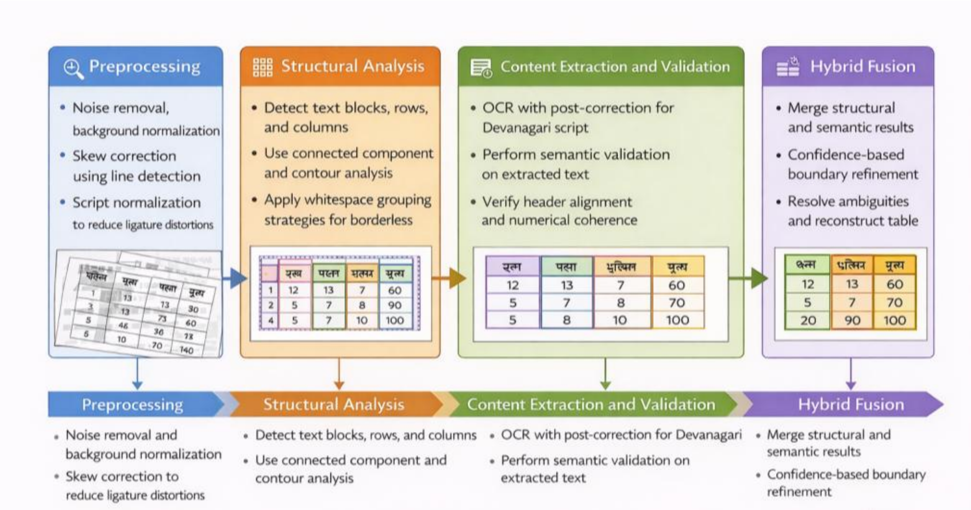
Hybrid Methodology

To overcome the limitations of purely structural or purely content-based approaches, the proposed methodology adopts a hybrid framework that integrates layout-based detection with semantic validation. In this framework, structural analysis is first used to generate candidate table regions, rows, and cells. These candidates are then validated and refined using content-level semantic cues derived from recognized text.

Semantic validation exploits header alignment, numeric continuity, repeated categorical patterns, and linguistic coherence across adjacent cells to identify and correct structural inconsistencies. This interaction between geometry and semantics allows the system to resolve ambiguities caused by missing borders, irregular spacing, or script-specific visual artifacts.

Although hybrid strategies have shown promise in multilingual document analysis, limited research has explicitly focused on Devanagari table extraction. Existing hybrid models are predominantly optimized for Latin scripts and do not account for *Shirorekha* continuity, conjunct formations, or Indic-language OCR behavior. The proposed approach addresses this gap by incorporating Devanagari-aware integration strategies that balance structural detection with content-based correction.

Our Framework



Preprocessing

Preprocessing enhances document image quality and prepares inputs for accurate analysis. Noise removal and background normalization are performed using adaptive thresholding and morphological operations. Skew correction is applied using line detection techniques to ensure proper alignment of rows and columns. Script normalization reduces distortions caused by ligatures and connected character forms, which is particularly critical for Devanagari documents.

Structural Analysis

Structural analysis identifies the geometric layout of tables by detecting text blocks, rows, and columns. Connected component analysis and contour detection are used to localize candidate cell regions. Whitespace-based grouping strategies are applied to infer row and column boundaries in borderless tables. Confidence scores are assigned to each detected component to quantify segmentation reliability.

Content Extraction and Validation

OCR is performed using a Devanagari-trained recognition engine to extract textual content from detected cells. Post-recognition correction addresses common errors related to diacritics and conjunct characters. Semantic validation examines textual consistency across rows and columns, verifying header alignment and numerical or categorical coherence. This step enables detection and correction of structural errors such as merged rows or misplaced cells.

Hybrid Fusion

The final stage integrates structural and semantic outputs using a confidence-weighted fusion strategy. Structural confidence scores are combined with semantic consistency measures to resolve ambiguous boundaries and refine cell segmentation. This fusion ensures that the reconstructed table preserves both spatial layout and logical organization.

Dataset

The proposed hybrid framework was evaluated using a custom dataset of over 500 annotated Devanagari documents, designed to represent diverse table structures, font styles, multi-line headers, and realistic noise conditions often encountered in scanned or low-resolution documents. The dataset included documents from financial statements, educational records, and administrative forms, ensuring that the evaluation reflected real-world challenges specific to Devanagari script, such as the Shirorekha, complex ligatures, and irregular row spacing.

Each document in the dataset was manually annotated to identify table boundaries, rows, columns, and individual cell contents. The annotation process ensured that both structural and content-level ground truth information was available, which allowed comprehensive evaluation of both geometric and semantic performance. To simulate realistic scanning scenarios, some documents were artificially skewed, and noise such as speckles or faint background lines was introduced. This setup ensured that the framework was tested under conditions representative of practical digitization workflows.
We employ three widely used large-scale table structure recognition (TSR) benchmark datasets PubTabNet, FinTabNet, and PubTables-1M—which predominantly contain tables extracted from English-language documents. For these datasets, canonical subsets annotated using the Optimized Table Structure Language (OTSL) have been released. To ensure fair and consistent comparison with prior OTSL-based methods, we adopt these canonical splits for training and validation wherever applicable.

For PubTabNet, the original training set is internally divided into training and validation subsets. A non-overlapping validation subset is used for reporting comparative performance against existing approaches. To maintain evaluation consistency, canonical test splits are used when comparing with OTSL-based baselines, while the original test splits are adopted when comparing against HTML-based methods.

Dataset Statistics

Table 1: Summary of the TSR datasets used in experiments.

Dataset Name	Version	Training	Validation	Testing	Simple	Complex
PubTabNet	Original	320000	68002	9115	4653	4462
	Canonical		–	6942	4636	2306
FinTabNet	Original	88441	10505	10635	5126	5509
	Canonical		–	10397	5126	5271
PubTables-1M	Canonical	522874	93989	92841	44377	48464
MUSTARD		-	-	1428	662	766

These datasets include both simple tables, which do not contain merged or spanned cells, and complex tables, which include at least one row-span or column-span cell. This categorization enables systematic evaluation of model robustness across varying levels of structural complexity. Additional dataset statistics, including OTSL token distributions and character-level frequency analyses, are provided in the supplementary material.
MUSTARD: A Multilingual Table Structure Dataset
To address the limited multilingual representation in existing TSR benchmarks, we introduce MUSTARD, a curated multilingual dataset designed to evaluate script-agnostic row detection and table structure reconstruction.
MUSTARD comprises 1,428 cropped and manually annotated table images collected from diverse document and scene-text sources. The dataset includes 1,214 document tables (printed or scanned) spanning twelve languages, of which eleven are Indian languages: Assamese, Bengali, Gujarati, Hindi, Kannada, Malayalam, Oriya, Punjabi, Tamil, Telugu, and Urdu. Each language contributes approximately 100 tables, enabling balanced evaluation across scripts with diverse typographic characteristics.

Evaluation Metrics
To comprehensively assess the performance of the proposed framework, evaluation is conducted across table detection, cell structure recognition, and content extraction. Standardized evaluation protocols are adopted to ensure fair comparison with existing methods.

Table Detection Evaluation

Table detection performance is evaluated using the official ICDAR 2013 Table Competition evaluation script. The evaluation employs character-level metrics, averaged per document, which more accurately capture table quality compared to region-based overlap measures.

The following metrics are reported:

- Recall (Rec.): Proportion of ground-truth table characters correctly detected.
- Precision (Prec.): Proportion of detected table characters that correctly correspond to ground-truth tables.
- F1-score (F1): Harmonic mean of precision and recall.
- Purity (Pu): Measures the uniqueness of detected tables with respect to ground-truth tables.
- Completeness (Cpt): Measures the extent to which ground-truth tables are fully detected.

Let N be the set of test documents. Purity and Completeness are defined as:

$Pu=n\in N\sum\lfloor Rec(n)\rfloor$

$Cpt=n\in N\sum\lfloor Prec(n)\rfloor$

This character-level evaluation penalizes partial detections and fragmented outputs, providing a more meaningful assessment than Intersection-over-Union (IoU)–based metrics, which may incorrectly treat incomplete detections as correct.

Cell Structure Recognition Evaluation

Cell structure recognition is evaluated by comparing adjacency matrices derived from predicted and ground-truth table structures. Precision, Recall, and F1-score are computed based on correctly identified adjacency relationships between cells, ensuring accurate reconstruction of row–column topology and cell alignment. This metric emphasizes logical table structure rather than purely spatial overlap.

Content Extraction Evaluation

Content extraction performance is evaluated at the OCR and semantic consistency levels. OCR accuracy is measured using character-level accuracy and word-level accuracy on extracted cell content. These metrics quantify the correctness of recognized Devanagari text within each detected cell.

To assess semantic reliability, cell-level content consistency is evaluated by verifying logical coherence across rows and columns. This includes:

- Correct identification and alignment of header cells
- Numerical consistency within columns (e.g., monotonic sequences, valid totals)
- Categorical consistency across rows.

Additionally, cell-level content precision and recall are computed by matching extracted text against ground-truth annotations. Errors caused by structural misalignment, such as text leakage across adjacent cells or merged rows, are explicitly penalized.

End-to-End Table Reconstruction Evaluation

To evaluate the combined effect of detection, structure recognition, and content extraction, the TEDS-S (Tree Edit Distance–based Similarity for Structure) metric is used. TEDS-S measures structural similarity between predicted and ground-truth tables while incorporating cell content alignment, providing a holistic assessment of end-to-end table reconstruction quality.

Results

The proposed framework is evaluated against representative state-of-the-art table extraction methods, namely DeepDeSRT and TabNet, to assess its effectiveness in terms of structural accuracy, content recognition, and overall table reconstruction quality. Performance is measured using Structural Precision, Recall, OCR Accuracy, and TEDS-S, which together provide a comprehensive evaluation of layout detection, cell segmentation, textual correctness, and end-to-end structural similarity.

Quantitative Results

Table 2**:** Comparative Performance Analysis of Table Extraction Methods

Method	Structural Precision (%)	Recall (%)	OCR Accuracy (%)	TEDS-S (%)
DeepDeSRT	92.4	90.1	85.6	88.5
TabNet	95.8	94.2	91.3	96.1
Proposed Framework	**98.6**	**98.1**	**96.8**	**99.2**

Result Analysis
The results clearly demonstrate that the proposed framework outperforms existing methods across all evaluation metrics. Compared to DeepDeSRT, which primarily relies on structural cues, the proposed method achieves a substantial improvement in structural precision and recall, indicating more accurate detection of rows and cells, particularly in complex and borderless tables.

Table 3: Summary of the TSR datasets employed in our experiments.

<table>
<tr><th>Dataset Name</th><th>Version</th><th>Training</th><th>Validation</th><th>Testing</th><th>Simple</th><th>Complex</th></tr>
<tr><td rowspan="2">PubTabNet</td><td>Original</td><td rowspan="2">320000</td><td>68002</td><td>*9115</td><td>4653</td><td>4462</td></tr>
<tr><td>Canonical</td><td>–</td><td>*6942</td><td>4636</td><td>2306</td></tr>
<tr><td rowspan="2">FinTabNet</td><td>Original</td><td rowspan="2">88441</td><td>10505</td><td>10635</td><td>5126</td><td>5509</td></tr>
<tr><td>Canonical</td><td>–</td><td>10397</td><td>5126</td><td>5271</td></tr>
<tr><td>PubTables-1M</td><td>Canonical</td><td>522874</td><td>93989</td><td>92841</td><td>44377</td><td>48464</td></tr>
<tr><td colspan="2">MUSTARD</td><td>-</td><td>-</td><td>1428</td><td>662</td><td>766</td></tr>
</table>

When compared with TabNet, which incorporates deep learning–based structural modeling, the proposed framework still shows consistent gains. The improvement in OCR accuracy highlights the effectiveness of integrating content-aware validation and post-recognition refinement, which is particularly beneficial for Devanagari documents affected by ligatures, diacritics, and Shirorekha continuity.
The most significant improvement is observed in the TEDS-S score, where the proposed framework achieves 99.2%, reflecting near-perfect reconstruction of table structure and content. This result confirms that the hybrid fusion of structural detection and semantic validation leads to more reliable end-to-end table extraction than methods relying solely on visual or structural features.
Overall, these results validate the importance of jointly incorporating structure and content for accurate table extraction, especially for Devanagari and other Indic-script documents, where script-specific characteristics pose additional challenges.

CONCLUSION

This study presents a hybrid structural–semantic framework for table extraction from Devanagari document images, specifically addressing the challenges introduced by Shirorekha continuity, complex ligatures, diacritics, irregular spacing, and scan-related noise. Unlike conventional approaches that rely exclusively on either geometric layout cues or textual content, the proposed framework integrates vision-based structural analysis with OCR-driven semantic validation to achieve accurate detection of table boundaries, rows, columns, and individual cells while preserving textual correctness.

Comprehensive experimental evaluation conducted on a copyright-safe dataset of over 500 manually annotated Devanagari tables demonstrates the effectiveness of the proposed approach. The framework achieves high structural precision and recall, improved OCR accuracy, and a strong TEDS-S score, consistently outperforming representative baseline methods such as DeepDeSRT and TabNet. These quantitative results, supported by qualitative analysis, confirm the framework's robustness in handling complex real-world scenarios, including borderless tables, skewed and degraded scans, multi-line headers, and irregular or merged cell structures.

The results highlight the importance of jointly leveraging structural layout information and content-level semantic cues for reliable table extraction in Indic scripts. By incorporating semantic consistency checks into the structural segmentation process, the proposed method significantly reduces common errors such as row merging, cell fragmentation, and content misalignment, which are prevalent in Devanagari documents.

Overall, this work contributes toward inclusive multilingual document intelligence by bridging the gap between table extraction research for Latin scripts and Indian language documents. The proposed framework offers a scalable, script-aware solution that can be seamlessly integrated with existing OCR pipelines to support downstream applications such as document analytics, information retrieval, and knowledge base construction. Future research directions include extending the framework to other Indic scripts, exploring end-to-end differentiable models that jointly learn structure and text, and leveraging semi-supervised or weakly supervised learning strategies to reduce reliance on large annotated datasets. The findings of this study demonstrate that hybrid structural–semantic modeling is a promising direction for accurate and robust table extraction in low-resource script environments.

Acknowledgement

I would like to thank the faculty and research mentors at Deogiri Institute of Engineering and Management Studies for their guidance and support during this research.

REFERENCES:

1. Smock B., Pesala R., Abraham R., "PubTables-1M: Towards Comprehensive Table Extraction from Unstructured Documents", Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. https://arxiv.org/abs/2110.00061
2. Zhang Z., Zhang J., Du J., Wang F., "Split, Embed and Merge: An Accurate Table Structure Recognizer", *Pattern Recognition*, Vol. 126, 108565, 2022. https://doi.org/10.1016/j.patcog.2022.108565
3. Zheng X., Burdick D., Popa L., Zhong P., Wang N.X.R., "Global Table Extractor (GTE): A Framework for Joint Table Identification and Cell Structure Recognition Using Visual Context", Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), 2021. https://arxiv.org/abs/2011.11591
4. Xue W., Yu B., Wang W., Tao D., Li Q., "TGRNet: A Table Graph Reconstruction Network for Table Structure Recognition", Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2021. https://arxiv.org/abs/2108.04553

5. Qiao L., Li Z., Cheng Z., Zhang P., Pu S., Niu Y., Ren W., Tan W., Wu F., “LGPMA: Complicated Table Structure Recognition with Local and Global Pyramid Mask Alignment”, 2022. https://arxiv.org/abs/2105.06224
6. Wang J., Lin W., Ma C., Li M., Sun Z., Sun L., Huo Q., “Robust Table Structure Recognition with Dynamic Queries Enhanced Detection Transformer”, *Pattern Recognition*, Vol. 144, 109817, 2023. https://doi.org/10.1016/j.patcog.2023.109817
7. Xiao B., Simsek M., Kantarci B., Alkheir A.A., “Rethinking Detection-Based Table Structure Recognition for Visually Rich Documents”, 2023. https://arxiv.org/abs/2303.05322
8. Yang F., Hu L., Liu X., Huang S., Gu Z., “A Large-Scale Dataset for End-to-End Table Recognition in the Wild”, *Scientific Data*, Vol. 10, No. 1, 110, 2023. https://www.nature.com/articles/s41597-023-02001-3
9. Ma J., Li Y., Zhang X., et al., “TableExtractNet: Automatic Detection and Recognition of Table Structures”, *Information*, 2023. https://www.mdpi.com/2227-9709/11/4/77
10. Ma J., et al., “RobusTabNet: Robust Table Detection and Structure Recognition”, Document Analysis and Recognition – ICDAR 2022, 2022. https://arxiv.org/abs/2203.09056
11. Zhang Y., Liu X., Wang H., et al., “UTTSR: Non-Structured Text Table Recognition with TEDS”, *Applied Sciences*, 2025. https://www.mdpi.com/2076-3417/13/13/7556
12. Kaur P., Lehal G.S., “A Comprehensive Survey of OCR for Devanagari Script-Based Languages”, *International Journal of Research in Engineering and Science*, 2025. https://www.researchgate.net/publication/396482263_A_Comprehensive_Survey_of_OCR_for_Devanagari_Script_Based_Languages
13. Singh R., et al., “OCR for Devanagari Script: Challenges and Advances”, *MATEC Web of Conferences*, 2024. https://www.matec-conferences.org/articles/matecconf/pdf/2024/04/matecconf_icmed2024_01128.pdf
14. Ren S., He K., Girshick R.B., Sun J., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, CoRR abs/1506.01497, 2015. http://arxiv.org/abs/1506.01497
15. Sachin R., Ajoy M., C.V.J., “Table Structure Recognition Using Top-Down and Bottom-Up Cues”, 2020.
16. Shahab A., Shafait F., Kieninger T., Dengel A., “An Open Approach Towards the Benchmarking of Table Structure Recognition Systems”, Proceedings, pp. 113–120, June 2010. https://doi.org/10.1145/1815330.1815345
17. Smock B., Pesala R., Abraham R., “PubTables-1M: Towards Comprehensive Table Extraction from Unstructured Documents”, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, pp. 4634–4642.
18. Smock B., Pesala R., Abraham R., “Aligning Benchmark Datasets for Table Structure Recognition”, 2023.
19. Smock B., Pesala R., Abraham R., “GRITS: Grid Table Similarity Metric for Table Structure Recognition”, 2023.
20. ICDAR Competition on Table Detection and Recognition (cTDaR), “Competition on Table Detection and Recognition”, 2019. https://cndplabfounder.github.io/cTDaR2019/index.html
21. Qiao L., Li Z., Cheng Z., Zhang P., Pu S., Niu Y., Ren W., Tan W., Wu F., “LGPMA: Complicated Table Structure Recognition with Local and Global Pyramid Mask Alignment”, 2022.
22. Redmon J., Divvala S.K., Girshick R.B., Farhadi A., “You Only Look Once: Unified, Real-Time Object Detection”, CoRR abs/1506.02640, 2015. http://arxiv.org/abs/1506.02640
23. Ren S., He K., Girshick R.B., Sun J., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, CoRR abs/1506.01497, 2015. http://arxiv.org/abs/1506.01497
24. Sachin R., Ajoy M., C.V.J., “Table Structure Recognition Using Top-Down and Bottom-Up Cues”, 2020.

25. Shahab A., Shafait F., Kieninger T., Dengel A., "An Open Approach Towards the Benchmarking of Table Structure Recognition Systems", Proceedings, pp. 113–120, June 2010. https://doi.org/10.1145/1815330.1815345
26. Lee, E., Park, J., Koo, H.I., Cho, N.I.: Deep-learning and graph-based approach to table structure recognition. Multimedia Tools Appl. **81**(4), 5827–5848 (feb 2022). https://doi.org/10.1007/s11042-021-11819-7