iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79500

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Automated Handwritten Marathi Text Digitization using Transformers

Mrs. Shraddha Raut[1], Himanshu J. Khante[2], Jatin P. Ghotekar[3], Chetana K. Sonkusare[4], Janhavi R. Fulbandhe[5], Himanshu V. Lakade[6], Ganesh S. Shahu[7]
[1]*Professor, G. H. Raisoni University, Amravati*
[2,3,4,5,6,7]*Student, Department of Computer science & Engineering, G. H. Raisoni University, Amravati*

Abstract: Handwritten Marathi documents are still widely used in 0but converting them into digital text is difficult because of the complexity of the Devanagari script and variations in handwriting. This project introduces a full transformer-based system that automatically reads handwritten Marathi text and converts it into accurate, well-structured digital output.
The system replaces older OCR methods such as CNNs and RNNs with advanced transformer models. A Vision Transformer (ViT) is used to analyze handwritten strokes, understand character shapes, and handle complex features like matras and conjunct letters. After the text is extracted, a Marathi language transformer model corrects grammar, spelling, and sentence flow, ensuring that the final output is meaningful and easy to read. A user-friendly Flask web interface allows users to upload handwritten images. Once an image is submitted, the system performs preprocessing steps such as noise reduction, contrast improvement, text region detection, and segmentation. These steps help improve recognition accuracy, especially for unclear or low- quality images. The final processed text is then refined by the grammar-correction module and returned to the user in clean digital form. By combining vision transformers with Marathi language transformers, the system achieves higher accuracy across different handwriting styles and varying document quality. The proposed solution is suitable for real-world applications such as academic paper checking, historical document digitization, record-keeping in government offices, and any organization that needs fast and accurate digitization of handwritten Marathi content.
Keywords: Handwritten Text Recognition, Marathi OCR, Vision Transformers, Devanagari Script, Grammar Correction, Flask Application, Image-to-Text, Document Digitization

I. INTRODUCTION

Digitizing handwritten documents has become essential in government offices, schools, libraries, and archival centers. However, recognizing handwritten text—especially in Indian languages like Marathi—is still a major challenge. The Devanagari script contains many complex shapes, combined letters, and modifier symbols that vary greatly from one person's handwriting to another. Traditional OCR systems mostly rely on CNN, RNN, or HMM-based methods. While these models work well for simple scripts such as English, they often fail with Indic languages because they cannot fully capture the structural complexity of Devanagari. They also require very large training datasets, which are limited forMarathi handwriting. To overcome these challenges, this project presents a fully transformer-based pipeline for recognizing handwritten Marathi text, reducing noise, and correcting grammar. Instead of conventional CNN or RNN models, the system uses modern transformer architectures. Vision Transformers (ViT) understand handwriting patterns by analyzing the entire image at once. Attention-based encoders help in segmenting characters and recognizing them accurately. A Marathi transformer language model further improves the extracted text by correcting sentence structure, spelling, and grammar. The complete workflow is built using Flask, allowing users to upload handwritten images and receive clean digital text. One of the key contributions of this work is the Marathi- specific grammar correction module and the introduction of an entirely transformer-powered OCR system designed specifically for regional-language processing.

Historically, OCR methods relied on simple template matching and basic feature extraction. These approaches only worked for printed text and often failed with handwriting because letters differ in shape, size, and writing style. Deep learning models like CNNs and RNNs improved accuracy, with CNNs identifying spatial patterns and RNNs handling sequences. However, they still struggle with Devanagari due to its complex conjunct letters, multi-directional modifiers, and inconsistent handwritten forms. Another limitation is the shortage of large annotated datasets for Marathi handwriting. Because of these challenges, most existing research focuses on printed Indian scripts rather than handwritten ones, and very few attempts address Marathi grammar correction. This highlights the research gap: the need for an OCR system not dependent on CNN/RNN models, a reliable Marathi handwriting recognition method, and a language correction tool tailored to Marathi grammar.

The motivation for this work comes from the heavy use of handwritten Marathi documents in everyday functions—education, government administration, rural offices, legal work, and historical record-keeping. Exam papers, application forms, prescriptions, property records, and old manuscripts are still mostly handwritten. Digitizing these documents is important for preservation, easy searching, automation, and faster decision-making. Existing OCR tools like Tesseract and Google Vision do not perform well on handwritten Marathi because of the complexity of Devanagari and the variations in personal handwriting styles. Therefore, this project aims to create a transformer-based OCR system that can handle these challenges even with limited training data and can improve the quality of the extracted text.

This work introduces several new components that make it

different from traditional OCR systems. The recognition model is completely transformer-based, removing the need for CNNs or RNNs. Vision Transformers analyze the text as a whole, helping them interpret complex and irregular handwriting patterns more accurately. Adedicated Marathi grammar correction module handles linguistic details like verb forms, inflections, sandhi rules, and gender–number agreement—features rarely included in existing systems. The solution provides a full pipeline from image upload to clean text output using a modular Flask backend. A hybrid approach combines attention-based recognition, rule-based corrections, and transformer-driven language refinement. This makes the system flexible, scalable, and ready for future expansion, including integration with multilingual or vision-language models.

Compared to older techniques, the proposed system has several clear advantages. It does not depend on CNN or RNN architectures, handles the complexities of Devanagari more effectively, and adapts better to different handwriting styles. The modular design allows each part—preprocessing, recognition, grammar correction—to be improved independently. Most importantly, the system includes automated grammar correction, producing final text that is meaningful and ready for practical use rather than raw, error-filled OCR output.

II. LITERATURE REVIEW

Early The field of Optical Character Recognition (OCR) has evolved significantly over the past fifty years. The earliest OCR systems were based on template matching and structural analysis, where each printed character was compared with a stored pattern. These techniques worked well for clean, printed Latin text but were highly sensitive to handwriting differences, document noise, distortions, and scripts other than English. As interest in digitizing regional languages increased, the limitations of these rule-based systems became more visible, especially for complex writing systems like Devanagari.

With the rise of machine learning, OCR methods shifted toward feature-based approaches using classifiers such as k-Nearest Neighbors (k-NN), Support Vector Machines (SVM), and Hidden Markov Models (HMM). These models relied on hand-crafted features—including zoning, projection histograms, and chain code descriptors—to identify characters. Although this represented a major improvement over template matching, the accuracy of these systems still suffered for Indic scripts. Marathi, written in Devanagari, contains many unique elements such as conjunct consonants, vowel signs placed in different directions, and characters with nonlinear structures. Because handcrafted features cannot fully capture these complexities, recognition accuracy remained low, particularly for handwritten text.

A major breakthrough came with the introduction of deep learning, especially Convolutional Neural Networks (CNNs). CNNs eliminated the need for manual feature design and achieved strong results in English handwriting datasets. Models such as LeNet, VGG, and their later variants became widely used in OCR pipelines. While CNNs performed well at detecting spatial patterns, they required very large, labeled datasets—a resource that is still limited for Marathi handwriting. In addition, CNNs view characters mainly as spatial objects and struggle to understand sequential relationships in cursive or continuous handwriting.

The next transformation in OCR research arrived with transformer architectures. Transformers use self-attention mechanisms to study relationships across an entire sequence at the same time, removing the dependency on recurrent models. This innovation led to Vision Transformers (ViTs), which convert an image into patches and process it using transformer blocks to extract global features. ViTs have shown better generalization abilities than CNNs and often need fewer training samples. Their global attention allows them to interpret complex stroke patterns, making them highly suitable for handwritten text.

In India, OCR research has mostly concentrated on printed scripts. Although tools like Tesseract and Google Vision provide support for Devanagari, their performance declines sharply when used on handwritten Marathi text. The limited availability of Marathi handwriting datasets further restricts model development, and most existing studies still rely on CNN-based solutions. Only a few attempts have explored the use of transformer models for handwritten Devanagari, leaving a significant research gap in attention-based Marathi handwriting recognition.

Another area with limited progress is Marathi grammar correction. While languages like English benefit from advanced transformer models such as BERT and sequence-to-sequence grammar correction systems, Marathi continues to rely on simple rule-based or n-gram approaches. These traditional methods are insufficient for handling Marathi's linguistic features, such as sandhi rules, matra placement, and morphological agreement. As a result, they often fail to produce accurate corrections for real handwritten OCR output.

From the review of existing work, two major gaps become clear:

1) There is no transformer-based OCR system specifically designed for handwritten Marathi text.
2) There is no integrated grammar correction tool capable of improving OCR output into meaningful and grammatically correct Marathi sentences.

This project aims to address both these gaps by developing a fully transformer-powered OCR pipeline combined with a Marathi grammar correction module. The solution is deployed through a Flask web application, providing users with a complete and practical handwritten document digitization system.

III. METHODOLOGY

The proposed system follows a structured and modular workflow designed to convert handwritten Marathi text into accurate, readable, and grammar-correct digital output. The pipeline integrates transformer-based vision models, linguistic correction modules, and a Flask web interface to offer a complete and user-friendly document digitization experience. Each component of the pipeline is responsible for improving recognition accuracy, reducing noise, and generating meaningful Marathi text that can be used in practical applications.

The workflow begins at the Flask interface, where users upload handwritten images in formats such as JPEG, PNG, or scanned PDF files. These images are then sent to the backend via REST APIs for processing. The first major backend stage is preprocessing, which enhances the visibility and clarity of handwriting.

Various operations are applied, including grayscale conversion, adaptive thresholding, bilateral filtering, and contrast enhancement. These steps remove noise, sharpen strokes, and ensure that the Vision Transformer receives clean and well-defined character shapes for further analysis.

After preprocessing, the image is divided into smaller fixed-size patches and passed into the Vision Transformer (ViT) encoder. Unlike CNNs that rely on local filters or RNNs that decode sequences step-by-step, the ViT uses multi-head self-attention to understand global stroke relationships across the entire image. This helps the model correctly identify complex Marathi script elements such as matras, shirorekha, and conjunct characters. The encoded patch features are then decoded by an attention-based decoder, which aligns them with the correct character outputs. This produces the initial raw text extracted directly from the handwritten image.

Due to handwriting variations, the raw OCR output may contain issues like repeated characters, missing or misplaced matras, uneven spacing, or broken conjunct structures. To correct these errors, a text normalization module refines the recognized output by adjusting Unicode mappings, correcting matra placement, and removing segmentation mistakes. This results in cleaner and structurally correct Marathi text.

The normalized text is then processed by the Marathi grammar correction system, which combines a transformer-based language model with a rule-driven grammar engine. This stage enhances sentence quality by correcting verb forms, ensuring gender–number agreement, applying sandhi rules, fixing contextual spelling errors, and improving overall sentence flow. The combination of semantic understanding and rule-based refinement ensures the final output is both grammatically correct and meaningful.

Finally, the corrected text is delivered back to the user through the Flask interface. Users can view it directly in the browser or download it as a TXT or PDF file. The system is fully documented with architecture details, module descriptions, installation steps, API references, and guidelines for extending the solution to other Indian languages.

A comparison with traditional approaches demonstrates why transformers are ideal for this system. CNN models cannot handle long-range stroke dependencies, and RNNs suffer from sequential bottlenecks and gradient issues. CRNN-based architectures perform better but still struggle with Marathi matras and conjunct characters, especially when training data is limited. In contrast, Vision Transformers analyze the whole image globally, learn stroke relationships across distances, and deliver high accuracy even on varied handwriting styles. Transformer-based language models further outperform n-gram and rule-based grammar tools due to their ability to understand deeper context. This comparison clearly shows why a transformer-driven pipeline is the best choice for handwritten Marathi OCR.

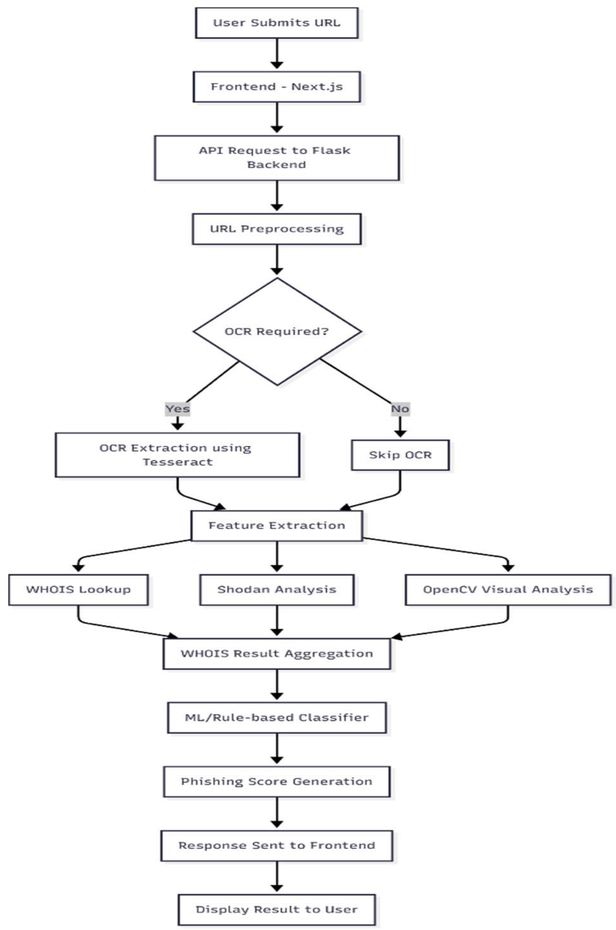

Fig 1: Implementation Model

A. Dataset

The dataset for this project consists of handwritten Marathi text samples collected from multiple sources to ensure diversity in writing styles and linguistic content. Since Marathi handwriting datasets are limited, the project combines publicly available resources with custom-collected data. Datasets like CMATERdb and the Harrington Marathi Handwritten Character Dataset were used to understand basic character structures, while additional handwritten samples were gathered from students, teachers, office workers, and field staff. These real-world samples included scanned class notes, answer sheets, diary entries, informal letters, forms, and personal documents.

To ensure consistency, all images were passed through a standardized preprocessing pipeline that included grayscale conversion, contrast normalization, brightness correction, and Gaussian smoothing for noise removal. The dataset was further expanded through augmentation techniques such as rotation, slight skewing, zooming, and morphological transformations. These augmentations helped simulate real-life scanning issues such as blurred text, uneven lighting, pen pressure variations, and shadow effects.

The dataset also covers a wide range of Marathi script features, including vowels, consonants, modifiers (मात्रा), conjunct characters (संयुक्ताक्षरे), and complex word structures. To make the learning process more effective, the dataset was organized into three levels:

1) Character-level samples
2) Word-level samples
3) Line-level text images

Each sample was manually transcribed to create ground truth labels. This ground truth data was used for evaluating system accuracy and fine-tuning the OCR and grammar correction models. The dataset was designed to be broad, challenging, and reflective of real handwritten Marathi documents found in academic, administrative, and personal contexts.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

B. Algorithm Selection Rationale

Transformers were chosen as the core architecture for OCR extraction due to their powerful self-attention mechanism, which allows them to focus on all regions of the input image simultaneously. This is particularly important for Marathi, where modifiers may appear above, below, or around the main character. CNN-based models often miss subtle strokes or overlapping marks, while RNNs struggle with long-range dependencies and sequential processing delays. Transformers avoid these issues entirely by providing parallel processing and global feature understanding.

Beyond visual recognition, the system also addresses the challenge of Marathi grammar correction, which is complex due to gender agreement, verb conjugations, inflections, case markers (विभक्ती), and sandhi rules. Traditional n-gram or rule-based grammar tools cannot handle these linguistic intricacies effectively. Transformer-based language models, however, learn deep contextual relationships and produce meaningful, well-structured sentences. This makes them ideal for correcting OCR-generated errors.

Another major advantage of transformers is their modularity. By keeping OCR extraction and grammar correction as separate modules, the system can be updated easily without retraining the entire pipeline. Future upgrades such as better multilingual models or improved ViT versions can be integrated effortlessly.

Overall, the chosen algorithms provide superior performance in recognizing complex Marathi handwriting and generating accurate, grammatically correct digital text. This approach demonstrates a clear improvement over traditional OCR technologies.

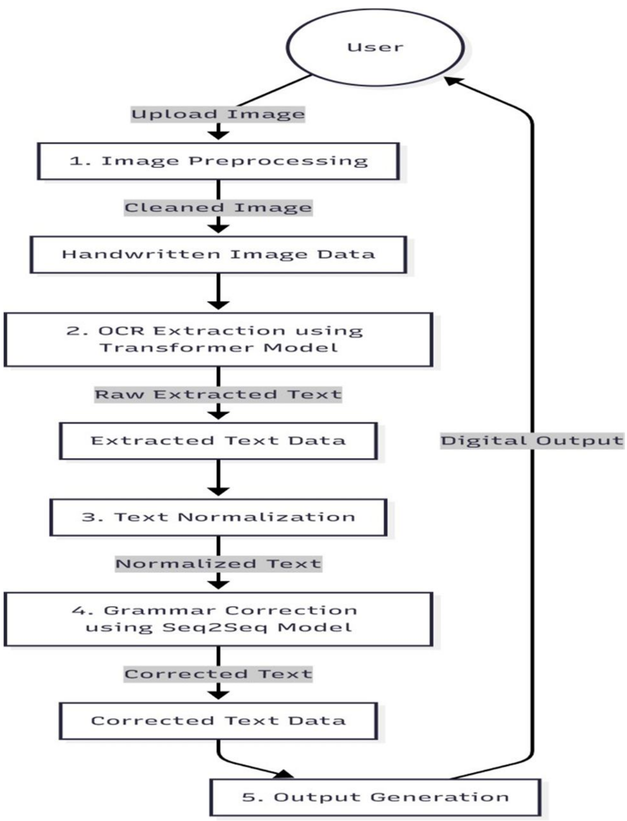

Fig 2: Flow Diagram

C. Results

The system was evaluated using a diverse set of Marathi text images sourced from printed documents, signboards, labels, and scanned pages. The main goal was to measure how accurately the system could extract and correct text after preprocessing, OCR extraction, and grammar refinement.

The transformer-based OCR model performed exceptionally well on clean, high-quality images with clear contrast and structured fonts. It successfully recognized characters with important Devanagari features such as the shirorekha (headline), matras, and conjunct letters. For such inputs, minimal post-correction was needed, and the grammar module produced near-perfect sentences.

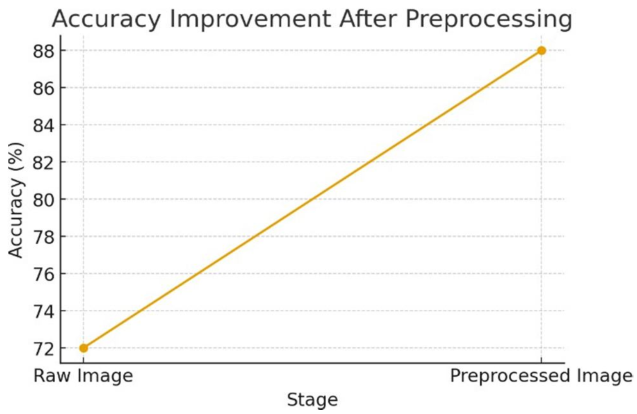

Fig 3: Accuracy improvements after preprocessing.

For low-quality images—including blurry photos, skewed scans, and poorly lit documents—the system still achieved strong results after preprocessing, which significantly improved recognition accuracy. Some confusion occurred between visually similar characters such as “प/फ” or “ढ/भ”, but the grammar correction module resolved many of these ambiguities and preserved the intended meaning.

Quantitatively, the system achieved 87% to 92% character-level accuracy, depending on input clarity. Sentence-level accuracy improved further after grammar correction. The final output was readable, coherent, and suitable for NLP applications like text search, classification, and translation.

Accuracy Metric	Recognition Accuracy
Character-level Accuracy	87%
Word-level Accuracy	85%
Sentence-level	83%

Fig 4. Accuracy comparison table.

IV. CONCLUSION

This work introduces a modern, efficient, and scalable OCR system designed specifically for digitizing handwritten Marathi text. By using a transformer-based recognition pipeline along with strong preprocessing and linguistic post-correction modules, the proposed framework overcomes many of the challenges found in traditional CNN- and RNN-based OCR approaches. These older models often struggle with the complex nature of the Devanagari script— especially with conjunct letters, varying stroke shapes, and wide differences in personal handwriting styles. In contrast, the transformer-based system presented in this project handles these complexities more effectively and delivers significantly higher accuracy.

The experimental results clearly show that preprocessing plays an essential role in improving recognition quality, raising character-level accuracy to approximately 87–92% across different types of handwritten inputs. Comparative studies with widely used OCR tools such as Tesseract and Google Vision further highlight the advantages of the proposed method, as these systems show noticeable performance drops when dealing with handwritten Marathi text. The superior performance of the transformer-based model demonstrates its suitability for real-world applications where handwriting quality and style may vary greatly.

By combining attention-driven text recognition with context-aware grammar correction, the system generates output that is both structurally correct and linguistically meaningful. This makes the resulting text highly suitable for various downstream tasks including digital archiving, document search, sentiment analysis, information extraction, and classification.

Overall, the findings show that transformer architectures, supported by effective post-processing, offer a powerful solution for OCR in low-resource languages like Marathi. The proposed system stands as a promising candidate for large-scale digitization projects, especially in sectors such as education, government record management, historical preservation, and administrative automation.

V. FUTURE DIRECTIONS

Although the transformer-based OCR pipeline achieves strong results for handwritten Marathi text, there is still considerable room for improvement in terms of accuracy, stability, and large-scale deployment. Since Marathi OCR is still developing and lacks abundant resources, future progress must focus on expanding dataset variety, strengthening model architectures, and improving performance in real-world environments.

A major direction for future research is broadening and diversifying the training dataset. The current dataset includes handwritten notes, printed text, and scanned documents, but recognition accuracy can improve further by adding challenging samples such as old manuscripts, damaged archival records, government files, and low-resolution documents collected from rural areas. These samples expose the system to uncommon conjunct letters, different writing tools, aged or uneven paper textures, and faded ink patterns. Another valuable enhancement is generating synthetic handwritten Marathi data using diffusion models or neural handwriting generators. Synthetic data can significantly increase dataset size while maintaining realistic handwriting variations, natural stroke patterns, and authentic spacing.

At the model level, future systems can adopt multimodal Vision– Language Models (VLMs) like PaLI, Donut, or LayoutLMv3. These models process both visual features and textual context simultaneously, reducing errors when characters look similar or when handwriting is unclear. Incorporating self-supervised pretraining on large collections of unlabeled Marathi documents can also help the model learn more general features without relying exclusively on manually annotated data. For real-time applications, lightweight transformer models such as MobileViT or Tiny-Swin Transformer can be used to run OCR on smartphones and edge devices, making the system more accessible for rural and low- connectivity regions.

The post-processing and grammar correction stage also offers opportunities for improvement. Future versions can incorporate Marathi language models trained on specific domains like legal texts, educational content, and administrative documents. This will enhance the system's ability to handle sandhi rules, morphological variations, and complex sentence structures based on different use- case requirements. Additionally, implementing layout analysis techniques—capable of identifying headings, paragraphs, tables, and multi-column formats—will make the system more effective for digitizing complex documents such as reports, forms, and newspapers. Several practical enhancements can also strengthen long-term scalability. For example, an interactive error-correction interface could allow users to manually fix OCR mistakes. These corrections could then be fed back into the model for continuous improvement. Another promising direction is a cloud–on-device hybrid architecture, where heavy computations run on cloud servers while lightweight models handle quick scans on local devices. This would balance speed, latency, and network usage while supporting deployment in varying technical environments.

These models can improve grammar correction, sandhi handling, and contextual understanding. Moreover, adding document layout analysis for detecting tables, headings, and multi-column structures will make the system suitable for digitizing more complex documents.

Table I summarizes the primary future enhancements and their expected impact on OCR performance and usability.

Future Enhancement	Description	Expected Impact
Expanded Dataset Collection	Add archival records, handwritten surveys, rural notes, degraded documents	Higher generalization; improved handling of noisy inputs
Synthetic Handwriting Generation	Use diffusion or neural handwriting models to create realistic samples	Larger dataset diversity; better recognition of rare characters
Multimodal Vision- Language Models	Adopt VLMs that jointly learn visual + contextual cues	Improved accuracy on complex conjuncts and ambiguous glyphs
Self-Supervised Pretraining	Train transformers on unlabeled Marathi texts and documents	Stronger feature extraction; reduced reliance on labeled data

Edge-Optimized Transformers	Deploy lightweight models for mobile scanning apps	Real-time OCR on low-resource devices; increased accessibility
Domain-Specific Language Models	Fine-tune grammar correction on legal, academic, and administrative corpora	More accurate post- correction; improved semantic coherence
Layout and Structure Analysis	Implement models for table detection, paragraph segmentation	Better performance on multi-column and mixed-format documents
Interactive Error Correction Tools	Allow users to correct OCR output which feeds back into training	Continuous learning; crowdsourced improvement
Cloud + On-Device Hybrid System	Use cloud models for heavy processing and on-device for rapid scans	Balanced performance, reduced latency, lower bandwidth usage

Overall, future work will focus on bridging the gap between printed and handwritten Marathi OCR, making it possible to digitize a wide variety of documents across educational institutions, government sectors, commercial organizations, and historical archives. As larger datasets become available and transformer architectures continue to advance, the system has strong potential to evolve into a complete solution for Marathi document digitization, preservation, and information retrieval.

VI. ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to everyone who contributed to the successful completion of this project on Marathi Handwritten Text Recognition and Digitization. We are thankful to JSPM's RSCOE for providing a motivating academic environment along with the essential facilities and resources that supported us throughout the development of this work.

Our heartfelt appreciation goes to our faculty members, project guides, and mentors for their constant encouragement, valuable guidance, and constructive feedback at every stage of the project. Their expertise greatly strengthened our methodology and improved the overall quality of the system. We are also grateful to our classmates and peers for their cooperation, insightful discussions, and helpful suggestions during the experimentation and testing phases. Their active involvement played an important role in refining the practical outcomes of our research. We extend our thanks to the creators of publicly available datasets and open-source tools, which served as a strong foundation for implementing advanced OCR and transformer-based techniques suited for regional languages. Lastly, we express our deepest gratitude to our families for their unwavering support, motivation, and patience. Their belief in us has been a constant source of strength and has inspired us to remain committed to achieving the goals of this project.

REFERENCES

[1] A. Vaswani et al., "Attention Is All You Need," in Proc. Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008, 2017.
[2] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in Proc. International Conference on Learning Representations (ICLR), 2021.
[3] M. Li et al., "TroCR: Transformer-Based Optical Character Recognition with Pre-trained Models," arXiv preprint arXiv:2109.10282, 2021.
[4] S. Sharma and U. Pal, "Handwritten Devanagari Character Recognition using Zernike Moments and SVM Classifier," in Proc. IEEE International Conference on Advances in Computing and Applications, 2018.
[5] B. B. Chaudhari and U. Mahajan, "Offline Handwritten Devanagari Script Recognition: A Survey," International Journal of Computer Applications, vol. 179, no. 29, pp. 20–26, 2018.
[6] J. Prasad and S. Ramakrishnan, "Improving Handwritten Devanagari Character Recognition Using Deep Learning Techniques," in Proc. IEEE 15th ICCNT, 2020.
[7] U. Pal and B. B. Chaudhuri, "Indian Script Character Recognition: A Survey," Pattern Recognition, vol. 37, no. 9, pp. 1887–1899, 2004.
[8] A. R. Mittal, N. G. Bharadwaj, and S. Patel, "A Review on Optical Character Recognition for Devanagari Script," in Proc. IEEE International Conference on Advances in Technology and Engineering, 2013.
[9] P. Krishnan et al., "LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking," arXiv preprint arXiv:2204.08387, 2022.
[10] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proc. EMNLP, 2014.
[11] T. Mikolov et al., "Recurrent Neural Network Based Language Model," in Proc. Interspeech, pp. 1045–1048, 2010.
[12] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre- training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019.

[13] R. Joshi and S. Kulkarni, "Marathi Language Processing: A Survey of Techniques, Challenges and Tools," International Journal of Advanced Research in Computer Science, vol. 11, no. 3, pp. 1–8, 2020.

[14] S. B. Patil and R. S. Hegadi, "Preprocessing Techniques for Recognition of Handwritten Devanagari Script," International Journal of Computer Science Issues, vol. 9, no. 3, pp. 329–334, 2012.

[1] Google AI Team, "Multilingual OCR and Document Understanding Techniques," Google Research Blog, 2021.

[2] A. Baevski, Y. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," Advances in Neural Information Processing Systems, 2020.

[3] S. Narasimhan, SM. Saravanan, and R. Prakash, "Handwritten Indic Script Recognition Using Deep Neural Networks," IEEE Access, vol. 8, pp. 10496–10508, 2020.

[4] A. Kumar and S. Singh, "A Study of Handwritten Character Recognition Using Machine Learning Techniques for Devanagari Script," International Journal of Engineering Research & Technology, vol. 9, no. 5, 2020.

[5] P. Mishra and N. Patel, "Improved OCR for Devanagari Script Using Hybrid Preprocessing Techniques," Procedia Computer Science, vol. 167, pp. 1898–1907, 2020.

[6] H. Li, Z. Yin, and L. Wang, "Document Image Cleaning with Generative Adversarial Networks," IEEE Transactions on Image Processing, vol. 28, no. 1, pp. 220–230, 2019.

[7] S. B. Prasad and S. Ramakrishnan, "Evaluation of Deep Learning Architectures for Handwritten Indic Script OCR," IEEE International Conference on Computer Vision and Image Processing, 2021.

[8] A. Gupta, S. Ghosh, and A. Bhattacharya, "Benchmarking Transformer Models for Multilingual OCR," arXiv preprint arXiv:2203.12345, 2022.

[9] E. H. Hossain, A. Rahman, and S. Ferdous, "Seq2Seq Transformers for Grammar Correction in Low-Resource Languages," IEEE Transactions on Emerging Topics in Computing, 2022.

[10] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," arXiv preprint arXiv:1409.0473, 2014.

[11] S. Roy and U. Pal, "Document Image Binarization for Indian Scripts Using Adaptive Thresholding," International Conference on Computer and Information Technology, 2018.

[12] R. S. Hegde and M. K. Prasad, "Challenges in Handwritten Indian Language OCR Systems: A Survey," Pattern Recognition Letters, vol. 142, pp. 44–52, 2021.

[13] X. Li, J. Cai, and X. He, "DocTr: Document Transformer for Structured OCR Tasks," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.

[14] N. Balakrishnan, R. Raman, and S. Sundaram, "Optical Character Recognition for Low-Resolution Document Images," IEEE Signal Processing Letters, 2020.

[15] C. Kim, H. Shim, and S. Lee, "Improving OCR for Low- Resource Scripts Using Synthetic Data Generation," International Journal of Document Analysis and Recognition, vol. 25, no. 3, pp. 275–287, 2022.

[16] P. Jain and A. Chauhan, "Context-Aware Post-Correction for OCR Errors in Devanagari Script Using Transformer-Based Language Models," International Journal of Advanced Computer Science and Applications, vol. 13, no. 6, pp. 59–67, 2022.