Seeing Justice Clearly: Handwritten Legal Document Translation with OCR and Vision-Language Models

Shubham Kumar Nigam[1,3*†], **Parjanya Aditya Shukla**[1*], **Noel Shallum**[2], **Arnab Bhattacharya**[1]

[1]Department of Computer Science, IIT Kanpur, Kanpur, India   [2]Symbiosis Law School, Pune, India
[3]University of Birmingham, Dubai, UAE

shubhamkumarnigam@gmail.com,   padityashukla26@gmail.com,   noelshallum@gmail.com,   arnabb@cse.iitk.ac.in

Abstract

Handwritten text recognition (HTR) and machine translation continue to pose significant challenges, particularly for low-resource languages like Marathi, which lack large digitized corpora and exhibit high variability in handwriting styles. The conventional approach to address this involves a two-stage pipeline: an OCR system extracts text from handwritten images, which is then translated into the target language using a machine translation model. In this work, we explore and compare the performance of traditional OCR-MT pipelines with Vision Large Language Models that aim to unify these stages and directly translate handwritten text images in a single, end-to-end step. Our motivation is grounded in the urgent need for scalable, accurate translation systems to digitize legal records such as FIRs, charge sheets, and witness statements in India's district and high courts. We evaluate both approaches on a curated dataset of handwritten Marathi legal documents, with the goal of enabling efficient legal document processing, even in low-resource environments. Our findings offer actionable insights toward building robust, edge-deployable solutions that enhance access to legal information for non-native speakers and legal professionals alike.

Code — https://github.com/anviksha-lab-iitk/SJC

Introduction

The Indian judiciary, one of the world's most complex legal systems, continues to face challenges in ensuring timely justice and efficient case handling. A major bottleneck lies in its persistent reliance on handwritten documentation at the grassroots level, such as district courts and police stations, where First Information Reports (FIRs), case diaries, witness statements, and court proceedings are still manually recorded. These documents are critical to criminal and civil proceedings, but their handwritten, unstructured nature makes them difficult to archive, search, and analyze. Variability in handwriting, language diversity, legal terminology, and the poor quality of scans all pose significant hurdles to digitization.

---

[*]These authors contributed equally.
[†]Corresponding Author

This work lays the foundation for building an automated legal document digitization pipeline aimed at enabling structured and accessible digital case records. Our focus is on automatically converting handwritten legal documents in Marathi into English through two distinct paradigms: (i) a modular OCR + Machine Translation (OCR-MT) pipeline, and (ii) direct translation using Vision Large Language Models (vLLMs). The objective is to assess their performance, robustness, and suitability for deployment in low-resource legal environments.

Traditional Optical Character Recognition (OCR) systems, such as Tesseract (Smith 2007), EasyOCR (JaidedAI 2020), and PaddleOCR (Li et al. 2022), follow a multi-stage pipeline involving text detection, segmentation, and recognition. These systems perform reasonably well for printed documents but often fail on handwritten legal content due to limited generalization, poor handwriting support, and a lack of layout-awareness. Moreover, they are not designed to work with low-resource languages like Marathi unless extensively retrained. Our work evaluates these tools on a custom dataset of handwritten Marathi legal documents and investigates their effectiveness when paired with modern MT models like IndicTrans2 (Gala et al. 2023) and Sarvam-1 (Sarvam AI 2024).

However, OCR-MT pipelines suffer from cascading errors; misrecognized words by OCR adversely affect translation quality. This motivates the use of vLLMs, which can jointly process image and text inputs, reducing dependency on rigid pipeline stages. Models such as Chitrarth (Khan et al. 2025), Ovis2 (Lu et al. 2024), and Maya (Alam et al. 2024) are capable of zero-shot visual reasoning and multilingual output generation. These models offer an attractive alternative for legal text digitization, particularly when working with noisy or incomplete inputs.

The key contributions of this work are:

- *OCR Evaluation:* We benchmark Tesseract, EasyOCR, and PaddleOCR on a curated dataset of handwritten Marathi legal documents.

- *MT Analysis:* Analyze translation performance using IndicTrans2 and Sarvam-1 models on OCR-extracted text.

- *vLLM Benchmarking:* We compare three vision-language models against OCR-MT pipelines, highlighting their ability to perform direct image-to-English translation.
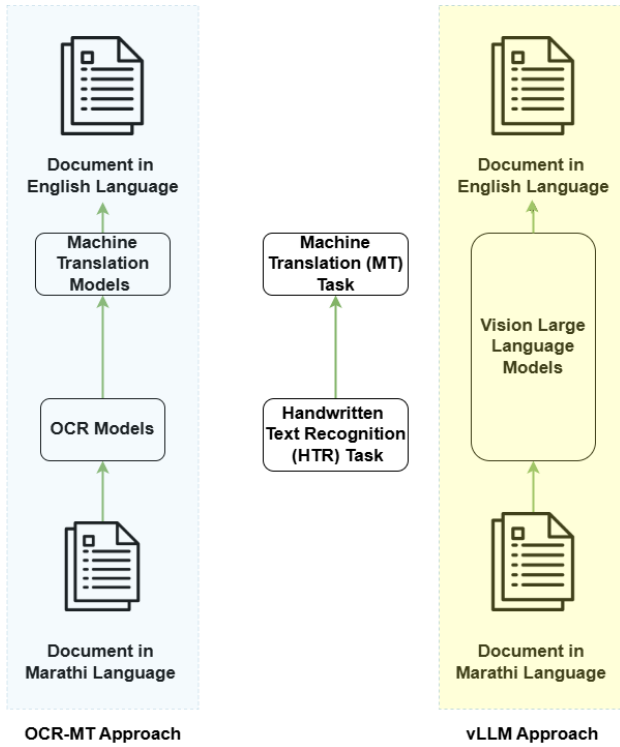
Figure 1: Comparison of OCR-MT and vLLM-based approaches for handwritten text translation. The OCR-MT pipeline decomposes the task into separate HTR and MT stages, whereas vLLMs unify the process into a single end-to-end step.

MT for Indian languages. Sarvam-1 (Sarvam AI 2024) is a 2B parameter model optimized for 10 Indian languages and English, demonstrating strong performance through careful data curation. IndicTrans2 (Gala et al. 2023), based on a transformer encoder-decoder architecture, supports all 22 scheduled languages. Other efforts include Anuvaad (Project Anuvaad 2025) and Nemotron-4-Mini-Hindi-4B (Joshi et al. 2025), the latter trained via continued pre-training for bilingual MT.

Recent vLLMs such as LLaMA 4 (Meta AI 2024), Qwen2.5-VL (Bai et al. 2025), GPT-4V (Yang et al. 2023), Gemini 2.0 Flash, and PaliGemma 2 (Steiner et al. 2024) demonstrate strong multimodal reasoning capabilities but often require high-end resources, making them unsuitable for deployment in low-resource legal infrastructures. Lightweight models like Chitrarth (Khan et al. 2025), Ovis (Lu et al. 2024), and Maya (Alam et al. 2024) balance efficiency and accuracy and are better suited for real-world deployment in district courts. Our work evaluates such models under zero-shot prompting for handwritten document translation. PLATTER (Kasuba et al. 2025) offers an end-to-end handwritten OCR framework with two-stage processing (handwritten text detection and recognition) and supports 10 Indian languages. TransDocAnalyser (Chakraborty, Harit, and Ghosh 2023) is tailored to legal F.I.R. documents, combining a FastRCNN+Vision Transformer encoder with a BERT-based decoder fine-tuned for legal vocabulary. Our work differs by focusing on translation from handwritten legal documents and comparing OCR-MT and vLLM approaches under a common benchmark.

Related Work

The literature relevant to our task spans four key areas: (1) advancements in Optical Character Recognition (OCR) systems, especially those capable of processing diverse document layouts; (2) developments in machine translation (MT) for Indian languages; (3) systems for handwritten and printed text extraction in Indian legal documents; and (4) progress in vLLMs that perform tasks such as image captioning, visual QA, and multimodal translation.

Recent OCR systems incorporate layout and spatial awareness for enhanced recognition. VISTA-OCR (Hamdi et al. 2025) introduces a generative, layout-aware OCR pipeline using an encoder-decoder framework. olmOCR (Poznanski et al. 2025) leverages document-anchoring and fine-tunes the Qwen2-VL-7B-Instruct vLLM to extract structured information. Another approach, +FIRST (Iwana et al. 2017), improves multimodal transcription by combining OCR outputs from full documents with image features of only the first page. Systems such as Nirnayak (Datta et al. 2024) and related pipelines (S et al. 2024) apply OCR for downstream tasks like translation and summarization in the Indian legal domain. However, their reliance on OCR introduces error propagation, limiting end-to-end accuracy. This motivates the use of vision-language models that can jointly reason over visual and textual modalities. Several models have advanced

Dataset

We utilize a custom dataset curated to reflect real-world legal document scenarios. The dataset comprises approximately 60 scanned PDF documents written in Marathi, collected from authentic legal sources. These documents vary in length and structure, including both single-page and multi-page entries. Each page in the dataset contains handwritten Marathi text, often interspersed with printed text. The documents also include diverse visual elements such as official stamps, seals, signatures, and structured tables, which introduce additional challenges for both text recognition and translation.

To establish a reliable ground truth, the Marathi text was manually translated into English by a team of two native Marathi speakers. These translations were subsequently reviewed by a legal language expert to ensure fidelity, contextual accuracy, and terminological consistency. The resulting high-quality annotations serve as reference outputs for evaluating the performance of both the OCR-MT pipeline and the vLLM approach discussed in this study.

An example page from the dataset is provided in the supplementary materials to illustrate the visual and textual characteristics of the documents. All private and sensitive information has been blurred to preserve confidentiality.

Proposed Methodology

We explore two strategies for translating handwritten Marathi legal documents into English: (1) modular OCR-based translation pipelines, and (2) direct end-to-end translation using Vision Large Language Models (vLLMs).

OCR-MT Pipelines

We construct six distinct OCR-MT pipelines by combining three OCR tools—Tesseract, EasyOCR, and PaddleOCR—with two state-of-the-art Indian language translation models: Sarvam-1 (Sarvam AI 2024) and IndicTrans2 (Gala et al. 2023). Each scanned Marathi document image is first processed by one of the OCR tools to extract textual content. This output is then passed to one of the MT models. This modular architecture allows a comparative analysis of how different OCR-MT pairings influence the final translation quality.

Vision-Language Models

We further evaluate three vision-language models—Chitrarth (Khan et al. 2025), Maya-8B (Alam et al. 2024), and Ovis2-34B (Lu et al. 2024)—for direct image-to-English translation without intermediate OCR steps. These models were tested in a zero-shot setting, guided by manually designed prompts tailored for legal document understanding.

Evaluation Protocols

We evaluate the OCR outputs using standard fidelity metrics: Character Error Rate (CER) (K et al. 2025) and Word Error Rate (WER) (Ali and Renals 2018). These quantify the textual fidelity of Marathi text extracted from the scanned documents. For translation quality, we conduct human evaluations across the following criteria

- **Fluency:** Grammatical correctness and naturalness of English output.
- **Adequacy:** Degree to which the translation preserves the original meaning.
- **Correctness:** Alignment with gold-standard human translations.

All outputs were evaluated by human annotators fluent in both Marathi and English, allowing us to identify the most effective and robust pipeline configurations.

Results and Analysis

OCR models exhibited notably better performance on printed text segments than handwritten ones. While printed segments were recognized with reasonable accuracy (see Figure 4), handwritten content often led to errors, such as omissions, misrecognized characters, or fragmented outputs. Figures 2 and 3 highlight typical OCR behaviors on handwritten digits and dates, while Figure 5 shows successful stamp extraction. Although EasyOCR consistently outperformed PaddleOCR and Tesseract, it still struggled with inconsistent handwriting styles.

As shown in Table 1, OCR errors significantly affected translation quality (see Appendix). The pipeline sometimes

Image	Extracted	English Output	Ground Truth
निर्यम	Niyam	Rule	Rule
गाव :	Gaav (Gaon)	Gaon	Village
गाव :	Gaav (Gaon)	Village	Village
फेरफराचा दिनांक	Ferfaracha Dinank	Date of Share Transfer	Date of Change

Table 1: OCR-MT pipeline: Extracted Marathi text (Latin transliteration) and corresponding English translations compared with human-annotated ground truth.

generated incorrect or misleading translations (e.g., mapping the Marathi word "Gaav" (transliterated from "gaon", meaning "village") to "Gaon" instead of "Village"), or dropped key information. This error propagation across pipeline stages is a core limitation of the OCR-MT approach.

Translation models also produced incoherent or mixed-language outputs when fed noisy OCR text. Some translations included trailing untranslated Marathi fragments. Human evaluation revealed loss of legal or factual content, and incomplete or awkward phrasing, particularly for longer sentences with complex legal semantics.

We next evaluated vLLMs, Chitrarth, Maya-8B, and Ovis2 variants, for direct image-to-text translation. Prompt engineering was critical: results improved significantly with detailed instructions. Chitrarth often failed to produce coherent outputs, while Maya-8B showed partial correctness in translation with rich prompts. Ovis2-34B (int4 quantized) and Ovis2-16B also performed inconsistently in zero-shot mode, suggesting a need for domain-specific fine-tuning or prompt-tuning.

Although vLLMs could interpret visual elements and page layout, they often failed to extract precise handwritten content. Generated translations sometimes focused on high-level descriptions of documents rather than their verbatim content. Human annotators rated vLLM outputs based on clarity and fidelity, concluding that while promising, these models currently lack the precision required for legal-grade document translation.

Qualitative Comparison of vLLM Outputs

Table 2 (Appendix) presents a qualitative comparison of the translated outputs generated by four vLLMs against the gold-standard human-annotated translation. The human-annotated translation captures the full context with high fidelity, including masked details of dates, registration numbers, survey numbers, and transaction clauses. In contrast, all four models failed to extract the actual semantics of handwritten content.

- **Chitrarth** produced a hallucinated summary about a meeting, with invented names, dates, and locations that were not present in the original document. This indicates its inability to ground visual input in real text.
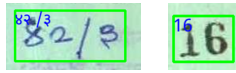- **Maya-8B** interpreted the input as a study guide, again

Figure 2: Accurate recognition of printed and handwritten Marathi numeric characters.
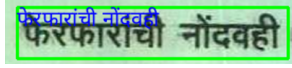

Figure 3: Incorrect extraction of handwritten dates.
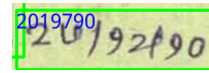

Figure 4: Correct extraction by OCR models for Marathi.

Figure 5: Correct extraction of Marathi stamp details.

showing a lack of alignment with the legal nature of the text. Its output was generic and irrelevant.

- **Ovis2-34B** demonstrated slightly better recognition. It provided partial translations of text such as Marathi ledger headers and dates. However, the content was still largely fabricated or misunderstood.

- **Ovis2-16B** performed relatively better by identifying some legal and financial cues (e.g., account numbers, names, locations). It also translated a few phrases and recognized structural layout, but its output lacked completeness and was partially incoherent.

This qualitative example reveals a consistent limitation across current vLLMs: while they are adept at interpreting general visual or structural elements, they fall short of extracting and translating complex handwritten legal content. Moreover, they tend to hallucinate plausible-sounding text when unable to recognize tokens, posing serious risks in high-stakes domains such as law. These findings motivate further efforts in fine-tuning, prompt design, and alignment mechanisms for vLLMs in legal NLP tasks.

Conclusion and Future Scope

This work compared traditional OCR–MT pipelines with end-to-end vision-language models (vLLMs) for translating handwritten Marathi legal documents into English. While OCR-based systems offer modularity and transparency, they suffer from significant error propagation, especially when handling noisy handwritten text. vLLMs provide a unified alternative that bypasses intermediate recognition steps, but current models still struggle with accurately interpreting complex handwriting and domain-specific legal terminology. Despite these limitations, our results highlight the promising potential of vLLMs for future legal digitization efforts. Their ability to reason over multimodal inputs opens a pathway toward more robust and scalable translation systems that do not rely solely on brittle OCR stages.

Future Scope. This study represents an early step in a broader research effort, and several important directions remain:

- **Fine-Tuning on Legal Handwritten Data:** Both OCR and vLLMs require domain-specific adaptation. We plan to fine-tune models on larger annotated datasets of handwritten legal documents.

- **Improved Prompting and Alignment:** Since vLLM performance varies heavily with prompt quality,

systematic prompt engineering and instruction alignment will be explored to reduce hallucinations and improve fidelity.

- **Hybrid OCR–vLLM Pipelines:** Combining OCR for structural cues with vLLMs for contextual translation may yield better accuracy than either approach alone.

- **Enhanced Evaluation Protocols:** We will develop evaluation metrics tailored to legal translation, focusing on factual consistency, terminology preservation, and completeness.

- **Model Efficiency and Deployment:** Real-world adoption requires lightweight models that function on district-level compute resources. Techniques such as quantization and distillation will be investigated.

- **Dataset Expansion:** We will continue curating a larger, diverse corpus of handwritten legal documents across states and formats to support model training and benchmarking.

Overall, while current models fall short of deployment-ready performance, the insights gained from this work lay a foundation for developing reliable multilingual legal AI systems capable of supporting digital transformation across the Indian judiciary.

References

Alam, N.; Kanjula, K. R.; Guthikonda, S.; Chung, T.; Vegesna, B. K. S.; Das, A.; Susevski, A.; Chan, R. S.-Y.; Uddin, S. M. I.; Islam, S. B.; Santhosh, R.; A, S.; Sharma, D.; Liu, C.; Chaturvedi, I.; Winata, G. I.; S, A.; Mukherjee, S.; and Aji, A. F. 2024. Maya: An Instruction Finetuned Multilingual Multimodal Model. arXiv:2412.07112.

Ali, A.; and Renals, S. 2018. Word Error Rate Estimation for Speech Recognition: e-WER. In Gurevych, I.; and Miyao, Y., eds., *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 20–24. Melbourne, Australia: Association for Computational Linguistics.

Bai, S.; Chen, K.; Liu, X.; Wang, J.; Ge, W.; Song, S.; Dang, K.; Wang, P.; Wang, S.; Tang, J.; Zhong, H.; Zhu, Y.; Yang, M.; Li, Z.; Wan, J.; Wang, P.; Ding, W.; Fu, Z.; Xu, Y.; Ye, J.; Zhang, X.; Xie, T.; Cheng, Z.; Zhang, H.; Yang, Z.; Xu, H.; and Lin, J. 2025. Qwen2.5-VL Technical Report. arXiv:2502.13923.

Chakraborty, S.; Harit, G.; and Ghosh, S. 2023. Trans-DocAnalyser: A Framework for Offline Semi-structured

Handwritten Document Analysis in the Legal Domain. arXiv:2306.02142.

Datta, S.; Chakraborty, A.; Bhattacharya, S.; Samui, A.; Chakraborty, A.; and Dasgupta, K. 2024. Nirnayaak: Revolutionizing Legal Research. In *2024 4th International Conference on Computer, Communication, Control & Information Technology (C3IT)*, 1–5.

Gala, J.; Chitale, P. A.; AK, R.; Gumma, V.; Doddapaneni, S.; Kumar, A.; Nawale, J.; Sujatha, A.; Puduppully, R.; Raghavan, V.; Kumar, P.; Khapra, M. M.; Dabre, R.; and Kunchukuttan, A. 2023. IndicTrans2: Towards High-Quality and Accessible Machine Translation Models for all 22 Scheduled Indian Languages. arXiv:2305.16307.

Hamdi, L.; Tamasna, A.; Boisson, P.; and Paquet, T. 2025. VISTA-OCR: Towards generative and interactive end to end OCR models. arXiv:2504.03621.

Iwana, B. K.; Rizvi, S. T. R.; Ahmed, S.; Dengel, A.; and Uchida, S. 2017. Judging a Book By its Cover. arXiv:1610.09204.

JaidedAI. 2020. EasyOCR: Ready-to-use OCR with 80+ supported languages. https://github.com/JaidedAI/EasyOCR. Accessed: 2025-05-18.

Joshi, R.; Singla, K.; Kamath, A.; Kalani, R.; Paul, R.; Vaidya, U.; Chauhan, S. S.; Wartikar, N.; and Long, E. 2025. Adapting Multilingual LLMs to Low-Resource Languages using Continued Pre-training and Synthetic Corpus. arXiv:2410.14815.

K, T. D.; James, J.; Gopinath, D. P.; and K, M. A. 2025. Advocating Character Error Rate for Multilingual ASR Evaluation. In Chiruzzo, L.; Ritter, A.; and Wang, L., eds., *Findings of the Association for Computational Linguistics: NAACL 2025*, 4926–4935. Albuquerque, New Mexico: Association for Computational Linguistics. ISBN 979-8-89176-195-7.

Kasuba, B. V.; Kudale, D.; Subramanian, V.; Chaudhuri, P.; and Ramakrishnan, G. 2025. PLATTER: A Page-Level Handwritten Text Recognition System for Indic Scripts. arXiv:2502.06172.

Khan, S.; Tarun, A.; Ravi, A.; Faraz, A.; Patidar, A.; Pokala, P. K.; Bhangare, A.; Kolla, R.; Khatri, C.; and Agarwal, S. 2025. Chitrarth: Bridging Vision and Language for a Billion People. arXiv:2502.15392.

Li, C.; Liu, W.; Guo, R.; Yin, X.; Jiang, K.; Du, Y.; Du, Y.; Zhu, L.; Lai, B.; Hu, X.; Yu, D.; and Ma, Y. 2022. PP-OCRv3: More Attempts for the Improvement of Ultra Lightweight OCR System. arXiv:2206.03001.

Lu, S.; Li, Y.; Chen, Q.-G.; Xu, Z.; Luo, W.; Zhang, K.; and Ye, H.-J. 2024. Ovis: Structural embedding alignment for multimodal large language model. *arXiv preprint arXiv:2405.20797*.

Meta AI. 2024. Llama 4: Advancing Open Multimodal Intelligence. Accessed: 2025-05-18.

Poznanski, J.; Borchardt, J.; Dunkelberger, J.; Huff, R.; Lin, D.; Rangapur, A.; Wilhelm, C.; Lo, K.; and Soldaini, L. 2025. olmOCR: Unlocking Trillions of Tokens in PDFs with Vision Language Models. arXiv:2502.18443.

Project Anuvaad. 2025. Anuvaad: Open Source Document Translation Platform for Indic Languages. https://anuvaad.sunbird.org/. Accessed: 2025-05-19.

S, M.; J, A. G.; Kumar, G.; Shetty, M. P.; and C, S. P. 2024. Kannada Legal Document Summarizer: A Survey. In *2024 Second International Conference on Advances in Information Technology (ICAIT)*, volume 1, 1–5.

Sarvam AI. 2024. Sarvam-1: Bridging the Language Gap with India's First Large Language Model. https://www.sarvam.ai/blogs/sarvam-1.

Smith, R. 2007. An Overview of the Tesseract OCR Engine. In *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, volume 2, 629–633.

Steiner, A.; Pinto, A. S.; Tschannen, M.; Keysers, D.; Wang, X.; Bitton, Y.; Gritsenko, A.; Minderer, M.; Sherbondy, A.; Long, S.; Qin, S.; Ingle, R.; Bugliarello, E.; Kazemzadeh, S.; Mesnard, T.; Alabdulmohsin, I.; Beyer, L.; and Zhai, X. 2024. PaliGemma 2: A Family of Versatile VLMs for Transfer. arXiv:2412.03555.

Yang, Z.; Li, L.; Lin, K.; Wang, J.; Lin, C.-C.; Liu, Z.; and Wang, L. 2023. The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision). arXiv:2309.17421.

Appendix

Model	Translated Text
Ground Truth	On perusal of Mutation Entry No. <masked> dated <masked>, it is recorded that by virtue of the Agreement for Sale dated <masked> bearing registration no. <masked> registered at the Offices of Sub-registrar of Assurances at <masked>... <masked> sold land bearing Survey No. <masked> for an area admeasuring (10 m X 81 m) 8 Ares out of 75.1 Ares in favor of <masked> and <masked> for a consideration of <masked> on certain terms and conditions contained therein...
Chitrarth	The document is written in Marathi and contains information about the details of a meeting. The text mentions the names of several people, including "Shri. K. K. Patil", "Shri. S. S. Patil", and "Shri. K. K. Patil". The document also contains information about the dates of the meeting...
Ovis2-34B (Int4)	The image is a page from an old Marathi ledger... The text at the top reads **"krushnaavari nidhidand"** which translates to *Krishnaavari Fine Ledger*. **Key details:** **1) Date:** 23/06/'94 appears multiple times. - Mentions "etch." and "57". Further text like "jath." suggests amounts paid. Additional details describe transactions and account settlements.
Maya-8B	The image displays a handwritten sheet of paper written in the Marathi language. It is a notebook page with various notes... The paper appears to be a study guide or a notebook...
Ovis2-16B	The document is a financial record in Marathi, dated 27-04. It appears to be an official ledger entry... Top Section: 1) *kramanchiti nondavane* — Serial Number Entry Place: Jalgaon Name & Account No.: Dattatraya Udaylal Desai... Amounts: INR 6 lakh =

Table 2: Qualitative comparison of translations generated by different vision-language models against the human-annotated ground truth.