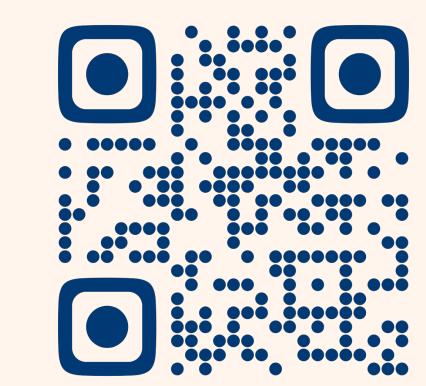


Seeing Justice Clearly: Handwritten Legal Document Translation with OCR and Vision-Language Models



Shubham Kumar Nigam^{1,3} Parjanya Aditya Shukla¹
Noel Shallum² Arnab Bhattacharya¹

¹IIT Kanpur, India

²Symbiosis Law School Pune, India

³University of Birmingham Dubai, UAE

Link to paper and code

Motivation

- Goal:** Engineer an AI-driven, edge-deployable solution for document digitization in legal domain.
- The Indian judiciary rely on First Information Reports (FIRs), case diaries, evidence documents, handwritten in native languages.
- Address the hurdles of low quality scans, handwriting variations, linguistic diversity and legal terminology with Vision Language Models (vLLMs).
- Enable scalable legal digitization and efficient case processing across Indian law courts.

Task Description

- Handwritten Text Recognition (Task A):**
 - Generate the text content from a scan of document handwritten in any low-resource language (like Marathi).
- Machine Translation (Task B):**
 - Generate an accurate translation of text content (from Task A) to any language.
- Emphasizes transparency and accurate interpretability of handwritten text for an unbaised judgment.
- Enables users to understand text handwritten in any Indian language.

Dataset Overview

- Custom Curated:** ~60 scanned PDFs of Land Mutation Records (Legal Domain) containing handwritten entries in Marathi language.
- Real-world dataset:** Derived from Land Records and Revenue Offices, Maharashtra.
- Document Complexity:** Documents vary in length (single- and multi-page) and contain handwritten and printed text, along with stamps, seals, signatures, and tables, reflecting real-world digitization challenges.
- Components:**
 1. NyayaScans – Collection of scanned PDFs of documents containing handwritten Marathi text.
 2. NyayaTranslations – High Quality annotation of document text in Marathi, English by legal language experts, that serves as ground truth for Handwritten Text Recognition and Machine Translation task.

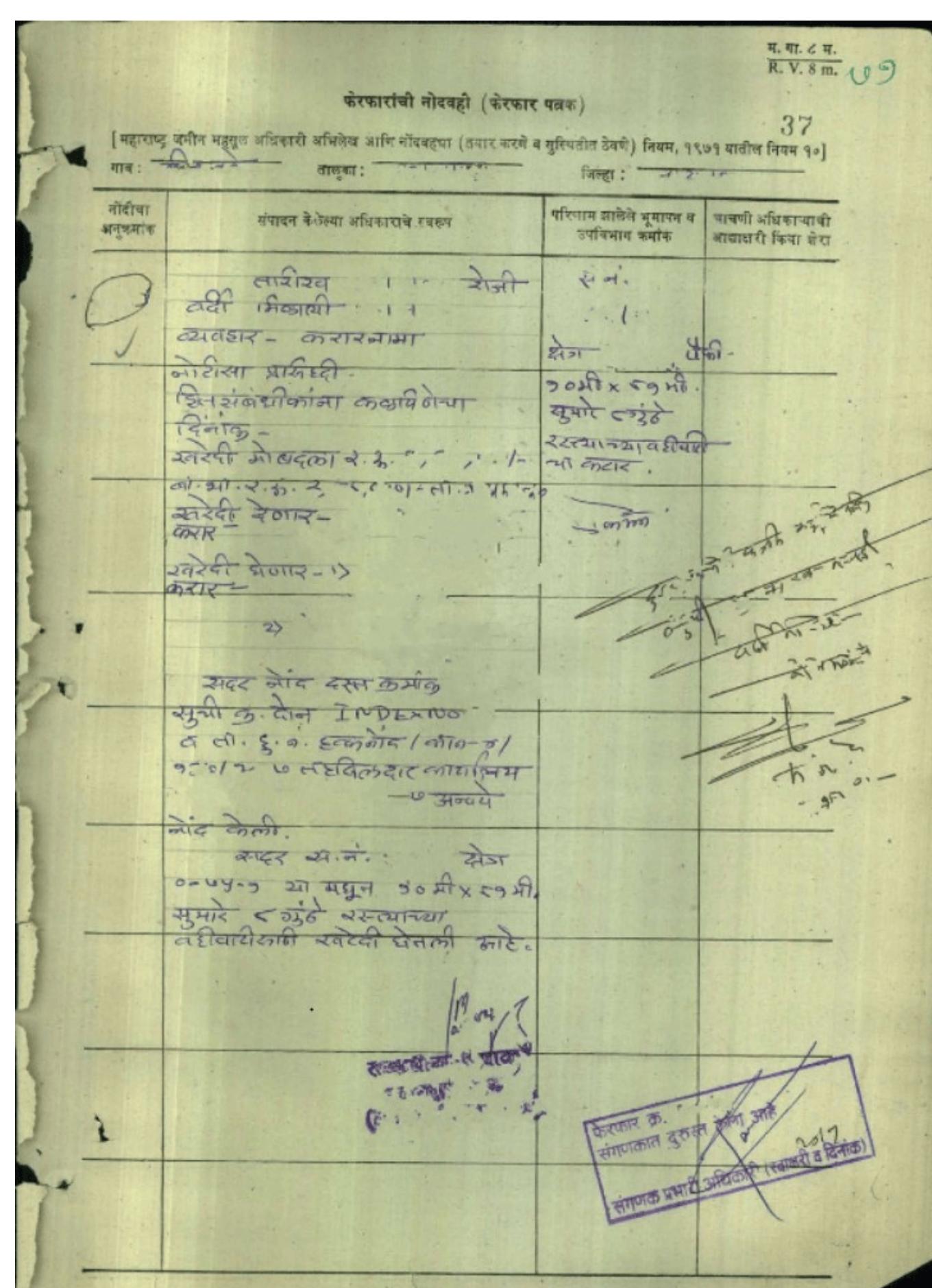
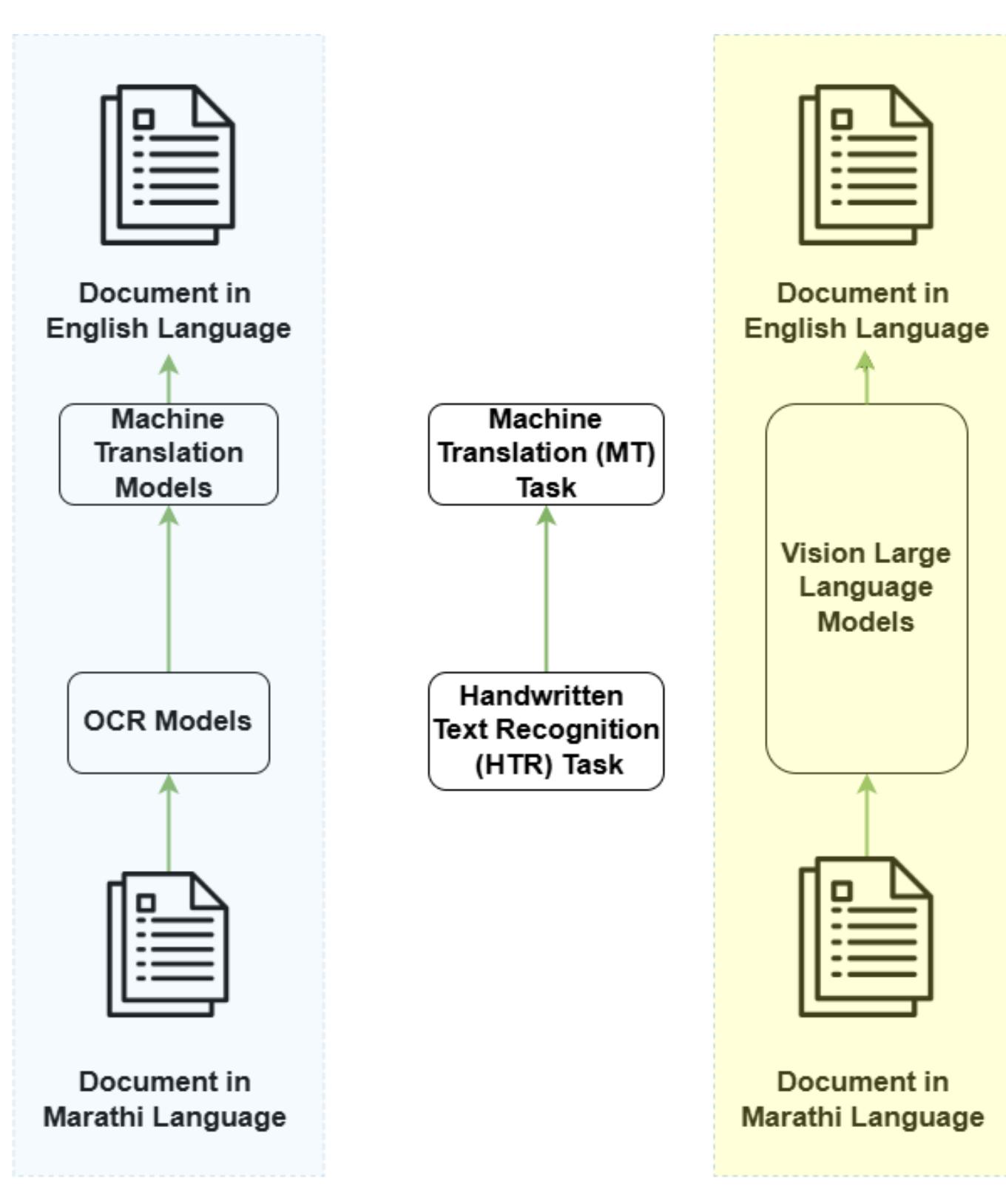


Figure 1. A Sample document from dataset.

Approaches and Models

Two different paradigms:

- OCR-MT:** 2 phased document translation using Optical Character Recognition (OCR) for handwritten text extraction and Machine Translation (MT) for translation to any language.
- Vision Large Language Models (vLLMs):** Single phased (direct) translation to any language using vLLMs in a Zero-Shot setting.
 - OCR Models:** Tesseract, EasyOCR, and PaddleOCR.
 - MT Models:** IndicTrans2 by AI4Bharat and Sarvam-1 by Sarvam.
 - Vision Large Language Models:** Chitrarth by Krutrim, Maya-8B and Ovis2 by AIDC-AI.



Results and Analysis

- Handwritten Text Recognition Performance:** Better performance on printed text in comparison to handwritten text. Errors such as omissions, misrecognized characters, or fragmented outputs are common for handwritten text.
- OCR Models Comparison:** EasyOCR consistently out-performed PaddleOCR and Tesseract.
- Machine Translation Performance:** Incoherent or mixed-language outputs, loss of legal or factual content, and incomplete phrasing for longer sentences.
- Evaluating vLLMs:**
 - Chitrarth produced a hallucinated summary about a meeting, with invented names, dates, and locations that were not present in the original document.
 - Maya-8B output was generic and irrelevant.
 - Ovis2-34B provided partial translations of text such as Marathi ledger headers and dates.
 - Ovis2-16B identified some legal and financial cues (e.g., account numbers, names, locations). It also translated a few phrases and recognized structural layout, but it provided incomplete and incoherent output.

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 "et.," 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 1. Qualitative comparison of translations generated by different vision-language models against the human-annotated ground truth.

Contributions and Impact

- Benchmarked Tesseract, EasyOCR, and PaddleOCR on a curated dataset of handwritten documents.
- Analyzed translation performance using IndicTrans2 and Sarvam-1 models on OCR extracted text.
- Evaluated performance of Vision Large Language Models (Chitrarth, Maya and Ovis) against OCR-MT pipelines

Limitations and Future Work

- Dataset Expansion:** Limitations pertaining to dataset size and scope - only 60 scanned legal documents in Marathi language.
- Domain Adaptation:** Models are evaluated in zero-shot settings; domain-specific pretraining and fine-tuning should improve performance.
- Faithfulness:** Reducing hallucinations via grounding and verification to ensure legal reliability.
- Evaluation Frameworks:** Developing legal-specific translation evaluation metrics.

References

- [1] Nahid Alam, Karthik Reddy Kanjula, Surya Guthikonda, Timothy Chung, Bala Krishna S Vegeesa, Abhipsha Das, Anthony Susevski, Ryan Sze-Yin Chan, S M Iftekhar Uddin, Shayekh Bin Islam, Roshan Santhosh, Sneha A, Drishi Sharma, Chen Liu, Isha Chaturvedi, Genta Indra Winata, Ashvanth, S, Snehanush Mukherjee, and Alham Fikri Aji. Maya: An instruction finetuned multilingual multimodal model, 2024.
- [2] Jay Gala, Pranjal A. Chitale, Raghavan AK, Varun Gumra, Sumanth Doddapaneni, Awasanth Kumar, Janki Nawale, Anupama Sujatha, Ratish Puduppully, Vivek Raghavan, Pratyush Kumar, Mitesh M. Khapra, Raj Dabre, and Anoop Kunchukuttan. Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled Indian languages, 2023.
- [3] Jaide AI. Easyocr: Ready-to-use ocr with 80+ supported languages. <https://github.com/JaideAI/EasyOCR>, 2020. Accessed: 2025-05-18.
- [4] Shaharukh Khan, Ayush Tarun, Abhinav Ravi, Ali Faraz, Akshat Patidar, Praveen Kumar Pokala, Anagha Bhangare, Raja Kolla, Chandra Khatri, and Shubham Agarwal. Chitrarth: Bridging vision and language for a billion people, 2025.
- [5] Shiyin Lu, Yang Li, Qing-Guo Chen, Zhao Xu, Weihua Luo, Kaifu Zhang, and Han-Jia Ye. Ovis: Structural embedding alignment for multimodal large language model. arXiv preprint arXiv:2405.20797, 2024.
- [6] Sarvam AI. Sarvam-1: Bridging the Language Gap with India's First Large Language Model. <https://www.sarvam.ai/blogs/sarvam-1>, 2024.

Acknowledgements: We express our sincere gratitude to BharatGen, India, for providing essential computational resources and hardware support that made this work possible. We also thank our collaborators, student research assistants, and the legal experts from partnering law colleges for their valuable contributions to annotation, evaluation, and the overall refinement of the project. Their collective effort and insights greatly strengthened this research.