

Forum:[Hack Forums](#)

Thread: [Building OCR Systems for Tamizhi and Kurdish Historical Documents](#)

Stanley - August 19, 2025, 6:20 pm

Building OCR Systems for Tamizhi and Kurdish Historical Documents

Developing OCR for ancient scripts like Tamizhi (Tamil-Brahmi) and Kurdish historical texts is uniquely challenging due to character complexity, noise in source materials, and the lack of specialized datasets. Recent research using AI models such as LSTM, CNN, and fine-tuned Tesseract systems shows promising results, with Tamizhi OCR achieving over 91% accuracy. While no Kurdish-specific OCR exists yet, leveraging pre-trained Arabic models offers a practical pathway. These findings highlight the importance of tailored datasets, advanced machine learning techniques, and ongoing research in preserving and digitizing historical documents.

<https://hackernoon.com/building-ocr-systems-for-tamizhi-and-kurdish-historical-documents?source=rss>

JoseMar - September 3, 2025, 8:05 am

(Aug 19, 2025 06:20 PM)Stanley Wrote: [\[CITE\]](#) Building OCR Systems for Tamizhi and Kurdish Historical Documents

Developing OCR for ancient scripts like Tamizhi (Tamil-Brahmi) and Kurdish historical texts is uniquely challenging due to character complexity, noise in source materials, and the lack of specialized datasets. Recent research using AI models such as LSTM, CNN, and fine-tuned Tesseract systems shows promising results, with Tamizhi OCR achieving over 91% accuracy. While no Kurdish-specific OCR exists yet, leveraging pre-trained Arabic models offers a practical pathway. These findings highlight the importance of tailored datasets, advanced machine learning techniques, and ongoing research in preserving and digitizing historical documents.

<https://hackernoon.com/building-ocr-systems-for-tamizhi-and-kurdish-historical-documents?source=rss>

Tackling ancient scripts with all their quirks... that's a huge challenge.

sharn - September 14, 2025, 9:39 am

Working with linguistic experts seems essential.

They know the nuances that machines can easily miss.