BHASHA SETU: A MODULAR MULTILINGUAL TRANSLATION SYSTEM FOR LOW RESOURCE AND MULTISCRIPT ENVIRONMENTS

Dev Shah[1*], Shikhiin Marudkar[2,], and Chinmay Raut[3]

[1,2,3]Department of Computer Engineering, SVKM's Dwarkadas J. Sanghvi College of Engineering, Mumbai, India*
devshah1682003@gmail.com mshikhiin@gmail.com chinmay.raut@djsce.ac.in

***Corresponding Author:** devshah1682003@gmail.com*

Abstract

Translation systems currently in use often fall woefully short in environments with spotty connectivity and diverse linguistic requirements largely because offline functionality and image-based content are frequently unsupported alongside low-resource languages spoken in India. Essential is a modular multilingual translation solution guaranteeing privacy and adaptability across many languages in various formats quite seamlessly. This work presents an amazing multilingual document translating system with image-based translating capability as well as offline and online translating capability. It mostly depends on neural machine translation to enable several languages smoothly and makes good use of optical character recognition (OCR) to extract text from images. The system uses external APIs to prioritize privacy by providing offline translating between English and Marathi without an internet connection, so enabling real-time translation. Its modular architecture guarantees effective translation paths and general compatibility with many formats, including PDF and picture files. These days, it readily and effectively negotiates language obstacles in many different fields. The paper closes with evaluating performance and making recommendations for future improvements including progressively using more sophisticated NLP models and including more eccentric language support.

Keywords: *optical character recognition(ORC) · offline translation · online translation · indic languages · GAN based image generation · transfer learning · low resource languages · text to image synthesis 200 · cultural and script rendering · subtitles translation.*

1 Introduction

These days, among many other sectors worldwide, the demand for effective cross-language communication has surged sharply in official domains including government, healthcare, education, and law. Millions of people view papers written in quite different languages every day, which limits their capacity to correctly grasp information. The challenge is especially clear in linguistically diverse nations like India, where many languages and dialects coexist rather vibrantly within the same area. Traditional translation methods usually fall short when confronted with region-specific jargon or technical terms in very remote settings. Translating very sensitive documents raises some major privacy concerns since many of the available solutions depend mostly on cloud architecture. Scanning documents presents further difficulties in low-tech environments or rural areas since translating usually requires OCR processing first. These days, our modular multilingual document translating system performs effectively both online and offline translation. These days, it generates multilingual fonts using the Django framework and quite seamlessly combines modern NLP components with neural machine translation. Three main modules make up the system: an offline module translating English to Marathi, an online module leveraging real-time cloud APIs, and an image-based module leveraging OCR. These days, it seamlessly offers flexible, user-centered translation over PDF, DOCX, and image files among other formats. The platform is especially meant to enable speakers of languages other than English interact in a variety of settings..

2 Literature Survey

With a move from Statistical Machine Translation mostly towards Neural Machine Translation paradigms, machine translation has made great advancement recently. Inspired by pioneering work by Johnson et al. [1], who presented a basic but effective method allowing a single neural network model to translate across many language pairs, innovations in multilingual NMT emerged rather early. This approach, using an artificial token denoting the target language, set the stage for zero-shot translation—where a model might translate between untrained language pairs. This creativity proved the viability of parameter sharing across multilingual systems and lessened the need for many customized models. Advancing this field, Conneau et al. [2] proposed XLM-R, a transformer-based multilingual language model trained on 100 languages using more than two terabytes of filtered CommonCrawl data. Particularly in low-resource languages like Swahili and Urdu, XLMR proved better than mBERT in cross-lingual tasks including natural language inference (XNLI) and question answering (MLQA). Large-scale pretraining thus greatly improved cross-lingual transfer without sacrificing performance in high-resource languages. Developed by Arivazhagan et al. [3] is a massively multilingual NMT model supporting 103 languages. Their efforts showed that multilingual parameter sharing not only cut computing expenses but also greatly raised translating quality for underrepresented languages. This has spurred more investigation on low-resource language translation. Focusing on discourse-level events including anaphora resolution and gender agreement, Bawden et al. [4] investigated the flaws of conventional sentencelevel MT systems. Their research underlined the necessity of translating systems including more general linguistic background to attain improved coherence and accuracy over several sentences. Particularly by means of the integration of optical character recognition (OCR) with neural machine translation, multimodal translation using image-based inputs has lately attracted great interest. Combining Tesseract-based OCR with multilingual models for accurate translation has shown great success in studies involving OCR for Indic scripts including Devanagari and Tamil [5]. Fu et al. [6] focused on how pretraining improves multilingual image-caption pairs. Their studies showed that visual grounding can greatly improve translation quality, especially in scene-text situations including several scripts and intricate backgrounds. Furthermore finding use in machine translation are generative adversarial networks (GANs). Based on text input, Zhang et al. [7] presented the Self-Attention GAN, able to create semantically consistent images with realism. Ramesh et al. [8] who highlighted the features of big-scale text to image generating systems such as DALL-E, which support semantic cross-modal mappings and improve multilingual translation with visual context, further supported this. Matviienko et al. [9] carefully examined Google Translate, DeepL, and Microsoft Translator among other online translating tools. Their results showed that although Google Translate provides broad language coverage, DeepL usually shines in maintaining minute linguistic nuances—especially in European languages—while other languages. These comparative analyses offer insightful information to help low-resource translating systems be improved. Supporting the semantic alignment of visual content on multilingual platforms, Sharma and Bansal [10] offered a thorough review of image generating techniques and their function in enhancing text to image synthesis. There have also lately surfaced several domain-specific translating systems. Designed for low-connectivity settings, Chaudhari and Kamat [11] proposed an offline translator using multilingual neural networks. Hybrid translating systems for Indian languages combining rule-based and neural techniques were first presented by Patil and Deshmukh [12]. With an eye toward document translation and classification, Patel [13] investigated the use of multilingual NLP methods in the Indian legal system. Examining GANs for Indiclanguage text to image generation, Tiwari [14] emphasized their applicability for image-supported translation. At last, Kumar et al. [15] showed how adversarial learning might be used to improve low-resource neural machine translation by leveraging multilingualism to raise translation quality in underrepresented languages.

3 Methodology

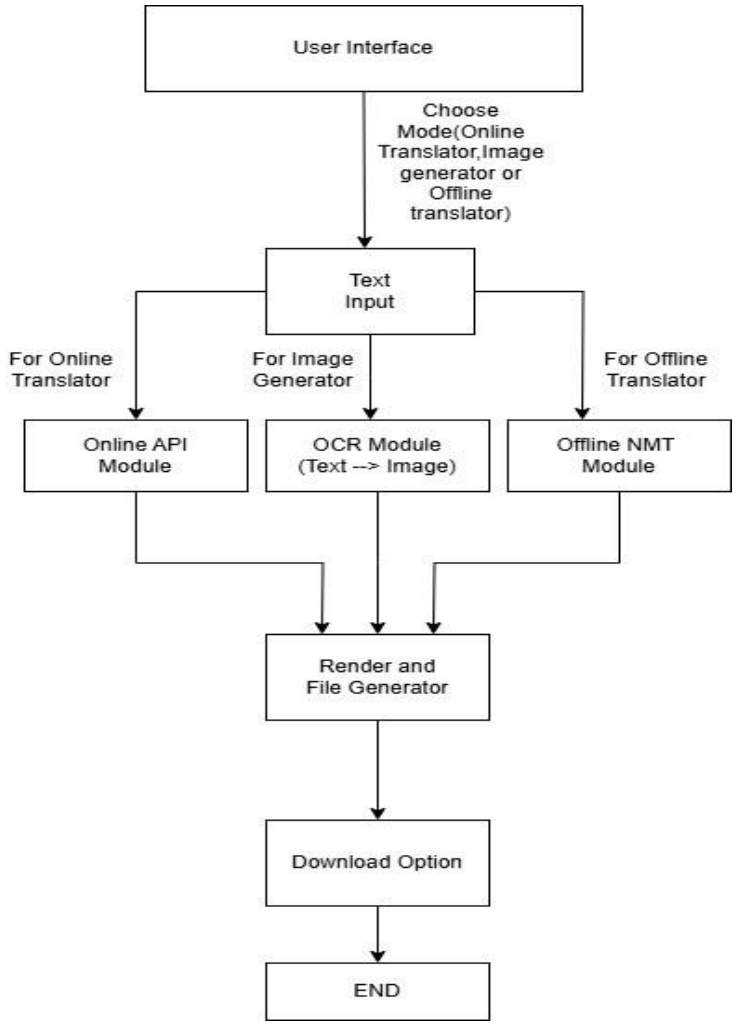

Figure1: Block Diagram

Three main components—online translation, offline translation, and picture production—make up a modular framework supporting this multilingual translating system. Though they run independently, modules are closely connected by standardized interfaces and a central middleware layer ensuring very effective data flow. The Online Translation Module uses APIs including DeepL, Google Translator, and Microsoft Translator for effective real-time cloud-based translations. These services offer great linguistic help very quickly and with very great accuracy by means of access to vast language models that are updated constantly. This module performs effectively in many network-enabled scenarios and supports many languages both internationally and regionally. To minimize privacy and connection issues, the Offline Translation Module now employs locally housed neural machine translation models. These days, transformer-based architectures including BERT variants and cryptic GPT models help to translate English-Marathi precisely. Models are trained on carefully chosen parallel datasets often including tokenization, normalizing, and laborious linguistic alignment techniques. Often these datasets include tokenization and normalizing. Especially in situations with limited resources, methods like data augmentation and transfer learning greatly improve performance. The Image Generation Module significantly enhances translating output by including pertinent nearby images. Transformer-based models and Generative Adversarial Networks allow very sophisticated capabilities to enable zero-shot text-to– image synthesis quite successfully. GANs trained on a range of datasets somewhat improve image realism while transformers create contextually relevant images by effectively accessing big picture-text corpora without image-text pair training. As it coordinates inter-module communication across several platforms and devices, the central middleware layer efficiently balances resource distribution. This architecture makes scalability in cloud environments simpler and allows rather flexible local deployment as well. Operating on separate, highly secured systems, offline components are perfect for environments with extremely sensitive data or no internet connection at all. Still of great relevance in design are data security and privacy. Under tight security, offline models operate subtly in far-off locations with highly limited user access and great encryption. Usually, safe APIs with encrypted channels sufficiently guard user data during online transactions. Black box and white box testing methods are rather widely and regularly applied to guarantee general system stability. In some sense, they offer the rather complete validation of internal logic and outward behavior spanning translation modules and picture generation. Future versions will include support for audio and video inputs in order to produce a flexible, user-centered translating system that can partially close linguistic gaps in a range of real-world scenarios.

4 Implementation

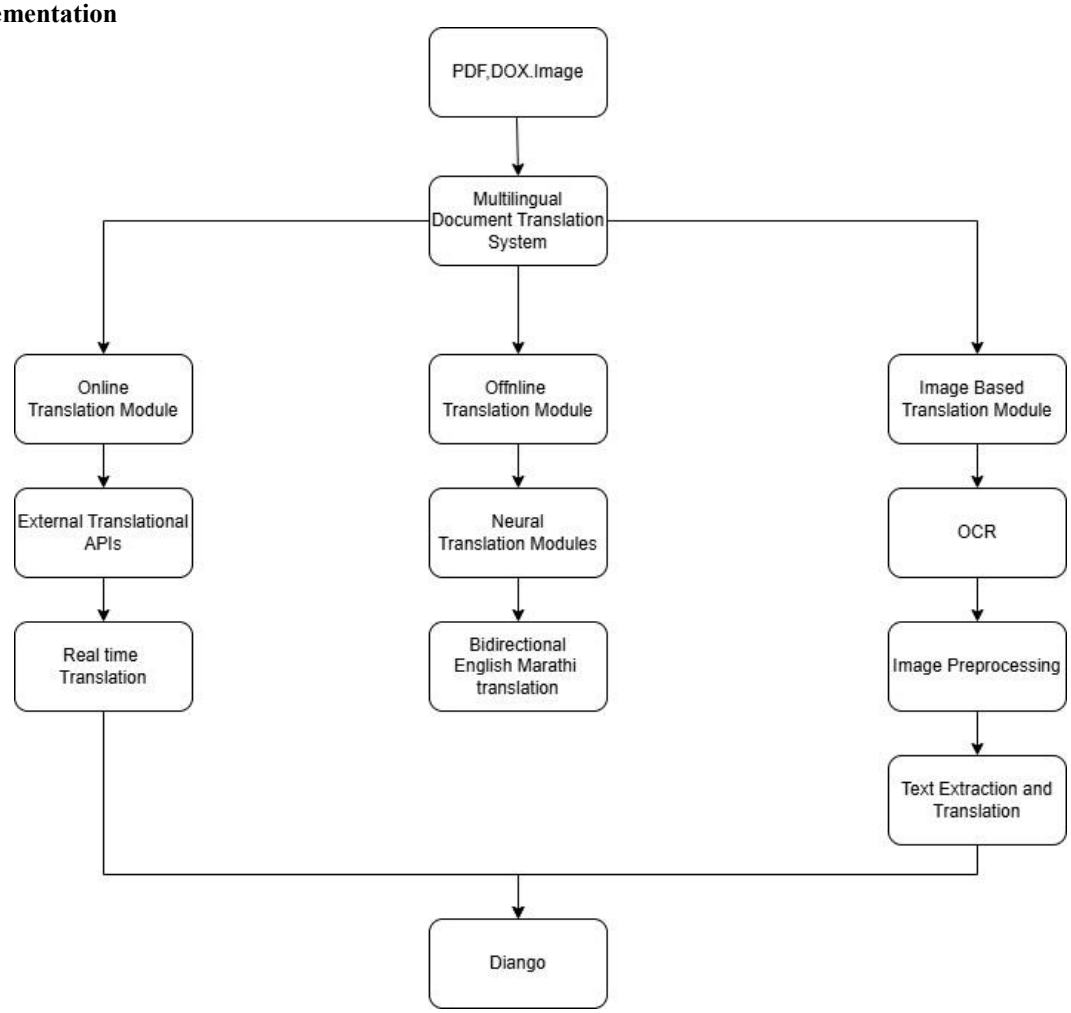

Figure2: Flowchart

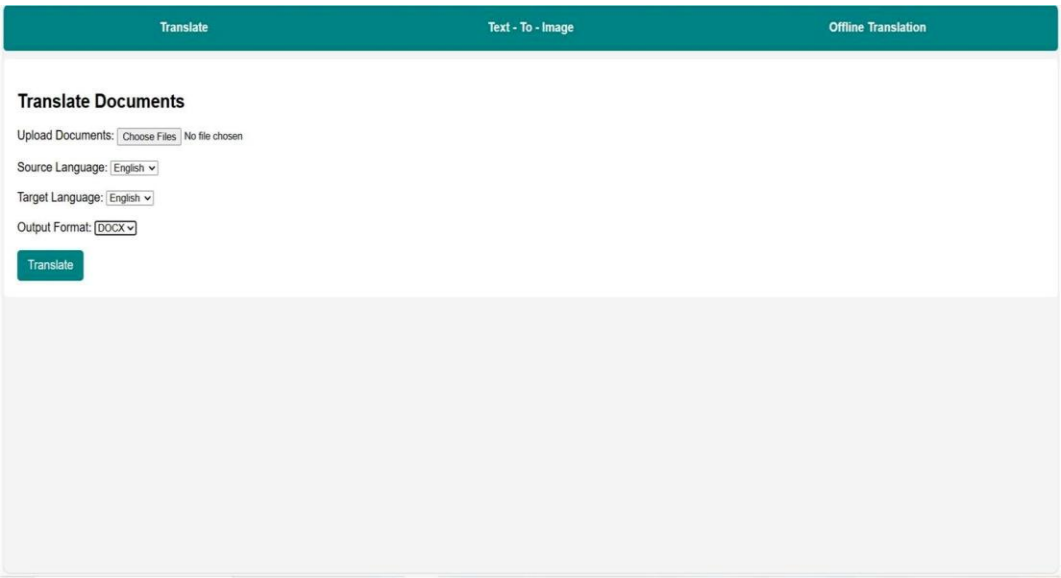

Figure3: Online Translation Interface

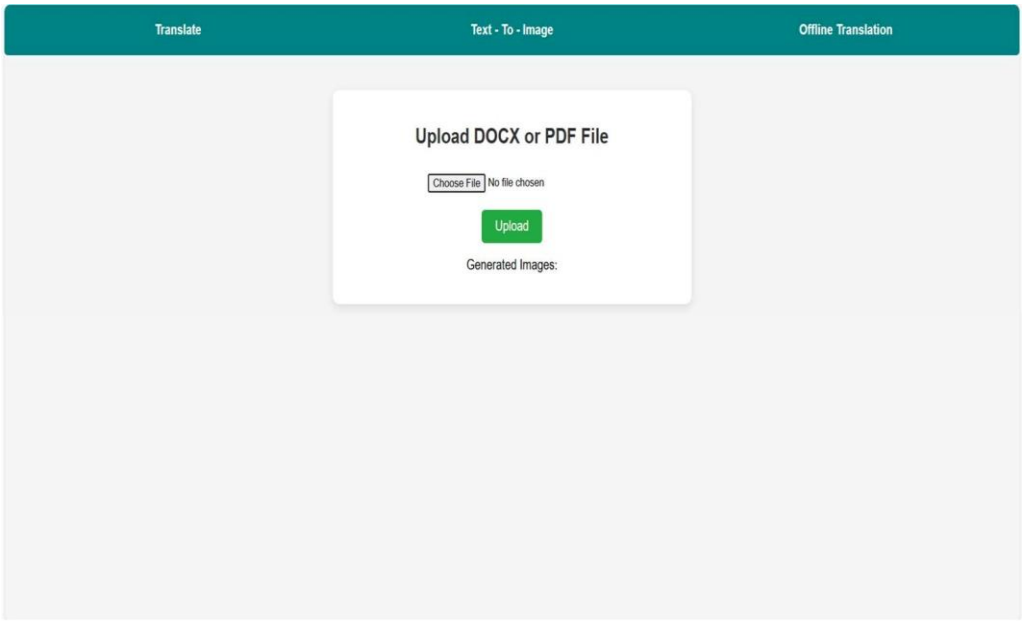

Figure4: Image Generator Interface

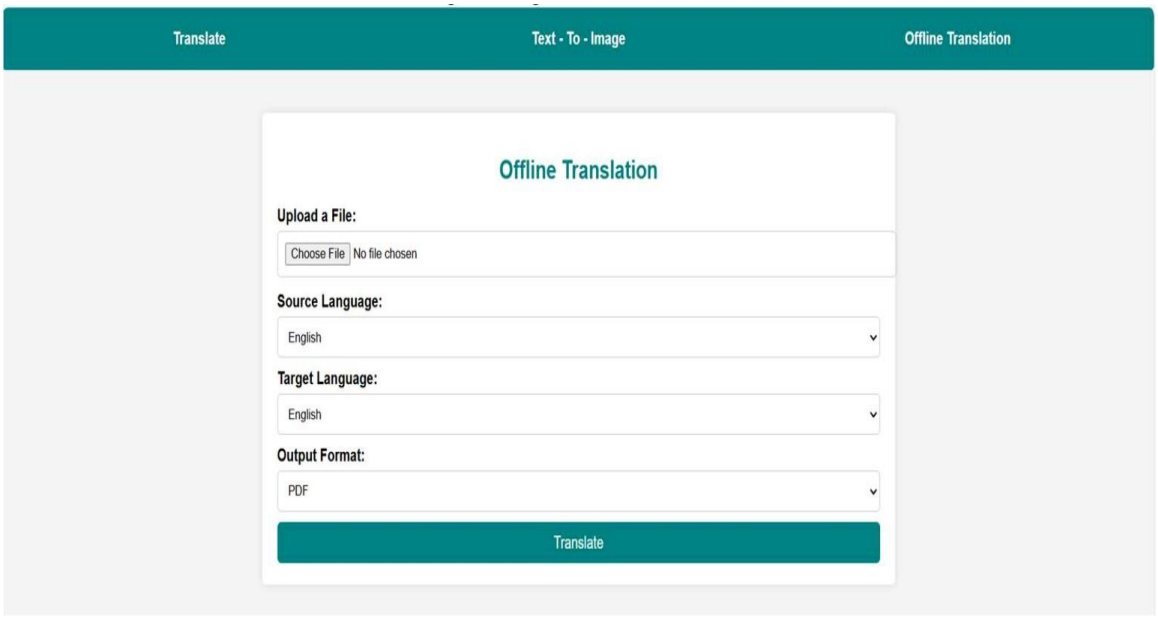

Figure5: Offline Translation Interface Table 1: Online Output Example

Input Text	Output (Marathi)	Output(Bengali)
On a quiet, rainy evening, Arjun was cleaning the attic of his ancestral home when he stumbled upon an old wooden chest. Covered in dust and cobwebs, it had clearly not been opened in years. Curiosity got the better of him, and he pried it open. Inside, he found a bundle of yellowed letters tied with a red ribbon. Among them was one addressed to his grandfather, written in an elegant but unfamiliar handwriting. As he carefully unfolded the fragile paper, Arjun realized it was a love letter. The writer, Meera, spoke of a deep affection for his grandfather and of a promise to meet him on a	शांत, पावसाळ्याच्या संध्याकाळ अजुर्न जुन्या लाकडी छातीवर अडखळत असताना आपल्या व डलोपा जत घराची अटारी साफ करीत होता. धूळ आ ण कोबवेबमध्ये झाकलेले, हे वषामध्ये पणे उघडले गेले नाही. कु तूहल त्याच्यापेक्षा चांगले झाला आ ण त्याने ते उघडले. आत, त्याला लाल रबनने बांधलेल्या पवळ्या अक्षरांचा एक बंडल सापडला. त्यापैक एकाने त्याच्या आजोबांना संबो धत के ले होते, जे एका मोहक परंतु अप र चत हस्तलेखनात ल हलेले होते. त्याने काळजीपूवर्क नाजूक कागद उलगडताच अजुर्नला समजले क ते	একিট শা , বষর্ার স য্ায় অজর্ন তার ৈপতৃক বাড়ির অয্াটিকিট পির ার করিছেলন যখন িতিন কাঠের পুরােনা বুেক েহঁাচট েখেয়িছেলন। ধুেলা এবং েকা েস াকা, এিট তই বছেরর পর বছর েখালা হয়িন। েকৗতহল তার েথেক আরও ভাল হেয় উঠল এবং েস এিটেক উ কের িদ- েয়িছল। িভতের, িতিন একিট লাল ফতা িদেয় বঁাধা হলুদ বেণর্ একিট বা ল েদখেত েপেলন। এর মেধয্ একিট িছল তঁার দাদােক সে াধন করা, এিট একিট মা জর্ত তেব অপিরিচত হ াক্ষর লেখিছল। িত-

certain day at an old railway station. But the letter had never been opened—his grandfather had never read it. Arjun's heart pounded as he wondered what had happened.	एक प्रेम पत्र आहे. मीरा या लेखकाने आपल्या आजोबांबद्दल आ ण जुन्या रेल्वे ानकात एका व श दवशीत्यालाभेटण्याच्याआ ासनाबद्दलमनापासून प्रेम के ले. परंतु हे पत्र कधीही उघडले नव्हते - त्याचे आजोबा हे कधीही वाचले नव्हते. काय घडले आहे याबद्दल आ यर्च कत झाल्यामुळे अजुर्नचे हृदय धडधडले.	িন যখন সাবধানতার সাে থ ভ রু কাগজিট কাশ কেরিছেলন, অজর্নু বুঝেত েপেরিছেলন েয এিট একিট ে েমর িচিঠ। েলখক মীরা তঁার দাদার িত গভীর ে েহর কথা বেলিছেলন এবং একিট পুরাে না েরলওেয় ে শেন একিট িনিদর্ িদেন তঁার সাে থ েদখা করার িত িত িদেয়িছেলন। তেব িচিঠিট কখনই েখালা হয়িন - তঁার দাদা কখনও তা পেড়ন িন। অজর্ে ু নর হৃদয় যখন ভাবিছল তখন কী ঘেটেছ তা ভাবিছল।

The Online Translation Module readily makes use of cloud-based translating services including Google Translate and DeepL via the Deep Translator Python module. Processing user-inputted text or uploaded files—such as DOCX and PDF—Python-docx and pdfplumber libraries extract content as needed. After selecting a cloud API, the acquired text is quickly sent for translation and returned either plain text or downloadable PDF form. This module is quite fit for time-sensitive applications mostly because of its fast response time and exact support for many widely spoken languages. Strict API rate restrictions and ongoing privacy concerns, among other external dependencies, however also call for constant internet connectivity. Examples of non-editable material the Image-Based Translation Module manages rather well these days are scanned documents and image-based PDFs. The Tesseract OCR engine detects text in many scripts, including handwritten input, quite remarkably. Using careful preprocessing techniques from past, grayscale conversion and binarization greatly raise recognition accuracy. Extracted text is carefully cleaned using regex and heuristic fixes before being sent to online or offline translating engines. Though performance is much influenced by image quality and formatting complexity, this module does a good job of translating historical archives and printed books. Designed for environments with limited internet access, such secure government buildings and rural schools, the offline translating module is It makes advantage of Meta's NLLB-200-distilled-600M model, which Hugging Face's Transformers library makes rather simple access possible using modern deep learning architectures. The system divides the material into reasonable chunks after extracting text from uploaded papers that cleanly fit inside the rather cryptic token constraints. Each chunk is first tokenized locally and then translocated before being painstakingly and precisely rebuilt into the finished output. While giving user anonymity top priority, this module supports translating between many languages, especially obscure Bengali dialects and Marathi. The first setup calls for downloading a model, installing dependencies, and painstakingly configuring an environment, so requiring several system resources. After translation, output is formatted into user-defined PDF or DOCX files with Unicode font integration guaranteeing very precisely rendering of many language scripts. Fonts like Noto Sans Devanagari and Noto Sans Gujarati guarantee rather accurate cultural representation and linguistic integrity in many contexts usually. Rendering engine painstakingly handles line wrapping and script display across many devices platforms with almost perfect accuracy and Table 2: Image Generator Output Example

Input Text	**Output Images**
On a quiet, rainy evening, Arjun was cleaning the attic of his ancestral home when he stumbled upon an old wooden chest. Covered in dust and cobwebs, it had clearly not been opened in years. Curiosity got the better of him, and he pried it open. Inside, he found a bundle of yellowed letters tied with a red ribbon. Among them was one addressed to his grandfather, written in an elegant but unfamiliar handwriting. As he carefully unfolded the fragile paper, Arjun realized it was a love letter. The writer, Meera, spoke of a deep affection for his grandfather and of a promise to meet him on a certain day at an old railway station. But the letter had never been opened—his grandfather had never read it. Arjun's heart pounded as he wondered what had happened.	 

Table 3: Offline Output Example

Input Text	Output (Marathi)
On a quiet, rainy evening, Arjun was cleaning the attic of his ancestral home when he stumbled upon an old wooden chest. Covered in dust and cobwebs, it had clearly not been opened in years. Curiosity got the better of him, and he pried it open. Inside, he found a bundle of yellowed letters tied with a red ribbon. Among them was one addressed to his grandfather, written in an elegant but unfamiliar handwriting. As he carefully unfolded the fragile paper, Arjun realized it was a love letter. The writer, Meera, spoke of a deep affection for his grandfather and of a promise to meet him on a certain day at an old railway station. But the letter had never been opened—his grandfather had never read it. Arjun's heart pounded as he wondered what had happened.	एक शांत, पावसाळ संध्याकाळ अजुर्न आपल्या पूवर्जांच्या घराच्या छतावर साफसफाई करत होता तेव्हा तो एका जुन्या लाकडी पेट वर पडला. धूळ आ ण कोळ ने झाकलेला, तो पणे वषार्नुवष उघडला गेला नव्हता. उत्सुकतेने त्याच्यावर वजय मळवला, आ ण त्याने ते उघडले. आत, त्याला लाल रबनने बांधलेले पवळे अक्षरे सापडले. त्यापैक एक त्याच्या आजोबाला संबो धत के लेला होता, ज्यामध्ये एक ल हलेला होता. मीरा यांनी आपल्या आजोबांबद्दल गहन प्रेम क्त के ले आ ण जुन्या रेल्वे ानकावर एक व श दवशी भेटण्याचे आ ासन दले. पण पत्र कधीही उघडले गेले नव्हते. आजोबांनी ते कधीही वाचले नव्हते. अजुर्नचे मन काय झाले हे वचारत धडकले.

flexibility somehow. While handling translating operations with custom storage handling, Django framework powers backend providing safe environment for file uploads and RESTful APIs. It somewhat successfully includes file validation tools and CSRF protection. File validation proceeds via rather indirect means using complex procedures. RESTful APIs respond quickly to user needs. Behind scenes, custom storage handling occurs under almost perfect security. Standard HTML CSS and Bootstrap help build user interfaces that are shockingly responsive and enable simple-ish document upload in many languages.

5 Conclusion

Developing multilingual translating systems marks a significant step toward removing language obstacles in geographically separated and globally linked regions. By allowing both offline and online translations, this platform guarantees accessibility independent of internet connection or different file formats. It is quite helpful in many real-world circumstances these days since it can translate a wide spectrum of information formats, including scanned papers and posters. Thanks to user-centered design, which also offers a range of translating techniques for total customizing, people can choose from a great variety of input languages and output modes. From official announcements and academic jargon to private records including unusual letters and quite long resumes, the system can handle a broad spectrum of needs. These days, multilingual knowledge is easily available to a larger audience since it reduces dependency on costly interpreters and proprietary technology. Mostly in bilingual surroundings, the system effectively promotes digital inclusion and offers a spectrum of educational, governmental, and social opportunities. It supports reasonably egalitarian access to knowledge, encourages involvement among many language groups, and remarkably fits to changing social demands. It is a relevant and major addition to inclusive, sustainable communication in modern society.

6 Future Scope

Future spoken language translation will be enabled by expanded support for voice recognition and speech synthesis making platform accessibility higher for visually impaired users. Real-time subtitle generation for videos and live streams boosts inclusivity significantly for non-native speakers and people with hearing impairments. Adding audio-visual features can further increase platform usefulness in public communication education and emergency response situations effectively. Crowdsourced datasets and transfer learning will boost digital inclusion remarkably for tribal languages and low-resource ones preserving linguistic diversity worldwide effectively. Enhancements are being implemented globally for individuals from diverse backgrounds making system usage remarkably easy and fairly straightforward overall.

References

1. Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N. et al. (2017). Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. *Transactions of the Association for Computational Linguistics*, 5, 339–351. Retrieved from https://arxiv.org/abs/1611.04558
2. Conneau, A., Lample, G., Ranzato, M., Denoyer, L., Jégou, H. (2020). Unsupervised Cross-lingual Representation Learning at Scale. In Proceedings of ACL. Retrieved from https://arxiv.org/abs/1911.02116
3. Arivazhagan, N., Bapna, A., Firat, O., Aharoni, R., Johnson, M., Macherey, W. (2019). Massively Multilingual Neural Machine Translation in the Wild: Findings and Challenges. arXiv preprint arXiv:1907.05019. Retrieved from https://arxiv.org/abs/1907.05019
4. Bawden, R., Sennrich, R., Birch, A., Haddow, B. (2018). Evaluating Discourse Phenomena in Neural Machine Translation. In Proceedings of NAACL-HLT. Retrieved from https://arxiv.org/abs/1711.00513
5. Paul, B., Rudrapal, D., Chakma, K., Jamatia, A. (2024). Multimodal Machine Translation Approaches for Indian Languages: A Comprehensive Survey. Journal of Universal Computer Science, 30(5), 694–717. https://doi.org/10.3897/jucs.109227
6. Fu, Z., Chen, Y., Liao, M., Yu, D., Fan, K., Li, C., Shi, X. (2024). Image-Text Multilingual Pretraining for Scene Text Translation. arXiv preprint arXiv:2503.04797. Retrieved from https://arxiv.org/abs/2503.04797
7. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A. (2019). Self-Attention Generative Adversarial Networks. arXiv preprint arXiv:1805.08318. Retrieved from https://arxiv.org/abs/1805.08318

8. Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., Sutskever, I. (2021). Zero-Shot Textto-Image Generation. arXiv preprint arXiv:2102.12092v2. Retrieved from https://arxiv.org/abs/2102.12092v2
9. Matviienko, V., Ashraf, A., Chatterjee, P. (2024). An Evaluation of Online Translation Services for LowResource Languages. arXiv preprint arXiv:2412.04351v2. Retrieved from https://arxiv.org/abs/2412.04351
10. Sharma, P., Bansal, R. (2023). Image Generation: A Review. International Journal of Computer Applications, 975(8887), 1–6.
11. Chaudhari, P., Kamat, M. (2023). Offline Translator Using Multilingual Neural Networks. International Research Journal of Modernization in Engineering Technology and Science, 5(4), 103–108
12. Patil, S., Deshmukh, K. (2023). Hybrid Translation Systems for Indian Languages: Integration of Rule-Based and Neural Approaches. International Journal of Computing and Digital Systems, 9(1), 75–80.
13. Patel, V. (2024). Multilingual NLP for Indian Legal Documents. arXiv preprint arXiv:2504.09570. Retrieved from https://arxiv.org/abs/2504.09570
14. Tiwari, D. (2024). GANs for Indic Language Text Image Generation. arXiv preprint arXiv:2504.11809. Retrieved from https://arxiv.org/abs/2504.11809
15. Kumar, A., Pratap, A., Singh, A. K. (2023). Exploiting Multilingualism in Low-resource Neural Machine Translation via Adversarial Learning. arXiv preprint arXiv:2303.18011. Retrieved from https://arxiv.org/abs/2303.18011