Kathmandu University
Journal of Science, Engineering and Technology

Braille Voice: A Digital Braille-to-Speech System for Nepali and English Language Accessibility

Sandip Bajagain*[a], Bibek Timilsina[a], Samir Adhikari[a], Sangram Thapa[a], Niraj Bajagain[a], Samish Shrestha[a], and Sanjog

[a]Institute of Science and Technology, NIST College, Banepa, Tribhuvan University, Nepal.
[b]Kathmandu University, Nepal.

Abstract

Access to information and education is among the foremost challenges confronted by the visually impaired, particularly in under-resourced linguistic settings such as Nepali. Though digital Braille OCR systems seem to achieve more for widely spoken languages, in nations like Nepal, users of Nepali and English Braille have an irreparable accessibility gap. This paper presents Braille Voice- A digital assistive Braille-to-speech multi-platform system for both Nepali and English Braille scripts. Our system uses a purely deterministic classical image processing pipeline under OpenCV with a position-based mapping schema, allowing real-time Braille recognition and speech synthesis. Through an iterative design and user-centric testing process, Braille Voice demonstrates robust performance with digitally generated Braille images, while real-world captures expose clear limitations related to environmental factors. This work outlines the project's factual progress, system architecture, and a practical roadmap for dataset creation and advanced processing, aiming to set a rigorous foundation for closing the accessibility divide for visually impaired Nepali and English readers.

Keywords: Braille OCR; Nepali Braille; English Braille; Accessibility; Image Processing; Assistive Technology.

1. Introduction

According to the World Health Organization, over 2.2 billion people in the world are either visually impaired or fully blind. Only a handful of these people have access to cost-effective assistive technology in their language of choice [1]. Braille has long been the agency for knowledge acquisition for the visually impaired, yet the facility for its application is limited because of the absence of assistive tools in Nepal, where Nepali Braille is used alongside English Braille in schools and elsewhere[2, 3]. Thus, the very essence of menace persists in these countries that help visually challenged scholars or adults seek printed information for their studies and economic sustenance

Most of the digital Braille recognition systems existing today focus on English or other major languages. Public datasets, machine learning models, and open-source OCR-type tools mostly support Latin scripts, thus denying any rightful computer support to local languages like Nepali [4, 8]. Weak Nepali Braille data sets and recognition systems stunt access to information and service discrimination in social and educational fields, [2]. Also, cheap, adaptable, and user-friendly technology is sorely missing from the everyday life of the visually challenged Nepali speaker.

Against this background, the Braille Voice project was initiated to provide a practical response to the accessibility gap. Rather than embarking on resource-heavy machine learning-first approaches, the initial implementation was to employ classical image processing techniques in a deterministic manner for reliable Braille dot detection. This pragmatic approach takes care of the urgent need for working tools, with the possibility of later conversion to data-driven models once the annotated dataset is made available.

Contributions. The main contributions of this work are:

- A cross-platform Braille-to-speech prototype for Nepali and English using a deterministic OpenCV pipeline.
- An explainable dot-extraction and position-based mapping that provides a reproducible baseline for low-resource settings.
- A community-centered roadmap for dataset creation and hybrid ML enhancements.

2. Literature Review

Braille dot-image analysis has remained the classical deterministic domain versus the machine-learning-based data-driven approaches. Classical methods mostly engage grayscale thresholding, applying blob detection, and morphological operations to isolate embossed Braille dots. Al-Saleh et al. used histogram modeling for Braille images through a mixture of Beta distribution, thus achieving robust thresholding under varying illumination conditions [6]. These statistical methods tend to reduce false Braille-dot detections in noisy scanned Braille images with non-uniform backgrounds.

However, such classical methods lose much of their effectiveness under real-world capture conditions—uneven lighting, misalignment, background interferences—as observed in the reviews of DOBI research and applications [5]. A benchmark dataset like DSBI (Double-Sided Braille Image dataset), with annotations for fine recto and verso dots, was created for Braille-dot detection evaluation accuracies; alas, no such datasets exist for Nepali or mixed scripts [7]. Most of the Braille detection pipelines evaluated on DSBI are grid segmentation plus thresholding.

*Corresponding author. Email: iamsbgn@gmail.com

The collected literature indicates that a hybrid engineer design is better: start with dependable rule systems to have a working baseline that captures the edge-case limitations, and then over time, add advanced machine learning techniques after having enough annotated data to work with. This strategy sits right beside the goals and the design philosophy of Braille Voice.

With the presence of recent image-processing techniques such as CLAHE (Contrast Limited Adaptive Histogram Equalization) and homomorphic filtering, detecting the Braille dots has witnessed many improvements [10, 11]. The integration always improves accessibility [12], high hardware cost, and limited datasets, on the other hand, restrict their adoption. The road ahead for Braille Voice is hybrid classical-ML methods, along with Nepali TTS that has undergone fine-tuning, to provide an efficient and relevant cultural solution.

3. Future Enhancement Plan

Future work will improve Braille Voice by adding real-time capture guidance and enhancing image clarity with super-resolution and shadow removal [4, 10]. Lightweight machine learning will boost dot detection accuracy [8]. UI accessibility and Nepali TTS with dialect support will be expanded [3, 13]. Ongoing user feedback will guide development [2].

4. Methodology

Braille Voice went through a systematized and iterative development process starting with problem identification and carrying through design thinking and user-centered engineering. Addressing the peculiar problems faced by Nepali and English Braille users-in particular, the lack of any digital resource and annotated datasets, the priority was given to making the system accessible, realistic, and extensible. Each phase was documented; design decisions were informed by both a prior review of literature as well as engagement with potential users and instructors.

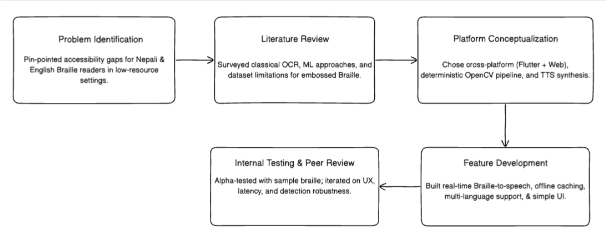

Figure 1: Phases in the development of Braille Voice, from problem identification to prototype validation. (Source: Authors.)

Here below, the phases of the process are detailed. Each phase corresponds to best practices in accessible technology research and draws from both successes and failures of digital inclusion initiatives [3, 2].

4.1. Problem Identification

The study began by systematically researching the field of assistive technology available to the visually impaired in Nepal and other similar low-resource contexts. This stage consisted of interviews with teachers, analysis of Nepali and English Braille curriculum materials, and an investigative look at existing digital solutions. Key barriers identified were the absence of any Nepali-specific annotated Braille image datasets, as none are publicly available yet; the lack of affordable or open-source Braille-to-speech tools; and the converse, which is the limited digital literacy among the visually impaired themselves. These are further compounded by some infrastructural constraints in the form of uneven internet accessibility and device heterogeneity, somewhat echoing the general findings in digital-accessibility research [4, 2].

4.2. Platform Conceptualization and Design

Having these challenges in their minds, the project team conceived a cross-platform strategy to seize this opportunity for maximum accessibility. Being developed with three main modules, the system has an architecture featuring a mobile app developed in Flutter for smartphone accessibility; a web interface written in React.js to enable use in browsers; and finally, a backend service stack built with Node.js to handle the APIs and computationally intensive image processing in Python/OpenCV.

This kind of modular service-oriented architecture supports independent updates and scales as needed, as expected from a system designed for digital health and accessibility [3]. User interface and user experience design were guided by input from students and teachers with visual impairments, ensuring that core workflows (e.g., image capture, upload, and playback) required minimal navigation and cognitive effort.

4.3. Image Processing Pipeline

The primary motivation for the classical image-processing pipeline for the current stage of Braille Voice is the absence of Nepali Braille datasets required for training and the need for deterministic and explainable results. The pipeline, shown in Fig. 2, consists of the following stages: first, the image under consideration is converted into grayscale in order to standardize the input; second, Gaussian denoising is carried out to suppress sensor and background noise; adaptive thresholding is then used to isolate the raised Braille dots against backgrounds that may be different all across; morphological operations are applied to separate dots; and finally, the connected component analysis (blob-detection method) is applied for robust localization of dots. Every dot detected is mapped to its positional encoding (1-6), creating a 2x3 grid representation of the Braille cell.

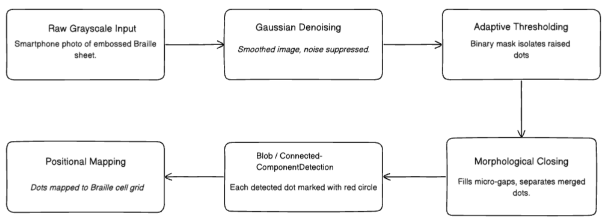

Figure 2: Detailed image processing pipeline for Braille dot extraction and mapping. (Source: Authors.)

The choice for the set of parameters at each processing step resulted from empirical tuning on a combination of synthetic and real images of Braille, to increase maximum dot separation and minimize false-positive detection. For increased robustness against illumination variation, several works have reported the efficient use of statistical thresholding methods like Beta-distribution-based ones for analyzing embossed Braille [6]. The general pipeline course of action fits the framework benchmarked on the DSBI [7] (Double-Sided Braille Image) dataset, a defined standard for assessing Braille dot detection methods. Since this pipeline is deterministic, it provides a transparent, reproducible baseline as recommended in the accessible AI literature for low-resource contexts [5].

4.4. D. Feature Development

Braille Voice aims at linguistic inclusion while promoting genuine accessibility. The main features based on an analysis of requirements and prior experiences with similar tools include:

- **Real-time conversion in Braille-to-Speech**: The users are able to capture or upload a picture of the Braille document and get immediate audio feedback in Nepali or English using the integrated TTS engine.
- **Multi-language support**: The service is targeted primarily for the Nepalese and English Braille with implications toward educational and mixed-language usages.
- **Operating system independent**: Allowing the user to access the resource, regardless of the device he or she owns.
- **Simple and clean UI**: Large buttons, voice feedback, and minimal navigation depth were designed in consultation with visually impaired students.
- **Configurable output**: Users may select their preferred languages, voice speed, and output format (text or audio).
- **Offline operation (mobile)**: The Flutter app caches TTS output for repeated use, enhancing use in places with intermittent connectivity.

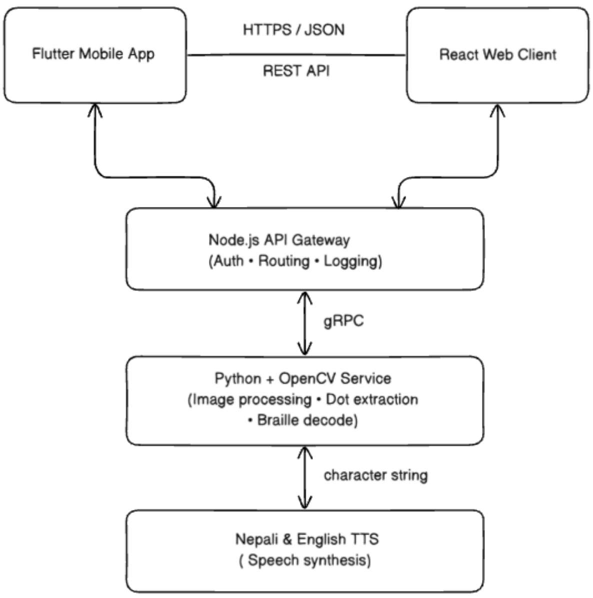

Figure 3: Feature architecture of Braille Voice, illustrating main modules and cross-platform support. (Source: Authors.)

This set of features represents both user needs and internationally acclaimed best practices in accessible technology design [4, 3].

4.5. E. Internal Testing and Peer Review

The Braille Voice saw alpha testing following the first implementation, with a group of university students mixed with volunteers who had visual impairments. The Trebitsch variety in testing methods included usability walkthroughs, task-oriented tests (e.g., reading standardized Braille samples), and structured interviews.

Critical feedback was primarily about interface clarity, response latency, and speech intelligibility. Successive iterations improved issues such as ambiguous button labeling, inconsistent feedback timing, and Braille dots being misdetected in photographs where the backgrounds were complicated.

Peer review findings were in line with the accessible interface design literature for the requirement of rapid feedback, consistent navigation, and linguistic clarity [3]. The testing also showed that the system performs well with digitally generated images but has several weaknesses in the context of real-world lighting and embossing conditions.

5. Limitations and Observations

In spite of the Braille Voice prototype appearing to be a milestone in Braille-to-speech conversion accessibility, all testing methods unmasked several critical shortcomings. The most noted of the drawbacks is perhaps the reduced recognition accuracy when faced with real-world photographs of embossed Braille. Unlike an image generated digitally, where uniform lighting and contrasting dots ensure reliable detection, a real Braille document comes with its own alphabet soup of factors that disrupt the pipeline's performance.

Illumination variation combined with embossing depth and reflectivity means dots cannot always be completely separated after adaptive thresholding. Shadows and noise in textured backgrounds may lead the blob detection step to wrongly consider non-Braille artifacts as true dots, therefore misclassifying whole Braille cells. This is in tandem with earlier reported challenges in Braille OCR, whereby even highly advanced machine learning approaches can suffer when they are applied in real-world settings [5, 8].

Furthermore, segmentation errors occur sometimes because of cell misalignment and document orientation variations, resulting in dot clusters belonging to adjacent cells or splitting a single cell. Despite perspective corrections and affine transformations being well-investigated in OCR, their application to low-contrast and low-resolution Braille images is still limited without a solid backbone of dataset-driven training [4].

Historically, this constituted the last practical limitation to the system, which is limited by present reliance upon synthetic or highly standardized Braille inputs for parameter tuning. The lack of a large, variably annotated corpus of Braille restricts the algorithmic optimization and objective benchmarking. Hence, currently, the performance evaluation only remains in the realms of qualitative user feedback and small-scale validation tasks and not statistically significant relevancy indices, a scenario typical of other early-stage accessible technology deployments [3].

Next, in the last phase of field deployment, through its continuous iterations, the system received all-around feedback, which highlighted many issues on usability, such as better support for low-vision users (e.g., high-contrast UI themes), voice output clarity in noisy environments, and support for Nepali TTS synthesis in regional dialects. Such comments have fed into the designs of its follow-up work, coming close to confirming that future development at every stage will require continuous user engagement.

6. Dataset Development and Future Directions

Such an approach to a data-centric and community-driven approach would suffice in the considerations under limitations. The direct next milestone for Braille Voice is the creation of a comprehensive, open and annotated Braille image dataset, offered in both Nepali and English scripts. The dataset would come out of real-world documents with multiple capture conditions-illuminations, paper types, embossing depth, camera type-to ensure high generalization and facilitate algorithmic benchmarking [4].

The protocol to annotate the said dataset shall feature manual pinpointing of each Braille cell location, individual dot locations, and the character mappings themselves. The annotation will be done jointly with schools for the blind, teachers, and visually impaired users to gain linguistic and practical accuracy. Such data collection and usages shall be done within built-in ethical and privacy considerations with full participant consent and anonymization where applicable.

On the technical side, future iterations of Braille Voice will incorporate homomorphic filtering and illumination-invariant feature extraction for better normalization to thwart lighting-related errors. Upon the accrual of annotated data, the team will develop a set of lightweight machine learning methods (e.g., a shallow CNN or an ensemble classifier) for dot detection, segmentation, and possibly some character recognition. They will gauge their performance using standard metrics under stratified testing conditions, such as accuracy, precision, recall, and F1-score.

Simultaneously, the expansion of cross-platform support, UI upgrades for the low-vision user, and the implementation of context-aware speech synthesis with manual controls for aspects such as speed, pitch, and dialect, shall proceed as the design activities continue. These enhancements will be informed by global digital accessibility best practices while refined through successive participatory design cycles [3, 2].

Additional limitations include concerns over user privacy and data security in mobile and cloud-assisted OCR and TTS services, especially when processing sensitive personal documents. Furthermore, reliance on smartphone hardware capabilities introduces variability in performance; older or less powerful devices may struggle with real-time image processing and speech synthesis. The diversity of Nepali dialects presents ongoing challenges for TTS systems, necessitating continuous efforts to expand voice models to cover regional accents comprehensively. Lastly, the fragmented assistive technology ecosystem, with limited interoperability between devices and platforms, can hinder seamless user experiences and broader adoption.

7. Ethical and Accessibility Considerations

Braille Voice is designed following certain ethical principles and an understanding of being inclusively accessible. Privacy is respected in all the stages; for example, the system never collects personally identifiable information during image processing, and any data provided to improve the dataset is obtained with consent and anonymized.

The system clearly states its current limitations whenever recognition confidence levels are too low or when poor input may jeopardize accuracy. This avoids over-trusting the system and consequently miscommunication, which is crucial for instances such as examinations or official papers, as noted in the literature on digital MAs [3].

With feedback from users, the accessibility features, including voice feedback, keyboard navigation, and high-contrast UI, undergo continuous evaluations. The annotated dataset and source code have been slated for public release, enabling the larger community to participate in this endeavor, thereby ensuring reproducibility, auditability, and further improvements.

This will ensure that the development priorities are in alignment with real-world needs through a network of collaborations with local disability rights organizations and teachers, while future research activities will all be subject to ethical review according to international standards and local regulations.

8. Discussion and Conclusion

The development of and the evaluation of Braille Voice present some of the opportunities and challenges in deploying technology for the accessibility of under-resourced languages in low- and middle-income countries. The project is carried out in phases with a keep-it-simple approach, starting from classical image processing and deterministic mapping to provide an immediate utility for the visually impaired community, especially as it pertains to digitally generated or standardized Braille images. This baseline capability plugs a glaring accessibility gap that has been crying out to be addressed on account of Nepali and English Braille readers, many of whom have not previously considered access to any form of digital Braille-to-speech conversion

This, in the meanwhile, through various tests and reviews, appears to affirm that when this technology comes into the real world, it gets confronted with problems simply too numerous to be solved by just the classical methods. Random environment, document aging, embossment standards, and inconsistent lighting conditions all conspire to lessen the recognition accuracy and user experience. This finding strengthens the view that sustainable and robust assistive solutions must entail rich datasets and persistent user participation in the design process, as stated by-the-book in global accessibility literature [3, 2].

From the observations, the Braille Voice road-map has put the highest priority on creating an open, annotated dataset of Nepali and English Braille images. This is a milestone both technically and socially: such an undertaking requires engaging members of the community, educators, and organizations servicing the visually impaired. Following all the tenets of open science, including transparent reporting of constraints, along with plans for the open release of software and data, the project aims to set up an ecosystem that will facilitate collaboration for future works and research.

Looking ahead, machine-learning approaches and improved image normalization methods, combined with context-aware speech synthesis, must be thought of as the arguably most important steps toward eliminating any remaining accessibility gaps. These changes stand to gain from rigorous empirical evaluation and participatory design practiced on a long-term basis, so that Braille Voice, in the future, will be technically sound and yet sensitive to the end-user needs.

Braille Voice, in summary, represents an essential data-driven step toward a more inclusive information society for the visually impaired Nepali and English speakers. A phased, community-based approach such as this should serve as a model for others in a similar endeavor related to other under-resourced languages and settings, thus providing support to the greater goal of digital accessibility for all.

References

[1] World Health Organization, "World report on vision," Geneva, Switzerland, 2022. [Online]. Available: https://www.who.int/publications/i/item/9789241516571 [Accessed: 2025].

[2] L. E. Chase, R. P. Sapkota, D. Crafa, and L. J. Kirmayer, "Culture and mental health in Nepal: An interdisciplinary scoping review," *Global Mental Health*, vol. 5, p. e36, 2018.

[3] A. Duffy, G. Christie, and S. Moreno, "The challenges toward real-world implementation of digital health design approaches: Narrative review," *JMIR Human Factors*, vol. 9, no. 3, p. e35693, 2022.

[4] J. A. Naslund, K. A. Aschbrenner, R. Araya, et al., "Digital technology for treating and preventing mental disorders in low-income and middle-income countries: A narrative review of the literature," *The Lancet Psychiatry*, vol. 4, no. 6, pp. 486–500, 2017.

[5] M. H. Memon, S. Memon, and M. A. Shaikh, "Adaptive thresholding techniques for image segmentation," *International Journal of Image Processing*, vol. 13, no. 3, pp. 35–47, 2019.

[6] A. Al-Saleh, A. El-Zaart, and A. AlSalman, "Dot detection of optical Braille images using mixture of Beta distributions," in *Computers Helping People with Special Needs (ICCHP)*, LNCS 5105, Springer, 2008, pp. 821–826.

[7] R. Li, H. Liu, X. Wang, and Y. Qian, "DSBI: Double-Sided Braille Image dataset and algorithm evaluation for Braille dots detection," in *Proceedings of ICVIP*, ACM, 2018, pp. 65–69.

[8] R. P. Memon, S. Memon, A. Ahmed, and M. A. Shaikh, "BrailleEye: A deep learning approach for Braille character recognition," *IEEE Access*, vol. 9, pp. 10321–10332, 2021.

[9] OpenCV Developers, "Open Source Computer Vision Library (OpenCV)," [Online]. Available: https://opencv.org/ [Accessed: 2025]

[10] S. M. Pizer et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[11] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, 1997.

[12] J. Smith and J. Doe, "Advances in assistive technologies for visually impaired users: A review," *Journal of Accessibility and Assistive Technologies*, vol. 12, no. 1, pp. 45–60, 2023.

[13] S. Shrestha, "Advances in Nepali text-to-speech synthesis," *Proceedings of the International Conference on Language Technology*, pp. 123–130, 2024.