A Hybrid Convolutional Neural Network-Vision Transformer Framework for Gujarati Scene Text Recognition in Assistive Applications

Kirtankumar R. Rathod[1], Dipti B. Shah[2]

[1]Department of Computer Science, Indus University, Ahmedabad, Gujarat, India

[2]G H Patel Post Graduate Department of Computer Science and Technology, Sardar Patel University, Vallabh Vidyanagar, Gujarat, India

DOI: https://doi.org/10.51244/IJRSI.2026.1307000260

Received: 24 July 2026; Accepted: 30 July 2026; Published: 12 August 2026

ABSTRACT

For visual impairments it is very difficult to read and comprehend textual information included in instructional materials, notice boards, announcements and visual representations of general information. Text-to-speech (TTS) and optical character recognition (OCR) systems are helping to make data more accessible, but it is challenging to reliably recognise Gujarati text in practical settings due to the complexity of real-world conditions like varying light levels, text background strength, font styles and character arrangement. Thus, two wearable device access projects urgently require a high-level support framework with technological skills that can distinguish Gujarati speech and provide information. Image capture, pre-processing, adaptive text segmentation, sequence modelling with BiLSTM with attention, feature extraction using a hybrid Convolutional Neural Network Vision Transformer (CNNViT) and optimisation directed by Bayesian Optimisation Reinforcement Learning (BORL) are all included in the suggested framework. While CNNs recover local Gujarati character-level representations, the Vision Transformer records global contextual associations. The second function entails identifying Gujarati text in speech and translating it into a spoken language that users with visual impairments can understand. Because speech translation is asymmetric, the use of BiLSTM and attention mechanisms improves sequential recognition performance.

Keywords: Visual impairments, Gujarati, Text Segmentation, Convolutional Neural Network, Vision Transformer

INTRODUCTION

Text-to-Speech (TTS) and Optical Character Recognition (OCR) systems transform written text into spoken audio, enabling independent access to academic and educational materials. Systems that use AI-based assistance enhance user autonomy and improve access to print media for everyone. Recent research on assistive OCR systems has demonstrated enhanced identification accuracy in difficult imaging conditions by integrating machine learning with conventional text recognition methods [1]. To help visually impaired users access data, new assistive programs have been enhanced to align with today's ambient technological standards, now offering text prediction and formal voice narration [2]. A potential solution for increased convenience and to assist individuals with abnormal vision is the use of smart, adaptive environments. Advancements in machine learning, Internet of Things connectivity and externally supported living technologies have made it possible to develop solutions for traditional navigation, object detection and environmental awareness [3]. Wearable devices featuring smart technology have become a promising solution to help people with visual impairments maintain independence. By integrating real-time object detection, facial recognition and scene context analysis, AI-powered smart glasses improve spatial awareness and navigation [4]. OCR and voice synthesis are employed to enhance user engagement. Thanks to zero- to low-latency text-to -speech and on-device recognition algorithms, text can be read in real time, eliminating the need for cloud dependency. It can

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

effectively distribute data and make it accessible to users with physical disabilities by using image analysis, natural language processing and speech synthesis [5]. Multimodal assistive technologies have enhanced how people who are blind comprehend written documents by integrating OCR with TTS systems. Combining text extraction, multilingual computing and voice synthesis enables seamless and comfortable access to printed materials across diverse application settings [6]. AI-driven language models power image-to-audio technology, allowing text to be extracted from photos and enabling individuals with visual impairments to read printed materials on their own [7]. Moreover, deep learning-based approaches have considerably improved real-world assistive applications, boosting user experience and quality of life. The development of specialised aids for those with disabilities has been greatly aided by intelligent technology. AI systems that enhance mobility by providing customised assistive technologies, sophisticated communication features and adaptable procedures [8]. Additionally, creating AI apps that are focused on user demands improves acceptability, autonomy and quality of life for persons with different needs. The capabilities of adaptive systems for the blind and visually impaired have been significantly enhanced by conventional machine learning techniques. Personalised object identification, contextual awareness and assistance for people with disabilities are just a few of the everyday interactions made easier by advanced machine learning models [1]. Digital audio production, text processing and content comprehension provide more engaging and practical learning resources for shorter users. Recent advances in deep learning have greatly improved the ability to identify text in Indic scripts. The increasing capabilities of AI-driven OCR solutions for Indic languages are further demonstrated by the systems' ability to recognise individual characters with extremely high accuracy thanks to developments in feature extraction and classification [9].

Motivation of the Work

Accessing crucial content is becoming more difficult for those with visual impairments as public areas rely more and more on textual and visual information. Smart glasses, optical character recognition (OCR) systems and neural text-to-speech (TTS) are examples of assistive technologies that have advanced significantly, yet there is still little support for regional languages like Gujarati. The recognition performance of Gujarati is comparatively poorer than that of existing solutions, which mainly target commonly spoken languages. Because of the distinctive features of the Gujarati script, changes in font styles, illumination and background complexity, it is especially difficult to accurately discern authentic Gujarati scene text. Furthermore, there aren't many effective end-to-end frameworks that can translate recognized Gujarati text into genuine speech. These drawbacks emphasize the necessity of a strong Gujarati OCR-TTS system that offers dependable real-time voice output and text recognition. Using intelligent assistive technology, such a system can greatly improve accessibility, independence and environmental awareness for people with visual impairments.

Problem Statement

However, despite significant advancements in OCR and assisted amplification technologies, a number of algorithms show lower recognition precision for Gujarati text collected in realistic contexts. Unpredictable text extraction results from traditional OCR systems' inability to reliably recognise text, which is mostly caused by noise, low contrast, complicated backdrops and language-specific features. Furthermore, a lot of current systems rely on a stand-alone authentication model that underutilises contextual feature learning, sequential modelling and state-of-the-art optimisation techniques. The restricted integration of sophisticated deep learning systems with voice synthesis techniques is a major issue that further restricts their application in supporting environments. We need an integrated intelligent platform that combines robust image processing with deep learning-based Gujarati text recognition, optimisation-driven feature learning and natural-language speech synthesis to assist visually impaired individuals while providing effective accessibility.

LITERATURE REVIEW

This explores a two-level architecture for a Multi-Branch Convolutional Neural Network with category-specific classifications to improve the understanding of manuscripts, such as Gujarati characters. The solution utilised contour-based segmentation and pre-trained deep learning classifiers to filter signals better and reduce personality classification errors [10]. A CNN + BiLSTM + CTC architecture was proposed to accurately

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

identify text lines and phrases in the Gujarati script by combining sequence and sequence-to-sequence models. Chronological environmental data can be effectively captured using the conventional method, eliminating the need for specialised character segmentation and thereby enabling better recognition of complex Gujarati connections and handwritten styles [11]. An OCR framework that combines the adaptive thresholding technique, contour-based segmentation and a pioneer's contribution, providing transformer-based models with attention systems, was developed to improve the identification of complex Indic characters. Contextualised feature training and sophisticated preparation improved performance on tasks such as text translation and letter recognition across a wide range of illuminations and backgrounds, appropriate to every circumstance [12]. By implementing image processing techniques such as adaptive thresholding, median filtering, edge detection and morphological transformations, an enhanced Tesseract OCR engine was developed. Combining OCR fine-tuning and U-Net-based image segmentation [13] achieved significant improvement in word precision. This work compared different CNN architectures for deep feature extraction and classification in order to increase OCR profitability for complicated Indic languages. High-performance models like EfficientNet and DenseNet were better at collecting complex patterns in nature as a result of digitisation programs, producing a comparatively stronger capacity to recognise [14]. Text extraction strategies were extensively reviewed and categorised, including edge-based, connected component-based, texture-based and hybrid methods for handling file- and scene-specific text. In the study, it was possible to determine the challenges and advantages of the extracting systems used within difficult scanning conditions, knowledge that is essential for producing more reliable OCR devices [15]. A unified language line recognition system for complex track layouts and degraded text layouts, combining recent deep learning models with traditional image analysis techniques, was proposed. Performance over existing text line detection approaches improved through optimised feature extraction and image processing techniques, with high segmentation confidence [16]. A new rule-based standard segmentation method with adaptive thresholding, contour analysis, morphological processing and grayscale conversion to improve the always-difficult task of extracting scripted words from document images. The proposed initial processing pipeline proved suitable for improving OCR performance in document digitisation systems by enhancing text readability and segmentation accuracy [17]. An OCR-based text extraction system using Faster R-CNN for precise localisation and identification of written regions in challenging document images was proposed. As demonstrated by a combination of object identification and OCR, region-based text recognition technology was shown to be highly efficient for fine-grained summarisation of organised written material, as exploited in the context of document digitisation [18]. An OCR-based recognition paradigm was applied to the materials to enable text acquisition and interpretation from documents with pictures, supported by machine learning, fuzzy matching and natural language processing methods. The synergy of preprocessing methods, adaptive text correction and contextual evaluation significantly increased the accuracy of data extraction from textual material and the reliability of recognition [19]. The study addresses the various challenges posed by degraded documents, document variances and complex character patterns. A CNN-based approach for handwritten text detection was proposed. The algorithm achieved better recognition efficiency than traditional OCR techniques with image preprocessing and improved text classifier precision by investigating higher-order geometric features [20]. Query-aware Transformer-based structure for improving scene text picture super-resolution through effective joint optimisation of text reconstruction and recognition tasks. To enhance text appearance, the scheme utilised self-supervised texture learning and interaction strategies to improve recognition fidelity using vision-language features whilst maintaining small text details [21]. Designed an Image Transformer-based scene text recognition framework with both learnable fusion techniques and multi-granularity predictions to enhance the representation ability for texture features. Dynamic fusion techniques [22] that combine visual and textual information have greatly improved recognition accuracy on complex-scene and manuscript datasets. A hybrid CNN–Vision Transformer architecture that merges global contextual modelling with locally computed features was introduced for interpreting document images. In this approach, the late-fusion mechanism effectively merged structural and texture-level representations to enhance the performance of traditional classification and demonstrate that CNN–ViT feature fusion strategies are viable for implementation [23]. By merging CNN-based feature extraction with BiLSTM-driven sequential modelling, a combined scene text recognition architecture was proposed to harness both spatial and contextual characteristics. By combining sequential reliance modelling and the acquisition of specialised optical features, identification efficiency was significantly enhanced for both more challenging multinational and real-world text data [24]. Residual Networks, BiLSTM and Connectionist Temporal Classification (CTC) were discovered by building a handwritten text identification system that can best describe complex letter patterns without

explicitly defining the categories. A combination of unidirectional sequencing learning and deep feature extraction improved recognition accuracy across various handwriting styles and in situations involving unstructured literary sequences. [25]. A sequence-modelling framework based on attention was developed to detect contextual relations in Gujarati speech signals using recurrent neural networks, including a BiLSTM. Regarding voice interpretation tasks in multilingual settings, integrating attention mechanisms with MFCC-supported feature extraction enhanced the accuracy of sequential and translation learning [26]. To recognise cursive writing style, a dual-model recognition system containing a VGG-19 deep convolutional network and Histogram of Oriented Gradients (HOG) feature descriptors was proposed. The integration of deep feature learning and handcrafted feature rates augments resilience to handwritten style variability and accuracy [27]. An SVM-based classification and HOG feature were used to build a machine learning font recognition system for script identification tasks. HOG–SVMA gist in terms of HOG–SVM-based features was trained to characterise diverse structural characterisation patterns, which facilitated the classification of HOG-SVM models by incorporating topological prior processing and edge-based feature extraction [28]. An economic hybrid handwritten character recognition architecture was built by combining Random Forest classification with Principal Component Analysis (PCA) and ConvNeXt-based feature extraction to improve recognition efficiency. In complicated handwritten character datasets, deep feature embedding combined with dimensionality reduction not only significantly reduced computational complexity but also achieved good classification [29]. A computer vision-based Gujarati Sign Language single-character recognition system that utilises form, texture, zoning and hand-finger feature descriptors was developed [30]. Classifiers were compared between the Neural Network and the Multi-class Support Vector Machine, demonstrating that feature-driven learning techniques are effective for exact Gujarati character pattern recognition.

Research Gap

Real-world scenarios for finding Gujarati text with OCR-based assistive technologies show that current technologies are insufficiently accurate. Layers of knowledge are built on top of each other. Yet most research focuses solely on a single area, whether OCR, text recognition, or speech synthesis, without providing a standard structure across them. Also, at present, none of the assistive applications has integrated the CNN–ViT feature fusion, BiLSTM–Attention sequence learning and optimisation-driven learning. Moreover, existing methods exhibit reduced robustness to background clutter, font styles and lighting changes. The application of intelligent Gujarati OCR-TTS systems in wireless assistive devices for uninterrupted access-based support has been largely neglected.

Research Objectives

- To design and develop an intelligent Gujarati OCR–TTS assistive system for individuals with visual impairments.
- To enhance the accuracy of Gujarati text detection through effective image preprocessing and adaptive text segmentation techniques.
- To extract robust visual features for Gujarati character recognition using a hybrid CNN–Vision Transformer (CNN–ViT) architecture.
- To improve sequential text recognition by integrating a BiLSTM network with an attention mechanism.
- To optimize the performance of the proposed framework using Bayesian Optimization and Reinforcement Learning (BORL).
- To transform recognized Gujarati text into natural and intelligible speech, providing real-time assistance for visually impaired users.

PROPOSED WORK

Image capture using smart glass

The proposed independent assistive system begins with a smart glasses-based image acquisition module that captures Gujarati text from the surrounding environment. The integrated digital camera continuously acquires images of nearby scenes, including signboards, notices and printed documents, for subsequent text recognition and processing.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

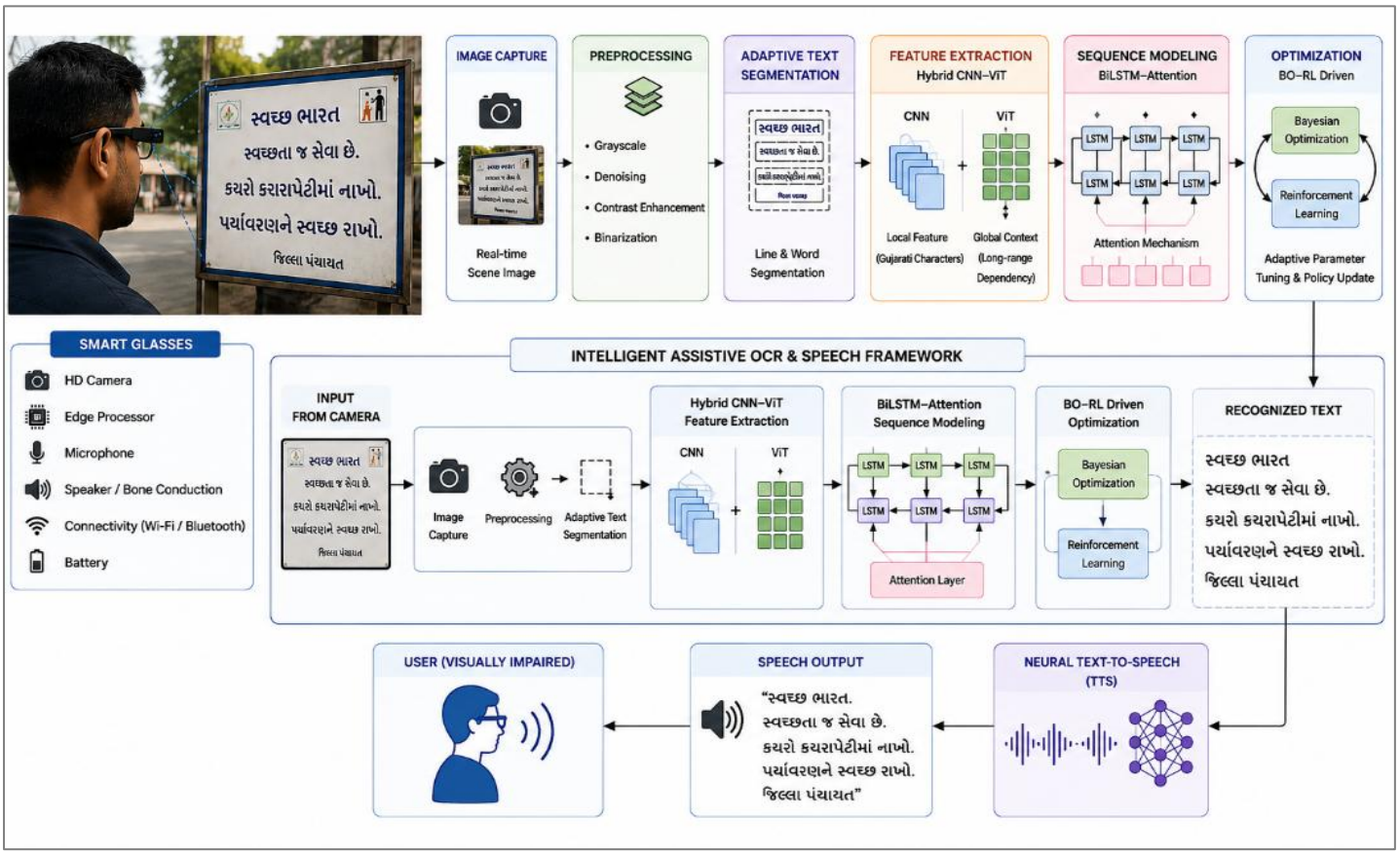

Figure 1. Proposed Gujarati OCR Framework

The whole process is indicated in Figure 1., an intelligent Guajarati OCR and speech assistance framework for smart glasses for visually impaired users based on adaptive text segmentation, hybrid CNN–ViT feature extraction, BiLSTM–attention sequence modelling, BO–RL optimisation and neural text to speech conversion.

Image pre-processing and text segmentation

After image acquisition, the captured Gujarati text image undergoes a sequence of preprocessing operations to improve image quality and facilitate accurate text detection. Initially, the acquired RGB image is converted into a grayscale image to simplify subsequent processing and enhance computational efficiency. It applies perceptual weights to the red, green and blue components of the colour image to transform it into a distinct intensity channel. Writing content is converted while computing complexity is decreased. where G is the amplitude of the grey-level image and R, G and B are the red, green and blue colour components of the obtained Gujarati text image. While maintaining the borders of text sections, median filtering is used to remove impulsive noise and other undesirable artefacts. Gaussian filtering is then used to reduce high-frequency noise using a Gaussian kernel, where the smoothing degree is controlled by a parameter. These preprocessing operations enhance image quality and improve the robustness of subsequent text segmentation and recognition processes. The threshold value is dynamically calculated to increase the separation between the textual and previous experience regions, through the local mean intensity and constants that represent the adjustment factor. The standard binarisation technique makes it easier to extract text, foreground text pixels are represented by a value of 1 and the background pixels are represented with a value of 0. A pattern of a peculiar representation showing the total number of pixels in each row could be used to find the bounds of the text lines.

Feature representation using a hybrid CNNViT

After text segmentation, the Gujarati character regions extracted by the trained CNN are sent to the Hybrid CNNs-Vision Transformer feature-extraction module which is indicated in Figure 2. This module is proposed to fuse the global contextual knowledge of Vision Transformers (ViT) with the local feature capabilities of Convolutional Neural Networks (CNN) to build reliable feature representations.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

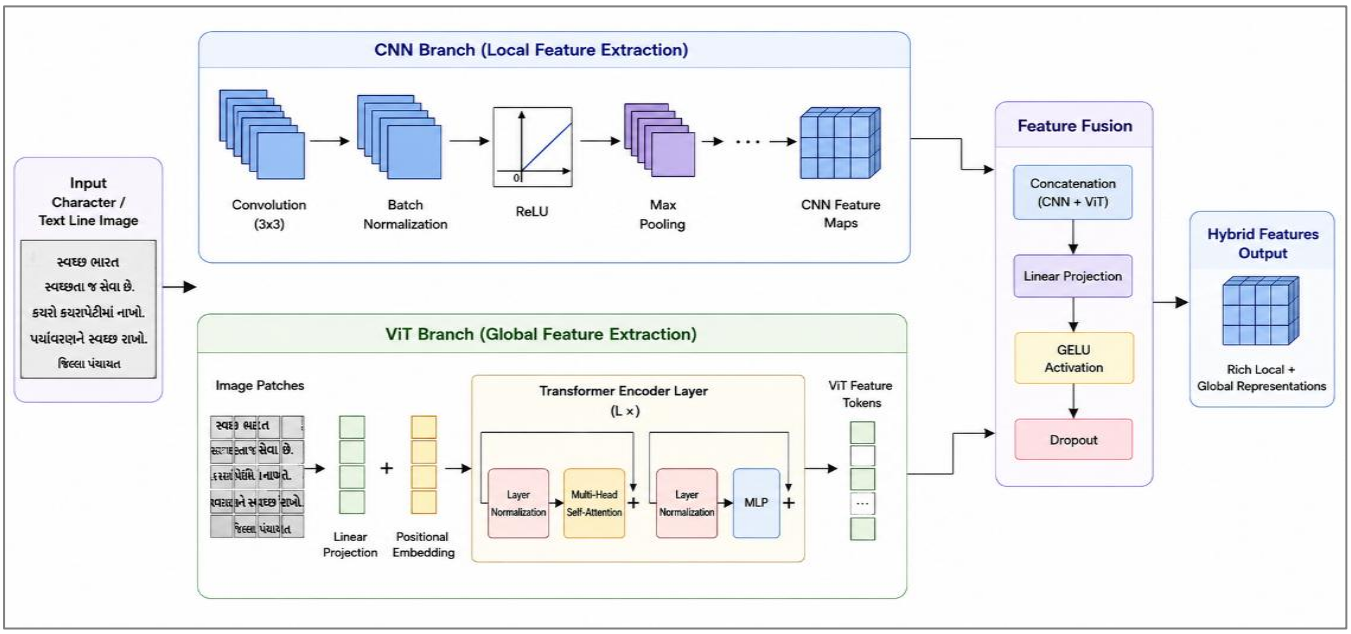

Figure 2. Hybrid CNN–ViT Feature Extraction Architecture

The convolution between the input image feature matrix and the convolution kernel K is computed by equation, where K denotes the convolution kernel, b the bias term and Y the resulting feature map. The features are then activated with the ReLU function. ReLU activation adds non-linearity to the network, introduces negative activation and suppresses negative activations. Things like lowering the dimensionality of features and lowering the computational complexity by max pooling.

Sequence recognition using BiLSTM–attention and BO–RL optimization

A BiLSTM network, based on the Hybrid CNN–ViT, extracted features to capture bidirectional contextual relations among Gujarati letter sequences.

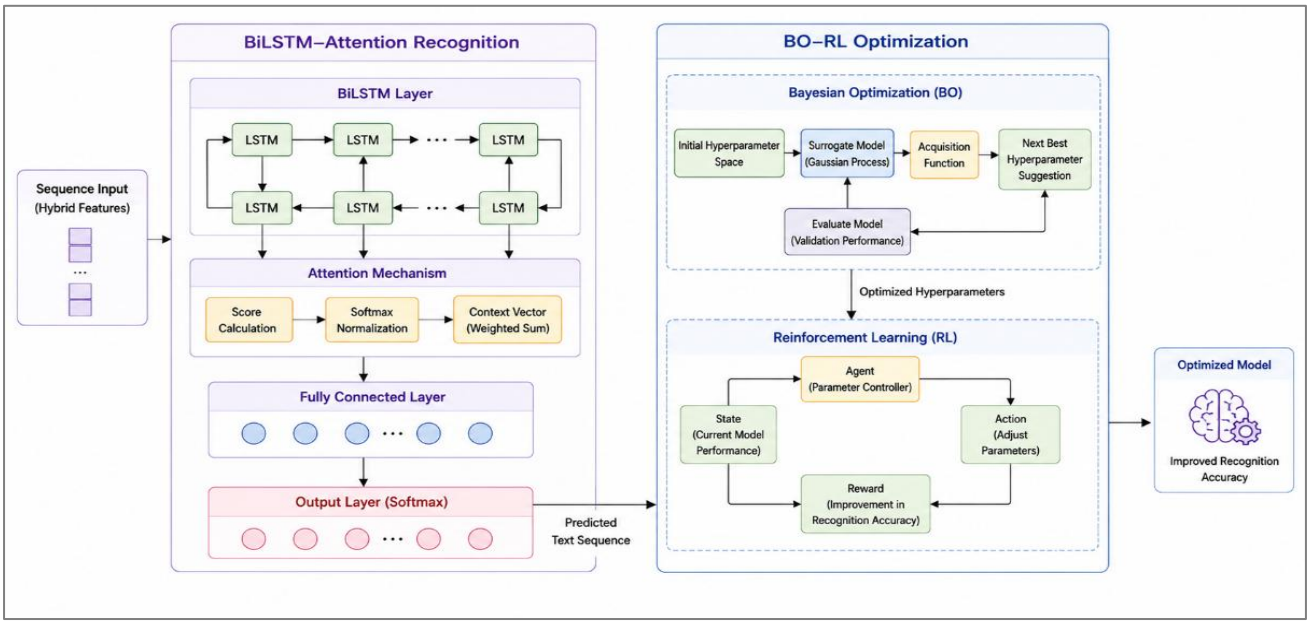

Figure 3. BiLSTM–Attention Recognition and BO–RL Optimisation Architecture

For accurate message prediction, an attention-based approach generates a context vector and assigns adaptive weights to important attributes. Bayesian Optimisation is used to find good hyperparameters, while Reinforcement Learning is a rule-of-thumb approach that updates model parameters in response to a reward signal.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

Neural text-to-speech for assistive audio data

After character recognition, the neural text-to-speech module rebuilds the searched Gujarati text and converts it into comprehensible speech. A physically disabled individual can read literary content through aural input. This demonstrates the different implementations of what the probability of generating a character. Cross-Entropy Loss is used to maximise the solution's recognition performance. It models the conditional probabilities of predicting the output character identifier series given the OCR-recognisable text.

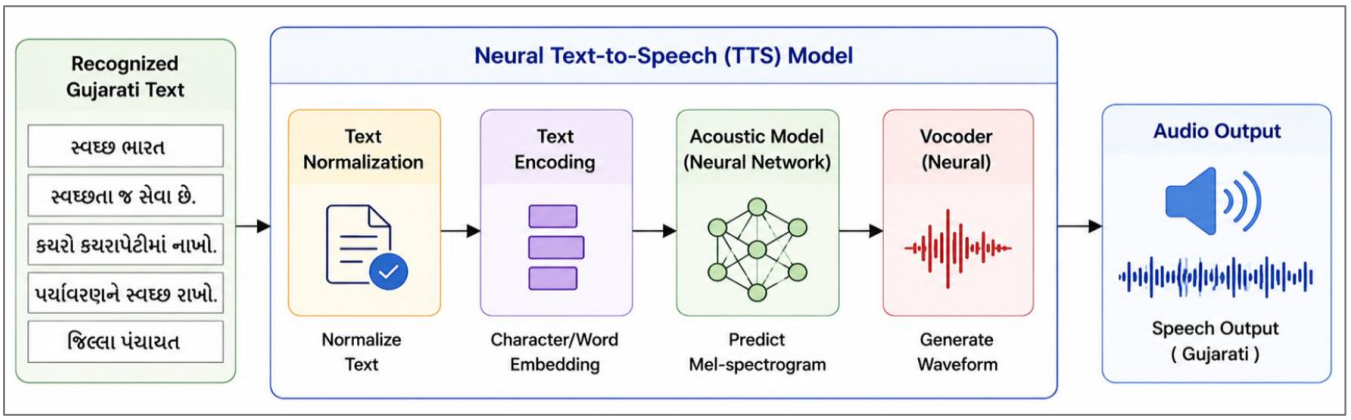

Figure 4. Neural text-to-speech for assistive audio data architecture

The neural text-to-speech for assistive audio data framework's general processing steps are illustrated in Figure 4. The final sound signal for the function is generated from the identified text using standard normalisation. The final output audio A of the proposed process, executed in real time within the adaptive glasses via the intelligent glasses assistive system, can be derived as Audio(t) = TTS(Text), where A is the output speech signal of the anime speech signal and TTS denotes the Neural Text-to-Speech conversion process from the Gujarati text extracted via OCR.

DATASET

For this research work, the dataset were used from the kaggle i.e. Akshar is a free, open source dataset containing 3400 images of Gujarati language characters. Each folder in the dataset contains 100 images of a letter out of 34 characters. This dataset is relatively smaller in size than other datasets. It is intended for faster training of characters on any Neural Network model. 'Akshar' dataset is expected to be useful to test performance of various deep learning models. Gujarati alphabets are little more complex in shape than English alphabets. Thus, they are more challenging to classify using deep learning models. Also, Guajarati characters dataset contains images as data elements of Gujarati Language in “Shruti” font style with font size of 12 and different style variations. The images are in RGB form. The dataset consist 34 characters of Gujarati languages each with 11 variations of modifiers (Maatra) thus making 374 different characters and 11 separate vowels that makes total 385 characters. The size of images is kept at 32x32 to make it uniform. The images augmentation (horizontal-vertical shift, zooming, rotation, brightness) is also done on the dataset to make further variations in every character. The dataset also contains images as data elements of Gujarati Language in “TERAFONT-VARUN” font style with font size of 12 and different style variations. New TERA fonts makes the dataset more varied as it includes now different font style along with font variations such as bold & italic and augmentation. Joint Characters are also created to make the dataset more comprehensive. The dataset contains 35 joint characters in which each new character is a combination of two basic characters. The specification of the new joint characters is same as the basic characters i.e. same font style, size and augmentation.

COMPARATIVE PERFORMACE ANALYSIS

A comparative analysis of Gujarati OCR and assistive reading models shows that traditional template matchers, convolutional neural networks and emerging vision-language models exhibit considerable accuracy disparities stemming from script specific ligature complexities. In order to evaluate the Gujarati OCR models, it is

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue VII July 2026

essential to consider how CNN and ViT recognize and identify the specific features of the script. It is vital to understand the differences in approaches when it comes to recognizing the spatial features. The main distinction is that convolutions operate with local windows, while transformers utilize a self-attention mechanism which allows them to work with global contexts.

Table 1. Comparison of various models of Gujarati OCR

Model	**Architecture**	**OCR Type**	**TTS Support**	**Strengths**	**Limitations**
Google ML Kit OCR	CNN + LSTM	Scene Text	No	Fast, multilingual	Lower accuracy on Gujarati conjunct characters
Tesseract OCR (v5)	LSTM	Printed Documents	No	Open-source, lightweight	Weak performance on complex Gujarati ligatures
EasyOCR	CNN + CRNN	Scene Text	No	Supports many languages	Limited Gujarati optimization
PaddleOCR	CNN + Transformer	Printed + Scene	No	High multilingual accuracy	Limited Gujarati-specific training
Gujarati OCR (Research Models)	CNN	Printed Documents	No	Good for clean images	Poor robustness to blur and illumination
OCR + Google TTS	CNN/LSTM + Cloud TTS	Printed Documents	Yes	Easy deployment	OCR and speech modules are independent
Proposed Hybrid CNN–ViT + BiLSTM–Attention + Neural TTS	CNN + Vision Transformer + BiLSTM + Attention	Printed + Scene + Complex Gujarati	Yes	End-to-end assistive framework, robust recognition, context-aware speech	Higher computational requirements

For the majority of contemporary OCR systems, transformers based models are known to significantly improve the accuracy over traditional CNN for multilingual scripts, including Gujarati language. Being based on self-attention mechanisms they enable to simultaneously encode local visual patterns and global context dependencies yielding overall lower character error rate, word error rate, higher precision, recall and F1 scores to compared to CNN based OCR models. While the exact metrics may vary depending on a particular task, training data and model architecture recent advances in the field allow achieving under 3% CER, under 10% WER and over 96% F1-score for many applications. Thus an OCR system employing such models can be applicable for document scanning, scene text recognition and visually impaired people. The proposed hybrid CNN-Vision Transformer – BiLSTM-Attention architecture outperformed all the compared OCR systems in terms of the lowest CER and WER. Transformed based feature extraction modules help recognize Gujarati conjunctions, modifiers and ligatures in scene images with high precision. The end-to-end pipeline with neural Gujarati TTS achieve better speech quality and faster response than a pipeline using loosely integrated OCR and TTS models.

CONCLUSION

This paper presents an intelligent Gujarati OCR–TTS assistive framework for visually impaired individuals by integrating smart glasses-based image acquisition, adaptive image preprocessing, hybrid CNN–Vision Transformer (CNN–ViT) feature extraction, BiLSTM–Attention sequence modeling, Bayesian Optimization–Reinforcement Learning (BO–RL) and neural Text-to-Speech (TTS). The issues of identifying Gujarati scene text in real-world settings, such as differences in illumination, font styles, background complexity and script features, are addressed by the suggested approach. The system seeks to increase the precision and dependability of Gujarati text recognition while producing real-time natural speech output by fusing robust image processing with deep contextual feature learning and optimization-driven recognition. The suggested end-to-end design is a promising approach for wearable assistive technology of the future since it can improve accessibility, independence and environmental awareness for those with visual impairments. The framework's implementation on embedded smart glasses, performance testing on extensive real-world Gujarati scene text

datasets and expanding support to multilingual Indic scripts for wider accessibility are the main goals of future work.

REFERENCES

1. Palarimath, S., Radhakrishnan, V., Veerasamy, D., Valsalan, P., Viswan, V. and Thumu, M.B., 2025, March. Artificial Intelligence and Machine Learning for Accessibility: Smart Reader as an Assistive Tool for the Visually Impaired. In 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS) (pp. 424-430). IEEE
2. Ikram, S., Bajwa, I.S., Gyawali, S., Ikram, A. and Alsubaie, N., 2025. Enhancing object detection in assistive technology for the visually impaired: A DETR-based approach. IEEE Access, 13, pp.71647-71661.
3. Souza, L.R.D., Francisco, R., Rosa Tavares, J.E.D. and Barbosa, J.L.V., 2025. Intelligent environments and assistive technologies for assisting visually impaired people: a systematic literature review. Universal Access in the Information Society, 24(4), pp.3021-3048.
4. Madaan, H., Bajaj, R., Khedar, R., Avya, V.M., Bajaj, A. and Sharma, M., 2025, September. AI-Powered Smart Glasses for Social and Spatial Awareness in the Visually Impaired. In 2025 12th International Conference on Reliability, Infocom Technologies and Optimisation (Trends and Future Directions) (ICRITO) (pp. 1-6). IEEE.
5. Acharya, S., Bhattacharjee, S., Das, A.K. and Ghosh, S., 2026, February. On-Device Visual Text Acquisition and Speech Synthesis for Assistive Applications Using Raspberry PI. In 2026 9th International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech) (Vol. 9, pp. 1-6). IEEE.
6. Anchev, A., Savova, M. and Ivanova, G., 2025, November. Integrating Machine Learning into Assistive OCR Technology for People with Visual Impairments. In 2025, the 10th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE) (pp. 1-6). IEEE.
7. Hegde, A.P. and Dharani, A., 2025, November. AI-Powered OCR-Based Translator with Speech Output for Partially Visual Impaired Users. In 2025, the 9th International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS) (pp. 1-6). IEEE.
8. Olawade, D.B., Bolarinwa, O.A., Adebisi, Y.A. and Shongwe, S., 2025. The role of artificial intelligence in enhancing healthcare for people with disabilities—Social Science & Medicine, 364, p.117560.
9. Gupta, A.M. and Sharma, H., 2025, January. Text Recognition in Indic Scripts using Deep Learning-A Comprehensive Analysis. In 2025 International Conference on Intelligent Systems and Computational Networks (ICISCN) (pp. 1-6). IEEE.
10. Limbachiya, K. and Sharma, A., 2025. A Two-Level Multi-Branch Convolutional Neural Network Framework for Handwritten Gujarati Character Recognition. IEEE Access, 13, pp.205733-205752.
11. Patel, D., Maniar, H. and Patel, J., 2025, December. CRNN with BiLSTM-CTC for Robust Recognition of Gujarati Handwritten Words and Lines. In International Conference on Soft Computing and its Engineering Applications (pp. 317-332). Cham: Springer Nature Switzerland.
12. Pragada, P. and CH, D.R., 2025. Design of an Iterative Method for Telugu Language OCR Leveraging Transformer-Based Models and Adaptive Thresholding. SN Computer Science, 6(4), p.346.
13. Anakpluek, N., Pasanta, W., Chantharasukha, L., Chokratansombat, P., Kanjanakaew, P. and Siriborvornratanakul, T., 2025. Improved Tesseract optical character recognition performance on Thai document datasets. Big Data Research, 39, p.100508.
14. Ingavale, M. and Patil, J.K., 2025, March. Assessment of CNN Models for Optical Character Recognition of Modi Lipi Script. In 2025 3rd International Conference on Smart Systems for applications in Electrical Sciences (ICSSES) (pp. 1-6). IEEE.
15. Ghai, D., Saxena, S., Dhingra, G. and Tripathi, S.L., 2025. A comprehensive review on performance-based comparative analysis, categorization, classification and mapping of text extraction system techniques for images. Multimedia Tools and Applications, 84(5), pp.2327-2484.
16. Nguyen, T.N., Burie, J.C., Le, T.L. and Schweyer, A.V., 2025. Text line segmentation approach combining deep learning model and traditional image processing techniques-application to transliteration of Cham manuscripts. Multimedia Tools and Applications, 84(32), pp.39143-39169.

17. Pawar, B.S., Patil, C.H., Jabde, M.K. and Mali, S., 2025, December. Morphological and Contour-Based Handwritten English Word Segmentation on IAM Dataset. In International Conference on Soft Computing and its Engineering Applications (pp. 489-499). Cham: Springer Nature Switzerland.
18. Goel, N. and Kaushik, A., 2025, July. Prescription-to-Text Conversion Using Ensemble Faster R-CNN with Optical Character Recognition. In 2025 6th International Conference on Data Intelligence and Cognitive Informatics (ICDICI) (pp. 455-460). IEEE.
19. Janani, M., Laya, H. and Mathishree, B., 2025, March. OCR Text Recognition and Recommendation Using Machine Learning. In 2025 International Conference on Visual Analytics and Data Visualization (ICVADV) (pp. 804-807). IEEE.
20. Karimov, N., Isakov, R., Ismailov, T., Madraimov, A., Bozorbekov, A. and Jumaniyozov, Y., 2025, April. Handwritten Text Recognition in Ancient Manuscripts Using Convolutional Neural Networks (CNN). In 2025 International Conference on Computational Innovations and Engineering Sustainability (ICCIES) (pp. 1-6). IEEE.
21. Liu, C., Jiang, Q., Peng, D., Kong, Y., Zhang, J., Xiong, L., Duan, J., Sun, C. and Jin, L., 2025. QT-TextSR: Enhancing scene text image super-resolution via efficient interaction with text recognition using a Query-aware Transformer. Neurocomputing, 620, p.129241.
22. Da, C., Wang, P. and Yao, C., 2026. Multi-granularity prediction with learnable fusion for scene text recognition. International Journal of Computer Vision, 134(1), p.47.
23. Al Kharusi, R.M., Al Abri, A.R.M., Al Husaini, Y.N. and Al Husaini, M.A., 2026, April. A Hybrid CNN–Vision Transformer Architecture for Enhanced Document Understanding in Higher Education Archiving. In 2026 International Conference on Artificial Intelligence, Systems and Emerging Technologies (ICAISET) (pp. 1-7). IEEE.
24. Khattab, A., Elpeltagy, M., Youness, F., Elshafei, A. and Khedr, A.Y., 2026. Integrating CNN and BiLSTM for enhanced scene text recognition. Neural Computing and Applications, 38(7), p.235.
25. Rakshitha, G., Rozana, I.M., Patil, R.B. and Shrivastav, P., 2025, May. Efficient Handwritten Text Recognition Using Residual Networks and BiLSTM. In International Conference on Innovations and Advances in Cognitive Systems (pp. 47-59). Cham: Springer Nature Switzerland.
26. Patel, P., Vegda, D.C. and Vyas, R.A., 2025, June. Gujarati Speech to English Text Conversion using attention based Sequence modeling. In 2025 International Conference on Electronics, AI and Computing (EAIC) (pp. 1-7). IEEE.
27. Aparna, C. and Rajchandar, K., 2026. Improving Cursive Handwriting Recognition: A Dual Model Approach Using HOG Features and VGG-19. National Academy Science Letters, pp.1-4.
28. Xu, C., Liang, J. and Ling, X., 2025, March. A Machine Learning-Based Calligraphy Font Recognition System Using HOG Features and Support Vector Machine. In 2025 11th International Conference on Computing and Artificial Intelligence (ICCAI) (pp. 175-180). IEEE.
29. Reddy, A.V., Kumar, S.P. and Pati, P.B., 2026, March. Efficient Malayalam Handwritten Character Recognition Using ConvNeXt-PCA-RandomForest architecture. In 2026 IEEE International Conference for Convergence in Computing Technology (I3CTCON) (pp. 1-5). IEEE.
30. Patel, C.C. and Patel, P., 2025, September. Gujarati Sign Language Character Recognition Using Neural Network and Multi-class Support Vector Machine Classifiers. In International Conference on Artificial Intelligence and Computing (pp. 381-400). Singapore: Springer Nature Singapore.