Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 7, No.4, October 2025, pp. 02504026-01 ~ 02504026-09
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v7i4.2380

"Demata 2.0": An On-Device AI Assistive Technology for the Visually Impaired Integrating YOLOv10 and OCR

Reza Febri Abadi[1*], Toni Yudha Pratama[1], Neti Asmiati[1], Ade Anggraini Kartika Devi[1], Joko Yuwono[2], Dwi Setia Permana[3], Febrian Alwan Bahrudin[1]

[1]Faculty of Teacher Training and Education, Universitas Sultan Ageng Tirtayasa, Serang City, Banten, 42117, Indonesia

[2]Special Education Program, Universitas Sebelas Maret, Surakarta, 57126, Indonesia

[3]Puslapdik (Center for Education Financing Services), Ministry of Primary and Secondary Education, Indonesia

*rezafebriabadi@untirta.ac.id

Abstract. Accessibility to printed materials and independent recognition of the environment remain key challenges for students with visual impairments. To address this issue, this study introduces Demata 2.0, a fully offline on device multimodal AI system. The system integrates Google ML Kit for Optical Character Recognition (OCR) and the YOLOv10 model via TensorFlow Lite for object detection. A mathematical distance algorithm in the RGB color space enables color identification. Evaluation showed that object detection achieved a mean average precision of 31.83%, with an average processing speed of 2–3 FPS. For OCR, the system recorded a Character Error Rate (CER) of 4.81% and a Word Error Rate (WER) of 10.71% on printed documents. The RGB algorithm also determined the closest possible color effectively. Overall, Demata 2.0 advances assistive technology by providing an efficient and practical blueprint for AI integration.

Keywords: assistive technology, on-device AI, optical character recognition, visual impairment, YOLOv10

(Received 2025-07-16, Revised 2025-09-22, Accepted 2025-10-22, Available Online by 2025-10-28)

1. Introduction

Although in the present day assistive technology has been much more developed, individuals who have visual impairments or are commonly called blind disabilities still face significant challenges in accessing printed information which is a crucial part of the teaching and learning process and academic activities in higher education [1]. Limitations in recognizing printed documents, such as textbooks, teaching materials, worksheets, administrative forms, correspondence documents, and scientific references, are an obstacle to their independence and involvement in academic activities [2]. This situation increasingly raises the urgent need for assistive technology that is more functional for the visually impaired.

Assistive technology refers to any object or tool that is modified or adapted to improve the functional abilities of individuals with disabilities [3]. The use of assistive technology has a significant impact on learning activities because it can increase engagement, foster learning motivation, and provide support that can be customized according to the needs of each individual as a user [4]. Nevertheless, although other reviews confirm that assistive technology for the visually impaired is indeed growing rapidly, there is still a gap between technological innovation and the fulfillment of specific needs [5]. Until now there is no system that simultaneously at one time is able to detect numbers, colors, and words in one system of integrated, portable, practical, adaptive to the user's context, and can be used offline.

Over time, computer vision and deep learning emerged as innovative breakthroughs to meet the complexity of the above. Computer vision methods are used in image processing and visual interpretation, while deep learning provides rapid detection with a high level of accuracy as shown by various recent studies [6–10]. For the needs of reading printed text, several studies have developed visual detection systems based on *Optical Character Recognition* (OCR) to recognize text, numerical, and visual objects with outputs in the form of audio or Braille [11–13]. OCR is used for a variety of character recognition by scanning objects and then converting them into a digital format with an audio-tactile interface that has proven to be an effective system for those with visual impairments [14]. Visual recognition systems are also seen in other fields of research on the application of *convolutional neural network (CNN) architectures,* such as DenseNet which can process images and classify objects with high accuracy as well as ResNet which can detect objects in real time and process images optimally [15,16]

On the basis of all the above exposures, we developed Demata 2.0 assistive technology, which is a multimodal software that is able to recognize colors, numbers, words, and objects in printed documents and read them through *a text-to-speech* (TTS) system with an on-device architecture that combines Google ML Kit and YOLOv10. This innovation is specifically aimed at the needs of blind lecturers and students in Indonesian universities. In general, this technological innovation was developed to fill the gap in offline operations, portability, and functional integration.

2. Methods

"Demata 2.0" is designed and developed based on the ADDIE (analysis, design, develop, implementation, and evaluation) model.

2.1 Analysis

The analysis was carried out by reviewing the relevant literature, observing the availability of technological resources, and identifying user needs through performance gaps which were seen based on the three main dimensions *of lack of resource, lack of motivation,* and *lack of knowledge* [17]. In addition, to gain a more comprehensive understanding, needs analysis is seen from the perspective of user acceptance by referring to the *Technology Acceptance Model* (TAM) framework including the perception of technology usefulness, *perceived ease of use,* and the level of user *acceptance of information technology* [18].

2.2 Design

The findings from the analysis stage are integrated into the design of the Demata 2.0 application. Important information, such as voice navigation requirements and scanning of printed objects or documents (including text, numbers, and colors) is the basis for determining the app's key features. It also includes independent access without depending on the internet connection. The design is broadly presented in the following block diagram.

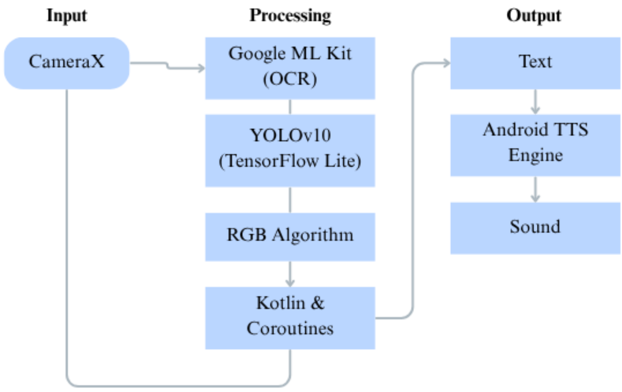

Figure 1. Diagram Block Demata 2.0

2.3 Development

The development of text and number detection uses pretrained OCR models from Google ML Kit that have been trained with a massive dataset by Google. Object detection using YOLOv10 models and backbones with a pretrained version of nano variants that have been trained with common datasets such as *Common Objects in Context* (COCO). Color detection uses a more accurate approach with a mathematical distance algorithm based on red, green, and blue (RGB). Text-to-speech technology relies on the built-in Android feature (Android TTS engine) with the consideration that it can fully operate without an internet connection. Optimization techniques to keep applications light and fast include the use of (1) *TensorFlow Lite* and YOLOv10 nano as lightweight models; (2) GPU delegate to speed up processing; and (3) Kotlin Coroutines to set a timer that automatically and efficiently triggers the color detection function every one second so that the detection process does not overload the main thread so that the application user interface remains smooth.

2.4 Implementation and Evaluation

The implementation stage aims to prepare for a real environment and engage users. Demata 2.0 passes the on-device test by looking at (1) the accuracy of object detection based on *the mean average precision* (mAP); (2) frames per second (FPS) processing speed; (3) OCR accuracy based on character error rate (CER) and word error rate (WER); and (4) color detection accuracy.

3. **Results and Discussion**

Demata 2.0 was developed using the foundation of user-centered design that puts user needs, expectations, and experiences as the core of design. Users are involved from the beginning in the development process through needs analysis. The results of the study [19,20] confirm that user participation from the beginning provides crucial feedback in the development process that not only improves quality and maximizes usefulness in academic and social contexts, but also minimizes rejection of the developed technology.

Based on the findings from the needs analysis, the majority of respondents stated the importance of the presence of assistive technology that is not only accessible, but can also be optimally functioned without relying on an internet connection for flexibility of use in various situations. Respondents need assistive technology that is able to support their independence in various academic activities, ranging from reading printed materials, understanding documents in the form of words or numbers, recognizing objects in the campus environment, to recognizing color information.

The following section explains how the core functionality of Demata 2.0 is optical character recognition, object detection with YOLOv10, and color identification using RGB algorithms.

3.1 Kinerja Optical Character Recognition

Optical Character Recognition is an important aspect of Demata 2.0 that allows users to convert physical text into sound output. OCR accuracy testing in Demata 2.0 was carried out on four different types of written documents, including handwriting. The results of the accuracy test are summarized in Table 1.

Table 1. OCR Testing on Writing

Document Type	Total Characters	Total Words	False Characters	The Wrong Word	CER (%)	WHO (%)
Printed documents	2950	420	142	45	4.81%	10.71%
Neat handwriting	1200	210	144	33	12.0%	15.71%
Bad handwriting	800	150	160	60	20.0%	40.0%
Creative-fonted ads	500	80	30	10	6.0%	12.5%

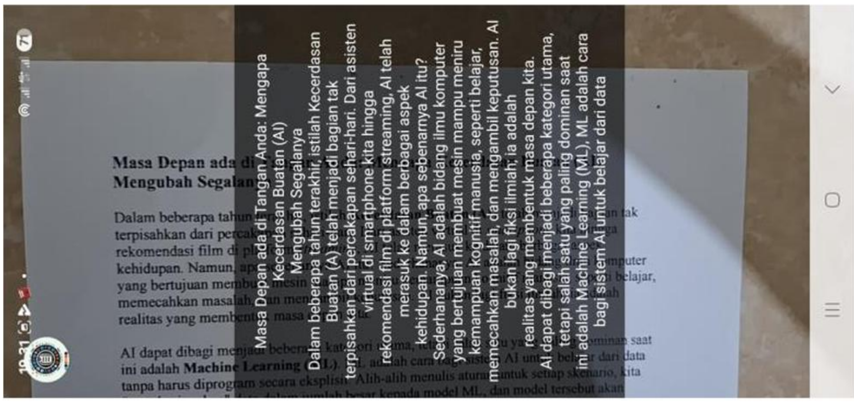

Figure 2. Example of Detection Results on Document

Based on testing, Demata 2.0 achieved a very good level of accuracy on printed documents that use computer fonts, as shown in Figure 2, with a *Character Error Rate* (CER) of 4.81% and *a Word Error Rate* (WER) of 10.71%. Moreover, even in creative fonts in ad text, the error rate is relatively low. This means that the OCR model of Google ML Kit shows adaptability to font style variations. However, as is common with all OCR technologies, OCR performance declines in unstructured writing, overly decorative fonts, and older typographic variations [21–23]. Accordingly, OCR performance showed a significant decrease in handwriting. Although neat handwriting can still be processed with a WER of 15.71%, OCR performance decreases drastically in poorly categorized handwriting with a CER of 20% and a WER of 40%. This is a consequence of Demata 2.0's strategic decision to choose the Google ML Kit model in device mode with considerations of lightweight, fast, and running without dependence on an internet connection.

3.2 Object Detection Performance with YOLOv10

Demata 2.0 uses the pretrained version of the YOLOv10 model. This model has been trained using common datasets such as *Common Objects in Context (COCO).* COCO is a common data set in a natural context containing more than 330,000 images with more than 1.5 million annotated objects [24]. Yolov10, which has been trained on the COCO dataset, makes the Demata 2.0 system more practical, stable, tested, and efficient on mobile devices. As shown in figure 3, YOLOv10 successfully detects sofa objects with a high degree of conformity. The performance of YOLOv10 on Demata 2.0 is seen from the main metric that shows average accuracy, namely the *mean average precision* (mAP) of 31.83%. Mean Average Precision is the result of a comparison between the application detection results and the ground truth of the dataset. The following is a summary table 2. of model performance across some of the categories of objects identified in the test.

Table 2. YOLOv10 Testing Summary of Various Objects

Category Object	**Detection Accuracy (AP)**	**Information**
sofa	90.48%	Excellent performance
remote	73.21%	Excellent performance
chair	53.84%	High accuracy
person	42.86%	Good accuracy
monitor	31.05%	Good accuracy
cupboard	7.93%	Low accuracy
faucet	1.39%	Very low accuracy

Figure 3. Example of Detection Results on Sofa Objects

Based on the test summary table, the YOLOv10 model on Demata 2.0 is able to detect everyday objects that are common and often appear in datasets, such as sofas and remotes with very high accuracy. However, for cabinet and faucet objects, they have low accuracy. This is a challenge because the complex background and small object size affect YOLO's performance, so in other studies, YOLO was modified and improved by adding modules [25–27]. Meanwhile, Demata 2.0 uses the standard YOLOv10 model, not the modified version.

Furthermore, in terms of processing speed (FPS), using the Samsung Galaxy A55 test device with 8 GB of RAM and Android 15 operating system, the processing speed of the measured model averaged 2-3 FPS. Although this number looks low in the logs, the app still runs very smoothly and object detection feels real-time. This happens because the object detection model runs on the device's CPU, not on the GPU. The time it takes to process a single frame of an image (inference time) is indeed longer (~498 milliseconds), but the app is designed to intelligently process video frames so that there is no lag.

3.3 Color Identification Performance via RGB Algorithm

Demata 2.0's performance in identifying colors using RGB algorithms was tested to evaluate how accurately the system was in detecting and classifying various color samples. The RGB algorithm works by comparing the detected colors with a standard color list of 200 colors present in the getColorCategory function. The following table 3. shows the results of color identification testing on various samples. Performance is assessed based on the comparison between the original color and the application detection results.

Table 3. Color Detection Accuracy

Color Sample	**Original Color**	**Application Detection Results**	**Accuracy Rate**
Sample A	Red	Brick	Highly Accurate
Sample B	Green	Dark Green	Highly Accurate
Sample C	Gray	Deep gray	Highly Accurate
Sample D	Sky Blue	Bright Sky Blue	Accurate
Sample E	Dark orange	Orange glows	Accurate
Sample F	Reddish brown	Brown	Accurate
Sample G	Light green	Light Green	Highly Accurate

Overall, the RGB algorithm on Demata 2.0 shows excellent and consistent performance in identifying a wide range of colors. Five of the seven samples (Samples A, B, C, F, and G) were detected with a very high level of accuracy. The system can recognize basic colors such as red, green, and gray very precisely, even to specific shades such as light green. The other two samples (Samples D and E) were successfully detected with an accurate level of accuracy. This means that the algorithm manages to find the closest color mathematically possible.

4. **Conclusion**

This research successfully developed Demata 2.0, an artificial intelligence-based assistive technology with a lightweight multimodal AI integration and fully on-device. The system is supported by nano YOLOV10 model optimization techniques, TensorFlow Lite, Google ML Kit for OCR, and RGB algorithms. The evaluation of the system showed effective performance with *a mean average precision* (mAP) of 31.83% for object detection and a low character error rate (CER) of 4.81% for printed documents.

Despite showing good performance, the study identified several challenges. The limitations of this system lie in the varying performance in object detection, especially objects that are less identified in the training dataset and objects with complex backgrounds or small sizes that appear at very low accuracy for faucets (1.39%) and cabinets (7.93%). OCR performance also showed a significant decrease in some unstructured handwriting. These limitations are a consideration of the strategic decision to use a lightweight pretrained model for on-device functionality.

Given the limitations as described, future research may focus on refining the YOLOv10 model with custom datasets. In addition, it needs to improve OCR with other architecture integrations. Furthermore, it can also optimize the GPU to increase FPS so that it potentially allows for faster video *frame* processing, while maintaining the right balance between power consumption and performance.

Acknowledgements
The development of the Demata 2.0 application was made possible through funding provided by the Ministry of Higher Education, Science, and Technology (contract number 111/C3/DT.05.00/PL/2025). We extend our sincere appreciation for their continued support in promoting inclusive education.

References

[1] Abadi RF, Pratama TY, Dewi CO. Penggunaan media ular tangga Braille dalam meningkatkan kemampuan mengenal bilangan angka 1-10 anak dengan hambatan penglihatan. Jurnal UNIK: Pendidikan Luar Biasa 2021;6:37. https://doi.org/10.30870/unik.v6i1.1186.3

[2] Maslahah S, Musayarah S, Alamsyah Sidik S, Febri Abadi R, Yhuda Pratama T, Tanjung Utami Y, et al. Pengembangan Modul Inovasi Pendidikan Berbasis Universal Design for Learning (UDL) yang Inklusif bagi Mahasiswa Disabilitas. Jurnal Unik: Pendidikan Luar Biasa 2023;8:1–8.

[3] Viner M, Singh A, Shaughnessy MF. Assistive technology to help students with disabilities. Research Anthology on Inclusive Practices for Educators and Administrators in Special Education, IGI Global; 2021, p. 579–600. https://doi.org/10.4018/978-1-6684-3670-7.ch033.

[4] Prystiananta NC, Noviyanti AI, Udhiyanasari KY. The Impact of Assistive Technologies in Enhancing English Learning Outcomes for Students with Disabilities: A Meta-Narrative Analysis. World Journal of English Language 2025;15:296–308. https://doi.org/10.5430/wjel.v15n2p296.

[5] Mashiata M, Ali T, Das P, Tasneem Z, Badal MFR, Sarker SK, et al. Towards assisting visually impaired individuals: A review on current status and future prospects. Biosens Bioelectron X 2022;12. https://doi.org/10.1016/j.biosx.2022.100265.

[6] Kela GP, Daga MG, Khandelwal R. Vision to Voice: An Advanced Blind Assistance System Integrating YOLOv3 and OCR Technologies for Enhanced Mobility. ICDT 2025 - 3rd International Conference on Disruptive Technologies, 2025, p. 1473 – 1476. https://doi.org/10.1109/ICDT63985.2025.10986482.

[7] Cho IJ, Park J, Bae H. A Computer Vision and Vibrohaptic Glove-Based Piano Learning System for the Visually Impaired. International Conference on Advanced Communication Technology, ICACT, 2025, p. 334 – 338. https://doi.org/10.23919/ICACT63878.2025.10936768.

[8] Bansal V, Joshi A, Aarsha VS, Shivali, Talwandi NS. VisionCane: A Comprehensive Review of an Intelligent Assistive Device for Object Detection and Real-Time Communication. Lecture Notes in Networks and Systems 2025;1408 LNNS:191 – 202. https://doi.org/10.1007/978-981-96-6297-5_15.

[9] Dube S, Bagde M, Bhagat J, Pardhi S, Ghosh M, Das D. Sound Vision: A Deep CNN Recursive Learning-Based Navigation Assistive Device for Divyang (Visually Impaired) Person. Lecture

Notes in Networks and Systems 2025;1200:523 – 533. https://doi.org/10.1007/978-981-97-9926-8_40.
[10] Ikram S, Sarwar Bajwa I, Gyawali S, Ikram A, Alsubaie N. Enhancing Object Detection in Assistive Technology for the Visually Impaired: A DETR-Based Approach. IEEE Access 2025;13:71647 – 71661. https://doi.org/10.1109/ACCESS.2025.3558370.
[11] Doore SA, Istrati D, Xu C, Qiu Y, Sarrazin A, Giudice NA. Images, Words, and Imagination: Accessible Descriptions to Support Blind and Low Vision Art Exploration and Engagement. J Imaging 2024;10. https://doi.org/10.3390/jimaging10010026.
[12] Reddy KK, Badam R, Alam S, Shuaib M. IoT-driven accessibility: A refreshable OCR-Braille solution for visually impaired and deaf-blind users through WSN. Journal of Economy and Technology 2024;2:128–37. https://doi.org/10.1016/j.ject.2024.04.007.
[13] Holanda GB, Souza JWM, Lima DA, Marinho LB, Girão AM, Bezerra Frota JB, et al. Development of OCR system on android platforms to aid reading with a refreshable braille display in real time. Measurement (Lond) 2018;120:150–68. https://doi.org/10.1016/j.measurement.2018.02.021.
[14] Sabu AM, Das AS. Proc. IEEE Conference on Emerging Devices and Smart Systems (ICEDSS 2018) : 2-3 March 2018, Mahendra Engineering College, Tamilnadu, India. [IEEE]; 2018.
[15] Salahuddin NS, Tarie FM, Saptariani T. Development of a Robotic System for Agricultural Pest Detection: A Case Study on Chili Plants. Advance Sustainable Science, Engineering and Technology 2025;7. https://doi.org/10.26877/asset.v7i1.1152.
[16] Zulhusni M, Sari CA, Rachmawanto EH. Implementation of DenseNet121 Architecture for Waste Type Classification. Advance Sustainable Science, Engineering and Technology 2024;6. https://doi.org/10.26877/asset.v6i3.673.
[17] Branch RM. Instructional Design: The ADDIE approach. Springer US; 2010. https://doi.org/10.1007/978-0-387-09506-6.
[18] Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Q 1989;13:319–39. https://doi.org/10.2307/249008.
[19] Guzman-Orth, Steinberg D;, Albee J;, Traci. English Learners Who Are Blind or Visually Impaired: A Participatory Design Approach to Enhancing Fairness and Validity for Language Testing Accommodations. n.d.
[20] Ntakolia C, Dimas G, Lakovidis DK. User-Centered System Design for Assisted Navigation of Visually Impaired Individuals in Outdoor Cultural Environments. Univers Access Inf Soc 2022;21:249–74. https://doi.org/https://doi.org/10.1007/s10209-020-00764-1.
[21] Seuret Mathias and van der Loop J and WN and MM and MJ and HT and CV. Combining OCR Models for Reading Early Modern Books. In: Fink Gernot A. and Jain R and KK and ZR, editor. Document Analysis and Recognition - ICDAR 2023, Cham: Springer Nature Switzerland; 2023, p. 342–57.
[22] Sadek J, Vlachidis A, Pickering V, Humbel M, Metilli D, Carine M, et al. Leveraging OCR and HTR cloud services towards data mobilisation of historical plant names. International Journal of Digital Humanities 2024;6:237–61. https://doi.org/10.1007/s42803-024-00091-4.
[23] Alghamdi M, Teahan W. Experimental evaluation of Arabic OCR systems. PSU Research Review 2017;1:229–41. https://doi.org/10.1108/PRR-05-2017-0026.
[24] Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft COCO: Common Objects in Context 2015. https://doi.org/https://doi.org/10.48550/arXiv.1405.0312.
[25] Hu M, Li Z, Yu J, Wan X, Tan H, Lin Z. Efficient-Lightweight YOLO: Improving Small Object Detection in YOLO for Aerial Images. Sensors 2023;23. https://doi.org/10.3390/s23146423.
[26] Zhou X, Jiang L, Hu C, Lei S, Zhang T, Mou X. YOLO-SASE: An Improved YOLO Algorithm for the Small Targets in Complex Backgrounds. Sensors 2022;22. https://doi.org/10.3390/s22124600.

[27] Mei J, Zhu W. BGF-YOLOv10: Small Object Detection Algorithm from Unmanned Aerial Vehicle Perspective Based on Improved YOLOv10. Sensors 2024;24. https://doi.org/10.3390/s24216911.