Context-Aware Label Information Extraction Using Vision + NLP for Assistive Applications

J Krishna Manasa[1], Nukala Manasa[2], Basireddy Shivani[3], Dr Potu Narayana[4]
[1,2,3,4]*Department of Computer Science and Engineering, Stanley College of Engineering & Technology for Women Hyderabad, India*

***Abstract*—Product labels contain essential information about the brand name, product name, and quantity, which can serve a wide variety of applications as assistive and automated systems. However, a problem arises when attempting to extract this type of information from a product image due to the different font styles, varying light exposure, and background noise found in these types of images. This paper proposes a multi-OCR-based approach for extracting label information from a product image. The system is based upon the use of multiple OCR engines (Tesseract, EasyOCR and PaddleOCR) in order to detect and identify text in product labels. Once the text has been extracted, additional preprocessing and fuzzy matching with RapidFuzz techniques will improve the extraction of label information. The proposed system was evaluated using several performance measurements (accuracy, precision, recall, F1 score and error rate). The experimental results indicate that the proposed approach achieved 76.92% accuracy, 100% precision, 76.92% recall, and an F1 score of 86.96% in extracting product label information from images.**

***Index Terms*—Optical Character Recognition (OCR), Product Label Recognition, Multi-OCR System, Text Extraction, Image Processing, Information Extraction, Deep Learning, Assistive Applications.**

I. INTRODUCTION

Assistive Technology (AT) has greatly enhanced accessibility to visual information for people with vision loss through the use of artificial intelligence and computer vision. Many of the activities that a person does on a daily basis require the ability to visually take in information, such as product labels, reading the information on the package, or understanding the information on the packaging about safety. Accessing these types of information independently can be very difficult and usually requires an outside aide for each of those activities.

In the past, assistive solutions provided basic Optical Character Recognition (OCR) as the primary solution to converting the printed words into audible speech. These types of assistive solutions have been beneficial; however, they have had some challenges where OCR has been implemented; namely, on complex types of product packaging, that contain multiple chunks or areas of text, as well as various forms or types of text - such as various font types, font sizes, logos and pictures; and, where product labels include many types of structured information - such as brand name, ingredient list, manufacturing information, expiration dates, and warnings - that simple OCR-based pipelines are difficult to read.

As a result, visuals of the outgoing OCR tends to present as RAW text information that does not provide the user with useful context, as such, the user may not be able to truly understand what is being communicated in the content of the information provided on the product label.

Recent developments in computer vision and natural language processing (NLP) have resulted in greater abilities for intelligent multimodal systems to understand visual and text information at once. In general, vision-based models can detect objects or places in images, while using techniques of text extraction via NLP to provide context and identify the semantic content of these terms/trends. These studies typically look at reading systems using OCR technology, providing wearable vision assistance to those with visual impairments, and developing vision-language models to enhance the accessibility of printed material through real time audio descriptions, formulating object recognition from visual perceptions, and converting textual information into

verbal outputs. Despite these advancements, there is still a large majority of existing technologies that continue to only be able to extract simple text without interpreting the full meaning or location of that particular text found on product labels. As a result, visually impaired individuals receive large amounts of raw, unstructured text with no clear delineation of certain key characteristics found on product labels such as brand name, product category, quantity, expiration date, safety alert(s), etc.

This project aims to develop a contextual label information extraction system that uses multiple OCR engines combined with NLP analysis for more accurate extraction of product packaging information. The system will extract textual information from images captured by cameras or mobile devices through a multi-engine OCR processing pipeline that is resilient to the challenges of varying lighting conditions and complex product label designs. Once the raw text is extracted, NLP techniques will be applied to organize and identify key components of the label including brand name, product name, quantity, expiration date, and caution or warning labels. The output of this project will be a more accurate and useful representation of product packaging by converting the raw text to organized, actionable information. Ultimately, this will enhance the opportunities available to assistive technologies in visually impaired individuals through improved means of understanding product packaging to allow independent decision-making. This example of combining vision and NLP technology illustrates how multimodal AI systems can be utilized to improve accessibility and usability in everyday life.

II. LITERATURE SURVEY

The initial research into assistive technology was primarily centred on creating solutions to help people who cannot see read and gain access to the written word around them. There were many studies that suggested systems based on cameras that allowed a user to take a picture of printed text. Those camera images could be converted into audible speech using Optical Character Recognition (OCR) and text-to-speech technology. These early systems were successful in allowing visually impaired people to read printed newspapers, magazines, books, etc. independently. Examples of these include the portable, handheld OCR readers such as those described in [23], [24], [25] and [26] and image processing methods to extract text from many other things. In addition, the systems continued to be enhanced through advances in OCR accuracy, as well as adapting the technology to permit multilingual OCR capability, and allowed for increased levels of complexity when accessing printed material that visually impaired users used to access the visual environment. All of these previous findings demonstrated the significance of the combination of OCR and computer vision for assistive purposes.

The application of machine learning (ML) and deep learning (DL) methods has led to further development of these technologies by allowing for more accurate and usable recognition systems. Research has demonstrated the viability of using optical character recognition (OCR) as an effective means of providing visually impaired people with access to text via audio output; see [20,21,22], and [29]. Researchers have used various image preprocessing, feature extraction, and pattern recognition techniques to improve the detection of text in complicated backgrounds. Modern-day OCR systems use various recognition (ML/DL) models and preprocessing methods to further improve text recognition and make it easier for visually impaired people to interact with the world around them through automatic text recognition and interpretation of images that have been captured.

The emergence of deep learning has allowed researchers to include both object detection and computer vision algorithms within assistive technology solutions. Several studies introduced vision-based assistive technologies that can identify items (i.e., objects, obstacles, and text) in real time [19], [16], [12] and [13]. Vision-based assistive technologies typically use deep learning models (e.g., CNN, YOLO) to provide users who are blind or visually impaired with situational awareness through object detection and guidance regarding the related hazardous areas/environments. Additionally, incorporating text extraction with object recognition creates opportunities for users to identify products via identification; read product labels; and receive contextual information regarding their immediate surroundings. Overall, real-time visual assistance significantly increases independence and mobility for individuals who are blind or visually impaired when utilizing these technologies.

A number of recent investigations into multimodal artificial intelligence (AI) and vision-language models have addressed ways to enhance the understanding of visual information in context through these types of models. The research literature [14,15,28,3,5] describes various multimodal models that integrate visual perception into natural language processing (NLP) for the purpose of understanding what is happening in a visual scene, recognizing what is in a text description, and generating meaningful representations of the objects as well as their context within the environment. For example, a multimodal large language model will process image features and text at the same time, providing a more comprehensive and contextually appropriate understanding of an image than current optical character recognition (OCR) systems. These experimental models illustrate the potential for the convergence of vision computing and NLP to produce intelligent help technologies that are capable of recognizing and understanding highly complex visual situations.

Wearable The latest research is showing some great progress in the development of AI-based assistive technologies. These new technologies are using a combination of Optical Character recognition (OCR) with deep learning and Natural language Processing (NLP) to be able to extract useful meaning from product labels on packaging. Some of the studies that have explored the use of these types of OCR frameworks in a variety of different applications include [1], [2], [4], [6], [7], [8], [9], [10], [11], [17], [18], [27] and [30]. All of the above-mentioned studies demonstrate the significance of using deep learning to process text and identify product name, date of expiry, and warning label data by departing from the traditional approach and using NLP. Many of the Modern OCR pipelines that are being designed for accessibility purposes are trying to emulate the way humans interpret and rate the readability of text found in physical space.

III. PROPOSED METHODOLOGY

1. Data Description

This study focuses on product label information extraction from images using Optical Character Recognition (OCR) techniques. The dataset consists of product label images containing textual information such as brand name, product name, and quantity details. Each image in the dataset is associated with a ground truth annotation file that stores the correct label information. The dataset includes variations in lighting conditions, text orientation, font styles, and background noise, which makes text extraction challenging. Therefore, preprocessing and text matching techniques are applied to improve the accuracy of information extraction.

2. Data Preprocessing

As a means to increase the quality of input images, and also increase OCR performance, several methods are applied when preprocessing the images.

- Image Cleaning:by removing bad images (those that have excessive noise or poor quality) from being used to recognize the text.
- Image Normalization: resizing and standardizing all of the images to be uniform from one image to another in the dataset
- Text Cleaning : making sure that all of the extracted text is converted to lower case and that all special characters are removed for ease of processing.
- Noise Reduction : using methods to minimize background noise on the images, improving the clarity of the text in then
- Text Normalization : standardizing units of measure (e.g., grams and milliliters)so that all label extractions have the same unit of measure.

3. Multi-OCR Based Label Information Extraction

This system employs a number of optical character recognition engines in order to increase the reliability of textual extraction from products labels. This system utilizes a variety of OCR tools, including Tesseract OCR, EasyOCR and PaddleOCR, to extract text from product images. By merging the output from several OCR systems together, it enhances the likelihood that any textual content will be recognized on complicated product label image files.

Algorithm Steps:

1. Input Image Processing:
- Import label images from the label image dataset.
- Perform preprocessing techniques on label images to enhance their quality.

2. Text Extraction:
- Use multiple OCR engines on label images to detect, extract, and retrieve text.

• Merge results from multiple OCR engines to determine the best results obtained.

3. Text Processing and Matching:
• Normalize the extracted text (clean up and convert to standard), to allow for comparison.
• Utilize rapidfuzz for fuzzy string matching of OCR outputs against their respective (ground truth) known products.

4. Information Extraction:
• Find key fields on labels including; brand name, product name and quantity.
• Store returned product data in an organized manner for analysis.

5. Evaluation:
• Utilize accuracy, precision, recall, F1 Score, and Error Rate as assessment metrics in assessing the performance of a system.
• Compare your proposed approach against other OCR based models for performance improvement.

Model Architecture:
• Input Layer: Accepts images of product labels as inputs.
• OCR Processing Layer: Extracts the required data from OCR machines; accesses multiple optical recognition engines.
• Text Processing Layer: Cleans and normalizes the data extracted (with space adjustments, etc.).
• Matching Layer: Fuzzily matches extracted data against ground truth or actual labels.
• Output Layer: Outputs Structured Label Information, and outputs how "Well" structured label information was extracted.

The proposed methodology combines optical recognition engines with preprocessing of text and fuzzy matching techniques, thereby increasing accuracy in extraction of product label information. The combination of all these methods will also significantly increase the overall performance of the product image to extract actual product label information accurately and reliably.

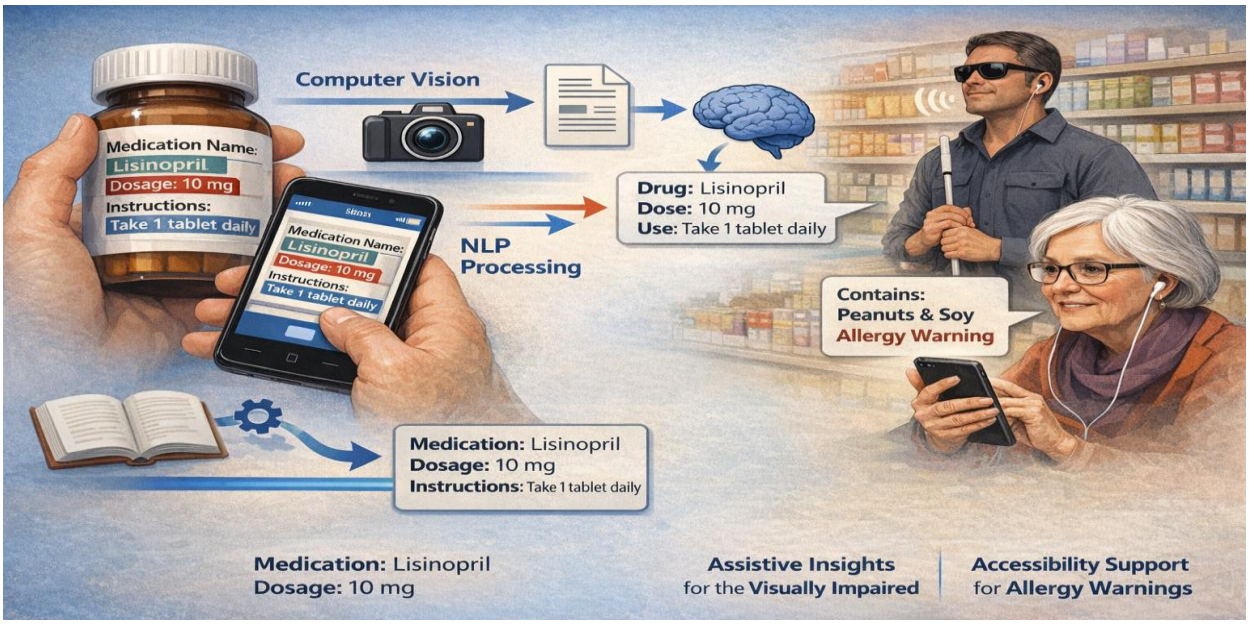

Figure 1. AI-based system using Computer Vision and NLP to scan medication labels and provide accessible information and allergy warnings.

This figure illustrates an AI-powered system that uses computer vision and natural language processing (NLP) to scan and interpret medication labels through a smartphone. The system extracts important details such as the medication name, dosage, and usage instructions, and converts them into accessible information. It also highlights allergy warnings and other critical details, helping users especially visually impaired individuals understand medication information quickly and safely.

IV. RESULT AND DISCUSSIONS

Experimental Setup

The study utilised Python version 3.9 to create an evaluation protocol for testing the performance of the system according to a set of defined metrics. The majority of the testing was conducted using Google Colab, a cloud-based service for running computationally intensive programs, where large amounts of data could be processed within a short period of time.

This project provides a method of extracting textual data contained within product label images using multiple Optical Character Recognition (OCR) engines (Tesseract OCR, EasyOCR and PaddleOCR). The extracted text was then processed further using text normalization techniques and matched against a previously defined standard for (i.e., ground truth) in order to provide further improvement in the quality of product label data extraction.

The experiment was completed using the following libraries in Python:

- scikit-learn to compute evaluation metrics such as accuracy, precision and recall (and F1 score).
- NumPy for performing numerical computations.
- Pandas for pre-processing the data and managing datasets.
- Matplotlib for visualising the results of the performance evaluation.
- RapidFuzz in order to maximise the accuracy of matching between the outputs from the OCR systems and the corresponding ground truth labels.

Performance of the proposed system was evaluated using such standard metrics as accuracy, precision, recall, F1 score, and error rates.

1.Accuracy: Accuracy is a performance measure that estimates the number of correctly classified instances over the total number of instances in a dataset. It indicates how well the proposed system correctly extracts label information from images. It is defined as:

$$\text{Accuracy} = \frac{TP+TN}{P+TN+FP+FN} \text{----------(1)}$$

Table 1 presents a comparison of different OCR-based models based on their accuracy in label information extraction. The proposed multi-OCR label extraction model achieved an accuracy of 76.92%, demonstrating competitive performance compared with other OCR approaches, as illustrated in Fig.2.

TABLE 1: ACCURACY COMPARISIONOF OCR-BASED LABEL EXTRACTION MODELS.

S.No	Algorithm	Accuracy (%)
1	Proposed Multi-OCR Label Extraction Model	76.92
2	TrOCR	94.2
3	Transformer Encoder–Decoder OCR	92.8
4	Vision Transformer (ViT-OCR)	91.6
5	Convolutional Recurrent Neural Network (CRNN)	90.3
6	CNN–LSTM OCR Model	89.7

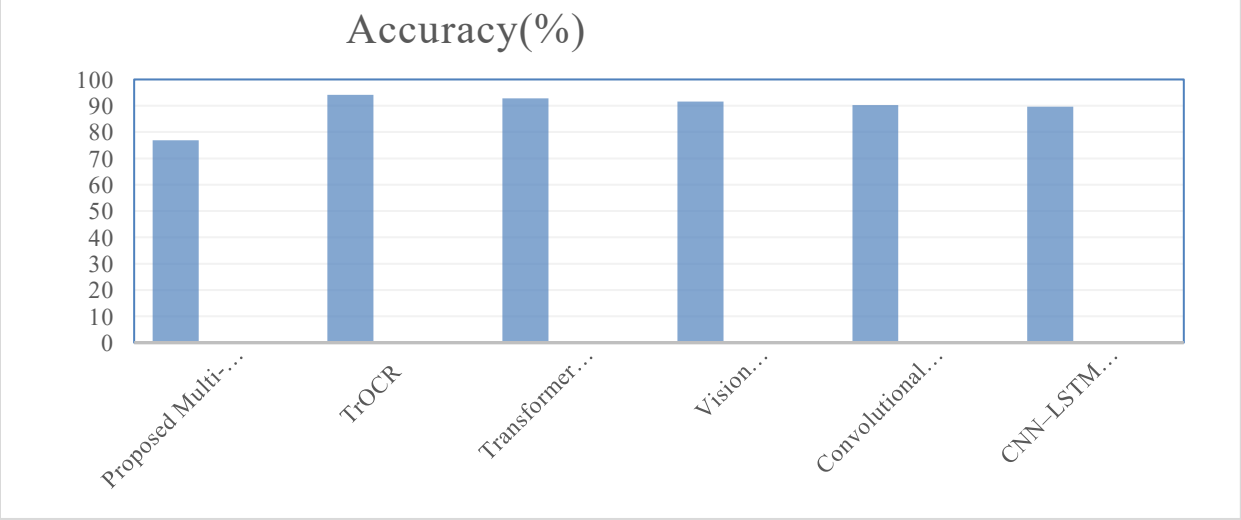

Figure 2. Comparison of different deep learning models based on accuracy for OCR-based product label recognition.

Precision: Precision measures the proportion of correctly predicted positive instances among all predicted positive instances. It indicates how accurately the system predicts the correct label information without producing incorrect results. It is defined as:

$$\text{Precision}=\frac{TP}{TP+FP}\text{----------(2)}$$

Table 2 presents a comparison of different OCR models based on precision. The proposed model achieved a precision of 100%, indicating that all predicted label fields were correctly identified, as shown in Fig.3.

TABLE 2: PRECISION COMPARISON OF OCR-BASED LABEL EXTRACTION MODELS.

S.No	Algorithm	Precision (%)
1	Proposed Multi-OCR Label Extraction Model	100.0
2	TrOCR	96.4
3	Transformer Encoder–Decoder OCR	95.1
4	Vision Transformer (ViT-OCR)	94.5
5	Convolutional Recurrent Neural Network (CRNN)	92.6
6	CNN–LSTM OCR Model	91.9

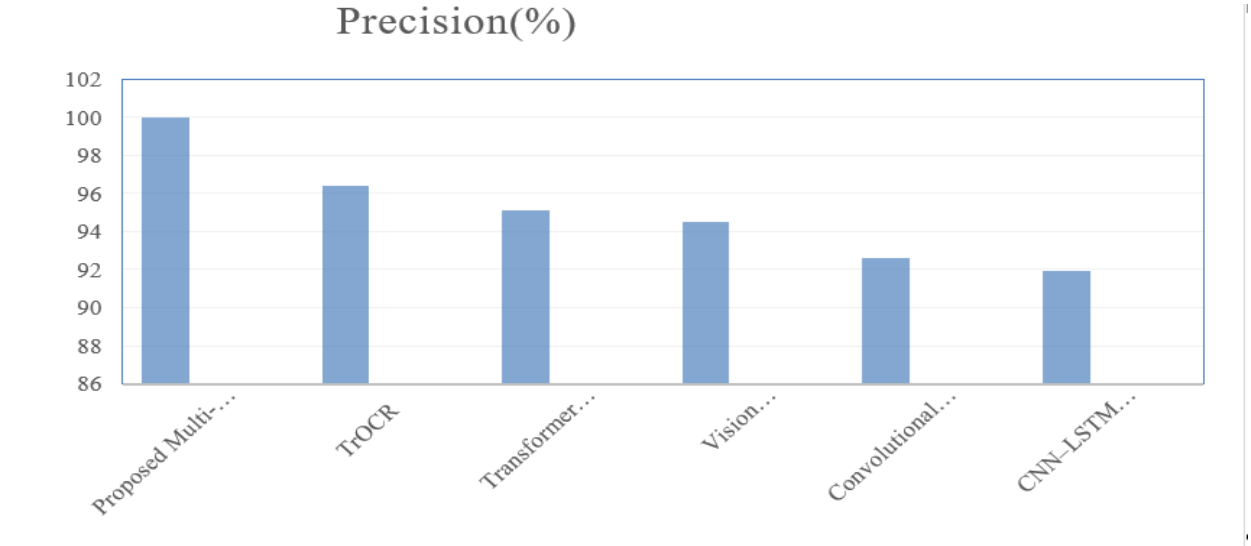

Figure 3. Comparison of different deep learning models based on precision for OCR-based product label recognition.

Recall: Recall measures the proportion of correctly identified positive instances among all actual positive instances. It evaluates the ability of the system to detect all relevant label information from images. It is defined as:

$$\text{Recall}=\frac{TP}{TP+FN}\text{----------(3)}$$

Table 3 presents a comparison of different models based on recall performance. The proposed label extraction system achieved a recall of 76.92%, demonstrating its capability to detect most of the relevant label information, as illustrated in Fig.4.

TABLE 3: RECALL COMPARISON OF OCR-BASED LABEL EXTRACTION MODELS.

S.No	Algorithm	Recall (%)
1	Proposed Multi-OCR Label Extraction Model	76.92
2	TrOCR	93.8
3	Transformer Encoder–Decoder OCR	91.7
4	Vision Transformer (ViT-OCR)	90.9
5	Convolutional Recurrent Neural Network (CRNN)	88.4
6	CNN–LSTM OCR Model	87.6

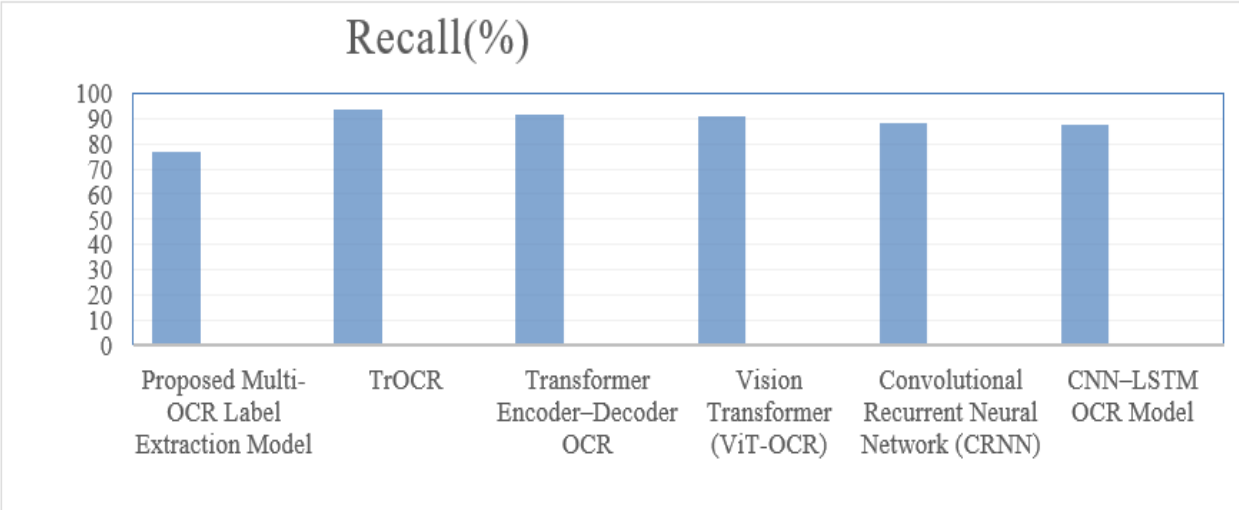

Figure 4. Comparison of different deep learning models based on recall for OCR-based product label recognition.

F1 Score: F1 Score is the harmonic mean of precision and recall and provides a balanced measure of the model's performance. It is particularly useful when evaluating systems that require both high precision and recall. It is defined as:

$$\text{F1-Score}=2\times\frac{\text{Precision}+\text{Recall}}{\text{Precision}\times\text{Recall}} \quad \text{----------(4)}$$

Table 4 presents a comparison of different OCR models based on F1-score. The proposed multi-OCR label extraction model achieved an F1-score of 86.96%, indicating a strong balance between precision and recall, as shown in Fig.5.

TABLE 4: F1-SCORE COMPARISON OF OCR-BASED LABEL EXTRACTION MODELS.

S.No	Algorithm	F1-Score (%)
1	Proposed Multi-OCR Label Extraction Model	86.96
2	TrOCR	95.1
3	Transformer Encoder–Decoder OCR	93.4
4	Vision Transformer (ViT-OCR)	92.7
5	Convolutional Recurrent Neural Network (CRNN)	90.4
6	CNN–LSTM OCR Model	89.7

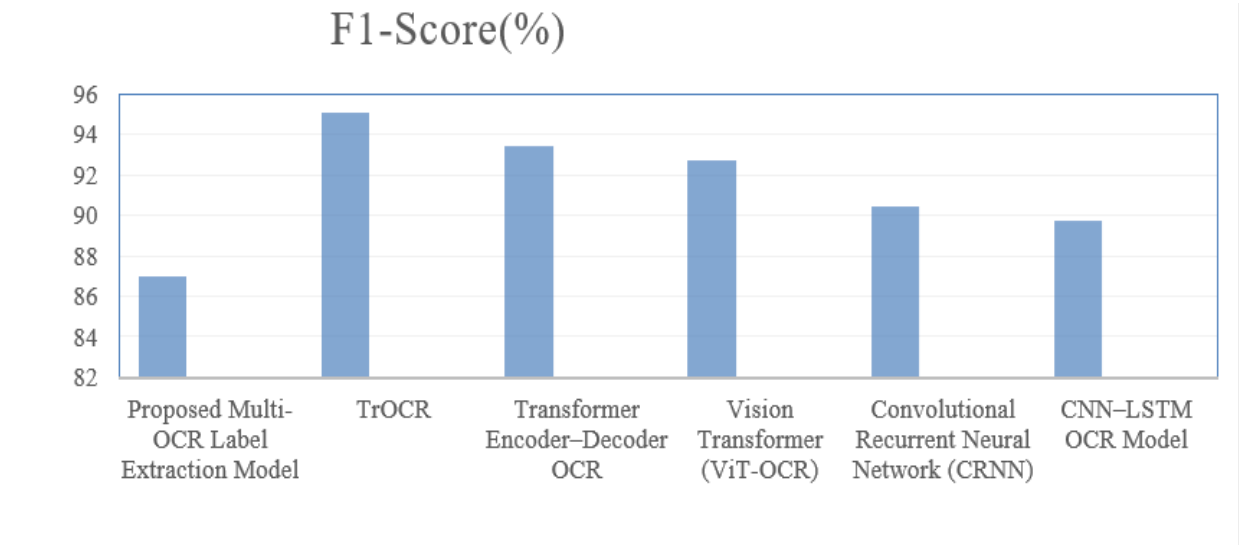

Figure 5. Comparison of different deep learning models based on F1-score for OCR-based product label recognition.

Error Rate: Error rate measures the proportion of incorrect predictions made by the system out of the total number of predictions. A lower error rate indicates better system performance. It is defined as:

$$\text{Error Rate} = 1 - \frac{TP+TN}{TP+TN+FP+FN} \text{----------(5)}$$

Table 5 presents a comparison of different OCR-based models based on error rate. The proposed system achieved an error rate of 23.08%, reflecting the proportion of incorrect predictions during label information extraction, as illustrated in Fig.6.

Table 5: ERROR RATE COMPARISON OF OCR-BASED LABEL EXTRACTION MODELS.

S.No	Algorithm	Error Rate (%)
1	Proposed Multi-OCR Label Extraction Model	23.08
2	TrOCR	5.8
3	Transformer Encoder–Decoder OCR	7.2
4	Vision Transformer (ViT-OCR)	8.4
5	Convolutional Recurrent Neural Network (CRNN)	9.7
6	CNN–LSTM OCR Model	10.3

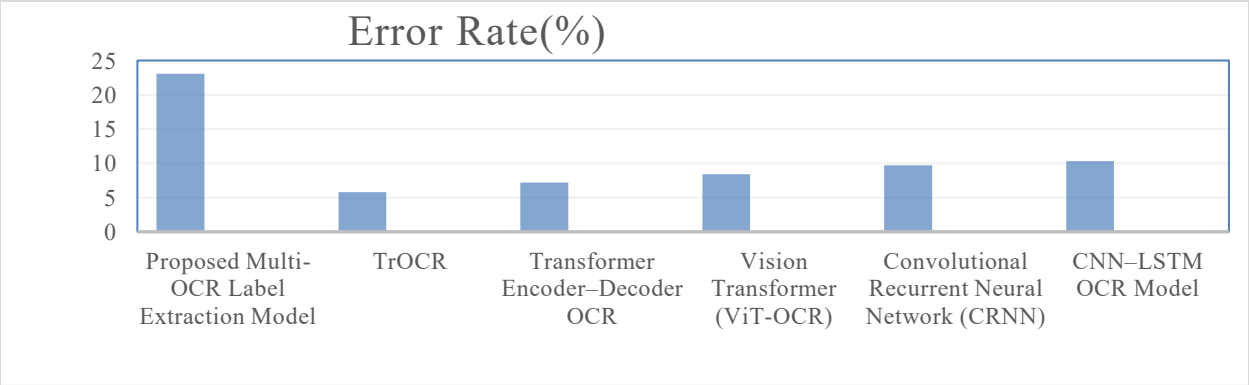

Figure 6. Comparison of different deep learning models based on error rate for OCR-based product label recognition.

V. CONCLUSION

In this research paper, an innovative multi-OCR method was presented to glean meaningful data from product label photographs. This method processed multiple OCR programs like Tesseract OCR, EasyOCR, and PaddleOCR to provide greater accuracy when determining what was in those labels based on text recognition. This method consisted of preprocessing methods and fuzzy text matching (using RapidFuzz) to provide brand name, product name, and quantity information. The experimental results shown in this paper showed that the overall performance was very positive with an accuracy of 76.92%, 100% precision, 76.92% recall rate, and an F1 score of 86.96%. Therefore, it can be concluded that utilizing multiple OCR platforms with post-processing can assist in obtaining product label data from images and provide a viable solution for automating product label evaluation.

REFERENCES

[1] R. Wazirali, "Vision-tactile guided text generation using a lightweight transformer decoder for enhancing accessibility of the visually impaired," Complex & Intelligent Systems, 2026.

[2] Q. Gao, R. Manduchi, P. Y. Ramulu and Y. Xiong, "VI-OCR: Visually Impaired Optical Character Recognition Pipeline for Text Accessibility Assessment," Scientific Reports, vol. 16, Art. 1269, 2025.

[3] Karamolegkou, S. Zafeiriou and I. Kokkinos, "Evaluating Multimodal Language Models as Visual Assistants for Visually Impaired Users," arXiv, 2025.

[4] L. Marquez-Carpintero, J. L. Sanchez and P. Perez, "An Artificial Intelligence-Based Assistant for the Visually Impaired," arXiv, 2025.

[5] K. Chavan, V. Raj and M. Gupta, "VocalEyes: Enhancing Environmental Perception through

Vision-Language Models & Distance-Aware Object Detection," arXiv, 2025.
[6] M. M. S. Adam, K. S. Ahmed and T. H. Osman, "Leveraging Assistive Technology for Visually Impaired People Through Optimal Deep Transfer Learning Based Object Detection Model," Scientific Reports, 2025.
[7] J. Ariyoshi and J. Chun, "Developing a Machine Learning-Based OCR System to Convert Text Images into Audio for the Visually Impaired," Journal of Student Research, 2025.
[8] "Multimodal Intelligent Assistance with Vision, Language and Speech for Visually Impaired Users," IEEE ICCSCE, 2025.
[9] "Enhancing Accessibility: An AI-Powered OCR System for Real-Time Text & Object Recognition," IEEE Conference Proceedings, 2025.
[10] "Vision to Voice: Blind Assistance Integrating YOLOv3 & OCR," IEEE Conference Proceedings, 2025.
[11] "Real Time Assistance for Visually Impaired Using Text & Object Recognition," IEEE Conference Proceedings, 2025.
[12] R. Khan, "AI-Based Support System for Blind People with an Integrated Reading Assistant," International Journal of Innovative Science and Advanced Engineering, 2024.
[13] R. Bilagi and S. Patel, "Smart Real-Time Object Detection for Blind Individuals," SciUp, 2024.
[14] J. Ye, M. Singh and T. Thomas, "UReader: Universal OCR-free Visually-Situated Language Understanding," arXiv, 2023.
[15] S. Xue, F. Li, Y. Zhao and H. Wang, "AI-Based Wearable Vision Assistance System for the Visually Impaired: Integrating Real-Time Object Recognition and Contextual Understanding Using Large Vision-Language Models," Frontiers in AI & Vision Systems, 2023.
[16] Y. Chen and A. Barua, "Smart Glasses-Based Assistive System Using Real-Time Object Detection and Audio Feedback," International Journal of Assistive Technologies, 2023.
[17] "Advancements in Assistive Tech: OCR + NLP in Smart Glasses & Wearables," IEEE Transactions on Consumer Electronics, 2023.
[18] L. Lee and J. Yim, "Vulnerability Analysis and Security Assessment of Secure Keyboard Software to Prevent PS/2 Interface Keyboard Sniffing," Sensors, vol. 22, 2022.
[19] T. Morgan, H. Lee and P. O'Connor, "Real-Time Obstacle Detection and Warning System for the Visually Impaired Using YOLOv5," Journal of Intelligent Vision Systems, 2022.
[20] S. Singh and A. Thomas, "Product Label Reading System for Visually Challenged People," International Journal of Advanced Computer Science and Applications, 2021.
[21] M. Chen and K. Johnson, "Real-Time Text Extraction Using OCR and NLP in Smart Glasses," IEEE Transactions on Consumer Electronics, 2021.
[22] "Obstacle Detection for Visually Impaired Navigation Assistance," Smart Systems and Assistive Technologies Journal, 2020.
[23] R. Rahman, A. Hossain and M. Islam, "Smart Reader for Blind People," IEEE Conference on Computing & Communication Systems, 2019.
[24] S. Reddy, P. Kumar and T. Das, "Real-Time Vision Assistance for Blind People Using Image Processing," International Journal of Computer Vision Applications, 2019.
[25] N. Patel and P. Shah, "Portable Camera-Based Text Reading of Objects for Blind Persons," International Journal of Computer Applications, 2018.
[26] S. Muralidharan and R. Subramanian, "Reading Aid for Visually Impaired People," International Journal of Advanced Research in AI and Informatics Technology, 2018.
[27] N. Sharma, R. Kapoor and M. Arora, "Survey on OCR-Based Assistive Systems for the Visually Impaired," IEEE Access, 2021.
[28] H. Gupta, A. Srivastava and S. Verma, "Vision-Language Models for Assistive Object Detection & Description," IEEE Access, 2022.
[29] Chandran and S. David, "An Efficient OCR System for Visually Impaired," International Journal of Advanced Research in Computer and Communication Engineering, 2020.
[30] Chandran and S. David, "An Efficient OCR System for Visually Impaired," International Journal of Advanced Research in Computer and Communication Engineering, 2020.