International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 44 No. 1 Nov. 2024, pp. 1-10
© 2024 Innovative Space of Scientific Research Journals
http://www.ijias.issr-journals.org/

Smart Learning Tool for Kids with Real-Time Image Classification

Moe Moe Zaw and Hla Hla Myint

Department of Information Technology Supporting and Maintenance, University of Computer Studies (Magway), Magway, Myanmar

ABSTRACT: The Smart Learning Tool for Kids with Real-time Image Classification is an AI-powered educational tool designed to assist young learners in recognizing and identifying objects through real-time image classification. The system captures images using a webcam, processes them through a Convolutional Neural Network (CNN) model and outputs the corresponding class label. It provides immediate audio feedback by pronouncing the class name in four languages: English, Myanmar, Thai and Chinese. The system aims to enhance kids' learning experience by engaging multiple senses—visual and auditory—that makes learning interactive and multilingual. The CNN model is trained with custom training data, enabling accurate classification of 12 object classes. This system serves as a smart and user-friendly tool for early childhood education.

KEYWORDS: Real-time Image Classification, Convolutional Neural Network (CNN), Multilingual Audio Feedback, Early Childhood Education, Interactive Learning, AI in Education.

1 INTRODUCTION

Technology is advancing rapidly, transforming many aspects of our lives, including education. The integration of modern technology into learning environments offers significant advantages. One major benefit is the ability to create more engaging and interactive learning experiences. Advanced tools and systems can make learning more enjoyable and effective by providing real-time feedback and personalized support.

Technology also supports cognitive growth by aiding in skills like recognizing objects and learning languages. Tools that support multiple languages and interactive materials help make education more accessible and inclusive for everyone regardless of their background.

According to [1], computer vision has become increasingly intriguing in recent years, particularly with the rise of self-driving cars. Generic object detection focuses on identifying and classifying objects within a single image, marking them with rectangular bounding boxes and indicating their confidence levels. Today, the advancement of object detection has been significantly enhanced by deep convolutional neural networks (CNNs), which have recently become the leading approach in object recognition research due to their exceptional performance on various challenging datasets, like ImageNet, gathered from the web.

In exploring the role of AI in education, [2] provides a comprehensive overview of how AI can enhance personalized learning and address ethical concerns like data privacy. Similarly, in [6], the authors focus on designing AI-driven educational systems that cater to the evolving needs of learners. Both books offer valuable insights into the integration of AI in educational environments, highlighting the potential and challenges of smart learning systems.

In [3], the project focuses on enhancing real-time object detection for autonomous navigation in fluvial environments by evaluating several deep learning algorithms, including Faster R-CNN, SSD, and various versions of YOLO. The study addresses the lack of relevant datasets by creating and openly sharing a dataset of 2,488 images with over 35,000 annotations across five classes: vessel, person, riverside, road signals, and infrastructure. Initial experiments show that these models can effectively detect and track objects in near real-time, proving their potential for autonomous vessels. Future work will involve expanding the dataset and improving detection accuracy through advanced mathematical methods to better define navigation areas.

In [4], they propose a model that fuses features from multiple layers of the CNN to leverage the discriminative power of both lower and higher layers. Utilizing the CIFAR-10 dataset and MatConvNet for efficient CPU training, our model aims to improve image recognition accuracy by incorporating more convolution layers and hidden neurons. This approach not only enhances object recognition but also paves the way for future advancements in real-time image recognition systems, demonstrating the significant potential of CNNs in advancing artificial intelligence and computer vision technologies.

In [5], This article explores Collaborative Interactive Machine Teaching (IMT), focusing on how groups of users can collectively structure the teaching process for image classification. It introduces TeachTOK, a web application that allows teams to curate data and train a model together. The study, involving ten participants divided into three competing teams, revealed collaborative patterns and insights into teaching strategies and user experiences with the application. The findings offer valuable implications for designing more interactive, collaborative, and participatory machine learning systems.

In [7], The review explores the evolution of CNNs from their early successes to their central role in the deep learning boom post-2012. It highlights their transformative impact on image classification, significant symbolic contributions to their popularity, and ongoing improvements and challenges based on a comprehensive analysis of over 300 publications. The review also touches on current trends and areas requiring further development.

In [8], the research introduces a novel custom CNN model, MedvCNN, and a hybrid model, MedvLSTM, for medical image classification. By integrating CNNs for feature extraction and LSTMs for handling sequential data, the study aims to improve classification accuracy and efficiency. Additionally, they present a real-time web-based AutoML framework to streamline the classification process. This work enhances the capabilities of medical image analysis through sophisticated neural network approaches. The study emphasizes the importance of thorough model evaluation and customization, and plans to advance further by incorporating neural architecture search (NAS).

In the field of deep learning, [9] offers a detailed exploration of how deep learning techniques, particularly convolutional neural networks (CNNs), are applied to computer vision tasks. [10] provides practical guidance on implementing deep learning models using Python, making it accessible for developers and researchers looking to leverage deep learning for various applications.

2 System Overview

This section provides an overview of the Smart Learning Aid for Kids with Real-time Image Classification. The system is designed to support early childhood learning by combining real-time object detection with interactive, multilingual audio feedback. The key features and object classes are outlined in the following subsections.

2.1 Language

The Smart Learning Aid for Kids with Real-time Image Classification processes live images captured by a webcam. The system preprocesses these images for accurate classification and outputs the object label with a confidence level. It also provides multilingual audio feedback in English, Myanmar, Thai and Chinese, enhancing both language learning and cognitive development. This approach integrates real-time object detection with interactive, auditory learning for young children.

2.2 Key Features

- Real-Time Object Detection: Objects are detected instantaneously as they appear in the webcam feed.
- Confidence Level: The system displays a confidence level percentage to indicate the accuracy of its predictions.
- Multilingual Audio Feedback: For every detected object, the system provides audio feedback in four different languages: English, Myanmar, Thai and Chinese.

2.3 Object Classes

The model is trained to recognize 12 different classes of objects commonly found in a kids' environment. These classes were chosen specifically for their relevance to young learners, ensuring that the system remains age-appropriate and stimulating.

3 Model Development

This section outlines the steps taken to develop the Convolutional Neural Network (CNN) model used for real-time object recognition in the Smart Learning Aid system. The process includes data collection, preprocessing, model architecture, training process, model integration and postprocessing output.

3.1 Data Collection And Preparation

In the data collection process, a total of 3750 images were gathered from Google Images and Bing Images. Of these, 3,000 images were allocated for training the model, while 750 images were reserved for testing. The collected images are distributed across the 12 target classes providing a balanced dataset for model training and testing.

3.2 Model Architecture

The system uses a CNN with several convolutional layers, pooling layers, and fully connected layers. Below is the detailed architecture:

3.2.1 Input Layer

The input shape for the Convolutional Neural Network (CNN) is specified as (64, 64, 3), indicating that the images fed into the model are resized to a resolution of 64x64 pixels. The "3" in the input shape represents the three color channels (RGB) of the images, which allows the network to process color information in addition to spatial features.

3.2.2 Convolutional Layers

The Convolutional Neural Network (CNN) employed in this study consists of a series of convolutional and pooling layers designed to extract and process features from images. The first convolutional layer utilizes 16 filters with a kernel size of (3, 3), enabling the model to learn various spatial features from the input images. This layer is followed by a MaxPooling layer with a pooling size of (2, 2), which effectively reduces the spatial dimensions of the feature maps by half, thereby decreasing computational complexity and helping to control overfitting.

The second convolutional layer employs 32 filters with the same kernel size of (3, 3), and also uses the ReLU (Rectified Linear Unit) activation function. After this layer, a second MaxPooling layer with a pooling size of (2, 2) further reduces the dimensions of the feature maps.

Following this, a third convolutional layer applies 64 filters with a kernel size of (3, 3), followed by a third MaxPooling layer with the same (2, 2) pooling size. Finally, a fourth convolutional layer uses 128 filters and is followed by a fourth MaxPooling layer, further refining the feature maps and reducing spatial dimensions.

3.2.3 Flatten Layer

This layer flattens the 2D feature maps from the convolutional layers into a 1D vector, so they can be passed to the fully connected layers.

3.2.4 Fully Connected Layers

The CNN includes a fully connected (dense) layer with 128 neurons, utilizing the ReLU activation function to enable the network to capture and learn complex patterns from the data. The final output layer consists of 12 neurons, each representing one of the classes. This output layer employs the Softmax activation function, which transforms the raw output values into probabilities, allowing the network to classify inputs into one of the 12 predefined categories.

Table 1. Parameter Setting of CNN

Layer Type	Number of Filters/Neurons	Kernel Size / Pooling Size	Activation Function
Convolutional	16	(3, 3)	ReLU
MaxPooling	-	(2, 2)	-
Convolutional	32	(3, 3)	ReLU
MaxPooling	-	(2, 2)	-
Convolutional	64	(3, 3)	ReLU
MaxPooling	-	(2, 2)	-
Convolutional	128	(3, 3)	ReLU
MaxPooling	-	(2, 2)	-
Flatten	-	-	-
Fully Connected	64	-	ReLU
Output	12	-	Softmax

3.3 Training Process

The system uses a CNN architecture for image classification with several convolutional layers, pooling layers, and fully connected layers. The model is trained using our own dataset. Then, the trained model is saved as an.h5 file.

Conversion to TensorFlow Lite

The convert.py is used to convert the trained.h5 model file to a TensorFlow Lite (.tflite) format.

3.4 Model Integration In The System

The.tflite mode is loaded in the system for image classification.

3.5 Post-Processing And Output

The image is labeled into predefined classes based on the model’s classification result. The system shows the class label and its confidence level. Then, the system provides audio feedback in different languages (English, Myanmar, Chinese and Thai) to the kids.

4 Implementation

The system is built using a CNN model for real-time object recognition, integrated with a simple GUI for child interaction. TensorFlow and OpenCV are used for image processing. The pre-recorded audio files are used multilingual feedback.

4.1 SYSTEM FLOW

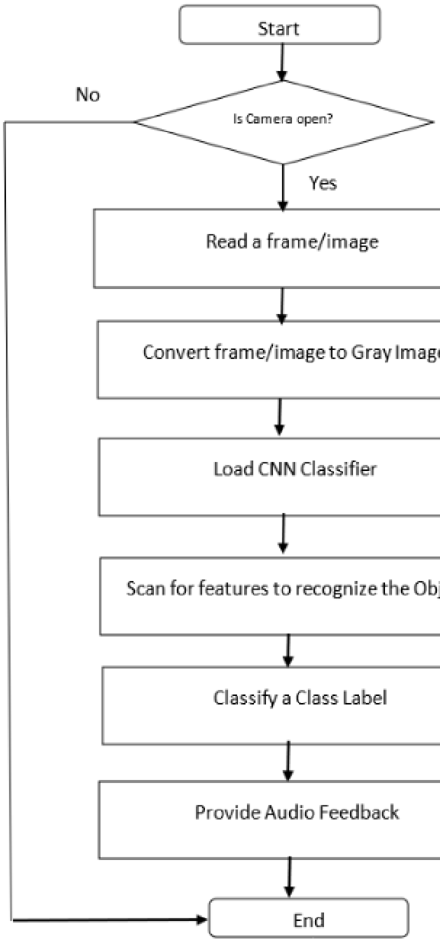


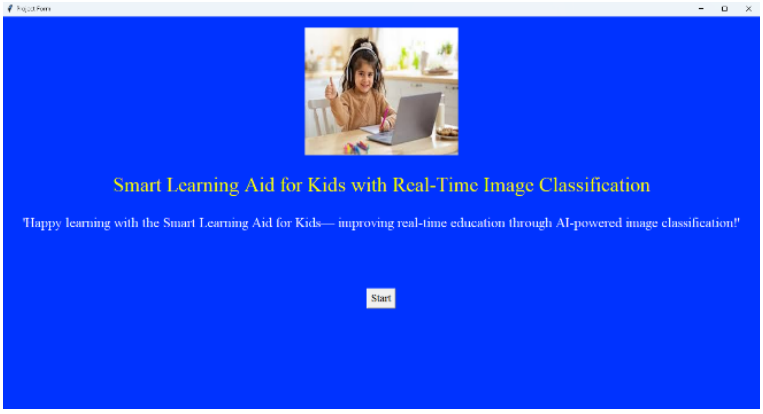
Fig. 1. System Flowchart

The webcam continuously captures frames, which are processed by the trained CNN model to identify objects. The system updates the screen with the recognized object label and the associated confidence level. Once an object is detected, the system uses pre-recorded audio files to announce the object's name in the five supported languages. This multilingual feature helps children develop language skills beyond their native language.

4.2 USER INTERFACE

The graphical user interface (GUI) is simple and child-friendly, with the following key elements:

- A live feed from the webcam.
- Detected object label and confidence score.
- Language selection buttons for switching between English, Myanmar, Thai and Chinese audio feedback.

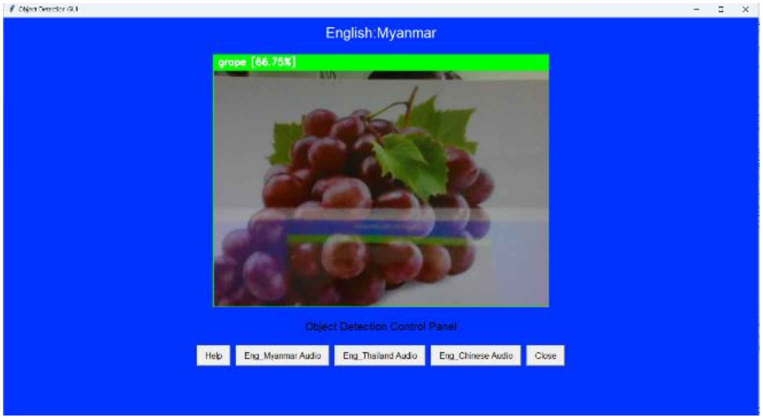


Project Form

Smart Learning Aid for Kids with Real-Time Image Classification

'Happy learning with the Smart Learning Aid for Kids— improving real-time education through AI-powered image classification!'

Fig. 2. Main Form of the System



Object Detection GUI

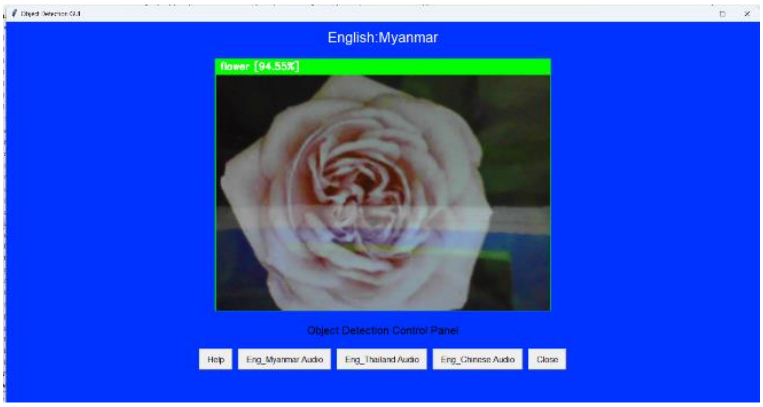
English:Myanmar

grape [66.75%]

Object Detection Control Panel

Fig. 3. Classification of Grape

Fig. 3 shows the detection of grape with 66.75 confidence level. It's pronounced as "grape စပျစ်သီး " in English-Myanmar language audio feedback.



Object Detection GUI

English:Myanmar

flower [94.55%]

Object Detection Control Panel

Fig. 4. Classification of Flower

Fig. 4 shows the detection of flower with 94.55 confidence level. It's pronounced as "flower ပန်း " in English Myanmar language audio feedback.

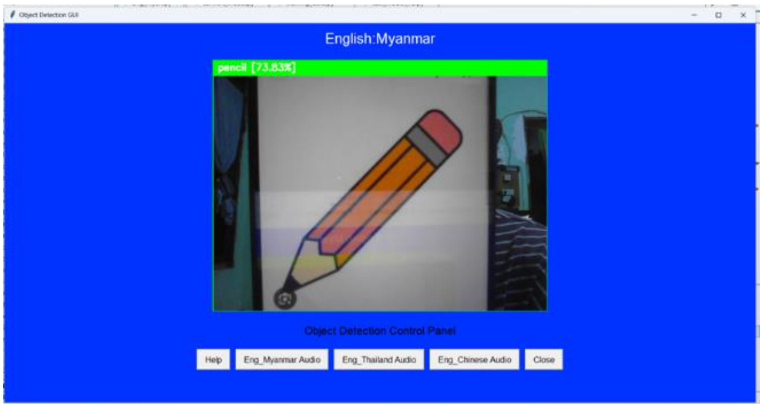


Fig. 5. Classification of Pencil

Fig. 5 shows the detection of pencil with 73.83 confidence level. It's pronounced as "pencil ခဲတံ " in English Myanmar language audio feedback.

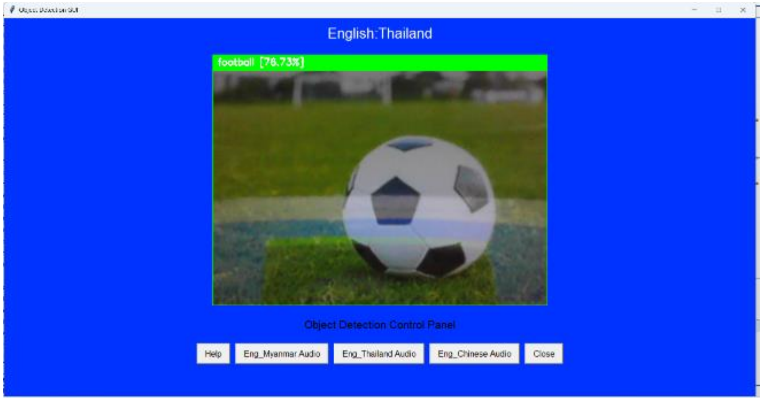


Fig. 6. Classification of Ball

Fig. 6 shows the detection of football with 76.73% confidence level. It's pronounced as "football ลูกบอล " in English Thai language audio feedback.

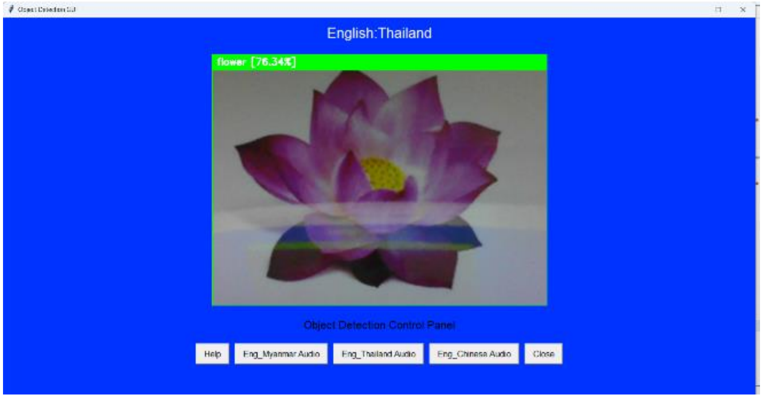


Fig. 7. Classification of Flower

Fig. 7 shows the detection of clock with 76.34% confidence level. It's pronounced as "flower ดอกไม้" in English Thai language audio feedback.

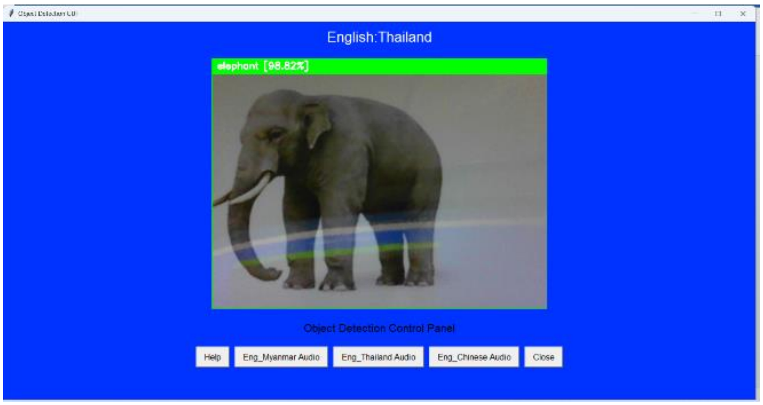


Fig. 8. Classification of Elephant

Fig. 8 shows the detection of elephant with 98.82% confidence level. It’s pronounced as “elephant ช้าง ” in English Thai language audio feedback.

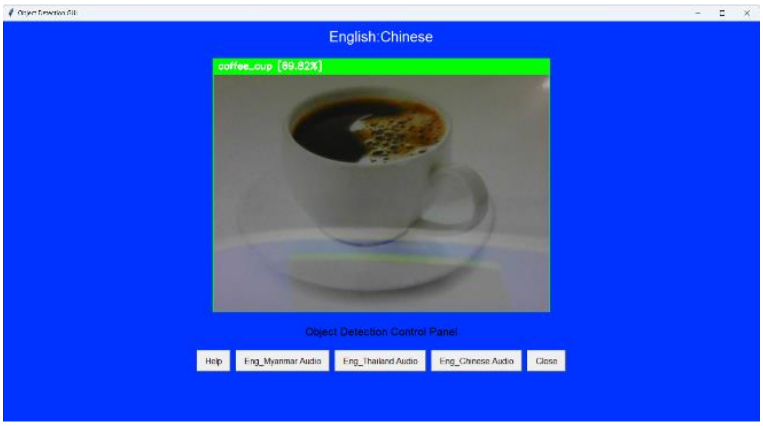


Fig. 9. Classification of Coffee_Cup

Fig. 9 shows the detection of coffee-cup with 69.82% confidence level. It’s pronounced as “coffee_cup 咖啡杯 ” in English Chinese language audio feedback.

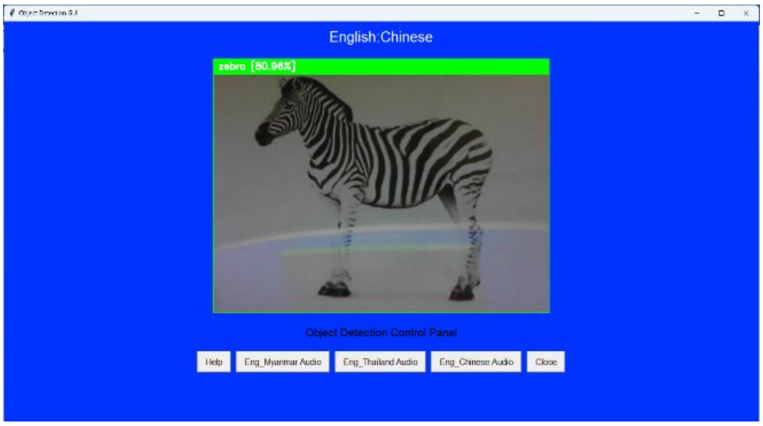


Fig. 10. Classification of Zebra

Fig. 10 shows the detection of zebra with 80.98% confidence level. It’s pronounced as “zebra 斑马 ” in English Chinese language audio feedback.

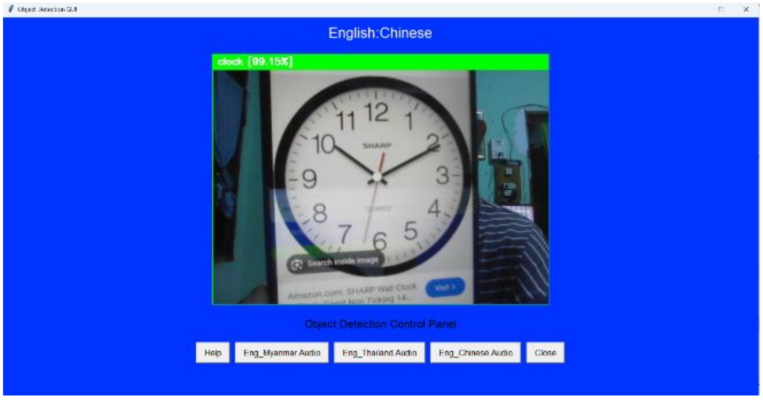


Fig. 11. Classification of Clock

Fig. 11 shows the detection of clock with 99.15% confidence level. It's pronounced as "clock 时钟" in English Chinese language audio feedback.

5 RESULTS AND EVALUATION

In this evaluation, we assess the performance of our image classification system designed to identify and classify objects into twelve distinct categories: clock, coffee cup, deer, elephant, flower, football, fork, grape, pencil, penguin, pig and zebra. The system was tested with a dataset comprising 750 images, from which key metrics such as Precision, Recall, and F1 Score.

5.1 ACCURACY AND PERFORMANCE

Table 2. Classification Results Overview

Metric	Formula	Value
True Positive (TP)	-	342
False Positive (FP)	-	46
False Negative (FN)	-	75
True Negative (TN)	-	287
Precision	$\frac{TP}{TP+FP}$	0.88 (88%)
Recall	$\frac{TP}{TP+FN}$	0.82 (82%)
F1 Score	$2 * \frac{Precision * Recall}{Precision + Recall}$	0.85 (85%)

The model was tested on 750 images, and the results showed a precision of 88%, meaning 88% of the predicted positive instances were correct. The recall was 82%, indicating the model successfully identified 82% of the actual positive instances. The F1 score, which balances precision and recall, was 85%, showing that the model is good at finding the correct positives and making fewer mistakes.

5.2 USER TESTING

User testing was conducted with a group of 20 children aged 4-7. The results indicated high engagement and positive feedback, with children showing a strong interest in interacting with the system and improving their object recognition and language skills.

6 CONCLUSION

The "Smart Learning Aid for Kids with Real-time Image Classification" is the integration of AI based technology into education and offers many benefits for learners, educators, institution and society. The system is a combination of playing and learning to support cognitive development in kids designed to enhance children's educational experiences by combining visual and auditory learning. It supports language acquisition, cognitive development, and technological literacy, making it an invaluable resource for young learners. Future work will include expanding the number of object classes, improving language pronunciation accuracy and adding more languages for feedback.

REFERENCES

[1] I.Salehin, M. S. Islam, N. Amin, M. A. Baten, S. M. Noman, M. Saifuzzaman, and S. Yazmyradov, «Real-Time Medical Image Classification with ML Framework and Dedicated CNN–LSTM Architecture,» *Journal of Sensors,* 2023.

[2] W. Rawat and Z. Wang, «Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review,» 2017.

[3] M. A. Hossain and M. S. A. Sajib, «Classification of Image using Convolutional Neural Network (CNN),» *Global Journal of Computer Science and Technology: D Neural & Artificial Intelligence,* vol 19, pp.12-18, 2024.

[4] B. Mohammadzadeh, J. Françoise, M. Gouiffès, and B. Caramiaux, «Studying Collaborative Interactive Machine Teaching in Image Classification,» *29th International Conference on Intelligent User Interfaces,* pp-195-208,2024.

[5] W. Hammedi, M. Ramirez-Martinez, P. Brunet, S. M. Senouci, and M. A. Messous, «Deep Learning-Based Real-time Object Detection in Inland Navigation,» 2023.

[6] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, «ImageNet: A large-scale hierarchical image database,» IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[7] B. du Boulay and A. Luckin, «Artificial Intelligence in Education: Promises and Implications for Teaching and Learning,» 2019.

[8] J. F. Pane and B. du Boulay, «Designing Smart Teaching and Learning Systems: Perspectives on the Future of Education,» 2021.

[9] R. Shanmugamani, Deep Learning for Computer Vision, 2018.

[10] I. Vasilev and D. Slater, *Python Deep Learning,* 2nd Ed, 2017.