Object Recognition and Verbalization Tool For Early Childhood Education

Project Work Synopsis

Submitted in the partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING
IN
COMPUTER SCIENCE WITH SPECIALIZATION IN
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Submitted by:

21BCS6729 DIPIN RAJ
21BCS6634 RASHAZ RAFEEQUE
21BCS6589 JEEVAN A.J
21BCS6272 RHISHITHA T.S

Under the Supervision of:

Ms. Kirti

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB

Oct, 2024

Abstract

This project is an innovative solution to early childhood learning that helps children get engaged to learn through games while adapting to their individual needs and learning styles. It lets toddlers navigate through the physical environment, identify objects in the vicinity in the blink of an eye and. describe them in words. This not only helps them expand their lexicon, but also increases their knowledge of the environment. This is through use of both active and intelligent interfaces that provide the child with personalized learning process depending on the child’s learning rates, preference and needs. It is always fun when children are challenged to participate in their learning, this is made possible through the involvement of games and other fun challenges which help their young brains to be provoked making learning more effective. The tool helps to design an environment that children of such age can touch an object, inquire, and get an immediate response. The given system of learning which makes a child learn through experiencing daily situations and cognitive activities rewires children’s approach to education making it engaging, valuable, and applicable. Moreover, the use of the object recognition and natural language processing makes the tool an easy-to-use intelligent educational assistant. In so doing, it reduces the gap between visual and auditory learning for the children, thus fostering an environment that not only gets the attention of the child, but also, utilizes their knowledge as the child learns. Such smooth incorporation of the technology in play aims help the children learn at the early stage in the school, and with ease since it can be fun, exciting, and very effective at the same time.

Table of Contents

1. INTRODUCTION

In recent years, there has been a considerable upward evolution in the setting that shapes early childhood by the incorporation of technology, especially AI and machine learning. The use of ordinary learning aids and techniques has slowly been changed to complicated apparatus such as the recognition of objects as well as the verbalization system that helps in teaching young children through interactive processes. Formally early childhood learning was characterized by the use of tangible objects and manipulation to pass on concepts. Objects like picture books, flash cards and baby toys are more appropriate in as much as they ensured both the visual and touch stimuli were incorporated in the learning process. Though these methods worked they failed to provide appropriate feedback and suit the learning style of the students.

AI brought into the usage more interactive and an effective education system compared to traditional methods. For instance, Convolutional Neural Networks (CNNs) have applied a significant boost in the improvement of the object recognition technology. These models are very effective in robots for image recognition due to their ability to categories the images based on several patterns and characteristics that are perceived in the field of vision. CNNs have been used in several fields mainly in the medical field or industries because of flexibility they have to reveal more features in data. The incorporation of the object recognition systems into the educational tools can therefore be considered as the next big step. These tools are capable of real time object identification and categorization through models such as CNNs, RCNNs, SSD, YOLO. Such capability facilitates development of tree-like learning spaces where the children can receive relevant and timely feedback through verbal descriptions in the context of the learning space.

The enhancement of the above educational tools by integrating game-based learning strategies sharpens it. With the help of game and challenge concepts, which are implemented with AI, the children engage more actively with content. Besides making learning an interesting process, it helps in the cognitive ability of children since they are able to solve problems as well as think logically. Despite the profound advancements witnessed in the educational system in the last few years, recent advancements in AI and Machine learning are still redefining the educative technology. Various current applications integrate features such as object recognition in real time with verbal feedback that is also real-time and changes depending on a person. All these improvements intended to contribute to the closing between the conventional model of learning approach and the innovative way of the highly technological learning environment. The shift from the mechanical educational instruments to brand new AI-based systems demonstrates the course for the improvement in the environments to learn: more individual, more stimulating, and more efficient. This is a huge step towards ensuring that technology in early childhood education develops as an effective tool in delivering meaningful and personalized learning for young children through the use of object recognition and technologies that make it possible to verbalize.

1.1 Problem Definition

The problem addressed in this study revolves around the need for enhancing early childhood education through interactive and personalized learning tools. Traditional educational methods often struggle to provide real-time, individualized feedback and adapt to the diverse learning styles and paces of young children. The challenge is to develop a system that not only recognizes objects within a child's environment but also provides immediate, contextually relevant verbal descriptions to support learning. The problem of enhancing early childhood education is compounded by the limitations of conventional teaching tools, which typically lack the capability to offer real-time, personalized feedback tailored to each child's unique learning style. Traditional methods often rely on static, one-size-fits-all approaches that do not address the individual needs and learning paces of young learners. This can lead to a less engaging educational experience, where children may struggle to stay motivated or fully grasp new concepts. As a result, there is a pressing need for innovative tools that can dynamically adapt to each child's learning requirements and provide immediate, meaningful feedback.

1.2 Problem Overview

The evolution of educational technology highlights a significant gap in the effectiveness of traditional learning methods for early childhood education. Conventional tools often rely on static interactions and lack the capability to offer personalized feedback or adapt to the individual needs of young learners. This limitation can hinder cognitive and language development, as children may not receive the necessary support to fully understand and engage with educational content. Recent advancements in AI and machine learning, particularly in the fields of object recognition and natural language processing, offer promising solutions to these challenges. Object recognition models, such as CNNs, RCNNs, SSDs, and YOLO, have demonstrated the ability to accurately identify and categorize objects in real time. When combined with text-to-speech technologies, these models can provide immediate, verbal descriptions of objects, enhancing interactive learning.

The problem, therefore, is to design and implement a system that integrates these advanced technologies to create an interactive learning tool for toddlers. This system must be capable of recognizing objects in a child's environment, delivering contextually relevant verbal feedback, and adapting to individual learning needs. By addressing these challenges, the proposed system aims to significantly improve early childhood education by making learning more engaging, personalized, and effective.

1.3 Hardware Specification

- 11th Gen Intel® i7-11800H @ 2.30GHz
- 16 GB RAM. 256GB SSD 1TB HDD

1.4 Software Specification

- Operating System: Windows with appropriate drivers for GPU.
- Development Environment: Pycharm/Jupyter/Google Colab
- Python: Programming language for implementing neural networks and TTS integration.
- Libraries and Tools:
  - TensorFlow / Keras / PyTorch for building and training CNN, RCNN, SSD, and YOLO models.
  - OpenCV: For image processing and real-time video capture.
  - gTTS (Google Text-to-Speech) or pyttsx3 for converting object labels to audio.
  - Pandas / NumPy for dataset manipulation.
  - Matplotlib / Seaborn for visualizing model performance.
- Dataset: COCO (Common Objects in Context) and Pascal VOC for training and validating object detection models.

2. LITERATURE SURVEY

2.1 Literature Review Summary

Year and Citation	Article/ Author	Technique	Source	Evaluation Parameter
Hanafi, H.F., Wong, K.T., Adnan, M.H.M., Selamat, A.Z., Zainuddin, N.A. and Lee Abdullah, M.F.N., 2021. Utilizing Animal Characters of a Mobile Augmented Reality (AR) Reading Kit to Improve Preschoolers' Reading Skills, Motivation, and Self-Learning: An Initial Study. *International Journal of Interactive Mobile Technologies*, *15*(24).	HF Hanafi, KT Wong, M.H.M Adnan, Selamat, A.Z., Zainuddin, N.A. and Lee Abdullah	Augmented Reality(AR), Object Recognition	Research Gate	AR technologies improve reading skills and motivation in preschoolers, directly supporting our research into enhancing early childhood education through real-time object recognition and verbalization tools.
Wu, Q., Wang, S., Cao, J., He, B., Yu, C. and Zheng, J., 2019. Object recognition-based second language learning educational robot system for Chinese preschool children. *IEEE Access*, *7*, pp.7301-7312.	Wu, Q., Wang, S., Cao, J., He, B., Yu, C. and Zheng	Object Recognition, Kinect Technology	IEEE	Supports the potential of real-time object recognition and verbalization tools to improve cognitive development and teaching methods for young children.
Qi, S., Ning, X., Yang, G., Zhang, L., Long, P., Cai, W. and Li, W., 2021. Review of multi-view 3D object recognition methods based on deep learning. *Displays*, *69*, p.102053.	Qi, S., Ning, X., Yang, G., Zhang, L., Long, P., Cai, W. and Li, W.	Verbalization, Object Recognition, Deep Learning	ScienceDirect	Offering insights into improving accuracy and processing time through view-based methods and CNNs.
Bazargani, J.S., Sadeghi-Niaraki, A., Rahimi, F., Abuhmed, T. and Choi, S.M., 2022. An iot-based approach for learning geometric shapes in early childhood. IEEE Access, 10, pp.130632-130641.	Bazargani, J.S., Sadeghi-Niaraki, A., Rahimi, F., Abuhmed, T. and Choi, S.M	IoT, Object Recognition.	IEEE	Demonstrates how IoT can enhance early childhood education by making learning interactive and enjoyable.

Rahiem, M.D., 2021. Storytelling in early childhood education: Time to go digital. *International Journal of Child Care and Education Policy*, *15*(1), p.4.	Maila D.H. Rahiem	Text-to-Video Technology	SPRINGLER LINK	Highlights the benefits of digital storytelling in enhancing language and cognitive skills in young children.
Zaini, N.A., Noor, S.F.M. and Wook, T.S.M.T., 2019. Evaluation of api interface design by applying cognitive walkthrough. *International Journal of Advanced Computer Science and Applications*, *10*(2).	Nur Atiqah Zaini, Siti Fadzilah Mat Noor, Tengku Siti Meriam Tengku Wook	API Technology, Cognitive Learning	ResearchGate	Demonstrating how to design and evaluate an educational tool for young children using interactive technology.
Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W. White. 2019. VERSE: Bridging Screen Readers and Voice Assistants for Enhanced Eyes-Free Web Search. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS'19). Association for Computing Machinery, New York, NY, USA, 414–426. https://doi.org/10.1145/3308561.3353773	Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W. White.	NLP, VUI, ML, AI Models, IoT, Cloud Computing, APIs	ACM Digital Library	VERSE prototype likely includes usability metrics, user satisfaction, accessibility improvements, task completion rates, and time efficiency in retrieving information
E. de la Guía, V. L. Camacho, L. Orozco-Barbosa, V. M. Brea Luján, V. M. R. Penichet and M. Lozano Pérez, "Introducing IoT and Wearable Technologies into Task-Based Language Learning for Young Children," in IEEE Transactions on Learning Technologies, vol. 9, no. 4, pp. 366-378, 1 Oct.-Dec. 2016, doi:10.109/TLT.2016.2557333.	E. de la Guía, V. L. Camacho, L. Orozco-Barbosa, V. M. Brea Luján, V. M. R. Penichet and M. Lozano Pérez	IoT	IEEE Org	Includes student engagement levels, instructor ease of use with wearable and IoT technologies, the effectiveness of realistic scenarios in enhancing communication skills, and accuracy of performance metrics collected during task-based activities.
Mevlüde Akdeniz, Fatih Özdinç, Maya: An artificial intelligence based smart toy for pre-school children, International Journal of Child-Computer Interaction, Volume 29, 2021, 100347, ISSN 2212-8689, https://doi.org/10.1016/j.ijcci.2021.100347.	Mevlüde Akdeniz, Fatih Özdinç	AI, Image Processing, NLP	ScienceDirect	Includes children's engagement levels with the smart toy, learning outcomes in concept development, usability and effectiveness based on teacher feedback, adaptability of the toy to individual learning paces.

Khondaker A. Mamun, Rahad Arman Nabid, Shehan Irteza Pranto, Saniyat Mushrat Lamim, Mohammad Masudur Rahman, Nabeel Mahammed, Mohammad Nurul Huda, Farhana Sarker, Rubaiya Rahtin Khan, Smart reception: An artificial intelligence driven bangla language based receptionist system employing speech, speaker, and face recognition for automating reception services, Engineering Applications of Artificial Intelligence, Volume 136, Part A, 2024, 108923, ISSN 0952-1976, https://doi.org/10.1016/j.engappai.2024.108923.	Khondaker A. Mamun, Rahad Arman Nabid, Shehan Irteza Pranto, Saniyat Mushrat Lamim, Mohammad Masudur Rahman, Nabeel Mahammed, Mohammad Nurul Huda, Farhana Sarker, Rubaiya Rahtin Khan	AI, Face Recognition, Speech Recognition, ASR, TTS	ScienceDirect	Include the accuracy of face and speaker recognition, the Word Error Rate (WER) for the ASR model, the Mean Opinion Score (MOS) for TTS, validation loss for the question-answering system, and overall user satisfaction rates from real-world testing among participants.
Amara, K., Boudjemila, C., Zenati, N., Djekoune, O., Aklil, D., & Kenoui, M. (2022). AR Computer-Assisted Learning for Children with ASD based on Hand Gesture and Voice Interaction. IETE Journal of Research, 69(12), 8659–8675. https://doi.org/10.1080/03772063.2022.2101554	Amara, K., Boudjemila, C., Zenati, N., Djekoune, O., Aklil, D., & Kenoui, M.	AR, Gesture Recognition, Voice Recognition	ResearchGate	include levels of engagement and concentration time of children using the AR-computer system, improvements in receptive vocabulary and social interaction, qualitative feedback from therapists and parents
arXiv:2401.05459 (cs) [Submitted on 10 Jan 2024 (v1), last revised 8 May 2024 (this version, v2)] Personal LLM Agents: Insights and Survey about the Capability, Efficiency and Security Yuanchun Li, Hao Wen, Weijun Wang, Xiangyu Li, Yizhen Yuan, Guohong Liu, Jiacheng Liu, Wenxing Xu, Xiang Wang, Yi Sun, Rui Kong, Yile Wang, Hanfei Geng, Jian Luan, Xuefeng Jin, Zilong Ye, Guanjing Xiong, Fan Zhang, Xiang Li, Mengwei Xu, Zhijun Li, Peng Li, Yang Liu, Ya-Qin Zhang, Yunxin Liu	Yuanchun Li, Hao Wen, Weijun Wang, Xiangyu Li, Yizhen Yuan, Guohong Liu, Jiacheng Liu, Wenxing Xu, Xiang Wang, Yi Sun, Rui Kong, Yile Wang, Hanfei Geng, Jian Luan, Xuefeng Jin, Zilong Ye, Guanjing Xiong, Fan Zhang, Xiang Li, Mengwei Xu, Zhijun Li, Peng Li, Yang Liu, Ya-Qin Zhang, Yunxin Liu	IPAs, LLMs, IoT	Arvix Org	The effectiveness of Personal LLM Agents in understanding user intent, the efficiency of task execution, user satisfaction levels, security

Qureshi, K.N., Kaiwartya, O., Jeon, G. and Piccialli, F., 2022. Neurocomputing for internet of things: object recognition and detection strategy. Neurocomputing, 485, pp.263-273.	Qureshi, K.N., Kaiwartya, O., Jeon, G. and Piccialli, F.	Object Recognition, IoT	ACM Digital Library	Accuracy, efficiency, robustness, scalability, and 5G integration.
Adarsh, P., Rathi, P. and Kumar, M., 2020, March. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. In 2020 6th international conference on advanced computing and communication systems (ICACCS) (pp. 687-694). IEEE.	Adarsh, P., Rathi, P. and Kumar, M.	Object Recognition	IEEE	Assess the trade-off between speed and accuracy for the single stage model.
Amit, Y., Felzenszwalb, P. and Girshick, R., 2021. Object detection. In Computer Vision: A Reference Guide (pp. 875-883). Cham: Springer International Publishing.	Amit, Y., Felzenszwalb, P. and Girshick, R.		ResearchGate	Use quantitative measures and contextualise the paper to the given domain.
Rahman, M.A. and Sadi, M.S., 2021. IoT enabled automated object recognition for the visually impaired. Computer methods and programs in biomedicine update, 1, p.100015.	Rahman, M.A. and Sadi, M.S.	IoT enabled Object Recognition	ScienceDirect	Higher performance of the object detection and object recognition models
Hussan, M.I., Saidulu, D., Anitha, P.T., Manikandan, A. and Naresh, P., 2022. Object Detection and recognition in real time using deep learning for visually Impaired people. International Journal of Electrical and Electronics Research, 10(2), pp.80-86.	Hussan, M.I., Saidulu, D., Anitha, P.T., Manikandan, A. and Naresh, P.	IoT enabled Object Recognition	IJEER	The integration of YOLOv3 for real-time object detection with gTTS for audio output.
Guravaiah, Koppala, Yarlagadda Sai Bhavadeesh, Peddi Shwejan, Allu Harsha Vardhan, and S. Lavanya. "Third eye: object recognition and speech generation for visually impaired." *Procedia Computer Science* 218 (2023): 1144-1155.	Guravaiah, Koppala, Yarlagadda Sai Bhavadeesh, Peddi Shwejan, Allu Harsha Vardhan, and S. Lavanya.	Object Recognition and Speech Generation	ScienceDirect	Evaluated on the basis of selection of audio generation library for the model.

Fitria, T.N., 2021, December. Artificial intelligence (AI) in education: Using AI tools for teaching and learning process. In *Prosiding Seminar Nasional & Call for Paper STIE AAS* (Vol. 4, No. 1, pp. 134-147).	Fitria, T.N	Deep learning models, CNNs, feature extraction, seed classification	PROSIDING SEMINAR NASIONAL ITB AAS INDONESIA TAHUN 2023	Underscores the importance of these metrics in ensuring the reliability and robustness of the classification models.
Ganesh, D., Kumar, M.S., Reddy, P.V., Kavitha, S. and Murthy, D.S., 2022. Implementation of AI Pop Bots and its allied Applications for Designing Efficient Curriculum in Early Childhood Education. *International Journal of Early Childhood Special Education*, *14*(3).	Ganesh, D., Kumar, M.S., Reddy, P.V., Kavitha, S. and Murthy, D.S	CNNs, 3D recognition, voxel-based, real-time recognition, verbal feedback.	Science Direct	Understanding feature fusion and robustness will enhance tool's recognition capability and reliability in diverse educational settings
Alam, A., 2022, April. A digital game-based learning approach for effective curriculum transaction for teaching-learning of artificial intelligence and machine learning. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 69-74). IEEE.	Alam, A	Game-based learning, digital games, AI/ML curriculum, interactive learning.	IEEE	Evaluates the success of this learning approach using several parameters, including student engagement, conceptual clarity, retention of information, and the ability to apply AI and ML knowledge.
Ng, D.T.K., Lee, M., Tan, R.J.Y., Hu, X., Downie, J.S., & Chu, S.K.W. (2023). A review of AI teaching and learning from 2000 to 2020. Education and Information Technologies, 28(7), 8445-8501. https://doi.org/10.1007/s10639-022-11312-3	Ng, D.T.K., Lee, M., Tan, R.J.Y., Hu, X., Downie, J.S., & Chu, S.K.W	Collaborative project-based learning, software development, robotics.	Ieeexplore.org	Analyzed studies based on learner types, teaching tools, and pedagogical approaches in AI education.
Lin, S.Y., Chien, S.Y., Hsiao, C.L., Hsia, C.H. and Chao, K.M., 2020. Enhancing computational thinking capability of preschool children by game-based smart toys. *Electronic Commerce Research and Applications*, *44*, p.101011.	Lin, S.Y., Chien, S.Y., Hsiao, C.L., Hsia, C.H. and Chao, K.M	Smart toys, tangible user interfaces (TUI), game-based learning.	Elsevier	Assessed learning outcomes based on cognitive performance, engagement in learning activities, and computational thinking skill development.

Devi, J.S., Sreedhar, M.B., Arulprakash, P., Kazi, K. and Radhakrishnan, R., 2022. A path towards child-centric Artificial Intelligence based Education. *International Journal of Early Childhood*, *14*(3), pp.9915-9922.	Devi, J.S., Sreedhar, M.B., Arulprakash, P., Kazi, K. and Radhakrishnan, R	Deep CNNs, medical image classification, diagnostic accuracy enhancement.	ResearchGate	Improvement in recognition accuracy, demonstrating the method's robustness and adaptability to complex scenarios.

2.2 Existing System

Contemporary systems are at the forefront of innovation, employing a diverse array of advanced technologies such as screen readers, voice assistants, IoT, wearable devices, AI, AR, and large language models. VERSE seamlessly integrates screen readers and voice assistants, significantly enhancing web accessibility for visually impaired users. Meanwhile, IoT and wearable technologies are revolutionizing task-based language learning by creating interactive and adaptive experiences for young children. AI-driven smart toys, like Maya, harness AI, NLP, and image processing to deliver personalized learning experiences, enabling preschoolers to master fundamental concepts through engaging play. Smart Reception systems showcase the transformative potential of AI in language processing, automating receptionist services in non-English languages like Bangla through speech, speaker, and face recognition technologies. AR-based learning systems, particularly designed for children with autism spectrum disorder (ASD), employ augmented reality, gesture recognition, and voice interaction to substantially improve engagement, learning, and social interaction. Finally, personal LLM agents leverage large language models to craft intelligent, personalized assistants, seamlessly integrating with user devices to enhance task management and interaction in educational contexts. Collectively, these systems underscore the vast potential of emerging technologies in advancing object recognition and verbalization skills in early childhood education, paving the way for more groundbreaking and effective learning tools.

2.3 Problem Formulation

The problem formulation centres around the development and customization of Conv3D neural network models for video classification tasks. In the realm of computer vision, understanding and interpreting video content pose unique challenges due to the temporal dependencies and spatial complexities inherent in video data. Conv3D architectures offer a powerful framework for capturing both spatial and temporal features, enabling effective analysis and classification of video sequences.

Key elements of the problem formulation include:

- Architecture selection: Finding the best Conv3D architecture (ResNet, pre-activation ResNet, etc.) based on data characteristics, computational resources, and task requirements.
- Parameter tuning: Optimizing model parameters like number of classes, sample size, and duration for performance and adaptability.
- Configuration settings: Determining optimal settings for the chosen architecture (depth, shortcut types, cardinality) for effective video processing.
- Efficiency and compatibility: Ensuring compatibility with GPUs for CUDA acceleration and designing efficient workflows for both training and inference.
- Scalability: Implementing data parallelism for efficient distribution of workload across multiple GPUs, addressing challenges of large datasets.
- Performance optimization: Seeking optimal performance and accuracy through customization, parameter tuning, and architectural refinement.

2.4 Proposed System

The proposed Interactive Learning Companion for Toddlers aims to revolutionize early education by seamlessly integrating cutting-edge object detection and natural language processing (NLP) to create an immersive and engaging learning experience. Leveraging multi-object detection powered by YOLOv7 and EfficientDet architectures, the system enables real-time recognition of multiple items, providing deep contextual understanding of objects within their environment. Dynamic speech output crafts contextual sentences that foster interactive dialogues, actively engaging toddlers in their learning journey. Personalization is at the system's core, with tailored learning profiles that adapt to each child's unique pace and interests, alongside customizable voice outputs to enhance engagement. The system also delivers multi-sensory feedback through visual cues, auditory responses, and haptic sensations, making the learning experience comprehensive and enjoyable. Gamified learning modules like "Find the Object" and "What's Missing?" motivate children through play while tracking progress and rewarding achievements. By focusing on advanced contextual learning, enhanced personalization, and multi-sensory engagement, this innovative system aims to significantly elevate the quality of early education, empowering toddlers to learn through interactive play and personalized guidance.

3. OBJECTIVES

- To create a robust object recognition tool that integrates multiple neural network architectures such as CNN, RCNN, SSD, and YOLO.
- To implement text-to-speech (TTS) functionality using libraries like gTTS or pyttsx3 to provide real-time audio descriptions of recognized objects.
- To optimize the system for real-time object recognition and audio output, ensuring that the tool responds quickly and accurately.

4. METHODOLOGY

The project aims to implement object recognition using various neural network architectures like –

- Convolutional Neural Network (CNN)
- Region-based Convolutional Neural Network (RCNN)
- Single Shot MultiBox Detector (SSD)
- You Only Look Once (YOLO)

The project also integrates **text-to-speech (TTS) technology** using libraries like gTTS or pyttsx3. These libraries convert the recognized object labels into spoken words, providing an auditory output that helps children learn about objects. The TTS component ensures that the system is interactive and accessible to young users.

The model is trained using datasets like COCO and Pascal VOC. The data is preprocessed, including steps like image resizing, data augmentation, and normalization, to enhance model performance. Each model will be trained and fine-tuned using various hyperparameters, with the best-performing model integrated into the final system. Since these are primarily focused on classification, the performance of these models will be evaluated based on metrics like **accuracy, precision, recall**, and F1-score. Inference time will also be measured to ensure that the system operates in real-time. The models will then by hyperparameter tuned to obtain enhanced performance resulting in higher accuracy and faster audio response.

5. EXPERIMENTAL SETUP

For the study, dataset consists of images from publicly available datasets like COCO and Pascal VOC, which provide labelled data for various objects in diverse environments. The dataset is divided into two parts for training and validation. The training data is used to train different neural network models such as CNN, RCNN, SSD, and YOLO, while the validation data is used to evaluate their performance.

To prepare the images for object recognition, they will be resized and augmented with techniques like rotation, flipping, and scaling to increase diversity. The images will also be normalized to ensure consistent input ranges. Each image will be passed through the neural network models, which will extract features like edges, textures, and object shapes.

The models will be trained using the prepared data, and their effectiveness will be assessed using the validation data. The trained models will then be integrated with a text-to-speech system, which converts recognized object labels into audio output.

A number of performance indicators, such as accuracy, precision, recall, F1-score, and audio response time, will be used to assess the effectiveness of the models and the overall system. Finally, the created system will be used for real time testing and will be optimized to meet specific requirements. The findings could demonstrate that the combined approach of using CNN, RCNN, SSD, and YOLO models, along with text-to-speech integration, can effectively recognize objects and provide real-time audio feedback to children.

6. CONCLUSION

In conclusion, the integration of AI and machine learning into early childhood education represents a significant leap forward from traditional learning methods. The conventional use of tangible objects like flashcards and toys, though effective in stimulating visual and tactile engagement, lacks the ability to provide immediate, personalized feedback and adapt to the individual learning pace of each child. The incorporation of technologies such as object recognition models, particularly CNNs, RCNNs, SSD, and YOLO, into educational tools offers an opportunity to bridge this gap. These models allow real-time identification and verbalization of objects, making the learning process more interactive and tailored to each child's needs, thus fostering a more effective and engaging educational experience.

The combination of object recognition systems with game-based learning strategies further enhances early childhood education by making the learning environment both stimulating and responsive. As children interact with their surroundings, receiving real-time verbal feedback, their cognitive abilities are nurtured in ways traditional methods cannot achieve. By utilizing AI-powered systems to personalize and adapt learning experiences, early education can evolve into a more dynamic and meaningful process that fosters problem-solving, logical thinking, and motivation in young learners. Ultimately, the shift toward AI-based educational tools holds the potential to significantly improve the quality of early childhood education by offering a more individualized and enriching learning environment.

REFERENCES

[1] Hanafi, H.F., Wong, K.T., Adnan, M.H.M., Selamat, A.Z., Zainuddin, N.A. and Lee Abdullah, M.F.N., 2021. Utilizing Animal Characters of a Mobile Augmented Reality (AR) Reading Kit to Improve Preschoolers' Reading Skills, Motivation, and Self-Learning: An Initial Study. *International Journal of Interactive Mobile Technologies*, *15*(24).

[2] Wu, Q., Wang, S., Cao, J., He, B., Yu, C. and Zheng, J., 2019. Object recognition-based second language learning educational robot system for Chinese preschool children. *IEEE Access*, *7*, pp.7301-7312.

[3] Qi, S., Ning, X., Yang, G., Zhang, L., Long, P., Cai, W. and Li, W., 2021. Review of multi-view 3D object recognition methods based on deep learning. *Displays*, *69*, p.102053.

[4] Bazargani, J.S., Sadeghi-Niaraki, A., Rahimi, F., Abuhmed, T. and Choi, S.M., 2022. An iot-based approach for learning geometric shapes in early childhood. *IEEE Access*, *10*, pp.130632-130641.

[5] Rahiem, M.D., 2021. Storytelling in early childhood education: Time to go digital. *International Journal of Child Care and Education Policy*, *15*(1), p.4.

[6] Zaini, N.A., Noor, S.F.M. and Wook, T.S.M.T., 2019. Evaluation of api interface design by applying cognitive walkthrough. *International Journal of Advanced Computer Science and Applications*, *10*(2).

[7] Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W. White. 2019. VERSE: Bridging Screen Readers and Voice Assistants for Enhanced Eyes-Free Web Search. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19). Association for Computing Machinery, New York, NY, USA, 414–426.

[8] E. de la Guía, V. L. Camacho, L. Orozco-Barbosa, V. M. Brea Luján, V. M. R. Penichet and M. Lozano Pérez, "Introducing IoT and Wearable Technologies into Task-Based Language Learning for Young Children," in IEEE Transactions on Learning Technologies, vol. 9, no. 4, pp. 366-378, 1 Oct.-Dec. 2016, doi: 10.1109/TLT.2016.2557333.

[9] Mevlüde Akdeniz, Fatih Özdinç, Maya: An artificial intelligence based smart toy for pre-school children, International Journal of Child-Computer Interaction, Volume 29, 2021, 100347, ISSN 2212-8689.

[10] Khondaker A. Mamun, Rahad Arman Nabid, Shehan Irteza Pranto, Saniyat Mushrat Lamim, Mohammad Masudur Rahman, Nabeel Mahammed, Mohammad Nurul Huda, Farhana Sarker, Rubaiya Rahtin Khan, Smart reception: An artificial intelligence driven bangla language-based receptionist system employing speech, speaker, and face recognition for automating reception services, Engineering Applications of Artificial Intelligence, Volume 136, Part A, 2024, 108923, ISSN 0952-1976

[11] Amara, K., Boudjemila, C., Zenati, N., Djekoune, O., Aklil, D., & Kenoui, M. (2022). AR Computer-Assisted Learning for Children with ASD based on Hand Gesture and Voice

Interaction. *IETE Journal of Research*, *69*(12), 8659–8675.

[12] : **arXiv:2401.05459** (cs) [Submitted on 10 Jan 2024 (v!), last revised 8 May 2024 (this version, v2)] Personal LLM Agents: Insights and Survey about the Capability, Efficiency and Security

[13] Devi, J.S., Sreedhar, M.B., Arulprakash, P., Kazi, K. and Radhakrishnan, R., 2022. A path towards child-centric Artificial Intelligence based Education. *International Journal of Early Childhood*, *14*(3), pp.9915-9922.

[14] Fitria, T.N., 2021, December. Artificial intelligence (AI) in education: Using AI tools for teaching and learning process. In *Prosiding Seminar Nasional & Call for Paper STIE AAS* (Vol. 4, No. 1, pp. 134-147).

[15] Ganesh, D., Kumar, M.S., Reddy, P.V., Kavitha, S. and Murthy, D.S., 2022. Implementation of AI Pop Bots and its allied Applications for Designing Efficient Curriculum in Early Childhood Education. *International Journal of Early Childhood Special Education*, *14*(3).

[16] Alam, A., 2022, April. A digital game based learning approach for effective curriculum transaction for teaching-learning of artificial intelligence and machine learning. In *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 69-74). IEEE.

[17] Ng, D.T.K., Lee, M., Tan, R.J.Y., Hu, X., Downie, J.S., & Chu, S.K.W. (2023). A review of AI teaching and learning from 2000 to 2020. Education and Information Technologies, 28(7), 8445-8501. https://doi.org/10.1007/s10639-022-11312-3

[18] Lin, S.Y., Chien, S.Y., Hsiao, C.L., Hsia, C.H. and Chao, K.M., 2020. Enhancing computational thinking capability of preschool children by game-based smart toys. *Electronic Commerce Research and Applications*, *44*, p.101011.

[19] Qureshi, K.N., Kaiwartya, O., Jeon, G. and Piccialli, F., 2022. Neurocomputing for internet of things: object recognition and detection strategy. Neurocomputing, 485, pp.263-273.

[20] Adarsh, P., Rathi, P. and Kumar, M., 2020, March. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. In 2020 6th international conference on advanced computing and communication systems (ICACCS) (pp. 687-694). IEEE.

[21] Amit, Y., Felzenszwalb, P. and Girshick, R., 2021. Object detection. In Computer Vision: A Reference Guide (pp. 875-883). Cham: Springer International Publishing.

[22] Rahman, M.A. and Sadi, M.S., 2021. IoT enabled automated object recognition for the visually impaired. Computer methods and programs in biomedicine update, 1, p.100015.

[23] Hussan, M.I., Saidulu, D., Anitha, P.T., Manikandan, A. and Naresh, P., 2022. Object Detection and recognition in real time using deep learning for visually Impaired people. International Journal of Electrical and Electronics Research, 10(2), pp.80-86.

[24] Guravaiah, Koppala, Yarlagadda Sai Bhavadeesh, Peddi Shwejan, Allu Harsha Vardhan, and S. Lavanya. "Third eye: object recognition and speech generation for visually impaired." *Procedia Computer Science* 218 (2023): 1144-1155.