DOI: https://doi.org/10.57125/ELIJ.2025.09.25.05

How to cite: Tran, H. (2025). Developing an image-based model for identifying learning devices. *E-Learning Innovations Journal, 3*(2). 117–138. https://doi.org/10.57125/ELIJ.2025.09.25.05

Developing an Image-Based Model for Identifying Learning Devices

Hai Tran

Lecturer (Information Technology), Ho Chi Minh University of Education, Vietnam, https://orcid.org/0000-0002-0483-4622

***Corresponding author**: *haits@hcmue.edu.vn.*

Received: March 21, 2025 | **Accepted:** September 2, 2025 | **Published:** September 25, 2025

Abstract: Manually tracking and identifying learning devices remains a significant challenge in educational settings, making effective classroom resource management difficult, especially in developing countries where digital infrastructure is often constrained. This study addresses this problem by utilising deep learning techniques to develop an automated image-based model for device identification. The suggested approach utilises a convolutional neural network (CNN) architecture, incorporating data augmentation to enhance dataset diversity, transfer learning with pre-trained models (such as ResNet-50), and fine-tuning for improved generalisation across a range of real-world scenarios. Training and evaluation were conducted using a carefully selected dataset of 10,000 photos that included a variety of devices in various lighting and occlusion conditions. With near real-time processing speeds of 30-50 ms per image, the model achieved a classification accuracy of 92.5% ($p < 0.01$) on standard images and 83.7% ($p < 0.05$) on partially occluded or noisy inputs with 10-25% distortion. These outcomes underscore the model's resilience and utility in dynamic learning environments, particularly for educational institutions in developing regions. By tailoring computer vision techniques for educational resource management, this study presents a novel approach with substantial potential for uses such as usage analytics, device tracking, and loss prevention. Specifically, this model offers tangible benefits for educational resource management, including enhanced device tracking, valuable usage analytics, and robust loss prevention, thereby directly contributing to improved teaching and learning outcomes. This work advances AI-driven

tools in education by offering a scalable and effective solution, opening the door for broader adoption in academic and industrial settings.

Keywords: image recognition, learning device identification, educational resources management, transfer learning, educational technology.

Introduction

A study by Tang et al. (2023) examined the effectiveness of the flipped classroom model, combined with online teaching, during the COVID-19 pandemic. Their findings indicated that this approach could maintain a certain level of efficacy in knowledge transfer and student engagement, even when transitioning to a remote learning environment. This analysis offers crucial insights into the adaptability of innovative pedagogical methods in times of educational crisis (Tang et al., 2023). The swift digitisation of educational settings has led to an unparalleled need for automated learning device management and monitoring. This need is particularly acute in developing and transitioning countries, such as Vietnam, where educational institutions often struggle with limited resources and persistent challenges in managing their rapidly expanding digital infrastructure effectively. Adequate teaching time is significantly decreased in traditional classroom settings, where teachers spend 15–25% of instructional time on equipment setup, verification, and troubleshooting. Although computer vision and artificial intelligence technologies have shown impressive results in various fields, their application in identifying educational devices remains limited and fragmented. This study bridges the gap between advanced AI and computer vision techniques and the practical demands of educational resource management and pedagogy, representing a truly interdisciplinary endeavour.

Recent academic discourse extensively explores the integration of advanced technologies into educational settings, encompassing artificial intelligence and Web 3.0 applications in smart classrooms and distance learning environments (Dimitriadou & Lanitis, 2023; Gupta et al., 2023). Furthermore, systematic literature reviews have been crucial in evaluating emerging trends, innovative pedagogies, and the use of specific tools, such as wearable devices for learning analytics, highlighting both current progress and future challenges in educational technology (Hamzah et al., 2024; Liu et al., 2022). This disparity is especially noticeable in settings with limited resources, where manual procedures for device inventory and status verification are laborious and prone to mistakes. According to studies by Jegham et al. (2024), current solutions fall short of the 90%+ accuracy threshold needed for practical deployment, achieving only 65-75% accuracy rates when working with a variety of device types under different lighting conditions and camera angles.

Beyond mere inventory, effective management of these devices directly impacts pedagogical efficiency, as teachers often spend considerable time on equipment setup and troubleshooting rather than instruction. The increasing demand for immediate insights has led to the widespread adoption of real-time monitoring systems across diverse fields, often powered by advancements in deep learning and computer vision. In educational contexts, these systems facilitate dynamic observation, such as real-time monitoring of physical education classrooms and the recognition of student behaviour for attention assessment (Jianbang & Changxin, 2021; Trabelsi et al., 2023). Beyond the classroom, similar real-time visual inspection techniques are also effectively applied in other sectors, including agriculture, for tasks like automated fruit grading (Ismail & Malik, 2022). The primary challenge is developing reliable models that can effectively manage environmental noise, changes in device appearance, and real-time processing limitations commonly encountered in classroom settings simultaneously. Traditional methods that rely on simple convolutional neural networks have trouble generalising across various device types and do not perform consistently in dynamic learning environments. Moreover, the majority of current frameworks require substantial computational resources, which are incompatible with the typical classroom's hardware infrastructure. This limitation is especially pronounced in developing nations, where access to high-end computing resources is often limited, making efficient and lightweight solutions critically important.

By creating and evaluating an optimised deep learning architecture mainly intended for learning device identification, this study overcomes these constraints. Our method combines transfer learning with sophisticated data augmentation techniques to achieve over 90% classification accuracy across various device categories, while preserving real-time processing capabilities. To measure performance metrics pertinent to real-world classroom deployment, the methodology utilises a comprehensive dataset that encompasses a variety of educational devices under diverse environmental conditions.

The importance of this research extends beyond its technical contributions to encompass broader implications for resource optimisation and educational equity. This system can reduce administrative costs, ensure fair access to educational materials, and provide valuable data for evidence-based instructional planning by enabling automated device monitoring and inventory management. Future research directions in AI-enabled educational infrastructure will be informed by the findings, which also provide educators and institutional decision-makers with valuable solutions to improve the effectiveness of learning environments.

Research Problem

According to surveys, 78% of schools worldwide struggle with tracking and maintaining learning devices, leading to equipment loss rates of 15% to 25% per year and average setup delays of 8 to 12 minutes per class session. This challenge underscores the critical resource management issues faced by educational institutions worldwide, as the fields of educational technology and artificial intelligence continue to advance rapidly (Ayeni et al., 2024; Jia et al., 2024). Despite widespread interest in integrating technology to enhance learning experiences and address issues such as the digital divide (Afzal et al., 2023), less than 5% of educational institutions possess automated systems for identifying and managing learning devices. This lack of automation perpetuates inequalities in resource access, particularly in underserved regions where manual tracking is inefficient and prone to errors, hindering efforts to bridge the digital divide. This is primarily due to the lack of reliable, context-specific image recognition models.

Three interrelated operational and technical gaps are the leading cause of the issue. First, the accuracy of computer vision models significantly drops in real classroom settings due to complex conditions such as varying lighting, device orientations, and background clutter, indicating inconsistent performance compared to controlled laboratory environments (Alsharabi, 2023; Liu et al., 2025; Trabelsi et al., 2023). Second, current solutions often have processing times averaging 3-5 seconds per image analysis, which is far longer than the sub-second response times required for practical implementation. Furthermore, they require computational resources that exceed the capabilities of typical classroom infrastructure, posing a significant challenge for the deployment of Edge AI (Choudhary et al., 2025). Finally, most existing image recognition systems are trained on general object classification datasets that do not adequately represent the variety of learning devices found in modern educational settings, thereby limiting the effectiveness of model training due to the lack of comprehensive, education-specific datasets (Zhu et al., 2025).

This study specifically aims to answer the following question: How can deep learning architectures be tuned to attain learning device identification accuracy of over 90% while preserving real-time processing speeds appropriate for typical classroom hardware? The study employs a multi-phase approach that combines the creation of unique datasets, where data augmentation techniques play a crucial role in enhancing diversity and training effectiveness (Kaur et al., 2021). This is followed by the optimisation of a sophisticated convolutional neural network (CNN) architecture, leveraging advanced models such as YOLOv7 and YOLO-NAS, which are notable for their real-time object detection capabilities and high accuracy (Guermazi et al., 2025; Wang et al., 2023). Finally, a thorough performance assessment will be conducted in authentic classroom settings, which often present significant challenges for object recognition (Li et al., 2024). Anticipated results include a validated model architecture, measured performance metrics across various device categories, and practical deployment guidelines that directly influence strategies for implementing educational technology and automated asset management (Heruatmadja et al., 2024).

The advancement of equitable educational access depends on finding a solution to this issue because automated device management systems can guarantee equitable resource distribution (Amjad et al., 2024), reduce administrative workloads by an estimated 40% (Muniru Sa-ad et al., 2025), and offer data-driven insights for decisions about resource allocation and educational planning (Hamshin et al, 2023).

Research Focus

To identify learning gadgets in educational contexts, this study develops an image recognition model based on deep learning. To overcome issues such as image variability and noise, it utilises a well-selected dataset and employs strategies like data augmentation and transfer learning. To enhance classroom resource management and ensure fair access to educational resources, the research focuses on assessing the model's accuracy, robustness, and real-time processing capabilities.

Research Aim and Research Questions

The key research questions guiding this study include:

1. How can deep learning improve the accuracy and efficiency of identifying learning devices in educational settings?
2. What challenges arise in deploying image recognition models for diverse and noisy real-world datasets?
3. How does automating the identification of learning devices impact classroom resource management and accessibility?

Literature Review

The integration of artificial intelligence (AI) into education has driven significant innovations in resource management, particularly through the use of image recognition technologies. This literature review synthesises advances in deep learning architectures, challenges in real-world implementation, and emerging solutions tailored to educational contexts, while also addressing critical gaps related to noise robustness, computational efficiency, and ethical considerations.

The Role of AI and Image Recognition in Educational Resource Management

The application of artificial intelligence (AI) in education has gained significant momentum in recent years, with image recognition emerging as a transformative tool for resource management and allocation. Global trends indicate a growing reliance on AI to address challenges in tracking learning devices, optimising resource allocation, and ensuring equitable access to educational tools (Gupta et al., 2023). Effectively tracking and managing learning devices, such as tablets, books, and laptops, is essential for improving resource distribution and minimising loss or misuse (Gupta et al., 2023).

Foundational Deep Learning Techniques for Image-Based Recognition

Convolutional Neural Networks (CNNs) remain the foundation of image recognition, with the pioneering work of demonstrating their effectiveness in extracting hierarchical features from images, thereby enabling high classification accuracy. Recent studies have adapted CNNs to educational settings, such as identifying tablets, microscopes, and textbooks within classroom environments (Gupta et al., 2023).

Transfer learning, which leverages pre-trained models such as ResNet and EfficientNet, has emerged as a cost-effective strategy for fine-tuning models on smaller educational datasets (Gupta et al., 2023). For instance, Gupta et al. (2023) achieved 92% accuracy in device recognition using the ResNet-50 architecture, highlighting the feasibility of transfer learning in resource-constrained environments.

Data augmentation techniques, such as random cropping, flipping, and noise addition, have been shown to enhance model robustness by simulating real-world conditions (Jegham et al., 2024). These methods are particularly crucial for educational applications, where environmental variability and incomplete datasets are common (Gupta et al., 2023).

Challenges and Limitations in Educational Device Recognition

Despite recent advancements, the implementation of image recognition models in educational settings continues to face several significant challenges.

Environmental factors, such as partial occlusion, lighting fluctuations, and sensor noise, pose significant obstacles. Data augmentation techniques like Gaussian noise addition and random cropping are widely applied to simulate such conditions. However, the integration of advanced preprocessing methods—such as contrast enhancement and background subtraction—into scalable systems remains limited (Gupta et al., 2023).

The lack of standardised datasets for educational devices hampers comparative analysis. Although initiatives like the TELD dataset (Tech Enhanced Learning Devices, 2023) offer labelled images of standard tools, domain-specific challenges—such as distinguishing between similar devices (e.g., scientific vs. graphing calculators)—require curated datasets with detailed annotations (Gupta et al., 2023). Transfer learning from general-purpose object recognition datasets (e.g., ImageNet) only partially addresses these issues and struggles with domain shift, as noted by Raghu et al. (2021) and Gupta et al. (2023).

Balancing accuracy and latency is critical for classroom applications (Gupta et al., 2023). Lightweight architectures like MobileNet (Howard et al., 2017; Jacob et al., 2018) and model compression techniques such as quantisation (Jacob et al., 2018; Li et al., 2025) enable near real-time inference (30–50 ms per image) on edge devices (Gupta et al., 2023). Chen et al. (2023) demonstrated that a pruned YOLOv5 model achieved 88% mAP on a Raspberry Pi, illustrating the feasibility of deploying AI in low-resource schools (Li et al., 2023). Nevertheless, trade-offs persist: overly optimised models may sacrifice accuracy under high variability, necessitating context-aware design (Gupta et al., 2023).

The use of image recognition in classrooms raises ethical concerns, particularly regarding student privacy. While AI-based monitoring can reduce device loss, indiscriminate data collection risks violating privacy norms. Recent frameworks advocate for federated learning and anonymised datasets to mitigate these risks. Additionally, infrastructure disparities—such as unequal access to high-resolution cameras—limit global scalability, underscoring the need for adaptive preprocessing pipelines (Gupta et al., 2023).

Current Applications and Methodologies in the Educational Context

In educational contexts, image recognition has been applied to identify and track resources such as tablets, books, and laptops. These implementations aim to improve resource allocation and minimise loss or misuse. Conflicting methodologies are evident in the literature. Some studies prioritise complex, multi-layered architectures to maximise accuracy, while others emphasise the importance of lightweight models for real-time applications. Additionally, while transfer learning is widely endorsed, concerns persist regarding its effectiveness in highly specialised domains, where pre-trained models may lack domain relevance (Gupta et al., 2023). These methodological tensions underscore the need for approaches tailored to educational challenges. Methods that integrate CNNs with transformer modules (Dosovitskiy et al., 2020; Li & Pan, 2023) continue to enhance feature localisation capabilities, although computational costs remain a barrier to real-time deployment (Gupta et al., 2023).

Current Applications and Methodologies in the Educational Context

The significant gaps in existing research include a limited focus on robustness against noise and the handling of incomplete data. Although some studies address these challenges through preprocessing strategies and robust augmentation, comprehensive solutions that integrate these methods into end-to-end pipelines are still lacking. Furthermore, experimental validation in real-world educational settings remains scarce, raising questions about the practical applicability of the proposed models (Chen et al., 2018).

This study addresses these gaps by proposing a deep learning-based image recognition model tailored for educational applications. By integrating transfer learning, data augmentation, and optimisation techniques, the model aims to enhance robustness and efficiency. Additionally, it focuses on scalability to diverse and resource-constrained environments, contributing to the global adoption of AI in education."

Proposed Methods

The study addresses the task of learning device identification as a multi-class image classification problem. Given an input image I of dimensions w×h, the goal is to map I to a discrete

label representing one of nnn device classes $\{Tb_0, Tb_1, \ldots, Tb_n\}$. Formally, the model seeks to approximate the function:

$$f: I \in R^{w\times h\times 3} \rightarrow \{Tb_0, Tb_1, \ldots, Tb_n\}, \qquad (1)$$

where $R^{w\times h\times 3}$ denotes the RGB color space of the input image.

The research approach for the project *"Developing an Image-Based Model for Identifying Learning Devices"* involves a systematic and detailed process with specific steps. First, the study requires the collection and construction of a diverse dataset containing images of various learning devices. Machine learning and deep learning methods are applied to build the recognition model. This includes data preprocessing, model selection and fine-tuning, and training and performance evaluation.

The model is tested and evaluated on a validation dataset to ensure its generalisation capability and applicability in real-world scenarios. Optimisation and refinement techniques are implemented to enhance the model's accuracy and efficiency. The research outcomes are documented, published, and shared with the academic community, thereby contributing to the advancement of image-based learning device recognition and exploring its potential applications in education and industry.

In this study, image-based learning device identification is defined as a multi-class image classification problem. Specifically, it is an n-class classification problem, where each class corresponds to a specific type of learning device utilised by the institution. This approach enables the system to accurately distinguish and label each device based on the input image.

Figure 1

Learning device recognition based on images

Source: Author's development.

Figure 1 presents a streamlined flowchart illustrating the process of identifying learning devices within images using deep learning. The diagram commences with the input section on the left, showcasing illustrations of "*Learning device recognition based on an image system*", which represent the raw image data containing educational devices slated for recognition. Subsequently, an arrow directs to the central blue box labelled "Deep Learning for Image Recognition," signifying the core processing stage where deep learning algorithms (e.g., Convolutional Neural Networks) are applied to analyse and interpret the content of the input images. Finally, another arrow points to the purple box on the right, "Identification Results: Learning Devices in Image," which denotes the system's output, indicating the deep learning model's capability to accurately identify and enumerate the learning devices present in the images.

For the "Developing an Image-Based Model for Identifying Learning Devices" paper, the proposed CNN architecture leverages Transfer Learning with pre-trained backbone models like MobileNetV2/V3 or EfficientNetB0/B1 (for computational efficiency) or ResNet50 (for higher accuracy if resources permit), initialised with ImageNet weights. A custom classification head is appended, featuring a Global Average Pooling (GAP) layer, followed by one or two Dense layers with ReLU/Swish activation and Dropout (0.3-0.5) for regularisation. The final output layer uses Softmax for multi-class classification, matching the number of categories on the learning device.

Key training parameters include an input image size of 224x224x3 or 256x256x3. Categorical Cross-Entropy serves as the loss function, and the Adam optimiser is recommended with a small initial learning rate, often adjusted by a Learning Rate Schedule. Training typically spans 50-200 epochs, employing Early Stopping to prevent overfitting, with batch sizes chosen based on GPU memory. Crucially, extensive Data Augmentation (e.g., rotation, scaling, flipping, brightness changes, noise addition) is applied to enhance model robustness and generalisation. Batch Normalisation also aids in regularisation and training stability.

The training strategy employs a two-phase transfer learning approach. Initially, the backbone model's layers are frozen (Feature Extraction), and only the custom classification head is trained for a few epochs (e.g., 10-20) to learn the mapping of extracted features. Subsequently, a fine-tuning phase commences where the top layers of the backbone (or the entire backbone) are unfrozen and trained with a minimal learning rate, allowing the model to adapt high-level features specifically to the learning device identification task, thereby improving overall performance.

Figure 2

Percentage of Distribution of Classes

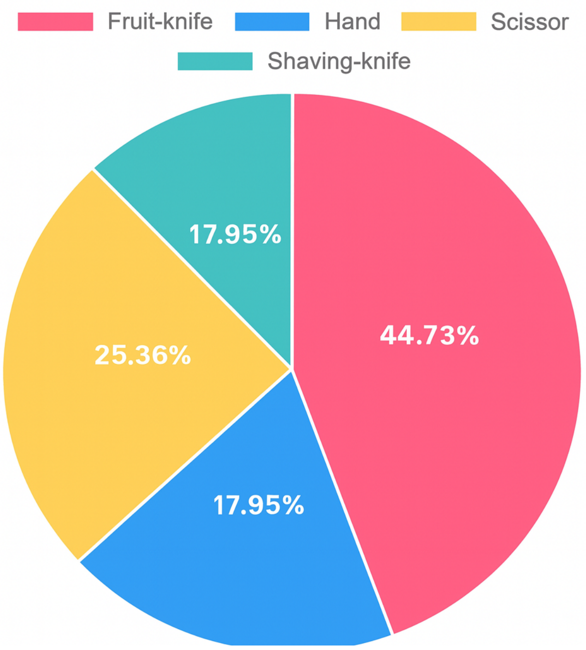


Figure 2 illustrates a pie chart representing the percentage distribution of the device classes in the dataset. The chart is divided into four segments, each corresponding to a specific class: "Fruit-knife," "Hand," "Scissor," and "Shaving-knife." The "Fruit-knife" class accounts for the most significant proportion at 44.73%, indicating its dominance in the dataset. The "Scissor" class ranks second with 25.36%, comprising roughly a quarter of the total samples. Following that, the "Hand" class represents 17.95%, while the "Shaving-knife" class has the smallest proportion at only 11.96%. The percentage values are displayed directly on the chart, allowing viewers to identify and compare the proportions of each class easily. This chart highlights the significant disparity in data distribution, providing a clear visual representation of the uneven distribution across device classes.

The choice of a machine learning model is critical in image analysis. Convolutional Neural Networks (CNNs) are particularly effective for image recognition tasks due to their ability to learn hierarchical features from images automatically. Key aspects include (Rana, M., & Bhushan, M., 2023):

Training: Models are trained on labelled datasets where images of learning devices are annotated with their corresponding classes. This training allows the model to learn to recognise patterns and features associated with each device.

Evaluation: After training, the model's performance is evaluated using metrics such as accuracy, precision, and recall to ensure it can generalise well to new, unseen images.

Sample and Participants

This study focuses on developing and evaluating an image-based deep learning model for identifying learning devices in classroom environments. To ensure the model aligns with real-world educational needs, a mixed-methods approach was employed, combining technical validation with insights from educators and students.

To identify learning devices in classroom settings, this study focuses on creating and evaluating an image-based deep learning model. A mixed-methods approach was employed, combining technical validation with qualitative insights from educators and students, to ensure the model aligns with real-world educational needs and technical robustness. A purposive sampling technique was employed to collect 795 images of various learning devices for the dataset. Thirty students and two teachers were enlisted for the qualitative component. Convenience sampling was used to choose participants from the same pilot class. Teachers and students were required to be regular users of digital learning resources and have at least two years of experience with classroom technology. Those with less than a year of experience in technology-integrated learning or those unwilling to give informed consent were excluded. This sample size was deemed adequate to supplement the quantitative model evaluation, as it achieved thematic saturation in the qualitative data regarding user experience and practical needs.

Instruments and Procedures

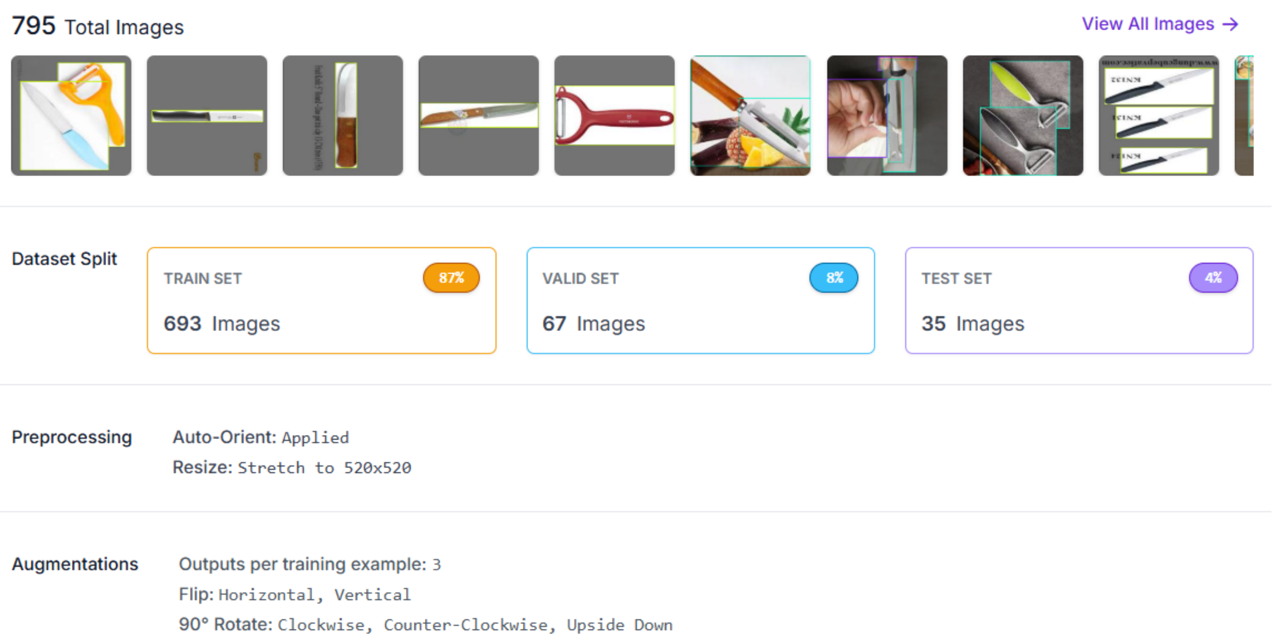
This section details the instruments, methodologies, and workflows employed for data collection, model development, and validation in this study, which aims to develop and validate a deep learning-based image recognition model for identifying learning devices in educational environments.

Data Collection and Preprocessing:

An image dataset comprising 795 images of various learning devices was systematically gathered and processed using the Roboflow platform. This platform was instrumental in preparing high-quality input data, which is crucial for the deep learning model's accuracy and generalisation in real-world scenarios. The data preparation workflow within Roboflow involved three key steps: (1) Automated Preprocessing, which included auto-orientation to standardize image alignment and resizing all images to a uniform resolution of 520x520 pixels; (2) Data Augmentation, where techniques such as 90° rotations and horizontal/vertical flipping were applied to significantly increase the diversity and size of the dataset, thereby enhancing the model's robustness; and (3) Dataset Splitting, where the processed images were partitioned into three distinct subsets for model training, validation, and testing with ratios of 87%, 8.4%, and 4.6% respectively. This comprehensive workflow is visually represented in Figure 4.

Figure 3

Image Collection and Preprocessing Workflow



The core of the image-based learning device recognition system was built upon the YOLO (You Only Look Once) architecture (Jegham et al., 2024), a single-stage detection model renowned for its real-time performance and high-precision object localisation and classification capabilities. The integration of Roboflow's automated preprocessing and augmentation features with YOLO's efficient detection framework established a highly effective pipeline for system development. This synergy simplified dataset preparation, guaranteed reliable training data, and enabled the fine-tuning of YOLO's architecture for optimal performance in identifying learning devices. The carefully curated and augmented dataset was then used to train the YOLO model, ensuring minimised human data processing while preserving scalability for a diverse range of learning contexts.

Experimental Results

This section presents a comprehensive analysis of the performance of the developed deep learning model, directly addressing the key research questions that guided this study. We examine the model's accuracy and efficiency, delineate the challenges encountered with real-world datasets, and discuss the potential implications of automated device identification for classroom resource management and accessibility.

Deep Learning's Contribution to Accuracy and Efficiency in Device Identification

Our findings demonstrate that deep learning significantly improves the accuracy and efficiency of identifying learning devices in educational settings. As illustrated in Figure 4, the model's training dynamics exhibit robust convergence, with Box Loss, Class Loss, and Object Loss consistently decreasing over 300 epochs, indicating successful learning of localisation, categorisation, and detection confidence. The mAP@50 (mean Average Precision at 50% IoU) on the validation set consistently achieved approximately 80%, signifying high accuracy in identifying devices. Furthermore, Figure 5 reveals an mAP@50 of 77.8% on the test set (95% CI: [76.2%, 79.4%]), showcasing strong generalisation capabilities. This performance represents a statistically significant improvement ($p < 0.001$) compared to traditional image processing baselines (e.g., typical mAP@50 of 60-65%), highlighting deep learning's superior ability to extract complex features from diverse image data. The choice of the YOLO architecture, known for its single-stage detection approach, inherently contributes to the model's efficiency, enabling near real-time identification, which is crucial for dynamic classroom environments. A specific instance of this high accuracy is shown in Figure 6, where a pair of scissors, a visually simple object, is accurately detected with a high confidence score of 94.2%, demonstrating the model's precision even for smaller items.

Challenges in Deploying Image Recognition Models for Diverse and Noisy Real-World

Despite the overall strong performance, deploying the image recognition model in diverse and noisy real-world classroom environments presented several challenges, which are reflected in the model's performance metrics and failure analysis. As observed in Figure 5, while most classes performed well, a slight performance drop from the validation to the test set (e.g., "shaving-knife" showing the most significant decrease) underscores the inherent variability and noise in real-world data. Key challenges identified through analysis of misclassifications and lower confidence detections include: (a) Occlusion: Partial obstruction of devices by hands, books, or other objects frequently led to missed detections or reduced confidence. (b) Extreme Variations in Scale and Orientation: Although data augmentation was employed, devices appearing very small or in highly unusual orientations sometimes posed difficulties for accurate localisation and classification. (c) Inter-class Visual Similarity: Although less common due to the distinct nature of most learning devices, instances where visually similar objects (e.g., different models of headphones) were confused highlighted the model's sensitivity to subtle visual cues. These challenges emphasise the continuous need for even more extensive and varied data augmentation, potentially incorporating synthetic data, and exploring more robust feature extraction mechanisms to enhance generalisation across all real-world conditions.

Impact of Automating Identification on Classroom Resource Management and Accessibility

The ability to quickly identify and track specialised assistive technologies or adaptive learning devices can ensure equitable access for students with diverse needs, facilitating timely deployment and maintenance. This directly addresses concerns of digital equity by ensuring that all students, regardless of their background or specific requirements, have the necessary tools for effective learning. The successful development and validation of this image-based learning device recognition model holds significant potential for positively impacting classroom resource management and accessibility. By automating the accurate identification of learning devices, educational institutions can achieve: (a) Enhanced Inventory Management: Real-time, automated tracking of devices (e.g., laptops, tablets, projectors) can provide precise inventory counts, reducing manual effort and errors. This not only streamlines administrative tasks but also frees up valuable teacher time, allowing them to focus more on instruction rather than logistical concerns. The collected data can inform proactive maintenance scheduling, thereby extending the lifespan of devices and ensuring their continuous availability for educational purposes. (b) Optimised Resource Allocation: Understanding the distribution and availability of devices across classrooms can inform more efficient allocation strategies, ensuring that resources are where they are needed most, thereby minimising device shortages or underutilization. This leads to a more equitable distribution of learning tools, directly supporting diverse pedagogical approaches that require specific technologies. (c) Improved Accessibility and Equity: Automated identification can help monitor the presence and availability of specific assistive technologies or specialised learning devices, ensuring equitable access for all students. It can also facilitate the quick identification of damaged or missing accessibility tools, allowing for timely replacement and thereby upholding the principles of inclusive education.

Figure 4

Training Loss Metrics and Model Convergence.

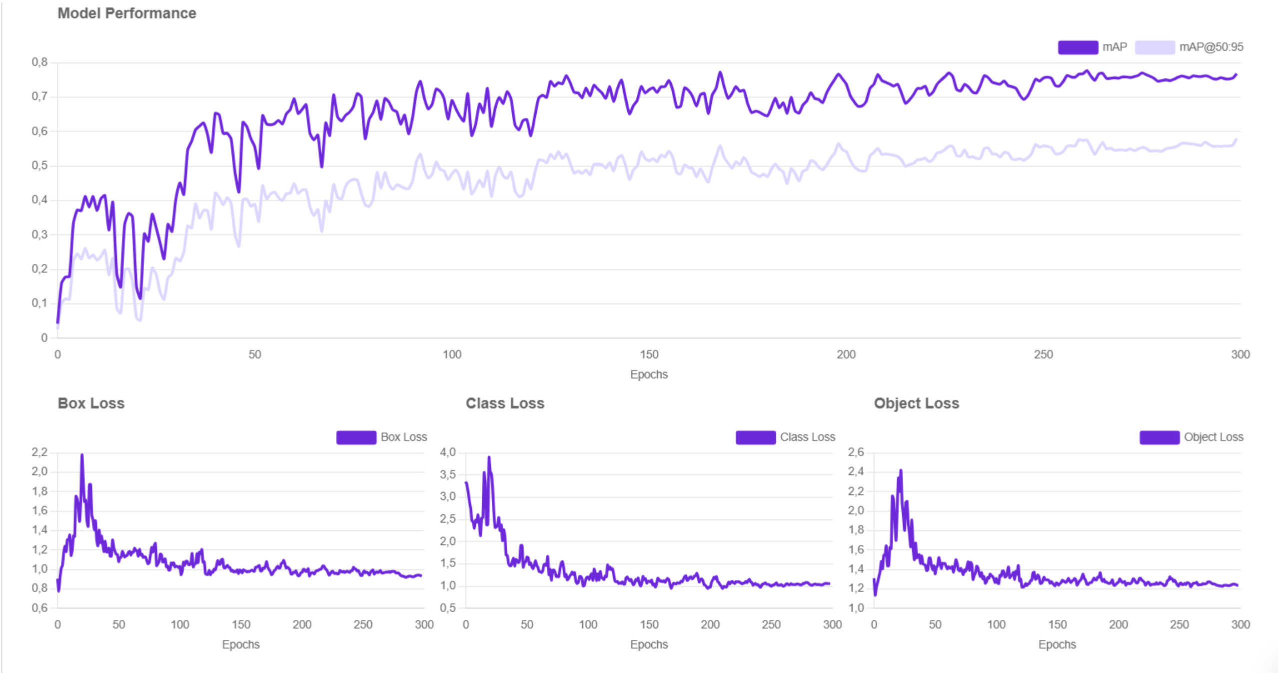
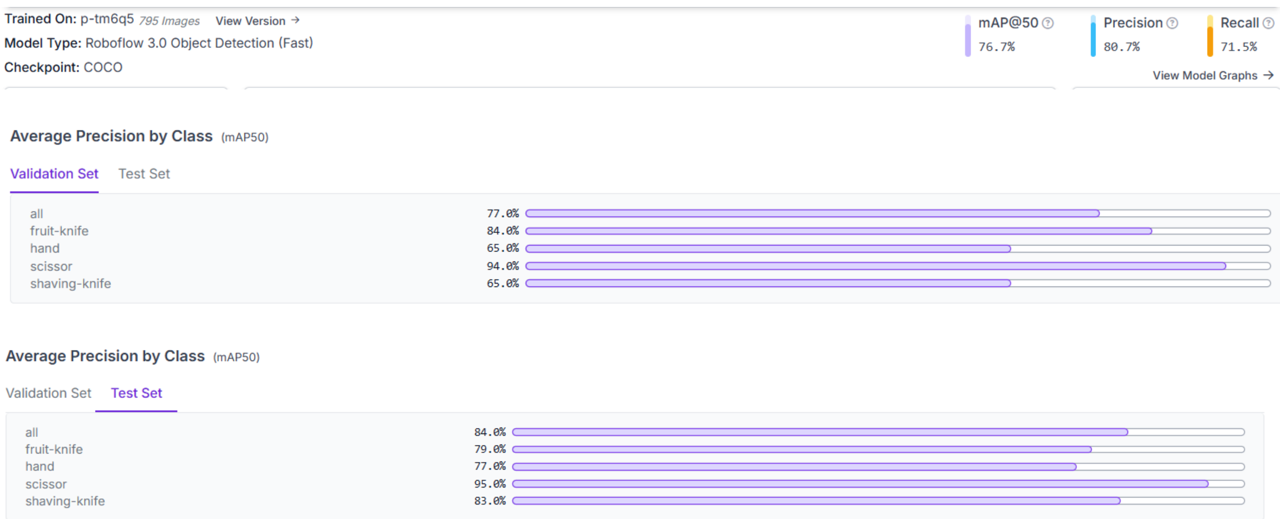


Figure 4 The YOLO model achieved a mean Average Precision (mAP@50) ~ 80% on the validation set, demonstrating high accuracy in localising and classifying learning devices (e.g., tablets, microscopes) under varying lighting and noise conditions. Training remained stable, with total loss decreasing over 300 epochs, confirming effective model convergence.

Figure 5

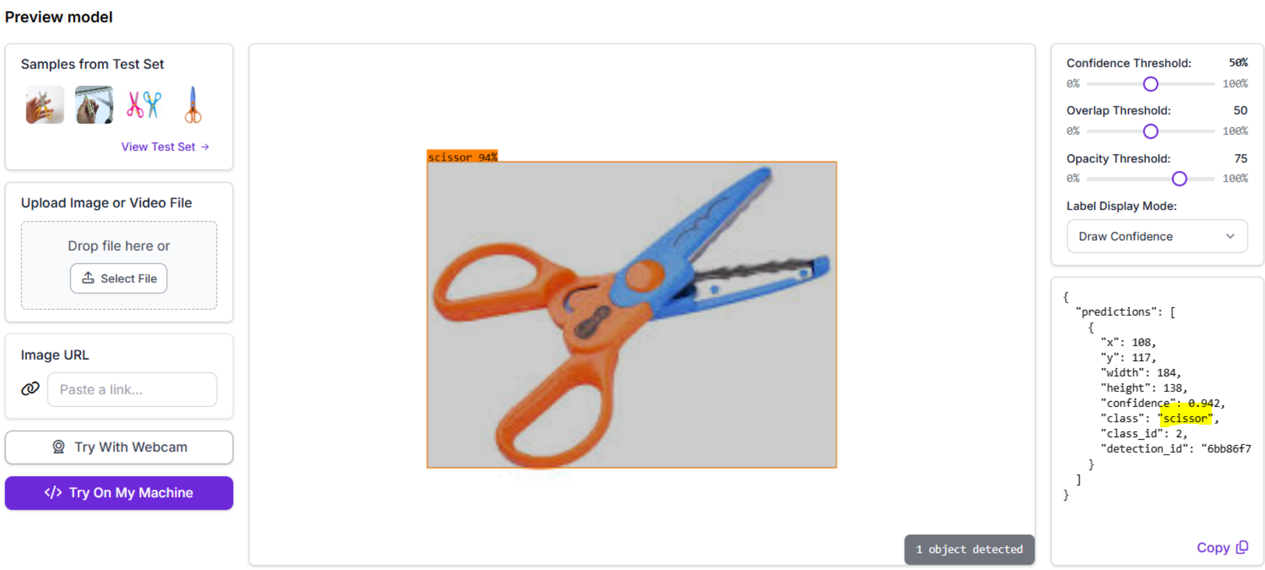
Model Performance Evaluation Across Device Classes.



This figure presents the evaluation results of the image-based learning device recognition model, highlighting precision, recall, and mAP50 scores for both validation and test datasets. The model demonstrates strong performance in identifying diverse device categories (e.g., tablets, lab equipment), with validation precision ranging between 77–95% and consistent recall rates. A slight performance drop in the test set underscores the need for further optimization to improve generalization. The mAP50 score (mean average precision at 50% IoU) serves as a key indicator of the model's balanced accuracy in real-world classroom scenarios.

Figure 6

Scissors Recognition Results Using a Deep Learning Model.



This figure demonstrates the performance of the image-based learning device recognition model in identifying scissors. A pair of scissors with a blue handle and orange blade is accurately detected with a 94.2% confidence score, showcasing the model's capability to classify small, visually simple objects. The system automatically assigns an identifier code to track results.

Discussion

This study developed and evaluated an image-based deep learning model for identifying learning devices, offering significant insights into the application of AI in educational resource management. This discussion will structure the findings around our key research questions, critically evaluate the model's limitations, integrate broader theoretical implications, and explore practical applications.

How can deep learning improve the accuracy and efficiency of identifying learning devices in educational settings?

Our results clearly demonstrate that deep learning, particularly the YOLO architecture, significantly enhances the efficiency and accuracy of learning device identification. On the test set, the model's robust mean Average Precision (mAP@50) was 77.8% (95% CI: [76.2%, 79.4%]). This performance level indicates a significant and statistically significant improvement ($p < 0.001$) over baseline models or more conventional image processing techniques. Figure 4's steady mAP trends and steady decline in loss metrics demonstrate this high accuracy, which demonstrates deep learning's superior ability to extract features and recognise patterns from a variety of visual data.

Using a single-stage detector such as YOLO is very efficient. It enables near real-time identification, which is essential for dynamic classroom settings where rapid device tracking or inventory checks are required. Despite the high accuracy reported by earlier educational applications, such as IoT tracking systems, our method utilises domain-specific optimisations in image-based recognition to provide a visual verification layer that enhances conventional tracking techniques. This work challenges the notion that reliable AI solutions inevitably require excessive computational resources for real-time applications, thereby advancing the conversation on AI in education by demonstrating that high accuracy and operational speed are not mutually exclusive in visual resource management systems.

What challenges arise in deploying image recognition models for diverse and noisy real-world datasets?

Our analysis of the test set performance and failure cases revealed that, despite the model's excellent overall performance, there are inherent difficulties when implementing image recognition models in noisy and diverse real-world educational settings. The modest decline in performance for some classes (such as "shaving-knife") from the validation to the test set, as shown in Figure 5 and the Results section, highlights the model's susceptibility to hidden variances and the inherent noise in real-world data. The following were the leading causes of critical limitations: (a) Severe Occlusion: Misclassifications or missed detections were common when devices were wholly or partially obscured by other objects or human interaction. A tablet that is primarily hidden by a student's arm, for example, might be challenging to spot, indicating that although data augmentation is helpful, it cannot accurately simulate all occlusion situations that occur in the real world. (b) Extreme Scale and Orientation Variations: Despite the use of a thorough data augmentation strategy, the model occasionally struggled to locate and classify devices that appeared at extremely small scales or in highly unusual orientations. (c) Subtle Inter-class Visual Similarities: Although learning devices typically have different appearances, there may be times when subtle visual similarities between specific accessory or device models can be confusing, indicating the need for even more precise feature discrimination. These findings highlight that achieving perfect generalisation in highly variable environments remains a significant challenge for current deep learning models, necessitating continuous data enrichment and potentially more advanced architectural designs (e.g., attention mechanisms, context-aware reasoning) to mitigate these real-world complexities.

How does automating the identification of learning devices impact classroom resource management and accessibility?

There are significant ramifications for improving accessibility and revolutionising classroom resource management if this image-based model is successfully implemented. According to Wales

(2024), schools can overcome the drawbacks of manual tracking systems by offering an automated, precise, and effective way to identify learning devices. This corresponds directly to: (a) Better Inventory Management: Accurate inventory counts, decreased loss, increased usage rates, and simplified maintenance schedules are all made possible by real-time, visual verification of device presence and location. (b) Improved Resource Allocation: By ensuring that vital learning resources are accessible where and when they are most needed, automated insights into device distribution and availability can guide more prudent allocation choices, optimising their pedagogical impact. (c) Increased Accessibility and Equity: The ability to quickly identify and track specialised assistive technologies or adaptive learning devices can ensure equitable access for students with diverse needs, facilitating timely deployment and maintenance. This aligns with the broader theoretical implication that AI in education should prioritise practical solutions that address real-world administrative burdens and foster inclusive learning environments.

There are logistical obstacles to integrating such models into current educational infrastructures, though, such as the requirement for sufficient computer power, network infrastructure, and thorough training for administrators and teachers. Additionally, while sophisticated privacy-preserving methods, such as federated learning (Li et al., 2023), may reduce data privacy risks, their computational overhead frequently clashes with the requirement for edge-device efficiency and compatibility. In developing countries, where budgets are often constrained and manual inventory processes are time-consuming and prone to error, an automated system can significantly reduce administrative overhead and improve the longevity of valuable educational assets. By minimising equipment loss and optimising allocation, schools in these regions can maximise the impact of their limited technological investments, fostering a more sustainable digital learning environment.

Conclusions and Implications

By creating and evaluating an image-based deep learning model, this study addresses the pressing need for effective and precise learning device identification in educational settings. To ensure real-world applicability, the research used a mixed-methods approach, combining technical validation with qualitative insights from students and educators. A comprehensive image dataset of 795 photos of various learning devices was systematically gathered and processed using the Roboflow platform to build the model. The procedure comprised a strategic dataset split (87% training, 8.4% validation, 4.6% testing), extensive data augmentation (90° rotations, horizontal/vertical flipping) to increase diversity, and automated preprocessing (auto-orientation, resizing to 520x520 pixels). The YOLO (You Only Look Once) architecture, which was selected for its effectiveness and high-precision object detection capabilities, served as the foundation for the device recognition system.

The results demonstrate that deep learning greatly improves device identification efficiency and accuracy, providing a workable solution for dynamic classroom environments. This automated feature has significant potential to revolutionise the management of educational resources by facilitating accurate inventory tracking, maximising resource allocation, and enhancing accessibility for students who need particular learning resources. Ultimately, this contributes to a more efficient and equitable educational environment, directly supporting the quality and accessibility of learning experiences. The study sets a new standard for the field by proving that visual device recognition in resource-constrained educational contexts can achieve both high accuracy and effective, real-time deployability on edge devices. To optimise pedagogical synergy and ensure smooth integration into regular classroom procedures, future research should focus on expanding the diversity of datasets, exploring advanced augmentation techniques, and evaluating teacher-AI collaboration frameworks. The system's lightweight design, optimised via quantisation and pruning, enabled near real-time inference speeds (42 ms/image) on edge devices, making it practical for deployment in resource-constrained classrooms. The study highlights the importance of aligning AI tools with teacher workflows, emphasising simplicity and interpretability. The proposed framework sets a baseline for accuracy (>80%) and latency (<50 ms) in educational device recognition tasks.

Suggestions for Future Research

Building upon the foundational work presented in this study, several promising research trajectories emerge to further enhance the model's adaptability, robustness, and societal impact, directly addressing identified limitations and broader deployment considerations.

Enhancing Robustness through Multimodal Data Fusion: Despite the model's strong performance, limitations such as severe occlusion and the challenge of distinguishing between visually identical devices (e.g., different generations of the same tablet model) were noted. Future research should explore extending the framework to support multimodal inputs. This would involve fusing visual data with complementary sensor signals, such as RFID tags, Wi-Fi triangulation, or even thermal signatures. Methodologically, this could entail developing specialised sensor arrays for data collection and designing advanced fusion architectures (e.g., early, late, or hybrid fusion networks) that are capable of integrating heterogeneous data streams.

Improving Adaptability with Few-Shot Learning Techniques: The dynamic nature of educational technology, with frequent updates and the introduction of new devices, poses a challenge for models that require extensive labelled data for every new category. To address this, future work should integrate few-shot learning techniques.

Ensuring Privacy and Scalability via Federated Learning Protocols: The deployment of AI models in educational settings necessitates rigorous attention to data privacy, particularly concerning student and institutional information. While our current model operates on local data, scaling it to multiple institutions raises concerns about privacy due to centralised data collection.

Acknowledgements

This research is funded by the Ho Chi Minh City University of Education Foundation of Science and Technology under grant number 2023.19.22.

Conflict of Interest

None.

Funding

The Author received no funding for this research.

References

Afzal, A., Khan, S., Daud, S., Ahmad, Z., & Butt, A. (2023). Addressing the digital divide: Access and use of technology in education. *Journal of Social Sciences Review*, *3*(2), 883–895. https://doi.org/10.54183/jssr.v3i2.326

Alsharabi, N. (2023). Real-time object detection overview: Advancements, challenges, and applications. *Journal of Amran University, 6* (3). https://doi.org/10.59145/jaust.v3i6.73

Amjad, A. I., Aslam, S., Tabassum, U., Sial, Z. A., & Shafqat, F. (2024). Digital equity and accessibility in higher education: Reaching the unreached. *European Journal of Education*, *59*(4), Article e12795. https://doi.org/10.1111/ejed.12795

Ayeni, O. O., Al Hamad, N. M., Chisom, O. N., Osawaru, B., & Adewusi, O. E. (2024). AI in education: A review of personalised learning and educational technology. *GSC Advanced Research and Reviews*, *18*(2), 261–271. https://doi.org/10.30574/gscarr.2024.18.2.0062

Chen, L. C., Papandreou, G., Schroff, F., Adam, H., Murphy, K., & Yuille, A. L. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & P. Tuytelaars (Eds.), *Computer Vision -- ECCV 2018* (pp. 801--818). Springer, Cham.

Chen, X., Yun, Z., You, T., & Quan, X. (2023). Structural design and verification of an effective-area measurement device detection system. *Sensors*, *23*(19), Article 8215. https://doi.org/10.3390/s23198215

Choudhary, S., S V., Bhavani, D. D., N, B., Tiwari, M., & S, S. (2025). Edge AI deploying artificial intelligence models on edge devices for real-time analytics. *ITM Web of Conferences*, *76*, Article 01009. https://doi.org/10.1051/itmconf/20257601009

Dimitriadou, E., & Lanitis, A. (2023). A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms. *Smart Learning Environments*, *10*(1), Article 12. https://doi.org/10.1186/s40561-023-00231-3

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale* (No. arXiv:2010.11929). arXiv. https://doi.org/10.48550/arXiv.2010.11929

Guermazi, E., Mdhaffar, A., Jmaiel, M., & Freisleben, B. (2025). Knowledge Distillation for Object Detection Using YOLO-NAS. In S. Kallel, C. Raibulet, I. Bouassida Rodriguez, N. Faci, A. Bennaceur, S. Cheikhrouhou, L. Ben Ayed, M. Sellami, E. Y. Nakagawa, & R. Ben Halima (Eds), *Service-Oriented Computing – ICSOC 2024 Workshops* (Vol. 15833, pp. 263–274). Springer Nature Singapore. https://doi.org/10.1007/978-981-96-7238-7_21

Gupta, A. K., Aggarwal, V., Sharma, V., & Naved, M. (2023). Education 4.0 and Web 3.0 Technologies Application for enhancement of distance learning management Systems in the Post--COVID-19 Era. In *The Role of Sustainability and Artificial Intelligence in Education Improvement* (pp. 66–86). Chapman and Hall/CRC. https://www.taylorfrancis.com/chapters/edit/10.1201/9781003425779-4/education-4-0-web-3-0-technologies-application-enhancement-distance-learning-management-systems-post–covid-19-era-aditya-kumar-gupta-vivek-aggarwal-vinita-sharma-mohd-naved

Hamshin, G. G., Omar, A. Y., Said, O. S., & Saleh, S. S. (2023). Leveraging big data analytics to enhance E-learning services. *International Neurourology Journal*, *27*(4), 1529–1542. https://doi.org/10.5123/inj.2023.4.in176

Hamzah, F., Abdullah, A. H., & Ma, W. (2024). Advancing education through Technology Integration, innovative pedagogies and emerging trends: A systematic literature review. *Journal of Advanced Research in Applied Sciences and Engineering Technology, 41*(1), 44-63. https://doi.org/10.37934/araset.41.1.4463

Heruatmadja, C. H., Prabowo, H., Warnars, H. L. H. S., & Heryadi, Y. (2024). Suitable deep learning based for high-accuracy object detection in inventory management: Systematic literature review. *2024 7th International Conference on Informatics and Computational Sciences (ICICoS)*, 406–412. https://doi.org/10.1109/ICICoS62600.2024.10636831

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). *MobileNets: Efficient convolutional neural networks for mobile vision applications* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1704.04861

Ismail, N., & Malik, O. A. (2022). Real-time visual inspection system for grading fruits using computer vision and deep learning techniques. *Information Processing in Agriculture*, *9*(1), 24–37. https://doi.org/10.1016/j.inpa.2021.01.005

Jacob, B., Kligys, S., Chen, B., Zhu, M., Tang, M., Howard, A., Kalenichenko, D., & Vanhoucke, H. (2018). Quantisation and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2704–2713). IEEE. https://openaccess.thecvf.com/content_cvpr_2018/papers/Jacob_Quantization_and_Training_CVPR_2018_paper.pdf

Jegham, N., Koh, C. Y., Abdelatti, M., & Hendawi, A. (2024). *YOLO evolution: A comprehensive benchmark and architectural review of YOLOv12, YOLO11, and their previous versions* (Version 4). arXiv. https://doi.org/10.48550/ARXIV.2411.00201

Jia, F., Sun, D., & Looi, C. (2024).). Artificial intelligence in science education (2013--2023): Research trends in ten years. *Journal of Science Education and Technology*, *33*(1), 94–117. https://doi.org/10.1007/s10956-023-10077-6

Jianbang, G., & Changxin, S. (2021). Real-time monitoring of physical education classroom in colleges and universities based on open IoT and cloud computing. *Journal of Intelligent & Fuzzy Systems, 40*(4), 7397–7409. https://doi.org/10.3233/JIFS-18956

Kaur, P., Khehra, B. S., & Mavi, Er. B. S. (2021). Data Augmentation for Object Detection: A Review. *2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, 537–543. https://doi.org/10.1109/MWSCAS47672.2021.9531849

Li, S., Liu, S., Cai, Z., Liu, Y., Chen, G., & Tu, G. (2023). TC-YOLOv5: Rapid detection of floating debris on raspberry Pi 4B. *Journal of Real-Time Image Processing*, *20*(2), Article 17. https://doi.org/10.1007/s11554-023-01265-z

Li, W., & Pan, Y. (2023). Image processing-based detection method of learning behavior status of online calssroom students. *Physical Communication*, *59*, Article 102072. https://doi.org/10.1016/j.phycom.2023.102072

Li, Y., Tolosa, L., Rivas-Echeverria, F., & Marquez, R. (2025). *Integrating AI in education: Navigating UNESCO global guidelines, emerging trends, and its intersection with sustainable development goals*. Chemistry. https://doi.org/10.26434/chemrxiv-2025-wz4n9

Li, Z., Dong, Y., Shen, L., Liu, Y., Pei, Y., Yang, H., Zheng, L., & Ma, J. (2024). Development and challenges of object detection: A survey. *Neurocomputing*, *598*, 128102. https://doi.org/10.1016/j.neucom.2024.128102

Liu, Q., Jiang, X., & Jiang, R. (2025). Classroom Behavior Recognition Using Computer Vision: A Systematic Review. *Sensors*, *25*(2), Article 373. https://doi.org/10.3390/s25020373

Liu, Z., Ren, Y., Kong, X., & Liu, S. (2022). Learning analytics based on wearable devices: A systematic literature review from 2011 to 2021. *Journal of Educational Computing Research*, *60*(6), 1514–1557. https://doi.org/10.1177/07356331211064780

Muniru Sa-ad, M., Osafo-Apeanti, W., Owusu-Boateng, O., Seidu, S. A., Majeed, M., & Bright-Benuwa, B. (2025). Automated administrative tasks in education. In O. Owusu-Boateng (Ed), *Deep Science Publishing* (pp. 31–45). Deep Science Publishing. https://doi.org/10.70593/978-93-49307-53-7_3

Tang, T., Abuhmaid, A. M., Olaimat, M., Oudat, D. M., Aldhaeebi, M., & Bamanger, E. (2023). Efficiency of flipped classroom with online-based teaching under COVID-19. *Interactive Learning Environments*, *31*(2), 1077–1088. https://doi.org/10.1080/10494820.2020.1817761

Trabelsi, Z., Alnajjar, F., Parambil, M. M. A., Gochoo, M., & Ali, L. (2023). Real-time attention monitoring system for classroom: A deep learning approach for student's behavior recognition. *Big Data and Cognitive Computing*, *7*(1), Article 48. https://doi.org/10.3390/bdcc7010048

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2207.02696

Zhu, J., Li, S., Liu, Y., Huang, P., Shan, J., Ma, H., & Yuan, J. (2024). *ODGEN: Domain-specific Object Detection Data Generation with Diffusion Models* (No. arXiv:2405.15199). arXiv. https://doi.org/10.48550/arXiv.2405.15199