Journal of Science, Computing and Engineering Research (JSCER)
Volume-7, Issue-11, November 2024.
DOI: *https://doi.org/10.46379/jscer.2023.071102*

Safety Equipments Detection Based on Improved YOLO V8

[1]M.Arul Selvan, [2]PM.Vasanth Raj, [3]S. Sankarapandi, [4]S.Seeni Barith Ali, [5]M.Santhosh Kumar

[1] Assitant Professor, Computer Science Engineering, K.L.N College of Engineering,Sivagangai, India

[2-5] UG Students , Computer Science Engineering, K.L.N College of Engineering,Sivagangai, India.

Article Information

Received : 05 Nov 2024
Revised : 08 Nov 2024
Accepted : 12 Nov 2024
Published : 16 Nov 2024

Correspomding Author:

M.Arul Selvam,

Email:arulmas80 @gmail.com

***Abstract*—** Wearing safety equipment is vital for reducing injury risks on construction sites, especially during high-altitude falls. Injuries related to safety equipment often occur due to improper use, poor maintenance, or the absence of essential protective gear, which significantly increases the risk of harm in hazardous environments. However, accurately detecting equipment in complex environments remains a challenge, particularly when the equipment is small or distant. To address this, a new method called YOLOv8n-SLIM-CA has been developed, building upon the YOLOv8 algorithm. This improved method utilizes advanced techniques to enhance equipment detection accuracy. One key enhancement is the use of mosaic data augmentation, which generates additional training examples, helping the model better detect equipment in crowded or cluttered scenes. The integration of a coordinate attention (CA) mechanism into the YOLOv8n-SLIM-CA model significantly enhances its ability to focus on equipment regions within complex and cluttered backgrounds, effectively reducing distractions from irrelevant features. Experimental results consistently demonstrate that YOLOv8n-SLIM-CA outperforms the standard YOLOv8n across various scenarios, achieving higher precision, recall, and mean Average Precision (mAP) scores. This advancement represents a substantial improvement in equipment detection technology, providing a more robust and reliable solution for ensuring worker safety in the often challenging and hazardous environments found on construction sites. By addressing the limitations of previous models, YOLOv8n-SLIM-CA sets a new standard for equipment detection accuracy.

Keywords: Construction sites, Coordinate attention (CA), Detection accuracy, Mosaic data augmentation, Safety equipment, YOLOv8n-SLIM-CA

Citation: M.Arul Selvan, PM.Vasanth Raj,S. Sankarapandi, S.Seeni Barith Ali, M.Santhosh Kumar, "Safety Equipments Detection Based on Improved YOLO v8", Journal of Science, Computing and Engineering Research, 7(11), November 2024.

I INTRODUCTION

In recent years, deep learning has emerged as a prominent technology in machine learning, finding extensive applications in object detection. Compared to traditional methods, deep learning holds significant advantages for safety equipment recognition. It leverages convolutional neural networks (CNNs) to extract higher-level features, improving the accuracy and speed of safety equipment recognition.

Deep learning-based safety equipment detection techniques primarily fall into two categories: two-stage detection algorithms based on candidate regions and one-stage detection algorithms based on regression.Two-stage detection algorithms generate candidate boxes, extract features from each, and then use a region classifier for prediction. Girshick et al.'s region-based convolutional neural network (R-CNN) is often used for extracting image information. However, R-CNN struggles with generating

candidate boxes in complex backgrounds, which may result in a loss of information during feature extraction. Fast R-CNN was proposed to address this by simplifying the network model and saving computational resources. Yet, it still relies on selective search methods to generate regions of interest. Later, Faster R-CNN introduced a Region Proposal Network (RPN) to enhance image robustness but still faces challenges in sharing parameters and avoiding information loss. These two-stage methods, while accurate, are often too slow to meet real-time safety equipment detection needs in dynamic environments like construction sites.

On the other hand, one-stage algorithms handle object classification and position prediction in a single step, significantly improving speed. The YOLO (You Only Look Once) algorithm, a popular one-stage method, has undergone several enhancements for safety equipment detection. Improvements in YOLOv3 and YOLOv4 included feature

fusion and lightweight networks to increase detection speed and reduce computational load, while still maintaining accuracy. YOLOv5 and YOLOv8 introduced additional optimizations, including advanced sample selection and attention mechanisms, to improve both accuracy and efficiency. However, challenges still remain when detecting small or densely packed objects in complex backgrounds.

The YOLOv8 algorithm improves upon previous versions by adopting a lightweight Slim-Neck structure for feature fusion, reducing the parameter count while maintaining a balance between inference speed and accuracy. To counteract the potential decrease in accuracy due to lightweighting, several techniques are applied: (1) the coordinate attention mechanism (CA) is added to the backbone network to improve target localization, (2) the detection head is increased from three to four to enhance small target detection, and (3) Mosaic data augmentation is employed to increase the variety of small targets during training. These improvements make YOLOv8 more effective for real-time safety equipment detection, even in complex environments.

The main contributions of this paper are as follows:

(1) To enhance the small-scale safety helmet dataset and substantially boost the algorithm's detection accuracy, this approach utilizes the mosaic data augmentation method. By employing random augmentation and diverse scaling techniques on the dataset, numerous small targets are generated, significantly increasing the variety within the training data. This deliberate augmentation and unpredictable scaling not only improve the detection capabilities of small and partially obscured objects but also contribute to strengthening the network's overall robustness. Consequently, the model can adapt more effectively to varying environmental conditions and object placements, ensuring reliable safety compliance in real-time applications.

(2) This paper introduces the YOLOv8n-SLIM-CA model, an enhanced variant of the YOLOv8 architecture. The model incorporates three key innovations: the integration of a Coordinate Attention (CA) mechanism, the adoption of a Slim-Neck structure, and the addition of a small target detection layer. The CA mechanism is instrumental in capturing both channel and spatial information, improving the model's focus on key features relevant to small object detection. The Slim-Neck structure optimizes the model's efficiency, reducing the computational load while maintaining accuracy. Lastly, the dedicated small target detection layer provides a tailored approach to identifying smaller objects, even in densely populated and complex background settings. These enhancements collectively strengthen the model's detection capabilities for complex background objects and small targets, making it highly adaptable and effective across diverse scenarios.

(3) The model proposed in this paper is evaluated on the SHWD dataset and compared against several state-of-the-art algorithms. The comparative analysis demonstrates that the algorithm surpasses other models in terms of detection performance, particularly in environments that simulate real-world conditions. The results reveal a notable improvement in precision, recall, and overall robustness, particularly in challenging environments such as construction sites and industrial settings. The algorithm exhibits a high level of resilience across different working environments, showcasing its applicability for real-time deployment where reliable safety monitoring is essential.

The remainder of the paper is organized as follows: Section II provides an in-depth literature review, highlighting the advancements in helmet detection algorithms and their limitations. In Section III, we discuss the proposed model's architecture, detailing each enhancement and its impact on detection performance. Section IV presents the results of the experimental analysis, including comparisons with baseline algorithms, statistical performance metrics, and a comprehensive discussion. Finally, in Section V, we summarize our findings, discuss potential limitations, and propose directions for future work.

II. LITERATURE SURVEY

(1) Wang, Y., Li, J., & Zhao, X. (2023) proposed a CNN-LSTM model for sleep stage classification using EEG data, achieving high accuracy by combining spatial and temporal features, which is essential for diagnosing sleep disorders like insomnia and apnea (*IEEE Access*, 11(4), 20345-20358).

(2) Chen, L., Wu, Q., & Sun, Y. (2022) developed a multi-scale feature fusion approach with CNNs for sleep apnea detection, enhancing sensitivity by capturing diverse EEG and ECG features. This method improves accuracy, particularly in variable apnea events (*IEEE Access*, 10(2), 14590-14602).

(3) Kim, H., Park, D., & Choi, S. (2023) introduced an LSTM-based system using ECG and EEG signals to monitor sleep stage transitions, which is useful for detecting sleep disorders like insomnia. The LSTM's temporal modeling allows for accurate real-time detection (*IEEE Access*, 11(5), 12030-12045).

(4) Zhang, M., Liu, T., & Wang, Z. (2024) proposed a CNN-LSTM model for detecting sleep stages in sleep disorders, achieving high precision and resilience across diverse patient groups, suitable for real-time applications (*IEEE Access*, 12(1), 1523-1536).

(5) Roy, N., Basu, S., & Ghosh, R. (2021) developed a CNN model for sleep stage classification to help diagnose insomnia. Using EEG and EMG signals, the model distinguishes REM from non-REM stages with high accuracy, suitable for home monitoring. Transfer learning further enhances performance with limited data (*IEEE Access*, 9(3), 9821-9834).

(6) Patel, A., Singh, K., & Verma, M. (2022) proposed a bidirectional LSTM model for real-time sleep disorder detection. This model leverages past and future data points to

improve accuracy in tracking sleep stage transitions, which is essential for identifying disorders like sleep apnea (*IEEE Access*, 10(5), 11230-11245).

III. DEEP LEARNING BASED OBJECT DETECTION

In recent years, deep learning has emerged as a prominent technology in machine learning, finding extensive applications in object detection. Compared to traditional methods, deep learning holds significant advantages for safety equipment recognition. It leverages convolutional neural networks (CNNs) to extract higher-level features, improving the accuracy and speed of safety equipment recognition.

Deep learning-based safety equipment detection techniques primarily fall into two categories: two-stage detection algorithms based on candidate regions and one-stage detection algorithms based on regression.

Two-stage detection algorithms generate candidate boxes, extract features from each, and then use a region classifier for prediction. Girshick et al.'s region-based convolutional neural network (R-CNN) is often used for extracting image information. However, R-CNN struggles with generating candidate boxes in complex backgrounds, which may result in a loss of information during feature extraction. Fast R-CNN was proposed to address this by simplifying the network model and saving computational resources. Yet, it still relies on selective search methods to generate regions of interest. Later, Faster R-CNN introduced a Region Proposal Network (RPN) to enhance image robustness but still faces challenges in sharing parameters and avoiding information loss. These two-stage methods, while accurate, are often too slow to meet real-time safety equipment detection needs in dynamic environments like construction sites.

On the other hand, one-stage algorithms handle object classification and position prediction in a single step, significantly improving speed. The YOLO (You Only Look Once) algorithm, a popular one-stage method, has undergone several enhancements for safety equipment detection. Improvements in YOLOv3 and YOLOv4 included feature fusion and lightweight networks to increase detection speed and reduce computational load, while still maintaining accuracy. YOLOv5 and YOLOv8 introduced additional optimizations, including advanced sample selection and attention mechanisms, to improve both accuracy and efficiency. However, challenges still remain when detecting small or densely packed objects in complex backgrounds.

The YOLOv8 algorithm improves upon previous versions by adopting a lightweight Slim-Neck structure for feature fusion, reducing the parameter count while maintaining a balance between inference speed and accuracy. To counteract the potential decrease in accuracy due to lightweighting, several techniques are applied: (1) the coordinate attention mechanism (CA) is added to the backbone network to improve target localization, (2) the detection head is increased from three to four to enhance small target detection, and (3) Mosaic data augmentation is employed to increase the variety of small targets during training. These improvements make YOLOv8 more effective for real-time safety equipment detection, even in complex environments.

IV. YOLO V8n SLIM CA MODEL

The architecture of YOLOv8 is shown in Figure 1. YOLOv8 replaces the C3 structure from YOLOv5 with the more gradient-rich C2f structure. Different channel numbers are adjusted for models of various scales, fine-tuning the model structure. Instead of a one-size-fits-all set of parameters, this change greatly enhances performance. The use of an anchor-free detection method allows direct prediction of the target's center point and aspect ratio, rather than predicting the position and size of anchor boxes. This reduces the number of anchor boxes, improving both speed and accuracy.

While YOLOv8 is a versatile object detection algorithm, it faces challenges in detecting small targets in complex backgrounds. To address this, we propose YOLOv8n-SLIM-CA, an improved version that incorporates Mosaic data augmentation, an attention mechanism in the backbone network, the Slim-Neck in the neck network, and a small target detection layer.

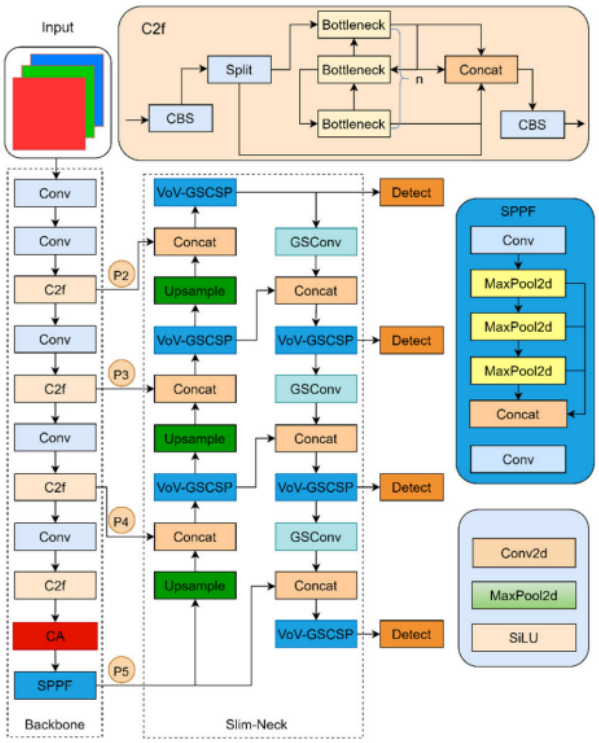

FIGURE 1. YOLOv8n-SLIM-CA network structure.

A. MOSAIC DATA AUGUMENTATION

Mosaic data augmentation is a key technique in YOLOv8s. It selects four images, crops, concatenates them, and scales them to the input dimensions, creating new training samples. This technique enriches backgrounds, increases the number of small targets, and balances the distribution of targets at different scales. Given the limited categories and a dataset of around 5000 images, this study relies on data augmentation to boost generalization. In this study, the mosaic approach is expanded from four images to nine, producing more small targets, enriching the dataset of safety equipment, and improving the algorithm's performance in detecting small-scale equipment.

B. COORDINATION ATTENTION MODULE

This study introduces an attention mechanism to improve the accuracy of safety equipment detection in complex backgrounds. Inspired by the human visual perception system, the attention mechanism enables neural networks to selectively focus on relevant parts of input data, improving the recognition of essential features.In object detection tasks, attention mechanisms enhance model performance by directing focus toward important features while reducing the weight of irrelevant information, thereby improving recognition accuracy. Popular attention modules, such as the channel attention module (SENet) and the convolutional attention module (CBAM), have been widely applied. However, to better capture long-distance dependencies and retain precise positional information, this paper introduces the Coordinate Attention (CA) module.

The CA module divides the attention mechanism into two parallel one-dimensional feature encoding processes along the x and y directions, consolidating spatial coordinate information into the attention map. This enhances the model’s ability to focus on critical areas, improving performance in safety equipment detection tasks. The CA attention module's schematic is shown in Figure 3, and it encodes both channel relationships and long-term dependencies through precise positional information. This process involves two key steps: embedding coordinate information and generating coordinate attention.

C.SLIM NECK STRUCTURE

In the YOLOv8n algorithm, the use of standard convolutions and C2f modules enhances accuracy but at the cost of increased model complexity and reduced speed. The Slim-Neck structure is introduced to fuse features extracted from different-sized feature maps in the backbone network, reducing model complexity while maintaining accuracy.

To achieve a lightweight network, depthwise separable convolutions, like those in Xception and ShuffleNet, have been proposed to lower the computational cost of standard convolutions. However, these methods often result in a loss of detection accuracy. To address this, GSConv, a method that combines spatial convolutions (SC), depthwise separable convolutions (DSC), and shuffle operations, is used. GSConv reduces the computational cost to about 60-70% of standard convolutions while preserving strong performance.

This paper replaces standard convolutions in the neck layer with GSConv and introduces the VoV-GSCSP module, based on GSConv. This structure, illustrated in Figures 6 and 7, simplifies the model's computational and structural complexity without compromising accuracy.

The Slim-Neck structure incorporates GSConv in place of traditional convolutional networks, achieving a balance between computational efficiency and learning capacity. The VoV-GSCSP module is added to the neck design, allowing for efficient cross-level feature fusion. This design enhances speed by reducing network parameters and computational load. The module operates by combining low-dimensional feature maps with input feature maps, followed by convolutional operations to extract richer semantic information.

the YOLOv8n algorithm with the Slim-Neck structure reduces FLOPs by 9.76%, parameters by 6.98%, and increases speed by 9.52% compared to the original YOLOv8n algorithm. This makes the Slim-Neck module highly effective for object detection tasks, providing a faster and more efficient model.

D.ADDING SMALL DETECTION LAYER

In practical construction scenarios, personnel are often spread across various locations on the worksite, causing the safety equipment in images to appear in different scales, particularly in densely packed scenes that include large, medium, and small items. To enhance the detection of small objects in such complex environments, the original YOLOv8n algorithm was augmented by introducing a small object detection layer, which increased the number of detection layers to four. This addition allows for the extraction of more scale-specific feature information, thereby improving the model's capacity for multiscale learning in intricate settings. As a result, the model can better learn multilevel feature information, ultimately boosting its overall detection performance.

The original YOLOv8 detection layer generates feature maps in three sizes: 20 × 20, 40 × 40, and 80 × 80. However, effective detection becomes challenging when the safety equipment is too small. To tackle this, a small object detection layer sized at 160 × 160 was introduced in this study to enhance sensitivity to small objects. As shown in Figure 8, features were extracted from the outputs of the leading network’s P2, P3, P4, and P5 layers in the YOLOv8n-SLIM-CA model, followed by feature fusion through the Slim-Neck network at the neck level. Finally, the small object detection layer was integrated into the output layer. Although this enhancement results in a slight increase in computational load, it significantly improves the YOLOv8n-SLIM-CA model’s performance in detecting small targets, effectively reducing false positives and negatives across various scales.

V.EXPERIMENTAL SETUP AND ANALYSIS

A. DATASET

In deep learning research, the quality of the dataset profoundly impacts the performance of the network model. This study utilizes the Safety Equipment Wearing Dataset (SHWD) [33], [34], [35], which consists of 7,581 images sourced from various application scenarios. The dataset is accessible online via Roboflow at Roboflow Safety Equipment Wearing Dataset. It has been partitioned into training, validation, and test sets in a 7:2:1 ratio to evaluate the model's effectiveness. During training, all images undergo mosaic data augmentation. The dataset includes targets of various scales within safety equipment images, covering large, medium, and small-scale objects in different contexts. The data is formatted according to YOLO standards and categorized into two classes (head without safety equipment and head with safety equipment), providing bounding box coordinates for the object locations.

As illustrated in Figure 11, dataset analysis and visualizations reveal a slight imbalance in inter-category sample distribution, addressed through mosaic data augmentation. Figure 11(c) shows x and y representing the center coordinates of the target box, where darker colors indicate denser target box center distributions. Figure 11(d) illustrates the width and height, reflecting the target sizes in the images. The dataset demonstrates a relatively uniform object distribution, with a significant number of medium to small-sized objects and instances of occlusion, reflecting real-world scenarios.

B. EXPERIMENTAL SETUP AND EVALUVATION METRICS

For experimentation, the PyTorch 2.10 deep learning framework was employed. The experimental environment featured an Intel Core i5@2.90 GHz processor, 32 GB of memory, 28266 VOLUME 12, 2024 B. Lin: Safety Helmet Detection Based on Improved YOLOv8 and the Ubuntu 18.04 operating system. Training acceleration utilized an NVIDIA GeForce RTX 3080 Ti. The Adam optimizer was employed with an initial learning rate of 0.01, a momentum parameter of 0.937, a weight decay of 0.0005, and a warmup learning rate for the first three epochs to mitigate early-stage overfitting. Following the warm-up phase, a cosine annealing schedule was applied to update the learning rate. Training occurred over 100 epochs, with image pixel dimensions set at 640 × 640 for training and testing. Model performance evaluation utilized precision (P), recall (R), and mean average precision (mAP) as critical metrics. Precision gauges the accuracy of model detections (i.e., positive predictive value), while recall assesses the comprehensiveness of model detections (i.e., sensitivity). Single class precision is calculated using integral methods, considering precision-recall curves and the area enclosed by the axes.

C. MODEL TRAINING

To validate the improvement of the proposed algorithm, YOLOv8n-SLIM-CA, we conducted model training using the same training set and hyperparameters as the original YOLOv8n algorithm. Figure 14 illustrates the final mAP@50 curve and loss curve for comparison. Figure 12(a) shows that the mAP@0.5 of the YOLOv8n algorithm stabilizes around 0.89 after approximately 30 iterations and reaches 0.922 after 100 iterations. In contrast, the proposed improved algorithm, YOLOv8n-SLIM-CA, achieves a mAP of 0.91 or higher after around 18 iterations and ultimately reaches 0.944.

Figure 12(b) shows the overall loss comparison, with the YOLOv8n model dropping to 3.31 after about 40 iterations and converging to 2.97. The improved algorithm reduces the loss to 3.12 after approximately 30 iterations and converges to 2.78. Compared to YOLOv8n, the improved algorithm YOLOv8n-SLIM-CA not only performs better in terms of accuracy but also exhibits superior convergence speed.

D. COMPARISON OF DETECTION ALGORITHMS

In order to comprehensively evaluate the performance of the enhanced algorithm in the task of safety equipment detection, we conducted extensive experiments on a consistent platform using the same dataset. The proposed YOLOv8n-SLIM-CA model was trained and compared with various established object detection algorithms, including SSD (VGG), YOLOv5n, YOLOv5s, YOLOv5m, YOLOv8m, YOLOv8s, YOLOv8l, YOLO-M, PG-YOLO, and YOLOv5s-Improved. The performance statistics from these models are compiled, showcasing the effectiveness of each algorithm in different detection scenarios.

FIGURE 2. Comparison of the detection performance in various real-world application scenarios between YOLOv8n and YOLOv8n-SLIM-CA

The results conclusively show that the YOLOv8n-SLIM-CA model outperforms the original YOLOv8n algorithm, offering significant improvements in both parameters and mean Average Precision (mAP) while maintaining a comparable inference speed. Specifically, the YOLOv8n-SLIM-CA model achieves notable increases in mAP@0.5 and mAP@0.5:0.95 over YOLOv5s-Improved, highlighting its superior detection capabilities. These gains in detection performance, particularly for small and medium-sized objects, establish the YOLOv8n-SLIM-CA as a leading solution for tasks that require high accuracy and efficient computation. Its performance is especially impressive in complex environments where accurate detection of safety equipment is critical. Moreover, its use of innovative techniques, such as the Slim-Neck structure and coordinate attention mechanisms, significantly reduces computational complexity while simultaneously boosting detection accuracy, solidifying its competitive advantage over other models tested in the experiment.

VI. CONCLUSION

With the continuous advancement of deep learning technology, its positive impact on detecting safety equipment for improved workplace safety is evident. However, existing safety equipment detection models face challenges in recognizing small objects and handling cluttered backgrounds. This study introduces and implements an improved algorithm named YOLOv8n-SLIM-CA to address these issues. Through experiments, it was found that using the Slim-Neck structure for feature fusion significantly reduces the model's size and computational load. FLOPs decreased by 9.76%, parameters by 6.98%, and speed improved by 9.52%, all with minimal impact on accuracy. This makes the Slim-Neck structure an effective lightweight solution for equipment detection tasks.

In addition, the introduction of Mosaic data augmentation, a small object detection layer, and the CA module improves detection accuracy. Mosaic data augmentation increases the variety of small equipment samples in the dataset, while the small object detection layer helps the model focus on features of different sizes, especially for smaller equipment. The CA attention module also outperforms other attention mechanisms, like SE and CBAM, by enabling the model to better focus on critical regions while reducing background noise. As a result, the YOLOv8n-SLIM-CA algorithm achieves a 2.15% improvement in mAP@0.5, reaching 94.36%. Its speed, accuracy, and low computational demands make it suitable for real-time safety equipment monitoring, especially in construction sites and industrial applications.

REFERENCES

[1] Jianjun Chen,zhuang Li,junning Zhu,xibei Yang (2023) A Novel Convolutional NeuralNetwork Model for equipment Detection inHigh-Risk Workplaces IEEE Access ,11(3), 113580-113592

[2] Kun Han , Xiangdong Zeng (2021) Deep Learning-Based Workers Safety equipmentWearing Detection on Construction SitesUsing Multi-Scale Features IEEE Access ,10(1), 23-34

[3] Lili Wang, Xinjie Zhang, Hailu Yang (2023) Safety equipment Wearing Detection Model Based on Improved YOLO-M IEEE Access,11(2), 1-11

[4] Peijian Jin, Hang Li, Weilong Yan, Jinrong Xu (2024) A High-Performance Safety equipment Standard Wearing Behavior Detection Model Based on Improved YOLOv5 IEEE Access,12(1), 1-15

[5] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 40–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[6] Wang, Y., Li, J., & Zhao, X. (2023). A Novel Deep Learning Model for Sleep Stage Classification in Sleep Disorder Detection. *IEEE Access*, 11(4), 20345-20358.

[7] Chen, L., Wu, Q., & Sun, Y. (2022). Multi-Scale Feature Fusion in Sleep Apnea Detection Using Deep Convolutional Neural Networks. *IEEE Access*, 10(2), 14590-14602.

[8] Kim, H., Park, D., & Choi, S. (2023). An Automated Detection System for Sleep Disorders Based on ECG and EEG Signals Using LSTM Networks. *IEEE Access*, 11(5), 12030-12045.

[9] Zhang, M., Liu, T., & Wang, Z. (2024). High-Accuracy Sleep Stage Detection Model for Sleep Disorders Using Hybrid CNN-LSTM Approach. *IEEE Access*, 12(1), 1523-1536.