Journal of Applied Data Sciences
Vol. 6, No. 3, September 2025, pp. 1968-1984

Development of A Deep Learning Model for Mental Health Classification and Early Screening through Draw a Person (DAP) Test Images

Nurasiah[1,*], Achmad Benny Mutiara[2], Tristyanti Yusnitasari[3], Anugriaty Indah Asmarany[4]

[1,2,3]*Faculty of Computer Science and Information Technology, Gunadarma University, Margonda Raya 100, Depok 16424, Indonesia*

[4]*Faculty of Psychology, Gunadarma University, Margonda Raya 100, Depok 16424, Indonesia*

(Received: December 30, 2024; Revised: February 1, 2025; Accepted: May 1, 2025; Available online: July 13, 2025)

Abstract

Mental health, as defined by the World Health Organization (WHO), is a fundamental aspect of overall well-being. The increasing complexity of modern society, coupled with rising levels of competition and stress, significantly impacts individuals' mental health. The DAP test is a psychological assessment tool that uses human figure drawings to gain insights into an individual's personality and mental condition. YOLO (You Only Look Once) is a deep learning algorithm based on Convolutional Neural Networks (CNNs) designed for real-time object detection. This study utilizes a DAP image dataset contributed by adolescents aged 12 to 16 years to develop a model for detecting and classifying objects in DAP images using the YOLOv8 algorithm. Optimal training results were achieved after 150 epochs, yielding a Precision of 0.821, Recall of 0.799, and mAP50 of 0.88. The model evaluation demonstrated an F1-Score of 0.78, indicating a balanced performance between Precision and Recall. Psychological analysis was conducted based on symptoms extracted from the characteristics of DAP images. Mental health conditions were classified according to severity levels consisting of minor, medium, and serious, based on weighted symptomatology derived from DAP image characteristics. The successful development of this model highlights its capability to classify various mental health conditions based on psychological analysis of DAP images. The findings suggest that mental health classification using DAP test images has the potential to support early screening and psychological assessment by providing an innovative and objective approach to identifying psychological indicators.

Keywords: Draw A Person Image, Mental Health, YOLO Algorithm

1. Introduction

According to the WHO, mental health is a crucial component of overall well-being. WHO defines health not merely as the absence of disease but as a state of complete physical, mental, and social well-being. The rapid advancement of modern society, coupled with rising competition and increasing pressure, has significantly impacted individuals' mental health. This growing pressure affects not only adults but also children and adolescents. The Indonesia National Adolescent Mental Health Survey (I-NAMHS) reports that approximately 34.9% (15.5 million) of Indonesian adolescents experience mental health issues, yet only 2.6% seek support or counseling services. The reluctance to seek professional help is largely influenced by societal stigma, as mental health remains a sensitive and often uncomfortable topic in Indonesia [1].

Mental health plays a vital role in achieving optimal development during adolescence. Early mental health screening can help reduce the risk of mental by identifying emotional and psychological issues at an early stage [2]. This screening process can be conducted using psychological assessment tools designed to evaluate symptoms, behavioral patterns, and factors influencing mental well-being. One such method is projective testing, which assesses mental health by exploring personality dynamics and internal emotions through image interpretation. An example of this is the DAP test, which provides insight into an individual's psychological state based on their drawings.

The DAP test is a psychological assessment tool that utilizes human figure drawings to explore an individual's personality and mental state. Known for its ease of use and broad applicability, the DAP test can be administered across various age groups, from children to adults. As a flexible tool in psychological evaluation, the DAP test provides deeper

***Corresponding author: Nurasiah (nurasiah@staff.gunadarma.ac.id)**
DOI: https://doi.org/10.47738/jads.v6i3.700

insights into an individual's character and mental well-being. Researchers have extensively studied its validity and reliability, leading to ongoing discussions about its effectiveness. The test's ability to offer valuable insights for practitioners, including psychologists and educators, has contributed to its popularity. Its objective, simple, and easily interpretable nature makes it a widely used tool in psychological assessments [3]. In the DAP test, human figure drawings are analyzed to uncover various aspects of personality and emotions, that may not be evident trough interviews or other psychological assessments. The DAP test is performed by examining specific elements in the image such as body proportions, facial details, expressions and the therapist can identify signs of stress, anxiety, depression, or other psychopathological mental [4].

Analyzing the results of the DAP test requires trained experts, as the assessment is conducted in person and manually evaluated. However, this process is time-consuming and susceptible to examiner subjectivity, as interpretations heavily rely on the evaluator’s perspective. These limitations become particularly concerning in light of the increasing prevalence of mental health disorders, which highlights the urgent need for greater mental health awareness [5] and exposes a critical shortage of qualified mental health professionals [6]. Automating psychological test analysis can help minimize subjectivity and improve the accuracy of evaluations, as well as produce mental health classifications for use as an initial screening tool. Computer vision, an advanced technology integrating cameras, software, and artificial intelligence (AI), enables automated recognition and interpretation of visual data. One key application of computer vision is object detection, which uses machine learning algorithms to classify and identify elements within an image. Several object detection techniques, such as R-CNN, Fast R-CNN, and YOLO, are widely recognized for their ability to detect objects rapidly and in real time [7]. This study aims to develop a DAP image detection model as an indicator of mental health classification, contributing to more objective and automated psychological assessments.

2. Related Work

Deep Learning is a branch of Machine Learning that is based on Artificial Neural Networks (ANN) and utilizes the concept of representation learning. Representation learning refers to a set of computational methods that enable machines to analyze raw data and transform it into representations that are easier for machines to understand [8], [9]. Deep learning-based object detection algorithms have the advantage of handling complex patterns and variations in images, providing significant improvements in object detection performance. The commonly used deep learning-based object detection algorithms consist of region-based Convolutional Neural Network (R-CNN), Fast Region-based Convolutional Neural Network (Fast R-CNN) and YOLO. YOLO is the most famous deep learning algorithm that produces much faster detection results [10].

YOLO method was developed to enhance the speed and efficiency of real-time object detection and was first introduced by Joseph Redmon and his team in 2016. Known for its high-performance real-time detection, YOLO is widely used in applications that require fast and accurate object recognition [11], [12]. YOLO is a deep learning algorithm based on CNNs, designed specifically for object detection. It can classify and detect objects with high speed and precision, making it one of the most efficient object detection models available. The researcher Yaseen conducted an in-depth investigation into the YOLOv8 algorithm, emphasizing its architectural innovations, training strategies, and performance characteristics. The study reveals that YOLOv8 constitutes a significant advancement in object detection technology, exhibiting substantial improvements over prior iterations. The algorithm demonstrates enhanced accuracy and faster detection performance across multi-scale objects, maintaining high precision and real-time capabilities across diverse hardware platforms. Moreover, the integration of a comprehensive Python package in YOLOv8 streamlines the processes of model training and deployment, thereby facilitating broader adoption and application in various real-world scenarios [13]. A study comparing FASTER R-CNN, YOLO, and Single Shot MultiBox Detection (SSD) found that while FASTER R-CNN achieved higher accuracy, it was 18 times slower than YOLO, and SSD was 5 times slower than YOLO, making them unsuitable for real-time applications [14].

Several studies have explored the use of image-based detection in psychological tests for mental health assessment. For example, researchers have classified DAP images to evaluate children's intellectual maturity [15]. Another study classified DAP images based on head feature detection alone [16]. Further research has expanded classification techniques by analyzing facial and body features, such as the mouth, teeth, and hands, to identify anger and aggression in children [17]. The Draw a Person in the Rain (DAPR) test has also been used to assess stressful experiences and

 ISSN 2723-6471

coping behaviors based on drawings of people and rain [18]. Additionally, the House-Tree-Person Test (HTP) has been analyzed using object detection techniques to classify mental health indicators [19].

3. Methodology

The object detection stage refers to the steps involved in classifying images using object detection techniques. The algorithm used for detecting images of individuals from the DAP test image utilizes the YOLOv8 algorithm. The stages of DAP image detection model are illustrated in figure 1.

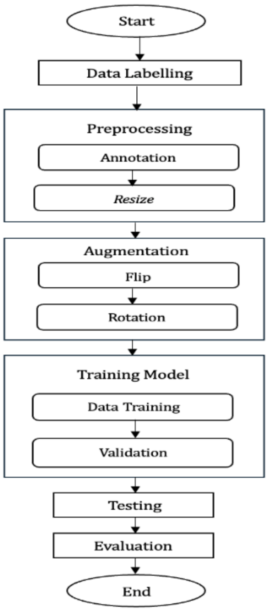


Figure 1. Stages of the DAP Image Object Detection Model.

Figure 1 outlines the stages involved in the object detection model to classify images of individuals. The stages include the model training preparation phase, which consists of data labeling, data preprocessing, and data augmentation. The subsequent phase is the model training stage, where the model is trained and validated during the training process. The final stage, after the model training is completed, involves the testing and evaluation of the model. The next stage involves the classification of potential mental health based on the identified characteristics of the DAP images, which are utilized as indicators of mental health issues through comprehensive psychological analysis using a rule-based approach.

3.1. Data Labeling

The DAP image dataset was collected from junior and high school students aged 12 to 16 years. Students were provided with unlined A4 HVS paper, HB pencils and instructed to draw a person, resulting in a total of 396 images. Data collection ethics encompasses the moral principles and standards that must be adhered to during the data collection process in research or information-gathering activities. The primary objective of data collection ethics is to ensure that data is gathered in a manner that is fair, respects individual privacy, safeguards participant rights, and prevents the misuse of data. In the data collection process, it is essential to ensure that participants fully understand the objectives of data collection and provide informed consent for their participation. In this study, data collection is conducted with the knowledge of the Guidance and Counseling teacher and under the supervision of a psychological expert. Confidentiality and data security are fundamental aspects of the data collection process, especially when handling sensitive information such as mental health. To protect participants privacy, data collection is carried out in an anonymous manner. This approach is implemented to prevent the disclosure of personal identities, ensuring that participants feel secure and protected throughout the data collection process.

Data labeling was conducted based on expert judgment from psychologist expert to identify image characteristics, and 265 images were selected for model development. Table 1 presents the characteristics of DAP images along with their corresponding class names for classifying psychological assessment indicators.

Table 1 describes the DAP images, which are categorized into 20 classes representing psychological assessment indicators based on expert judgment. These classes include large image, small image, left side, right side, bottom side, big head, small head, long hair, blank face, big mouth, mouth omitted, big eyes, small eyes, blank eyes, nose omitted, thick and short neck, thin and long neck, neck omitted, arm omitted and leg without foot.

Table 1. DAP Image Feature

Feature	Class	Feature	Class
Size	Large Image	Mouth	Big Mouth
Size	Small Image	Mouth	Mouth Omitted
Position	Left Side	Eyes	Big Eyes
Position	Right Side	Eyes	Small Eyes
Position	Bottom Side	Eyes	Blank Eyes
Head	Big Head	Nose	Nose Omitted
Head	Small Head	Neck	Thick and Short Neck
Hair	Long Hair	Neck	Thin and Long Neck
Face	Blank Face	Neck	Neck Omitted

The identification of the 20 characteristics was conducted by creating bounding boxes around the drawing objects, which included the following image classes: person, shoulder, normal nose, normal leg, leg without foot, head, neck, normal arm, arm omitted, eyes, blank eyes, mouth, normal hair, long hair, stickman, and blank face. Table 2 provides a representation of image classes used to determine mental health indicators.

Table 2. Representation of Image Classes as Mental Health Indicator

Image Class	Class Indicator
Person	Large Image, Small Image, Left Side, Right Side, Bottom
Shoulder	Big Head, Small Head
Head	Big Head, Small Head
Normal Nose	Nose Omitted
Normal Leg	Normal Leg
Leg without Foot	Leg without Foot
Neck	Thick and Short Neck, Thin and Long Neck, Neck Omitted
Normal Arm	Normal Arm
Arm Omitted	Arm Omitted
Eyes	Big Eyes, Small Eyes
Blank Eyes	Blank Eyes
Mouth	Big Mouth, Mouth Omitted
Normal Hair	Normal Hair
Long Hair	Long Hair
Stickman	Small Image, Big Head, Neck Omitted
Blank Face	Blank Face

3.2. Preprocessing

Preprocessing is a critical stage in data processing conducted before utilizing the data for model training and testing. This phase focuses on cleaning, organizing, and transforming the data to ensure its compatibility with the intended model. Effective preprocessing improves data quality by eliminating irrelevant or duplicate information, addressing missing values, and standardizing data formats. By applying appropriate preprocessing techniques, the resulting model can achieve higher accuracy, enhanced efficiency, and improved generalization to new data.

3.2.1. Annotation

Annotation of the DAP image is performed using bounding boxes, applied to each body part. The annotated classes include person, shoulder, nose, leg, leg without foot, head, neck, arm, arm omitted, eyes, blank eyes, mouth, hair, long hair, stickman, and blank face. The annotations of body parts in the DAP image are as shown in figure 2 and figure 3.

Figure 2 represents the results of annotations on DAP images, which include the classes of person, normal hair, eye, normal nose, mouth, neck, shoulder, normal arms, and normal leg. Figure 3 provides a detailed representation of the annotations applied to the DAP images for the stickman class, highlighting the specific features and characteristics identified within this class for the purpose of psychological analysis.

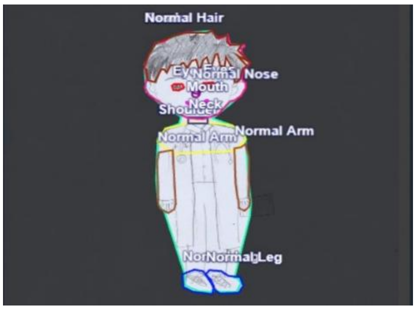


Figure 2. Annotation DAP Image

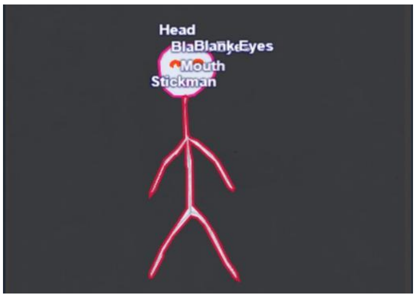


Figure 3. Annotation DAP Image for Stickman Class

Stickman is a term used to describe a simple illustration of a human figure, typically drawn using straight lines for the body, arms, and legs, and a small circle for the head. In the context of the DAP (Draw-a-Person) test, a stickman refers to a human figure drawn in a very basic and minimalistic style, which can provide insights into an individual's psychological state or perceptions.

3.2.2. Resize

At the resize stage, the images are adjusted to ensure uniformity in the drawings. The DAP image data used for training the detection model is resized to 640 × 640 pixels. The resize function in the YOLO algorithm standardizes the input image size, ensuring consistency across all images processed by the model. This resizing step is crucial because YOLO operates by dividing the image into a fixed grid, with each grid cell responsible for detecting objects. By resizing the image to 640 x 640 pixels, the computational load is reduced, which enables faster inference and lower memory usage, while maintaining a balance between detection accuracy and processing speed. Additionally, resizing facilitates the model's scalability, allowing it to handle input images of varying sizes while ensuring optimal performance in object detection tasks.

3.3. Augmentation

Data augmentation is a technique employed to expand the size and diversity of a dataset by applying specific transformations to the original data. The primary objectives of data augmentation are to improve the model's generalization capacity, reduce the risk of overfitting, and simulate potential real-world variations that may arise in unseen data. Augmentation is applied only to the training data to enhance data variability without affecting the validity of the validation and test data. The augmentation techniques employed include flipping (horizontal and vertical), as shown in figure 4; rotation (90° clockwise, 90° counterclockwise, and upside down), as shown in figure 5; angle rotation within a range -15° to the left and +15° to the right as shown in figure 6.

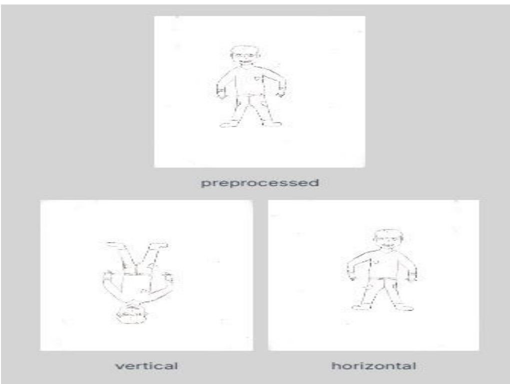


Figure 4. Flipping Augmentation

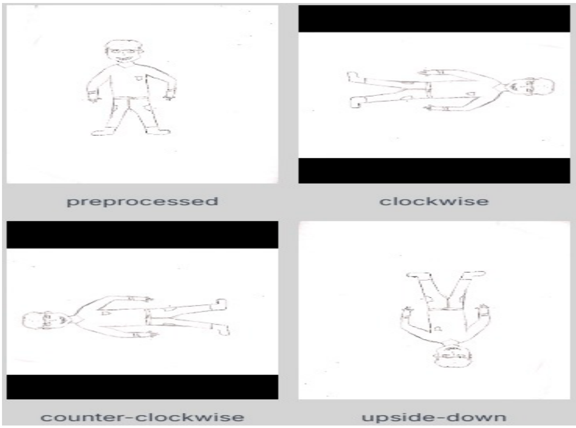


Figure 5. 900 Rotation Augmentation

 ISSN 2723-6471

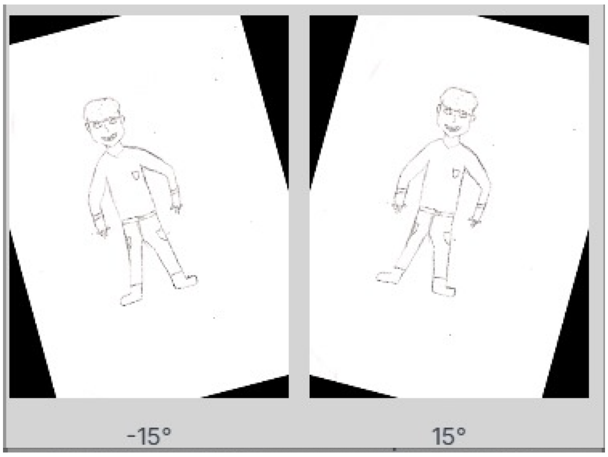


Figure 6. Angle Rotation Augmentation at -150 and +150

The augmentation process is performed using Roboflow after the dataset is divided into three subsets: training, validation, and testing data. Augmentation was performed three times the original dataset size. Augmentation is applied only to the training data to preserve the accuracy of the evaluation process and to mitigate the risk of data leakage. In this study, two data split scenarios were implemented, consisting of training (70%), validation (15%), testing (15%) and training (80%), validation (10%), testing (10%), as shown in table 3.

Table 3. Split DAP Image Datasets Before and After Augmentation

Data Split	Ratio (%)	Number of Samples (Before)	Number of Samples (After)
Training	70%	187	560
Validation	15%	39	39
Testing	15%	39	39
Training	80%	213	639
Validation	10%	26	26
Testing	10%	26	26

The number of data, originally 265, after the augmentation process, with a training dataset split of 70%, validation of 15%, and testing of 15%, is 638. When the training dataset split is 80%, validation 10%, and testing 10%, the total increases to 691.

3.4. Training Model

The training stage for detecting and classifying DAP begins with the design of dataset distribution. The dataset is divided into training data, validation data, and testing data. The dataset used in the training process is the dataset after the augmentation process. The model training was conducted using two dataset split scenarios. In the first scenario, the dataset was divided into 560 samples (70%) for training, 39 samples (15%) for validation, and 39 samples (15%) for testing. In the second scenario, the dataset was divided into 639 samples (80%) for training, 39 samples (10%) for validation, and 39 samples (10%) for testing. The details of the dataset split are presented in table 4.

Table 4. DAP Datasets

Scenario	Training Data	Validation Data	Testing Data
1	560	39	39
2	639	26	26

The object detection model for DAPs is implemented using the YOLOv8 algorithm. The training process for the object detection model involves multiple epochs, allowing the model to learn more complex patterns and improve its performance. An epoch refers to one complete pass through the entire training dataset. During each epoch, the model processes every sample in the dataset once to update the weights based on the computed errors. The training process aims to enable the model to learn the patterns and characteristics of the data, thus allowing it to make accurate predictions or classifications. The final outcome of the training process is a trained model, ready to be applied to new data.

3.5. Evaluation Model

The evaluation stage is conducted to ensure that the model can generalize effectively and produce accurate results in detecting and classifying DAP characteristics. The performance of the detection and classification model is measured using evaluation metrics such as precision, recall, F1-score and the confusion matrix.

3.6. Classification of Object Detection in DAPs

The classification of 20 types of DAP characteristics is performed using the object detection model that has been successfully developed. The object detection process is based on the identification and localization of objects within bounding boxes. The DAP characteristics related to size, such as large, small, big mouth, big eye, and small eye, are classified based on ratio values. The characteristics of large and smalls are classified according to the person class. The ratio values are obtained by grouping DAPs into categories of large, small, and normal-sized persons based on expert judgment, followed by the calculation of ratio values using the object detection model, based on the bounding box area of the person class. DAPs with big eye, small eye, and normal eye are grouped according to expert judgment, and the ratio values are calculated based on the bounding box area of the eye class using the object detection model.

Similarly, DAPs with big mouth, small mouth, and normal mouth are grouped based on expert judgment, and the corresponding ratio values are determined based on the bounding box area of the mouth class. A decision tree model is then employed to classify the categories of large, small, big eye, small eye, and big mouth based on the calculated ratio values. The decision tree results show that a person is classified as small if the ratio value is ≤ 0.106, as normal if the ratio value is ≥ 0.106 and ≤ 0.214, and as large if the ratio value is > 0.214. The ratio value for small eye is ≤ 0.010, for normal eye is > 0.010 and ≤ 0.016, and for big eye is > 0.016. The ratio value for small mouth is ≤ 0.017, for normal mouth is > 0.017 and ≤ 0.033, and for big mouths is > 0.033.

The classification of neck characteristics, including thin and long neck, thick and short neck, and neck omitted, is determined based on the ratio of the bounding box width to its height for the neck class, with the normal person, determined through expert judgment, serving as the reference. Specifically, if the width of the neck bounding box is less than or equal to 0.90 times its height, the neck is classified as thin and long neck. If the width is greater than 0.90 but less than or equal to 1.10 times the height, it is classified as a normal neck. Otherwise, it is classified as a thick and short neck. If the neck bounding box is not detected, it is classified under the neck omitted category.

The position of the DAP, whether on the left, right, or bottom, is categorized based on the bounding box of the person class. The center position of the bounding box is calculated from the width (xcenter = x2 - x1) and the height (ycenter = y2 - y1) of the bounding box. If the xcenter is less than half the width of the paper, the is positioned on the left side; if the xcenter is less than two-thirds of the width of the paper, the is centered; otherwise, the is positioned on the right side. If the ycenter is greater than or equal to half the height of the paper, the is positioned at the bottom.

The classification of large and small heads is based on the bounding boxes of the head and shoulder classes, with the normal person determined through expert judgment, serving as the reference. If the width of the head bounding box is less than 0.75 times the width of the shoulder bounding box, it is classified as a small head. If the width of the head bounding box is greater than 0.75 but less than 1.25 times the width of the shoulder bounding box, it is classified as a normal head. If the width of the head bounding box is greater than 1.25 times the width of the shoulder bounding box, it is classified as a big head.

3.7. Psychological Analysis

Psychological analysis is conducted based on the detected characteristics of the DAP as an indicator of potential mental health. The types of mental health with their corresponding codes are presented in table 5.

Table 5. Codes of Mental Health

Mental Health	Code
Aggression	A
Social Anxiety Disorder	B
Depression	C
Interpersonal Avoidance	D
Self-esteem	E
Emotional Instability	F
Seeking Affection	G
Inferiority Complex	H
Regression	I

Table 6 presents the names of various types of mental health that will be classified based on the characteristics of DAPs. The mental health include aggression, social anxiety disorder, depression, interpersonal avoidance, self-esteem, emotional instability, seeking affection, inferiority complex, and regression [20]. The names of the symptoms of mental health are listed in table 6. There are 20 DAP characteristics that serve as indicators symptoms of mental health.

Table 6. Symptoms of Mental Health Code

Symptoms of Mental Health	Code	Symptoms of Mental Health	Code
Large	G1	Mouth Omitted	G11
Small	G2	Big Eyes	G12
Left Side	G3	Small Eyes	G13
Right Side	G4	Blank Eyes	G14
Bottom Side	G5	Nose Omitted	G15
Big Head	G6	Thick and Short Neck	G16
Small Head	G7	Thin and Long Neck	G17
Long Hair	G8	Neck Omitted	G18
Blank Face	G9	Arm Omitted	G19
Big Mouth	G10	Leg without Foot	G20

3.7.1. Mental Health Classification Rules

Psychological analysis is defined as the automated evaluation of s by identifying detected characteristics associated with mental health. This process utilizes a rule-based approach to systematically generate the resulting representations of the s. The rules for determining the type of mental health condition are based on symptoms identified through the characteristics of the DAP, as referenced in [19] and expert validation, as presented in table 7.

Table 7. Mental Health Classification Rules

Rule	Conditional Rule
1	If G1, G4, G6, G10, G16, G18 then Aggression
2	If G2, G4, G5, G9, G11, G12, G13, G15, G17, G20 then Social Anxiety Disorder
3	If G2, G5, G8, G9, G19, G20 then Depression
4	If G2, G9, G19 then Interpersonal Avoidance
5	If G2, G5, G7, G9, G10, G17, G19, G20 then Self-esteem
6	If G2, G4, G5, G7, G8, G9, G10, G11, G12, G14, G19, G20 then Emotional Instability
7	If G3, G6, G8, G9, G10, G17 then Seeking Affection
8	If G2, G3, G5, G7, G17 then Inferiority Complex
9	If G2, G6, G7, G8, G9, G10, G18 then Regression

There are nine rules corresponding to nine types of mental health, each classified based on the characteristics observed in DAP s. For example, in Rule 1, if the DAP depicts a large, position in right side, big head, big mouth, thick and short neck, neck omitted are indicate aggression.

3.7.2. Weighting of Mental Health Symptoms

Weighting is the process of assigning values to symptoms of mental health to reflect their level of importance, priority, or influence. These weights help differentiate the contribution of each symptom, ensuring that more significant symptoms have a greater impact than less significant ones. The weighting process is conducted by expert psychologists. Table 8 presents the assigned values or weights for each symptom based on the classification rules for mental health.

Table 8. Weighting of Mental Health Symptoms

Mental Health Issue	Symptom Code and Weight (Expert Judgment)	Max Total Weight
Aggression	G1=2, G4=1, G6=3, G10=3, G16=3, G18=3	15
Social Anxiety Disorder	G2=2, G4=1, G5=1, G9=3, G11=3, G12=2, G13=2, G15=3, G17=3, G20=3	23
Depression	G2=2, G5=1, G8=3, G9=3, G19=3, G20=3	15
Interpersonal Avoidance	G2=2, G9=3, G19=3	8
Self-Esteem	G2=2, G5=1, G7=3, G9=3, G10=3, G17=2, G19=3, G20=3	20
Emotional Instability	G2=2, G4=1, G5=1, G7=3, G8=2, G9=3, G10=3, G11=3, G12=3, G14=3, G19=3, G20=3	30
Seeking Affection	G3=2, G6=3, G8=2, G9=3, G10=3, G17=3	16
Inferiority Complex	G2=2, G3=1, G5=1, G7=3, G17=3	10
Regression	G2=2, G6=3, G7=3, G8=2, G9=3, G10=3, G18=3	19

The symptom weighting process is determined based on expert judgment. A score of 3 is given when all symptom classes are identified in the DAP for each mental health issue; a score of 2 is given if only one symptom class is not present; and a score of 1 is assigned when more than one symptom class is not identified. The details of the symptom classes can be found in table 6. These weights are used to assess the severity of mental health based on the severity level framework, using the following formula (1).

$$\text{Percentage} = (\text{Symptom Score}) / (\text{Maximum Weight}) \times 100\ \% \tag{1}$$

The severity of a mental health is determined by the percentage obtained from comparing the symptom weight score to the maximum possible weight [21]. The following are the criteria for determining the severity of mental health based on the interpretation of the maximum score and its corresponding severity level. The severity levels of mental health are classified as follows: Minor for a score between 50% and 69%, Medium for a score between 70% and 89%, and Serious for a score of 90% or higher.

4. Results and Discussion

This section presents the research findings and provides a comprehensive analysis of the obtained results. It includes the training of the DAP detection model, the performance evaluation of the DAP detection model, the interpretation of detection accuracy metrics, and the discussion of their implications in mental health classification. The results are analyzed based on key evaluation metrics such as precision, recall, mAP50 and mAP 50-95 to assess the effectiveness of the proposed model.

4.1. Detection Training Model

The training process for the DAP detection model was conducted to classify various categories, including shoulders, nose, leg, leg without foot, head, neck, arm, arm omitted, eye, blank eye, mouth, person, hair, long hair, stickman and blank face. These classifications serve as key indicators in mental health classification. Training Model was conducted using two different data split scenarios. The outcomes of model training using a dataset split of 70% for training, 15% for validation, and 15% for testing are presented in table 9.

ISSN 2723-6471

Table 9. Training Results of the Model

Class		Instance	P	R	mAP50	mAP50-95
All	39	417	0.747	0.814	0.837	0.630
Shoulder	38	38	0.761	0.921	0.912	0.614
Normal Nose	29	30	0.652	0.600	0.653	0.333
Normal Leg	24	33	0.589	0.636	0.638	0.561
Leg without Foot	3	4	1.000	0.807	0.995	0.737
Head	39	39	0.897	0.974	0.968	0.912
Neck	37	37	0.677	0.811	0.798	0.525
Normal Arm	21	38	0.820	0.947	0.940	0.732
Arm Omitted	9	18	0.558	0.444	0.507	0.341
Eyes	27	54	0.730	0.685	0.728	0.430
Blank Eye	8	16	0.487	0.562	0.498	0.319
Mouth	34	34	0.788	0.873	0.878	0.519
Person	38	38	0.979	1.000	0.995	0.979
Normal Hair	16	16	0.745	0.875	0.950	0.776
Long Hair	18	18	0.828	0.944	0.943	0.800
Stickman	1	1	0.435	1.000	0.995	0.697
Blank Face	3	3	1.000	0.937	0.995	0.804

The model training results presented in table 9 correspond to training with 150 epochs, yielding a Precision (P) of 0.747, Recall (R) of 0.814, mAP50 of 0.837, and mAP 50-95 of 0.63 for all classes. The outcomes of model training using a dataset split of 80% for training, 10% for validation, 10% for testing and training epoch 150 are presented in Table 10. The overall detection results for all classes were as follows: Precision (P) of 0.739, Recall (R) of 0.809, mAP50 of 0.828, and mAP 50-95 of 0.623.

Based on the model training results with 100, 150, and 200 epochs, the 150-epoch model achieved the best performance, with the highest mAP50 value of 0.837 and 0.88. This indicates that 150 epochs provided the optimal balance between precision, recall, and overall detection accuracy for DAP detection. Table 10 and table 11 provides a comparative evaluation of model performance under two dataset split strategies: 70% training with 30% validation and testing, and 80% training with 10% validation and 10% testing. Although the overall differences are not statistically significant, the mAP score for the 80% training split (0.88) exceeds that of the 70% split (0.837). Accordingly, the model trained with the 80:10:10 distribution is identified as the most effective for DAP object detection tasks.

Table 10. Training Results of the Model

Class		Instance	P	R	mAP50	mAP50-95
All	26	280	0.821	0.799	0.880	0.687
Shoulder	25	25	0.794	0.794	0.955	0.652
Normal Nose	20	20	0.878	0.878	0.797	0.397
Normal Leg	15	19	0.564	0.564	0.641	0.591
Leg without Foot	3	4	1.000	1.000	0.995	0.872
Head	26	26	0.967	0.967	0.995	0.938
Neck	25	25	0.725	0.725	0.786	0.528
Normal Arm	13	25	0.796	0.796	0.955	0.801
Arm Omitted	9	18	1.000	1.000	0.698	0.463
Eyes	18	36	0.886	0.886	0.765	0.474
Blank Eye	4	8	0.448	0.448	0.686	0.543

Mouth	23	23	0.876	0.876	0.902	0.602
Person	25	25	0.974	0.974	0.995	0.983
Normal Hair	12	12	0.785	0.785	0.930	0.784
Long Hair	10	10	0.889	0.880	0.995	0.923
Stickman	1	1	0.579	0.570	0.995	0.597
Blank Face	3	3	1.000	1.000	0.995	0.845
All	26	280	0.821	0.799	0.880	0.687
Shoulder	25	25	0.794	0.794	0.955	0.652

Table 11. Training Results of the 150-Epoch Model

Data Split Ratio	Epoch	Training Time	mAP50	Preprocess	Speed Inference	Loss	Postprocess
70%:30%	150	0.871 hours	0.837	0.2ms	19.4ms	0.0ms	1.2ms
80%:20%	150	0.952 hours	0.88	0.2ms	20ms	0.0ms	1.3ms

4.2. Evaluation Results of the Model

The evaluation results of the model training are presented through the Precision confidence curve, Recall confidence curve, F1-Score curve and Confusion Matrix. The Precision confidence curve for evaluation model is illustrated in figure 7.

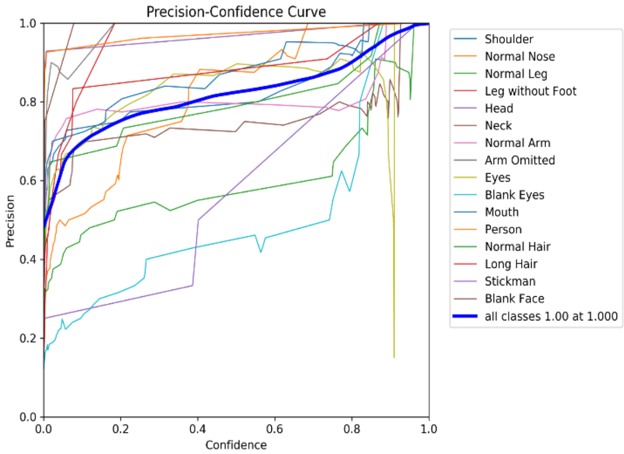


Figure 7. Precision Confidence Curve.

The figure 7 presented illustrates a Precision-Confidence Curve, serving as a quantitative measure to evaluate the classification performance of the model by plotting precision as a function of the predicted confidence scores across multiple object classes. The bold blue curve depicts the macro-averaged precision computed across all classes, achieving a perfect precision score of 1.00 at the maximum confidence threshold, thereby indicating that the model's predictions at full certainty are completely accurate. The majority of the class-specific curves demonstrate a positive correlation between prediction confidence and precision, suggesting that the model is generally well-calibrated and reliable across a wide range of confidence levels. However, certain classes, such as Normal Nose, Blank Eyes, and Stickman, exhibit lower or more fluctuating precision, particularly within intermediate confidence intervals, highlighting potential areas for further model refinement and calibration improvements. The Recall confidence curve for evluation model is illustrated in figure 8.

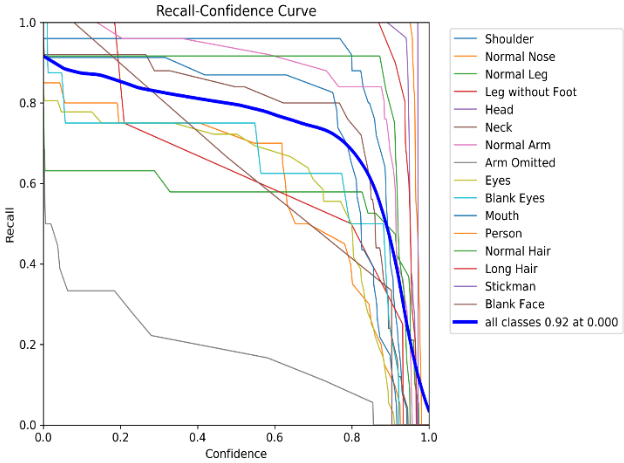


Figure 8. Recall Confidence Curve.

The figure illustrates the Recall-Confidence Curve, which depicts the relationship between recall values and the confidence levels of a classification model across various object classes. The model generally demonstrates strong detection capabilities at low to moderate confidence levels but experiences a significant drop in recall when only high-confidence predictions are retained. The bold blue curve represents the macro-averaged recall across all classes, indicating that the overall recall is approximately 0.92 at a confidence level of 0 and declines sharply once the confidence exceeds 0.7. Certain classes, such as Blank Face and Long Hair, maintain relatively high recall across various confidence thresholds, while others, including Arm Omitted and Blank Eyes, show low or highly variable recall performance. The F1-confidence curve for evaluation model is illustrated in figure 9.

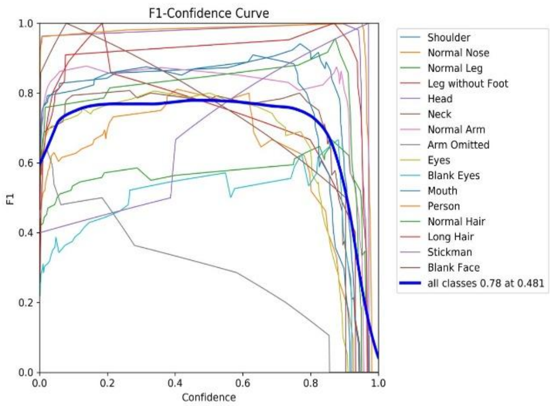


Figure 9. F1-Confidence Curve.

The F1-Score or F1-Confidence Curve in figure 9 presents the model evaluation by illustrating the relationship between the F1-Score and confidence value for each class. Each colored line represents the F1-Score variation with confidence for a specific class, while the bold blue line indicates the average performance across all classes. According to the curve, the highest F1-Score achieved is 0.77 at a confidence value of 0.358. This confidence value provides the best balance between Precision and Recall for all classes. The evaluation of the classification model in the DAP is also presented in the Confusion Matrix in figure 10.

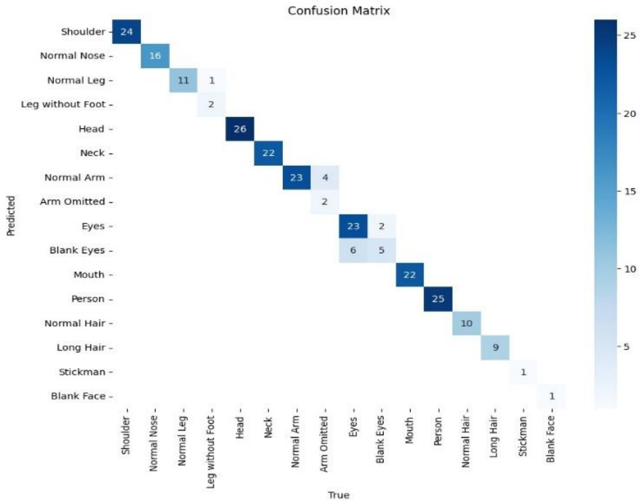


Figure 10. Confusion Matrix Curve

The Confusion Matrix curve illustrates how the model classifies data into different classes in the DAP. The detection results in the DAP, based on the Confusion Matrix from model training, are presented in table 12. Table 12 presents the classification results of DAP s as shown in the Confusion Matrix. Model testing demonstrates high accuracy in correctly identifying in DAP s.

Table 12. Training Results of the 150-Epoch Model

Class	Number of s	Number of s Detected
Shoulder	24	24
Nose	16	16
Normal Leg	12	11
Leg without Foot	2	2
Head	26	26
Neck	22	22
Arm	27	23
Arm Omitted	2	2
Eyes	25	23
Blank Eyes	11	5
Mouth	22	22
Person	25	25
Hair	10	10
Long Hair	9	9
Stickman	1	1
Blank Face	1	1

4.3. Results of Detection and Classification DAP s

The detection results from the DAP used the YOLOv8 algorithm which was used as a symptom for the classification of types of mental health. The detected characteristics from the DAP s include various features such as the size of the (large , small), position within the (left, right, bottom), head size (big head, small head), and specific facial features (blank face, big mouth, mouth omitted, large eye, small eye, blank eye, nose omitted), neck characteristics (thick and short neck, thin and long neck, neck omitted), long hair, arm omitted, and leg without foot. The detection results obtained using the YOLOv8 algorithm as shown in figure 11. The classification of mental health is determined based on the symptom characteristics observed in DAP s and analyzed according to established rules. The results of the detection and classification of s of people based on expert judgment are presented in table 13.

 ISSN 2723-6471

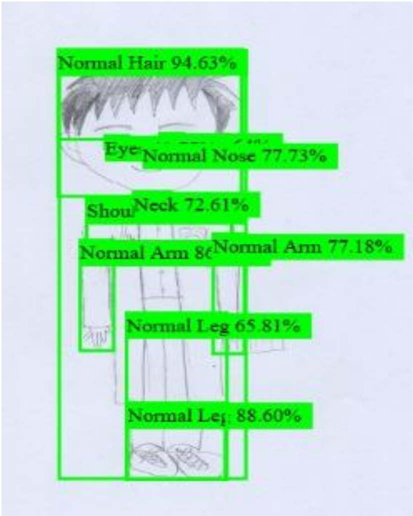
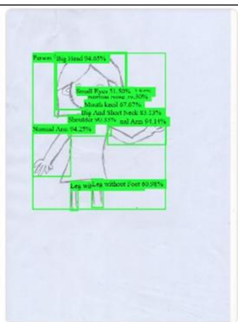


Figure 11. The Results of DAP   Detection

Table 13 illustrates the results of person   detection, which partially correspond with the expert judgment. The detection outcomes, generated by the object detection model trained for this task, align with the expert's predictions.

Table 13. Expert Judgment Results for Partial Detection

Image	Detected Category	Expert Judgment
 	Leg without Foot Normal Nose Thin and Long Neck Big Mouth Normal Head Normal Hair Normal Eyes Arm Omitted	Valid
 	Normal Arm Normal Head Normal leg Normal Nose Normal Mouth Large Thick and Short Neck Long Hair Normal Eyes	Valid
 	Big Head Thick and Short Neck Leg without Foot Small Mouth Large Normal Arm Long Hair Small Eyes Normal Nose	Valid

4.4. Results of Psychological Analysis

The classification of mental health is determined based on the symptom characteristics found in DAP s and analyzed according to the established rules. The results of detecting aggression in individuals, as indicated by their DAP s, are presented in figure 12.

Figure 12. The Results of the Mental Health Classification.

The detection results of individuals in figure 12, exhibiting symptoms large (G1), large heads (G6), long hair (G8), large mouths (G10), small eyes (G13), thick and short necks (G16), are identified based on the characteristics observed in DAP s. These symptoms are determined through a weighted calculation of the detected characteristics. The results of the psychological analysis for mental health classification, based on the weighted scores of these characteristics, are presented in table 14.

Table 14. Weighting of detected symptoms

Mental Health	Detected Symptoms	Total Weight	Weighting Result (%)
Aggression	G1 (2), G6 (3), G10 (3), G16 (3)	11	73.3%
Social Anxiety Disorder	G13 (2)	2	8.7%
Depression	G8 (3)	3	20.0%
Interpersonal Avoidance	–	–	–
Self-Esteem	G10 (3)	3	15.0%
Emotional Instability	G8 (3), G10 (2)	5	16.7%
Seeking Affection	G6 (3), G8 (2), G10 (3)	8	50.0%
Inferiority Complex	–	–	–
Regression	G6 (3), G8 (2), G10 (3)	8	42.1%

The symptoms identified in the characteristics are analyzed according to predefined rules to determine the type of mental health and calculate its corresponding weight. Based on the weighting results, the following percentages were obtained: Aggression (73.3%), Social Anxiety Disorder (8.7%), Depression (20%), Interpersonal Avoidance (not classified), Self-Esteem (15%), Emotional Instability (16.7%), Seeking Affection (50%), Inferiority Complex (not classified), and Regression (41.1%). Considering the weight of the final indicates medium aggression and minor Seeking Affection. The results of the psychological analysis have been validated by expert psychologists and mental health nursing specialists.

5. Conclusion

A model for detecting and classifying objects in DAP s has been developed using the YOLOv8 algorithm. The model was trained with 100, 150, and 200 epochs, using a batch size of 32. The best detection performance was achieved at 150 epochs, yielding a Precision of 0.821, a Recall of 0.799, and a mAP50 of 0.88. The mAP50 value of 88% indicates that the model has strong object detection capabilities and is suitable for implementation in an application system. The model evaluation resulted in an F1-Score of 0.78 (78%), demonstrating a good balance between Precision and Recall. This model successfully classifies DAP detection results for psychological analysis, determining both the type and severity level of mental health based on establish rules. The classified mental health include Aggression, Social Anxiety Disorder, Depression, Interpersonal Avoidance, Self-Esteem, Emotional Instability, Seeking Affection, Inferiority Complex, and Regression.

The resulting object detection model is capable of detecting objects and classifying symptoms of mental health within milliseconds. These findings indicate that, by utilizing an artificial intelligence-based system, the Draw a Person psychological assessment test can be projected rapidly. However, the model's accuracy needs to be improved by increasing the amount of training data, particularly data involving biased cases such as blank faces, blank eyes, and arm omitted. To improve model accuracy, data augmentation can be applied in greater detail, particularly in terms of lighting conditions, such as adjustments to brightness, contrast, and illumination variability.

The results of the mental health classification in this study do not constitute a final of mental health conditions. These outcomes are intended solely for preliminary mental health screening and may serve as a reference for determining subsequent steps in the treatment of mental health using more comprehensive assessment tools. Furthermore, the mental health classification model based on DAP feature extraction cannot be generalized, as the training data were limited to individuals in the adolescent age group.

6. Declarations

6.1. Author Contributions

Conceptualization: N., A.B.M., T.Y., and A.I.A.; Methodology: A.B.M.; Software: N.; Validation: N., A.B.M., and A.I.A.; Formal Analysis: N., A.B.M., and A.I.A.; Investigation: N.; Resources: A.B.M.; Data Curation: A.B.M.; Writing Original Draft Preparation: N., A.B.M., and A.I.A.; Writing Review and Editing: A.B.M., N., and A.I.A.; Visualization: N.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. D. Astuti, M.J. Heriyanto, W. R. Desvita, Rokhmayanti, S. K. W. Hastuti, B. B. K. Utami and A. Azka, "Mental Health Screening for University Students in the Special Region of Yogyakarta," *Journal of Epidemiology and Public Health*, vol. 9, no. 3, pp. 343-353, 2024

[2] P. Jurovaty and S. Demuthova, "Draw a Person: Intelectual Ability Test for Chilldren, Adolescents, And Adults-the potential for Screening Diagnostic of Selected Cognitive Abilities," *Global Journal of Arts Humanity and Social Sciences*, vol. 3, no. 3, pp. 313-323, 2023.

[3] O. Rakhmanov and S. Dane, "Effect of the age and gender on the reliability of draw-a-person test," *Journal of Research in Medical and Dental Science*, vol. 8, no. 5, pp.151-158, 2020.

[4] K. laumanns, J.J. Eilers, and E. Kasten, "The Draw-a-Person Test in Neuropsychological Assessment," *Journal of Medical-Clinical*, vol. 8, no. 5, pp.1-13, 2024.

[5] A. Auliya and T.A. Hopeman, "Is it true that student's mental health is a problem? A Literature Study of the Causes of Mental Health Problems," *Journal of Belaindika*, vol. 6, no. 1, pp. 69-77, 2024.

[6] H. Susanti, H. Brooks, I. Yulia, H.D. Windarwati, E. Yuliastuti, Hasniah and B.A. Keliat, "An Exploration of the Indonesian lay mental health workers (cadres) experiences in performing their roles in community mental health services: a qualitative study," *Int J Ment Health Syst,* vol. 8, no. 3, pp. 1-13, 2024, doi: 10.1186/s13033-024-00622-0.

[7] P. Malhotra and E. Garg, "Object Detection Techniques: A Comparison," *2020 7th International Conference on Smart Structures and nSystems (ICCS),* Chennai, India, vol. 7, no. Sept., pp. 1-4, 2020, doi: 10.1109/ICSSS49621.2020.9202254.

[8] B. Beltzung, M. Pele, J.P. Renoult, and C. Sueur, "Deep Learning for studying drawing behavior: A Review," *Frontiers in Psychology,* vo. 14, no. 992541, pp. 1-13, 2023, doi: 10.3389/fpsyg.2023.992541.

[9] N. Dhariwal, N. Sengupta, M. Madiajagan, K.K. Patro, P.L. Kumari, N.A. Samee, R. Tadeusiewicz, P. Plawiak and A.J. Prakash, "A pilot study on AI-driven approaches for classification of mental health disorders," *Frontiers in Human Neuroscience*, vol. 18, no. 1376338, pp. 1-17, 2024, doi: 10.3389/fnhum.2024.1376338.

[10] M.A. Hameed and Z.A. Khalaf, "A survey study in Object Detection: A Comprehensive Analysis of Traditional and State-of-the-Art Approaches," *Journal of Basrah Researches (Sciences),* vol. 50, no. 1, pp. 46-60, 2024.

[11] M. Hussain and R. Khanam, "In-Depth Review of YOLOv1 to YOLOv10 Variants for Enhanced Photovoltaic Defect Detection," *Solar*, vol. 4, no. 3, pp. 351-386, 2024.

[12] P. Jiang, D. Ergu, F. Liu, Y. Cai and B. Ma, "A Review of YOLO Algorithm Developments," *Procedia Computer Science*, vol. 199, no. 1, pp. 1066-1073, 2022

[13] A. Alshammari and M.E. Alshammari, "Emotional Facial Expression Detection using YOLOv8," *Engineering, Technology and Applied Science Research (ETASR)*, vol. 14, no. 5, pp. 16619-16623, 2024

[14] V.P. Lysechko, B.I. Sadovnykov, O.M. Komar, and O.S. Zhuchenko, "A Research of The Latest Approaches to Visual Recognition and Classification," *Radio Electronics, Computer Science, Control,* vol.1, no. 1, pp. 140-146, 2024, doi: 10.15588/1607-3274-2024-1-13.

[15] O. Rakhmanov, N.A. Nnanna, and S.A. Adeshina, "Experimentation on Hand Drawn Sketches by Children to Classify Draw a Person Test in Phsychology," *The Thirty-Third International FLAIRS Conference (FLAIRS-33),* Miami, USA, May 2020, vol. 2020, no. May, pp. 329-334, 2020.

[16] S. Widiyanto and J.W. Abuhasan, "Implementation the Convolutional Neural Network Method for Classification the Draw-A-Person Test," 2020 Fifth International Conference on Informatics and Computing (ICIC), Gorontalo, Indonesia, November vol. 2020, no. 1, pp.1-6, 2020, doi: 10.1109/ICIC50835.2020.9288651.

[17] A. Alshahrani, M.M. Almatrafi, J.I. Mustafa, L.S. Albaqami, and R.A. Aljabri, "A Children's Psychological and Mental Health Detection Model by Drawing Analysis based on Computer Vision and Deep Learning," *Engineering, Technology and Applied Science Research,* vol.14, no. 4, pp. 15533-15540, 2024, doi: :10.48084/etasr.7812.

[18] J. Kang, J. Kim, M. Yang, C. Park, T. Kim, H. Song, and J. Han, "SceneDAPR: A Scene-Level Free-Hand Drawing Dataset for Web-based Psychological Drawing Assessment," *In Proceedings of the ACM Web Conference 2024 (WWW '24), Association for Computing Machinery, New York, USA,* May, vol. 2024, no. 1, pp. 4630 – 4641, 2024.

[19] M. Lee, Y. Kim, and Y. Kim, "Generating psychological analysis tables for children's drawings using deep learning," *Data and Knowledge Engineering,* vol. 149, no. 102266, pp.1-19, 2024

[20] American Psychiatric Association, *Diagnosistic and Statistical ManuaL of Mental Disorders*, 5th ed. Arlington, VA: American Psychiatric Association, 2013.

[21] G. Barbalat, D. Bergh and J.J. Kossakowski, "Outcome Measurement in mental health services: insight from symptom networks," *BMC Psychiatry,* vol. 19, no. 1, pp. 1-9, 2019, doi: 10.1186/s12888-019-2175-7.