Cureus Journal of Computer Science

Open Access Original Article

Assessment of Children's Dental Fear and Anxiety Using Deep Learning Techniques

Rojin Adabdokht [1] , Ida Kornerup [1] , Fabiana Almeida [1] , Hollis Lai [1]

1. Dentistry, Mike Petryk School of Dentistry, University of Alberta, Edmonton, CAN

Corresponding author: Hollis Lai, hollis1@ualberta.ca

Received 03/22/2025
Review began 03/25/2025
Review ended 05/30/2025
Published 06/07/2025

DOI: https://doi.org/10.7759/s44389-025-03812-0

Abstract

Background

About one-third of children face dental fear and anxiety, jeopardizing treatment outcomes. This study develops deep learning algorithms to automate drawing analysis, addressing the limitations of subjective and time-intensive manual assessments.

Methods

Two deep learning systems were developed. In the pilot study, drawings were manually labeled using the Child Drawing: Hospital score sheet. The first system used a two-step classification, focusing on human figures. A ResNet-18 model categorized drawings into high or low-average anxiety, while YOLO-v5 detected human figures, and YOLO-v8 classified those in low-average anxiety drawings. The second approach evaluated layout characteristics using pixel value calculations and scored color attributes with k-means clustering. A two-step deep learning process assessed human figure features: YOLO-v5 detected human figures and recognized faces, while YOLO-v8 classified faces as happy or unhappy. Final anxiety levels were determined by the combined feature scores. Accuracy and precision were reported for classifiers and object detectors, respectively.

Results

The first approach showed promising results, with the ResNet-18 classifier achieving a testing accuracy of 0.93 (95% CI: 0.878-0.982), the object detector achieving precision of 0.64 (95% CI: 0.531-0.749), and across five-fold cross-validation, the final classifier achieving a mean test accuracy of 0.66 (95% CI: 0.627, 0.6966). Confidence intervals were computed using the t-distribution with 4 degrees of freedom. Since the testing set was too small, a normal distribution assumption for the testing set was unrealistic. Therefore, a five-fold cross-validation was performed. The second approach achieved a final accuracy of 0.76, with individual features scoring a moderate accuracy of around 0.60.

Conclusion

This study demonstrates the potential of artificial intelligence to automate children’s dental fear and anxiety assessments through drawing analysis. Practical Implications: The developed algorithms can be used to screen children's dental anxiety, aiding clinicians in managing anxiety effectively, understanding patient concerns, and improving communication with children with linguistic or cognitive barriers.

Categories: AI applications, Algorithm Design, Image Processing and Analysis
Keywords: dental anxiety, pediatric dentistry, deep learning, drawing, machine learning, medical image analysis

Introduction

Dental anxiety is a prevalent incident among young children [1-3]. About one-third of young children experience Dental Fear and Anxiety (DFA) [4]. It can negatively impact children's emotions and dental behaviour, inflicting treatment avoidance and uncooperation during the dental visit, making treatment delivery difficult, if not impossible [5-8]. Dental professionals need to be aware of pediatric patients’ DFA to maximize effective delivery of care.

Self-reported assessment tools are the most widely used method for collecting patient input on DFA. Young children's linguistic and cognitive limitations render interviews inapplicable to pediatric dental patients [9-12]. Alternatively, drawing analysis provides a non-verbal and open-ended way for children to express emotions when verbal communication falls short [12-16]. Initially created for anxiety assessment through drawing analysis in hospitalized children [16], the Children Drawing: Hospital (CD: H) score sheet was adapted for dental research by Aminabadi et al. [3]. Since then, many studies have demonstrated a strong correlation between CD: H results and other anxiety assessment methods, including behavioural tools [12,13,17].

Although CD: H provides useful information on children's DFA, it has limited application for daily practice [13,14,17]. Drawing interpretation is a highly subjective process, heavily reliant on the individual interpreter [15]. This subjectivity introduces potential rater biases, affecting the accuracy of the analysis [12,13]. Additionally, CD: H includes 23 questions, making it challenging to collect immediate feedback and limiting its application in daily practice [12,13].

Automated systems can reduce the assessment's time and subjectivity [15,18-20]. Artificial intelligence (AI) algorithms showed the potential to recognize features different from manual evaluation when evaluating children's drawings, which can be useful for overall drawing assessment [19,21]. In their review of deep learning (DL) for studying drawing behaviour, Beltzung et al. [15] highlighted that neural networks, a subtype of DL, are the most promising tool for analyzing drawings.

DL helps reduce human biases, such as confirmation bias [20]. This can influence traditional methods that rely on preselected features aligned with the designer's beliefs. Instead, DL processes nearly unaltered data, minimizing such biases [19,20,22]. While DL models have been applied to various drawing analyses, no research, to the authors' knowledge, has yet automated the assessment of children's dental anxiety through their drawings. Therefore, this study aims to introduce automated systems for evaluating children's drawings, for DFA assessment.

Materials And Methods

This secondary data analysis study was approved by the University of Alberta's Board of Ethics with IRB number Pro00126348.

Datasets

Four datasets were used for model development and assessment (Table *1*). A) The primary dataset for this study consists of 63 drawings previously collected for DFA assessment [12]. B) Twenty-eight drawings, previously collected for DFA evaluation as a part of another prior study [17]. For the rest of this document, the combination of these two datasets, which included drawings from 5- to 10-year-old children pre- and post-operatively, will be referred to as the A&B dataset. In both of these studies, children were asked to draw their experience at the dentist using a controlled set of materials, including 8 colours of crayons and a white sheet. C) A collection of hand-sketched human figures by 4- to 11-year-old children, collected for the Draw-a-Person-Test (DAPT) to evaluate children's mental age [23]. This dataset originally included 952 hand-sketched human figures, from which, 200 were randomly selected and included in the current study. D) A total of 1,225 AI-generated drawings with a prompt base generation approach. These drawings were purposefully generated using OpenAI - DALL.E 2 - text-to-image generator. These drawings were generated to include human figures in a childish drawing style.

This approach was employed to address the challenges of obtaining sufficiently large datasets, particularly when dealing with human-created artwork for training data-driven algorithms [24]. The prompts were engineered to imply the generation of images that share common features with children's drawings. These features include human figures with six body parts (head, trunk, two hands, and two legs) and facial expressions (eyes and mouth). By the age of 5, children are able to draw such human figures [13]. Images were included regardless of whether they contained all six body parts. This decision was based on the understanding that children may omit certain body parts in their drawings. For instance, children who have experienced abuse or feel insecure may omit their hands in their drawings. Similarly, the omission of eyes may reflect a reluctance to engage with others [25,26].

Dataset	Image Type	Context	Number of Images	Reason for Selection
A	Clinical – Hand-drawn	Children's drawings collected post-operatively in dental settings	63	Collected according to CD:H protocols in a clinical dental context.
B	Clinical – Hand-drawn	Children's drawings collected both pre- and post-operatively	28	Collected according to CD:H protocols in a clinical dental context.
C	Clinical – Hand-drawn	Children's drawings for psychological evaluations	200	Used to train algorithms due to its depiction of human figures, which share global visual features with children's dental drawings.
D	Digital – AI-generated	Artificial images simulating children's drawings	1,225	Developed to supplement limited clinical data. Chosen due to data scarcity, high collection costs, and the assumption that they share key global features with clinical data.

TABLE 1: Detailed description of datasets

CD:H, Children Drawing: Hospital

For the training and evaluation of each DL algorithm in this study, a random split of the combination of all or subsets of the mentioned datasets was utilized (Tables 2 and 3).

	Dataset	Split				Augmentation Techniques	Architecture	Hyperparameters
			Training		Testing			
			BA	AA				
High-Low/average classifier	A&B dataset n=92 (low=28, average=34, high=30)	Low-Average	80% (n=50)	174	10% (n=6)	Vertical Flip, Horizontal Flip	ResNet-18	Batch size= 8, Learning rate=0.001, Epochs=20, Loss Function=BCE, Optimizer=Adam, Early Stopping=patience of 6
		High	80% (n=24)	168	10% (n=3)	Vertical Flip, Horizontal Flip, Clockwise Rotations 15 and 90, Counter-Clockwise Rotations 15 and 90		
Human figure detector	C (DAPT, n=200), D (AI generated, n=1225), Total= 1425	-	80% (n=1128)	3,381	75 images of A&B dataset	Vertical Flip, Horizontal Flip, Clockwise Rotation 15, 45, and 90, Counter-Clockwise Rotations 15, 45, and 90	YOLO-v5	Batch size=16, Learning rate=0.001, Epochs=20, Loss Function=GIOU and BCE, Optimizer=Adam Early, Stopping=No
Low-average classifier	Human figures present in the low-average group of the A&B dataset (n=63)	-	70% (n=44)	132	10% (n=6)	Horizontal Flip, Clockwise Rotations 15 and 90, Counter-Clockwise Rotations 15 and 90	YOLO-v8	Batch size=16, Learning rate=0.001, Epochs=30, Loss Function=BCE, Optimizer=Adam, Early Stopping=No

TABLE 2: An overview of the task-specific datasets, preprocessing steps, data augmentation techniques, and the hyperparameters used for each deep learning algorithm.

BA, Before Augmentation; AA, After Augmentation; BCE, Binary Cross Entropy; DAPT, Draw-a-Person-Test

	Dataset	Split				Preprocessing	Augmentation Techniques	Architecture	Hyperparameters
		Training		Validation	Testing				
		BA	AA						
Face Recognition	C (DAPT, n=108), D (AI-generated, n=800) Total = 908	70% (n=636)	1908	20% (n=162)	10% (91)	Resized to 640 × 640 gray-scaled	Horizontal Flip, Clockwise Rotations 15 and 90, Counter-Clockwise Rotations 15 and 90, 2.5 px blurring	YOLO-v5	Batch size=16, Learning rate=0.001, Epochs=20, Loss Function=BCE, Optimizer=Adam Early Stopping=No
Facial Emotion Classifier	A&B dataset (n=141), D (AI-generated, n=147), Total = 288 (happy: 150 unhappy: 138)	70% (n=202)	606	20% (n=58)	10% (n=28)	Resized to 640 × 640	Horizontal Flip, Clockwise Rotations 15 and 90, Counter-Clockwise Rotations 15 and 90	YOLO-v8	Batch size=16, Learning rate=0.001, Epochs=20, Loss Function=BCE, Optimizer=Adam, Early Stopping=No

TABLE 3: Dataset splits for face recognition and classification algorithms.

BA, Before Augmentation; AA, After Augmentation; BCE, Binary Cross Entropy; DAPT, Draw-a-Person-Test

Annotation: The A&B dataset was first manually labelled as low, average, and high levels of anxiety based on the CD: H score sheet. Region of interest (ROI), defined as the human figures, was annotated using the Roboflow annotation tool [27] by drawing the smallest BBox that completely enclosed each figure, regardless of whether the body was entirely or partially visible in the image drawn.

Two-step classification with noise reduction (Pilot)

Drawings were first binarily classified into low-average and high levels of anxiety using a pre-trained ResNet-18 [28] architecture. Then human figures presented in the former were detected using a YOLO-v5 [29] and classified as low and average by a YOLO-v8 [30] algorithm, pre-trained on ImageNet. This two-step classification was based on the information human figures in children's drawings provide (Figure *1*). The restriction or absence of human figures is an indication of fear and emotional distress [16,25,26]. Therefore, the drawings showing a high level of anxiety, containing restricted or no human figure, were identified first. It is noteworthy that while the algorithms used in the current study were mainly YOLO models, the first classifier was decided to be a ResNet because of the limited ability of YOLO to recognize small objects [31]. The performance of the second classifier was evaluated by five-fold cross-validation.

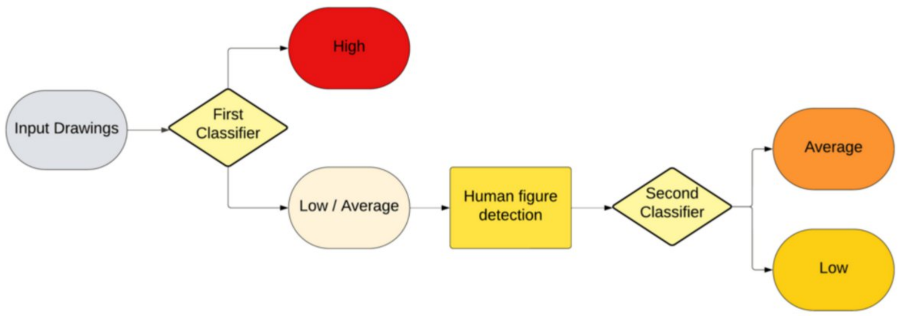

FIGURE 1: The two-step classification approach with noise reduction.

Drawings first classify binarily to low/average and high anxiety levels. The human figures from the low/average group were detected and further classified into low and average anxiety levels.

Feature extraction algorithm

Images are a high-dimensional data modality [32], which makes it challenging to identify the specific features used for classification and to comprehend the decision-making process of DL systems. The explainable AI can significantly improve clinical acceptance and adoption in daily practice [33]. The feature extraction algorithm, which is based on clinical deductive reasoning [33,34], was developed to increase interpretability and reduce the dependency of the algorithm's training on the number of clinical inputs. In this approach, individual features were extracted and scored, and the final decision on the DFA level was made based on the sum of scores for all individual features. These features were selected to evaluate layout characteristics, colour attributes, and human-figure-related characteristics.

In this system, first, the layout characteristics and colour specifics were measured. Then, the ROI (human figures) were detected similar to the pilot method. After scoring the size and length, the faces were recognized and classified based on their emotional expression. Figure *2* illustrates the workflow of this approach. The scoring system manual can be found in Appendix 1.

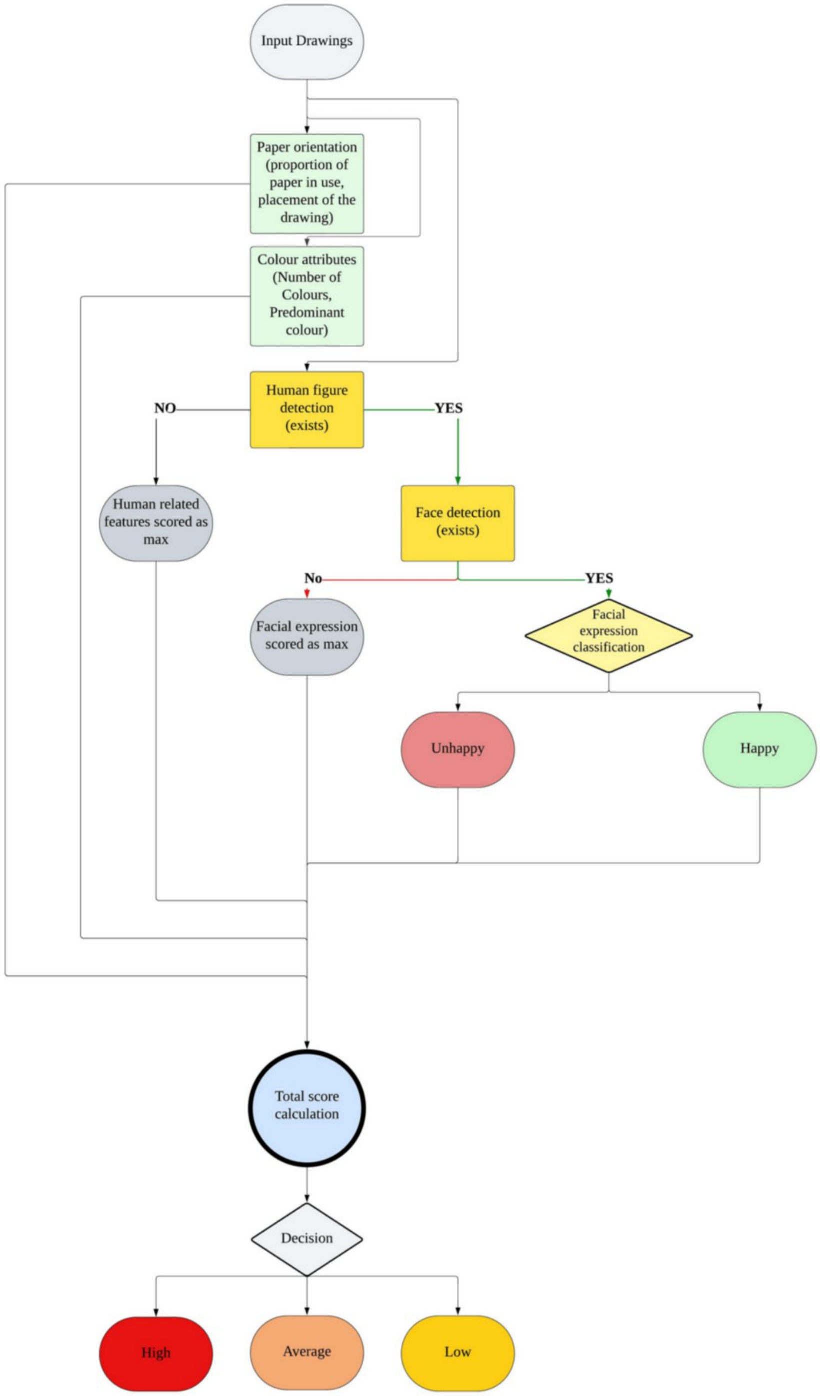

FIGURE 2: The workflow of the feature scoring system.

Layout Characteristics (proportion of paper in use, placement of the drawing on the paper): To recognize the parts of the image with drawing, the image was first converted to grayscale and denoised using a Gaussian blur filter. Then, the edges were detected using the Canny algorithm [35], resulting in a binary image outlining the contours of separate parts of the drawing. The bounding box (BBox) coordinates for each part were initially determined. Contours with a surface area smaller than 10,000 pixels were considered noise. Then, the minimum and maximum X and Y values were identified to create a BBox surrounding the drawing. Ultimately, the dimensions and coordinates of the final BBox were used for scoring layout features.

Colour Attributes (number of colours used, predominant colour): A k-means clustering algorithm was employed to group RGB pixel values into 10 prominent clusters based on their proximity (k = 10 was selected to account for the 8 possible colours, the white background, and an additional cluster for variability). The algorithm excluded white (background), if detected as the first prominent colour, and assigned the score based on the second detected colour.

Eight predefined categories - red, purple, blue, orange, green, yellow, brown, and black - were established, each associated with its closest RGB values. The number of colours was calculated by counting how many predefined colour categories appeared in the detected list.

Human Figure Detection (length of the human, size of the human compared to the environment): The same ROI detection algorithm as the pilot algorithm was used. The features were scored based on the detected BBox. If no human figures were detected, all related features would score 10.

Face Detection: The faces of the ROIs were annotated using BBox for training a face recognition algorithm (YOLO-v5). If no faces were detected, the facial emotion feature would score 10.

Facial Emotion Classification: A YOLO-v8 was trained to classify faces into happy and unhappy categories. Happy faces were defined as those displaying laughter, a smile, or a half-smile, while all other faces were classified as unhappy, including scared, confused, sad, and crying. If multiple faces were recognized, a score of 3 was assigned if happy faces were greater in number than unhappy faces, while a score of 8 was given if unhappy faces were greater in number. If the numbers were equal, a score of 5 was assigned (Appendix 1).

Table *3* presents the summary of the datasets, preprocessing steps, and hyperparameters for each DL algorithm.

If the sum of scores for all features was below 30, the drawing was classified as "low". Scores above 47 were classified as "high", while scores between 30 and 47 were categorized as "average".

Results

Two-step classification with noise reduction (Pilot)

Dataset Characteristics: The A&B dataset included 106 drawings, 63 from dataset A and 43 from B. All drawings from A were obtained post-operatively, while the B dataset included 24 post-operative and 19 pre-operative drawings. From these, 91 were included in the current study. The exclusion was based on the irrelevance of the drawings to the initial prompts provided at the time of data collection. After manual classification, 30 drawings showed a high, 33 showed an average, and 28 showed a low level of anxiety.

First Classifier: After augmentation, the training set increased to 216 for the low-average and 210 for the high group. The first classifier demonstrated 0.90 training accuracy and 0.93 testing accuracy.

ROI Detection: The ROI detector was trained on 1,425 images including at least one human figure. The dataset was split randomly into an 80% training (n = 1,128) and a 20% validation (n = 297). The detector achieved a precision of 0.89 and a recall of 0.94 when tested on an unseen subset of the original dataset.

For external validity, the algorithm was tested on the A&B dataset. This included 75 images with 137 instances of human figures. The algorithm demonstrated a precision of 0.64 and a recall of 0.65.

Second Classifier: The low-average class of the A&B dataset included 63 instances of human figures, with 35 average and 28 low anxiety images. The drawings were randomly divided into 70% training, 20% validation, and 10% testing sets. After augmentation, the number of images in the training set reached 132.

The algorithm achieved 0.67 training accuracy. Furthermore, across five-fold cross-validation, the model achieved a mean test accuracy of 0.66 (95% CI: 0.6278, 0.6966). Confidence intervals were computed using the t-distribution with 4 degrees of freedom. Since the testing set was too small, a normal distribution assumption for the testing set was unrealistic. Therefore, a five-fold cross-validation was performed. Notably, the training loss consistently declined from 0.74 in the first epoch to 0.17 by the final epoch. However, the validation loss remained relatively stable, fluctuating between 0.72 and 0.73 throughout the training process (Figure *3*).

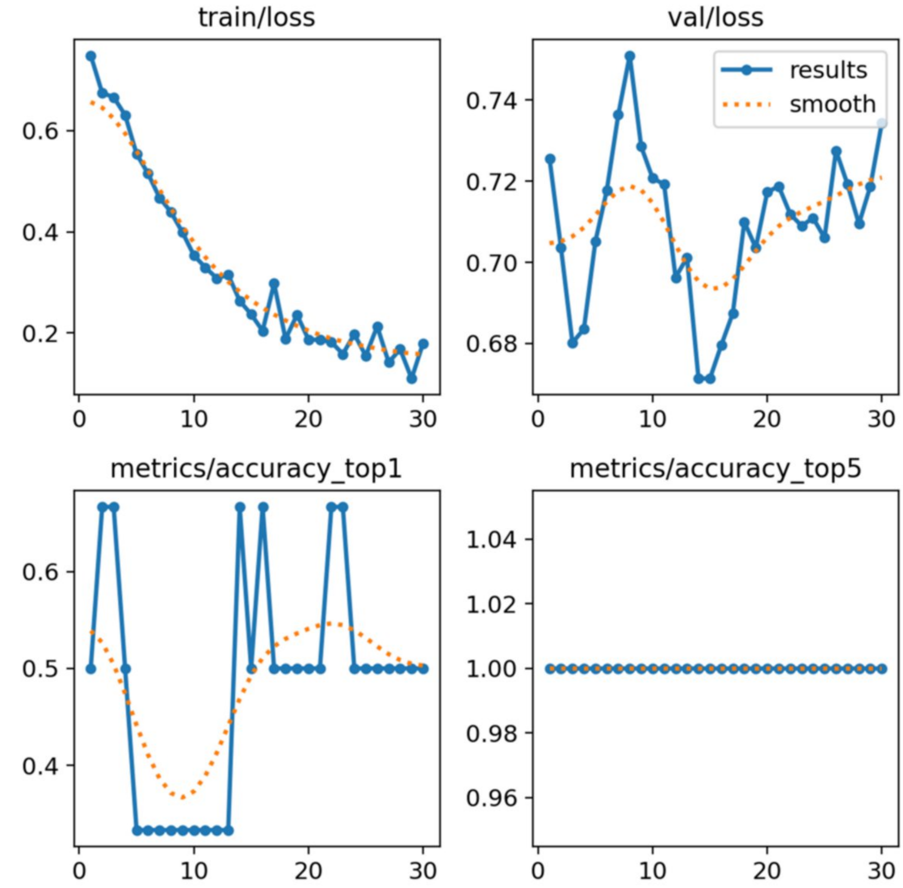

FIGURE 3: Training and validation loss during 20 epochs.

The fluctuation in the validation loss can indicate that the algorithm fails to capture the repetitive patterns and common features in the training set.

Feature extraction algorithm

Face Recognition: A total of 908 faces cropped from the C and D datasets were included. The dataset was split into 85% training (n = 776, n = 2327 after augmentation), 10% validation (n = 88), and 5% testing (n = 44). This algorithm was tested on 124 faces from dataset A. Testing on a test subset of the original dataset showed precision and recall of 0.87 and 0.90, respectively. When tested on dataset A, the metrics fell to the precision of 0.73 and recall of 0.58.

Facial Emotion Classification: A total of 288 faces from the A&B and C datasets were included. The dataset was split into 70% (n = 201, n = 605 after augmentation) training, 20% (n = 58) validation, and 10% (n = 29) testing. This algorithm showed 0.79 training and 0.67 testing accuracy. Although the overall accuracy of the facial emotion classification algorithm was 0.76, the model was successful in recognizing happy faces in 0.93 of the instances. However, what dropped the overall accuracy was the unhappy class, where the model failed to correctly classify the images 0.33 of the times.

Pipeline Evaluation: Overall, from 63 total drawings, 34 showed an average, 10 showed a high, and 19 showed a low level of anxiety. The algorithm suggested 44 drawings showing average, 7 showing high, and 12 showing low levels of anxiety. It demonstrated 0.76 accuracy in DFA classification. The F1-score for class “low” was 0.73, 0.80 for “average”, and 0.63 for class “high”. The Cohen Kappa test demonstrated the agreement between the rater, and the algorithm was 0.70 (Tables *4* and 5).

Item	Proportion of Paper in Use	Placement of the Drawing on the Paper	Colour Predominance	Number of Colours	Length of the Person	Size of the Person to the Environment	Facial Emotion	Total
Accuracy	0.64	0.66	0.65	0.72	0.61	0.58	0.57	0.76

TABLE 4: This table presents the accuracy of the feature scoring systems for each feature classification, along with the overall system performance.

		Manual Scoring						
		Low	Average	High	Total	Precision	Recall	F1-Score
Algorithm	Low	11	1	0	12	0.61	0.92	0.73
	Average	7	31	6	44	0.93	0.7	0.8
	High	0	1	6	7	0.5	0.86	0.63
	Total	18	33	12	63	0.68	0.82	0.72

TABLE 5: This table displays the confusion matrix, comparing the manual classification results with those obtained from the algorithm.

Discussion

The assessment of children's DFA through drawing analysis has been extensively studied [12-14,17]. However, interpreting drawings is a highly subjective process that heavily relies on the individual interpreter. While assessing low-level features might not seem too difficult, evaluating higher-level features, like assessing human figures, presents a fundamental challenge due to its dependence on the rater's judgment [15,21,36]. Furthermore, drawing collection and interpreter training are time-consuming processes. This can be a barrier to implementing this method in daily clinical practice [25,36]. The present study explored two approaches for automating the children's DFA assessment through drawing analysis, aiming to reduce the process’s subjectivity and time required for the rater training and drawing analysis.

The ROI detector algorithm demonstrated approximately 0.90 precision during training, which is aligned with the results of the previous studies [37]. However, when tested on the A&B dataset, the precision dropped to about 0.65. At first glance, this decline might imply an overfitted algorithm. However, this is unlikely, as the trained model achieved similar precision during testing on an unseen subset of the larger dataset from which the training set was derived. A more plausible explanation for the precision drop is that the model captured features during training that differed from those in the A&B dataset. To better understand the domain shift between AI-generated and children’s drawings of human figures, we conducted a low-level feature analysis. Feature maps were extracted from the early convolutional layers of a pre-trained ResNet-18 model, focusing on activations associated with each input channel. These early layers capture fundamental visual attributes such as edges, textures, and stroke intensity. We then quantified the distributional differences between the two datasets using the Kolmogorov-Smirnov (KS) test. The results revealed statistically significant discrepancies (KS-statistics > 0.66, $p < 0.001$), indicating that the model perceives markedly different low-level patterns in the AI-generated images compared to those drawn by children.

While both datasets depict human figures, they do so using distinct visual features. AI-generated drawings tend to be more consistent, geometrically regular, and stylized, whereas children’s drawings are often more abstract, expressive, and variable in execution. These differences suggest that the model trained solely on AI-generated data may struggle to generalize to hand-drawn examples. This low-level mismatch likely contributes to the observed performance drop.

Additionally, the inherent variability in artwork created by different individuals likely contributed to the increased variability in the testing set. The automatic detection of figures in abstract hand sketch art is complicated because of high variability, making it challenging for deep learning models to detect human figures in all styles of drawings [38-40]. The key point is that despite the declined precision, the results of the current study are comparable to and even higher than previous studies that attempted to detect objects in abstract artwork. Previous studies applied a pre-trained VGG19 deep learning algorithm to detect objects in digital fine arts. Although their model was successful in detecting objects like boats, showing an accuracy of 0.71 in testing, they could achieve a training accuracy of only 0.51 and a testing accuracy of 0.39 for detecting cows [24]. Similarly, in their study, Tran E. used residual deep convolutional neural networks to classify hand sketches. While features like noses and zebras, which have repetitive patterns of black and white, showed perfect accuracy, the model was less successful in categorizing pandas, achieving only 0.06 accuracy. This lower accuracy was due to the varied presentations of pandas in the hand sketches, which increased dataset variability and prevented the model from capturing the repetitive patterns effectively [39]. Overall, the pilot phase of the current study demonstrated the potential of DL algorithms to automate the classification of children's drawings based on their DFA levels. Similarly, Imran et al. investigated different approaches for artistic style classification. Their findings indicated that the accuracy of a two-phase classification approach outperformed single-step classification. In the two-phase method, patches of the artwork were extracted and used alongside the entire image to train and validate the classifiers. This approach achieved an accuracy of 0.78 in classifying artwork based on style [38].

The feature extraction method suggested varied levels of accuracy in scoring individual features. Despite the moderate accuracy of individual feature scoring, their combination demonstrated a classification F1-score of 0.72, which can be better understood through fuzzy logic. Fuzzy logic formalizes human reasoning and rational decision-making in contexts where information is imperfect. One of its key contributions is its powerful capability to precisely define what is inherently imprecise. By allowing variables to have a range of values and degrees of membership, rather than fixed binary states, fuzzy logic enables more flexible decision-making processes, closely mimicking human reasoning and decision-making. Particularly, in scenarios where distinctions like high or low anxiety levels lack clear boundaries. In this concept, various combinations of imprecise features can converge to a precise decision [41,42]. Similarly, in the feature extraction method, each imprecise feature contributes partially to the decision-making process. Different fuzzy rules, based on various combinations of features, determined the DFA through drawing analysis with high precision. The other noteworthy finding of the current study is the accuracy of the facial emotion classification algorithm. The confusion matrix indicated that the classification accuracy for happy faces was over 0.90, while unhappy faces were misclassified 0.33 times. This difference can be explained by the variability within each class. The happy class included faces with a smile or a half smile. On the other hand, the unhappy class included a wider variety of faces without a smile, including sad, frowning, confusion, straight mouth, and crying. The higher variability in the unhappy class makes it more complicated for the algorithm to learn from the data [38].

Limitations

The algorithm was tested on only 12 high-anxiety drawings, which could contribute to the low accuracy of this class. It is recommended that future studies assess and validate the feature extraction approach on a wider clinical dataset with less class imbalance and for generalizability.

Data augmentation through generative AI is a widely used technique to enhance the generalizability, adaptability, and accuracy of machine learning algorithms, particularly deep neural networks, when the real data is limited or imbalanced . However, there are several important limitations and potential biases to consider, particularly in the context of our study [43].

One key limitation is the nature of the AI-generated images themselves. While real children's drawings are typically scanned hand-drawn images saved as digital files, AI-generated samples are produced entirely through digital means. This difference can affect visual features such as sketch quality, texture, and stroke irregularities, characteristics that may hold value in understanding a child's emotional state.

Furthermore, prompt-based generative methods, while useful for producing data tailored to specific needs, can introduce bias. By defining specific prompts, we guide the generation process in a way that might overfit the data to certain expectations or aesthetics, potentially reducing the model's ability to generalize to naturally drawn samples.

Importantly, the AI-generated drawings do not inherently capture the emotional or psychological context associated with dental anxiety. These synthetic images lack the spontaneous, expressive qualities that are often present in real children's drawings created under emotional stress. As a result, it was essential to limit the use of AI-generated data to feature-level training only. These synthetic drawings were not used for emotion-related labeling or decision-making to avoid misleading conclusions about a child's psychological state.

The drawings were labelled by only one annotator; to set a more reliable reference standard, the inter-annotator study is recommended.

Conclusions

Overall, this study showed the potential of DL algorithms in automating children's DFA assessment through drawing analysis. This method has the potential to be used for dental anxiety screening in children and, when combined with behavioral assessments, can offer a comprehensive evaluation of both internal and external aspects of a child's anxiety. For clinical implementation, the next steps involve validating the tool across diverse pediatric populations in real-world dental settings through clinical trials and qualitative studies to ensure generalizability and adaptability to the dental workflows.

Appendices

Score	1	2	3	4	5	6	7	8	9	10
Proportion	> 3/4		1/2 - 3/4			1/4 - 1/2		1/8 - 1/4		< 1/8
Placement of the drawing on the paper	Proportion = 1 OR Center of drawing within the central	Drawing covers one-third of paper (horizontal/vertical)	Drawing occupies right half of paper	Drawing occupies left half of paper	Drawing occupies lower half of paper	Drawing occupies upper half of paper	Drawing in bottom right or central right third	Drawing in lower left corner or left middle third	Drawing in upper right corner	Drawing in upper left corner
Predominant Colour			yellow or green		orange, purple and blue			orange, purple and blue		
Number of colours	7 or 8		5 or 6 colours		3 or 4 colours			one or two colors		
Length of the person	> 7/10		5/10 - 7/10	3/10 - 5/10		1/10 - 3/10		< 1/10		No human figure
Size of the person to the environment	> 7/10		5/10 - 7/10	3/10 - 5/10		1/10 - 3/10		< 1/10		No human figure
Facial emotion classification			Smiling or laughing		Equal number of happy and unhappy faces			Any expression but smiling or laughing		No face

FIGURE 4: Manual for algorithm scoring.

The green area represents low levels of anxiety, the orange area indicates average levels, and the red area corresponds to high levels of anxiety.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Rojin Adabdokht, Hollis Lai, Ida Kornerup

Acquisition, analysis, or interpretation of data: Rojin Adabdokht, Hollis Lai, Fabiana Almeida

Drafting of the manuscript: Rojin Adabdokht

Critical review of the manuscript for important intellectual content: Rojin Adabdokht, Hollis Lai, Ida Kornerup, Fabiana Almeida

Supervision: Hollis Lai, Ida Kornerup, Fabiana Almeida

Disclosures

Human subjects: Consent was obtained or waived by all participants in this study. University of Alberta, Health Research Ethics Board issued approval Pro00126348. This study was approved by the University of Alberta Health Research Ethics Board (HREB), under protocol number Pro00126348. All research procedures were conducted in accordance with the ethical guidelines and regulations set forth by the HREB. . **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** Rojin Adabdokht declare(s) a grant from University of Alberta, Mike Petryk School of Dentistry. This project was funded by University of Alberta through Oral Health Community Engagement Fund (OHCEF). **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

Data are available on reasonable request.

References

1. Locker D, Liddell A, Dempster L, Shapiro D: Age of onset of dental anxiety. Journal of Dental Research. 1999, 78:790-796. 10.1177/00220345990780031201
2. Appukuttan DP: Strategies to manage patients with dental anxiety and dental phobia: Literature review. Clinical Cosmetic and Investigational Dentistry. 2016, 8:35-50. 10.2147/ccide.s63626
3. Aminabadi NA, Ghoreishizadeh A, Ghoreishizadeh M, Oskouei SG: Can drawing be considered a projective measure for children's distress in paediatric dentistry?. International Journal of Paediatric Dentistry. 2011, 21:1-12. 10.1111/j.1365-263x.2010.01072.x
4. Sun IG, Chu CH, Lo ECM, Duangthip D: Global prevalence of early childhood dental fear and anxiety: A systematic review and meta-analysis. Journal of Dentistry. 2024, 142:104841. 10.1016/j.jdent.2024.104841
5. Dahlander A, Soares F, Grindefjord M, Dahllöf G: Factors associated with dental fear and anxiety in children aged 7 to 9 years. Dentistry Journal. 2019, 7:68. 10.3390/dj7030068
6. Carrillo-Díaz M, Migueláñez-Medrán BC, Nieto-Moraleda C, Romero-Maroto M, González-Olmo MJ: How can we reduce dental fear in children? The importance of the first dental visit. Children. 2021, 8:1167. 10.3390/children8121167
7. Chi SI: What is the gold standard of the dental anxiety scale?. Journal of Dental Anesthesia and Pain Medicine. 2023, 23:193-212. 10.17245/jdapm.2023.23.4.193
8. Chiu A, Falk A, Walkup JT: Anxiety disorders among children and adolescents. FOCUS. 2016, 14:26-33. 10.1176/appi.focus.20150029
9. Ali DA: Patient satisfaction in dental healthcare centers. European Journal of Dentistry. 2016, 10:309-314. 10.4103/1305-7456.184147
10. Yon MJY, Chen KJ, Gao SS, Duangthip D, Lo ECM, Chu CH: An introduction to assessing dental fear and anxiety in children. Healthcare. 2020, 8:86. 10.3390/healthcare8020086
11. Eli I, Uziel N, Baht R, Kleinhauz M: Antecedents of dental anxiety: learned responses versus personality traits. Community Dentistry and Oral Epidemiology. 1997, 25:233-237. 10.1111/j.1600-0528.1997.tb00932.x
12. Ortiz S, Yoon M, Gibson M, Kornerup I, Zeinabadi MS, Lai H: Children's anxiety levels and their perspectives on dental experiences in students' clinical evaluation. International Journal of Clinical Pediatric Dentistry. 2023, 16:S206-S212. 10.5005/jp-journals-10005-2620
13. Torriani DD, Goettems ML, Cademartori MG, Fernandez RR, Bussoletti DM: Representation of dental care and oral health in children's drawings. British Dental Journal. 2014, 216:E26. 10.1038/sj.bdj.2014.545
14. Onur SG, Altin KT, Yurtseven BD, Haznedaroglu E, Sandalli N: Children's drawing as a measurement of dental anxiety in paediatric dentistry. International Journal of Paediatric Dentistry. 2020, 30:666-675. 10.1111/ipd.12657
15. Beltzung B, Pelé M, Renoult JP, Sueur C: Deep learning for studying drawing behavior: A review. Frontiers in Psychology. 2023, 14:992541. 10.3389/fpsyg.2023.992541
16. Clatworthy S, Simon K, Tiedeman M: Child drawing: Hospital manual. Journal of Pediatric Nursing. 1999, 14:10-18. 10.1016/s0882-5963(99)80055-4
17. Baghdadi ZD, Jbara S, Muhajarine N: Children's drawing as a projective measure to understand their experiences of dental treatment under general anesthesia. Children. 2020, 7:73. 10.3390/children7070073
18. Ossowska A, Kusiak A, Świetlik D: Artificial intelligence in dentistry—Narrative review. International Journal of Environmental Research and Public Health. 2022, 19:3449. 10.3390/ijerph19063449
19. Yamashita R, Nishio M, Do RKG, Togashi K: Convolutional neural networks: An overview and application in radiology. Insights into Imaging. 2018, 9:611-629. 10.1007/s13244-018-0639-9
20. Fan W, Zhang J, Wang N, Li J, Hu L: The application of deep learning on CBCT in dentistry. Diagnostics. 2023, 13:2056. 10.3390/diagnostics13122056
21. Jensen CA, Sumanthiran D, Kirkorian HL, Travers BG, Rosengren KS, Rogers TT: Human perception and machine vision reveal rich latent structure in human figure drawings. Frontiers in Psychology. 2023, 14:1029808. 10.3389/fpsyg.2023.1029808
22. Gili T, Di Carlo G, Capuani S, Auconi P, Caldarelli G, Polimeni A: Complexity and data mining in dental research: A network medicine perspective on interceptive orthodontics. Orthodontics and Craniofacial Research. 2021, 24:16-25. 10.1111/ocr.12520
23. Rakhmanov O, Nnanna N, Adeshina S: Experimentation on hand-drawn sketches by children to classify Draw-a-Person test images in psychology. The Thirty-Third International FLAIRS Conference (FLAIRS-33). 2020, 329-334.
24. Smirnov S, Eguizabal A: Deep learning for object detection in fine-art paintings. Metrology for Archaeology and Cultural Heritage (MetroArchaeo), Cassino, Italy. 2018, 45-49. 10.1109/MetroArchaeo43810.2018.9089828
25. Farokhi M, Hashemi M: The analysis of children's drawings: Social, emotional, physical, and psychological aspects. Procedia - Social and Behavioral Sciences. 2011, 30:2219-2224. 10.1016/j.sbspro.2011.10.433
26. Di Leo JH: Children's Drawings as Diagnostic Aids. Routledge, Taylor & Francis Group, London; 1973.
27. Everything you need to build and deploy computer vision applications. (2022). Accessed: March 2023: https://roboflow.com/.
28. Zhang X, Wang Y, Zhang N, Xu D, Chen B: Research on scene classification method of high-resolution remote sensing images based on RFPNet. Applied Sciences. 2019, 9:2028. 10.3390/app9102028
29. Jocher G, Chaurasia A, Stoken A, et al.: ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation (v7.0). Zenodo. (2022). https://doi.org/10.5281/zenodo.7347926.
30. Jocher G, Chaurasia A, Qiu J. (2023). https://docs.ultralytics.com/models/yolov8/.
31. Diwan T, Anirudh G, Tembhurne JV: Object detection using YOLO: Challenges, architectural successors, datasets and applications. Multimedia Tools and Applications. 2023, 82:9243-9275. 10.1007/s11042-022-13644-y
32. Selmy HA, Mohamed HK, Medhat W: Big data analytics deep learning techniques and applications: A survey. Information Systems. 2024, 120:102318-102318. 10.1016/j.is.2023.102318
33. Xu Q, Xie W, Liao B, et al.: Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: A systematic review. Journal of Healthcare Engineering. 2023, 9919269:13. 10.1155/2023/9919269

34. Shin HS: Reasoning processes in clinical reasoning: From the perspective of cognitive psychology. Korean Journal of Medical Education. 2019, 31:299-308. 10.3946/kjme.2019.140
35. Canny J: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1986, 8:679-698,. 10.1109/TPAMI.1986.4767851
36. Tanaka JW, Taylor M: Object categories and expertise: Is the basic level in the eye of the beholder?. Cognitive Psychology. 1991, 23:457-482. 10.1016/0010-0285(91)90016-h
37. Sudha S, Jayanthi KB, Rajasekaran C, Sunder T: Segmentation of RoI in medical images using CNN - A comparative study. TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON). 2019, 767-771. 10.1109/TENCON.2019.8929648
38. Imran S, Naqvi RA, Sajid M, Malik TS, Ullah S, Moqurrab SA, Yon DK: Artistic style recognition: Combining deep and shallow neural networks for painting classification. Mathematics. 2023, 11:4564. 10.3390/math11224564
39. Lu W, Tran E: Free-hand sketch recognition classification. 2017, 1-7.
40. Westlake N, Cai H, Hall P: Detecting people in artwork with CNNs. Computer Vision - ECCV 2016 Workshops. Hua G, Jégou H (ed): Springer, Cham; 2016. 9913:825-841. 10.1007/978-3-319-46604-0_57
41. Mago VK, Mago A, Sharma P, Mago J: Fuzzy logic based expert system for the treatment of mobile tooth. Software Tools and Algorithms for Biological Systems. Arabnia H, Tran QN (ed): Springer, New York, NY; 2011. 696:607-614. 10.1007/978-1-4419-7046-6_62
42. Zadeh LA: Is there a need for fuzzy logic?. Information Sciences. 2008, 178:2751-2779. 10.1016/j.ins.2008.02.012
43. Goyal M, Mahmoud QH: A systematic review of synthetic data generation techniques using generative AI. Electronics. 2024, 13:3509. 10.3390/electronics13173509