Artificial Intelligence Based Psychological Detection of HTP Drawings

MSc Research Project
Artificial Intelligence

Munevver Irem Hatipoglu
Student ID: x23154861

School of Computing
National College of Ireland

Supervisor: Kislay Raj

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Munevver Irem Hatipoglu
Student ID:	x23154861
Programme:	Artificial Intelligence
Year:	2024
Module:	MSc Research Project
Supervisor:	Kislay Raj
Submission Due Date:	12/08/2024
Project Title:	Artificial Intelligence Based Psychological Detection of HTP Drawings
Word Count:	6727
Page Count:	23

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	M. Irem Hatipoglu
Date:	9th August 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

National College of Ireland

AI Acknowledgement Supplement

MSCAI Thesis

Student Name	Munevver Irem Hatipoglu
Course	Artificial Intelligence
Date	9th August 2024

This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click here.

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool
Quillbot	A paraphrasing and summarizing tool.	Clickable link to the Tool

Description of AI Usage

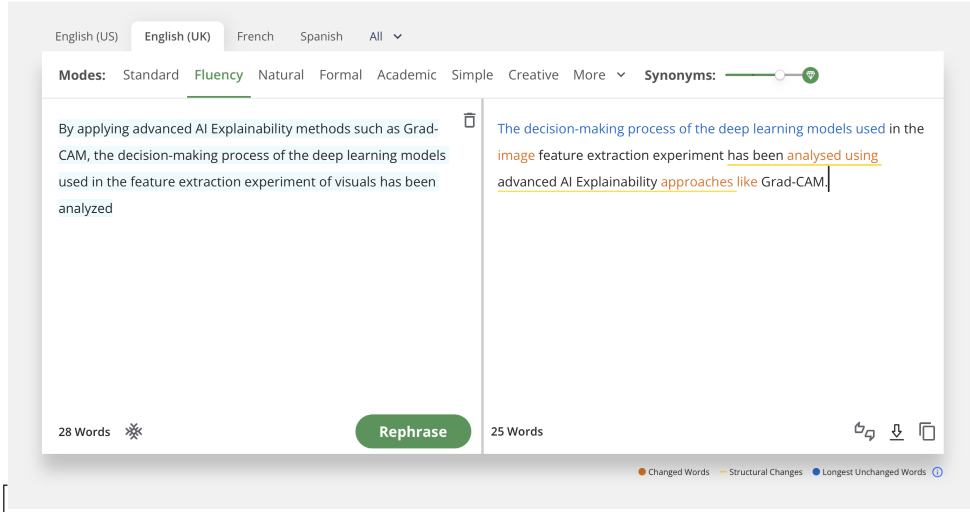
This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. One table should be used for each tool used.

Quillbot AI	
Description of use	Since my native language is different, sometimes I write sentences in that language first and then translate them to English. This tool assisted in the revision of phrases in order to satisfy required formality of the research document.
Sample prompt	At the end of the drawing step, questions were asked to the children about their drawings.
Sample response	After completing the drawings, the children were asked questions about their artwork.

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Additional Evidence for Quillbot:



Artificial Intelligence Based Psychological Detection of HTP Drawings

Munevver Irem Hatipoglu
x23154861

Abstract

Art therapy, unlike traditional psychotherapy approaches, is an approach that allows therapists to gain insight into the unconscious world of their patients by providing nonverbal communication. The House-Tree-Person (HTP) test, a drawing-based art therapy approach, is a significant technique used by experts, particularly on children, to provide analysis of familial bonds, interpersonal interactions, and children's self-perceptions. This test, in which professionals analyse psychological responses by analysing different elements in drawings and rating children's responses to specific questions about these drawings, is likely to suffer from mistakes. Developments in the artificial intelligence show promise in eliminating these errors and automating the process to ensure that children in need of psychological assistance reach the necessary treatment as quickly as possible. This paper thoroughly covers the fundamental principles of our study, which include using children's HTP drawings to detect mental disorders such as depression and anxiety, using advanced data augmentation methods, evaluating fine-tuned deep learning models for feature extraction from images, and training machine learning models with these extracted features to assess psychological scores. The results of the feature extraction experiment have shown that the fine-tuned VGG16 model has the highest accuracy with 97%. Using this model, the image features were extracted in terms of depression and anxiety. While the SVM model showed the best performance with a 79% accuracy performance in the depression detection experiment, Random Forest with a 68% accuracy in the anxiety detection experiment was the best performing model among others.

Index Terms- Art therapy, House-Tree-Person test, Machine learning, Transfer Learning, Deep Learning, Data Augmentation, Explainable AI

1 Introduction

One of the hardest concepts in the world to understand is human psychology, which is shaped by a variety of factors and reflected in the different ways that people behave. People may suffer from a variety of psychological disorders that might have crucial impact on their daily lives, such as anxiety, depression, post-traumatic stress disorder, obsessive-compulsive disorder, and post-traumatic stress disorder, which can be caused by factors such as childhood, genes, or traumatic experiences. Art therapy is one of the most important methods for analysing these psychological disorders (Du et al.; 2024). As Case and Dalley (2014) highlights, unlike traditional psychotherapy techniques, which depend on conversation between the therapist and the patient, art therapy focuses on artwork as the main way of communication. This therapy technique plays a very important role in understanding the inner world of the patients, since people often unconsciously express their thoughts and feelings through the art they produce (Malchiodi; 2020). As a result, art therapy is often used to analyse the mental health of children with limited language skills or individuals who refuse to communicate because of traumatic events they witnessed (Lee et al.; 2024). Drawing, clay sculpting, and painting, are a few of the many artistic approaches that are used for assessing and treating psychological issues in children, according to research by Eaton et al. (2007). Among these, drawing is the most frequently employed and and produces the most promising outcomes. According to Lee et al. (2024), several fundamental drawing tests are used to analyse children's psychological features, including the Bender Gestalt Test (BGT), the Draw a Person test (DAP), the Person Picking Apples from the Tree (PPAT), and the House-Tree-Person (HTP) test. The Bender Gestalt Test (BGT) offers details about cognitive problems and disabilities related to development, while the Draw a Person (DAP) test focuses on predicting a child's a sense of self. Besides all these, the House-Tree-Person (HTP) test which developed by Buck (1948), is more popular and representative than the other tests since it provides a thorough examination of family dynamics, personality traits, and interpersonal relationships, making it an effective tool for identifying the source of psychological problems.

Although the HTP test, which requires patients to draw a house, tree, and person while answering questions provided by a therapist who assesses the responses, is often used due to its ease of application and promise for diagnosing and treating psychological conditions, it has numerous limitations (Lee et al.; 2024). As Kim et al. (2021) highlights in their research, although this test is easy to perform, it is time-consuming and its results may vary depending on the experience and subjectivity of the practising therapist. In other words, this approach, based solely on the therapist's manual assessment, is quite open to errors. These limitations point out the necessity for advances that improve the accuracy and objectivity of psychological assessments. As technology develops, artificial intelligence techniques have become an essential part of our daily lives, offering potential solutions for automating many areas such as health, education, and trade while minimising errors. Given this situation, training and testing artificial intelligence models with HTP drawings offers an important chance for enhancing the accuracy of mental health assessments by overcoming challenges such as human subjectivity, limited material resources, and long evaluation periods.

Given the restrictions of traditional techniques to analyse the HTP test drawings, as well as the potential benefits of artificial intelligence in addressing these issues, the purpose of this research is to explore how machine learning algorithms may be implemented to analyse House-Tree-Person (HTP) test drawings and detect children's psychological diseases such as depression and anxiety. In this research, we used several pre-trained transfer learning networks to detect psychological diseases in the HTP test sketches, offering the following contributions:

- To avoid mistakes that may occur during the automatic classification of HTP images drawn by children with limited motor skills, enhanced labelling and annotation tools have been used.

- The unbalanced distribution in the dataset has been addressed by using an advanced data augmentation approach.

- To efficiently classify HTP drawings and extract features from images in the dataset, this research compared the performance of multiple pre-trained and fine-tuned CNN models to select the best performing model.
- Advanced Explainable AI techniques utilised, including Grad-CAM, to analyse the decision-making processes of deep learning models in feature extraction experiments.
- The performance of multiple machine learning models trained with the extracted features was compared in detecting depression, and anxiety scores of HTP images which assessed by psychologists.

The structure of the remaining sections of this research paper are as follows: Section 2 discusses the related works to detect psychological states of children by using House-Tree-Person (HTP) drawings to train AI models. Section 3 presents the proposed method for detecting the anxiety and depression in children's house, tree and person sketches, while the Section 4 explains the design specifications. The models' implementation is addressed in Section 5, before proceeding to the results and evaluation in Section 6, and the conclusion in Section 7.

2 Related Work

This section provides a detailed explanation of the HTP test, points out the possible advantages of using artificial intelligence models, and discusses previous studies on using these models to analyse the HTP test.

2.1 The House-Tree-Person (HTP) Test

The House-Tree-Person (HTP) test is a popular psychological assessment tool that was created by Buck (1948). This test is often administered to children because they may have limited communication skills, making it challenging for adults to fully understand their expressions. The HTP test is intended to evaluate characteristics of a person's inner world, family relations, and cognitive development by analysing three basic drawings: a house, a tree, and a person (Lee et al.; 2024). Various elements in these drawings have different psychological meanings (Buck; 1948). For instance, house full of windows often shows that the child is social and outgoing. On the other hand, a house with no or very few and small windows could mean that the child is introverted and close to interaction. This art therapy method, typically applied to children, consists of two main parts (Buck; 1948). In the first part, therapists ask their patients to draw a house, a tree, and a person. Once the drawings are complete, the second stage begins, where therapists conduct a detailed analysis through a series of structured questions outlined by Buck (1948). These questions help to gain deeper understanding into the child's ideas and emotions. Based on the child's responses, the therapist gives scores and analyses the drawings, taking into consideration all of the collected data.

2.2 AI Algorithms for HTP Test Analysis

Although the HTP technique provides valuable insights into patients' inner worlds, the assessment of these drawings is largely subjective and relies on the personal analysis of therapists, as highlighted by Kim et al. (2021). Considering the latest developments in AI models and their promising outcomes in the early diagnosis of many different diseases, using artificial intelligence to automate the analysis of HTP drawings could be extremely helpful in the early detection of children's psychological issues. This technique could remove errors due by external factors or therapist biases, resulting in more accurate and timely findings. It allows for the early detection of mental health conditions such as anxiety and depression, ensuring that children receive the treatment they require as soon as possible.

While the use of artificial intelligence models for the early diagnosis of plenty of illnesses has been established for a long time, using AI models to HTP test analysis was only recently proposed by Kim et al. (2021). In this study, which consists of three main parts such as Object Classification, Psychological Feature Detection, and Caption Generation, only house and tree drawings were used due to the dataset limitations. Kim et al. (2021) aimed to automate the processes of image classification in the HTP drawings, with a focus on detecting objects that represent unique psychological characteristics outlined by Buck (1948). In this research, which emphasises the potential benefits of using deep learning and transfer learning approaches with HTP test drawings, 70% accuracy performance was obtained using the ResNet50 model.

Similarly, Pan et al. (2022) have proposed an artificial intelligence approach consisting of two main parts, feature extraction and image classification, in order to automatically analyse the HTP test drawings. They used a special data set consisting of more than 3000 HTP images obtained from the psychology department of the College of Control Science and Engineering China University of Petroleum to identify holistic features such as size, position and shadow in the drawings. When they compared the performances of the two-layer CNN, VGG16, ResNet50, Vision Transformer (VIT), normal SVM model and their own proposed SVM-FEA model with special parameters, they found that the proposed model had the highest accuracy rate with 93% (Pan et al.; 2022).

Salar et al. (2023), similar to Pan et al. (2022), have taken two major methods to their research: image classification and psychological feature extraction. Salar et al. (2023), used HTP test images provided by the Faculty of Psychology at Istanbul Bilgi University in their research. Similarly, they employed computational neural network (CNN) and machine learning techniques in their studies just like Kim et al. (2021). In this project, which uses the ResNet152 model, various performance problems caused by the limited data set were tried to be prevented by increasing the data with images obtained from the QuickDraw[1] website dataset. The study, which focuses on psychological meanings provided through features such as pen pressure and object size, achieved an image classification performance of 67%.

Another research that emphasises the importance of the object detection approach is the Lee et al. (2024). This study, which uses a dataset gathered from the Goyang City Datathon[2] consisting of children's drawings of home, tree, man, and woman, applies the convolutional neural network (CNN) technique, as did previous studies. The study examined the performance of the YOLOv5, EfficientNet, SSD, and Faster R-CNN models, and showed that the Faster R-CNN model had the greatest accuracy in assessing psychological features of home, tree, and human drawings, with an accuracy rate ranging from 92% to 95%.

Table 1 shows a chronological analysis of projects that use House-Tree-Person (HTP)

[1]QuickDraw!: `https://quickdraw.withgoogle.com/data`

[2]Goyang City Datathon: `http://datathon.smilework.kr/`

test drawings for detecting mental health disorders in children. This table highlights the use of the different artificial intelligence approaches described in depth in this section. While the projects analysed showed considerable progress in integrating HTP test images with AI models, there are still significant needs in this field. For example, using only house, tree, and person drawings in these projects could limit the results' use in real life. Making decisions in daily life according to the results of the systems using data collected without being examined by psychologists may affect the access of many children to the necessary treatment.

To sum up, analysing the House-Tree-Person (HTP) drawing test, which is widely used by therapists for assessing children's psychological problems, requires artificial intelligence approaches to address the subjectivity of therapists and minimise errors caused by environmental factors. Although the HTP test was first proposed in the 1950s (Buck; 1948), there has been limited research in combination of AI models and the HTP test analysis (Kim et al.; 2021). A comprehensive review of related work reveals the need for a specialised dataset including HTP test drawings, which should be developed in collaboration with qualified therapists to ensure effective analysis using AI models. For example, Kim et al. (2021) and Lee et al. (2024) experienced performance challenges as a result of using datasets that were either unverified for psychological relevance or not reviewed by professionals, whilst Pan et al. (2022) and Salar et al. (2023) suffered from limitations by small dataset sizes despite being edited by experts. The literature review also emphasises the importance of improving AI algorithms' performance in order to provide prompt and accurate therapy for psychiatric diseases.

Author	Dataset Source	Model	Accuracy
Kim et al. (2021)	Google Image Data	CNN (ResNet50)	70%
Pan et al. (2022)	College of Control Science and Engineering China University of Petroleum	SVM-FEA	93%
Salar et al. (2023)	Istanbul Bilgi University & QuickDraw	CNN (ResNet152)	67%
Lee et al. (2024)	Goyang City Datathon	CNN (YOLOv5, Efficient-Det, SSD, Faster R-CNN)	House (Faster R-CNN) 95%, Tree (Faster R-CNN)92%, Person (Faster R-CNN) 95%

Table 1: Chronological Ordered Comparative Analysis of the Psychological Disorder Detection by The HTP Test Drawings

Furthermore, careful selection of model parameters is critical for optimising efficient algorithm. On the other hand, as with any artificial intelligence project that uses datasets created with data gathered from people, the significance of the data collecting procedure, storage of data, and model architecture in terms of ethical considerations is extremely important. When creating data sets, the necessary permissions must be obtained from people or parents of underage individuals and the data should not be shared with others. Addressing all of the problems discussed in the ethical way will significantly advance the automated interpretation of the HTP test with artificial intelligence models and help many children suffering from psychological problems to access the necessary treatment as soon as possible.

3 Methodology

This part addresses dataset definition, pre-processing, data augmentation, feature extraction and deep learning models. Figure 1 illustrates the proposed flowchart for the AI-based psychological detection by using HTP test drawings. As can be understood from the figure, this project consists of three main parts: Image Classification, Feature Extraction and Psychological Detection.

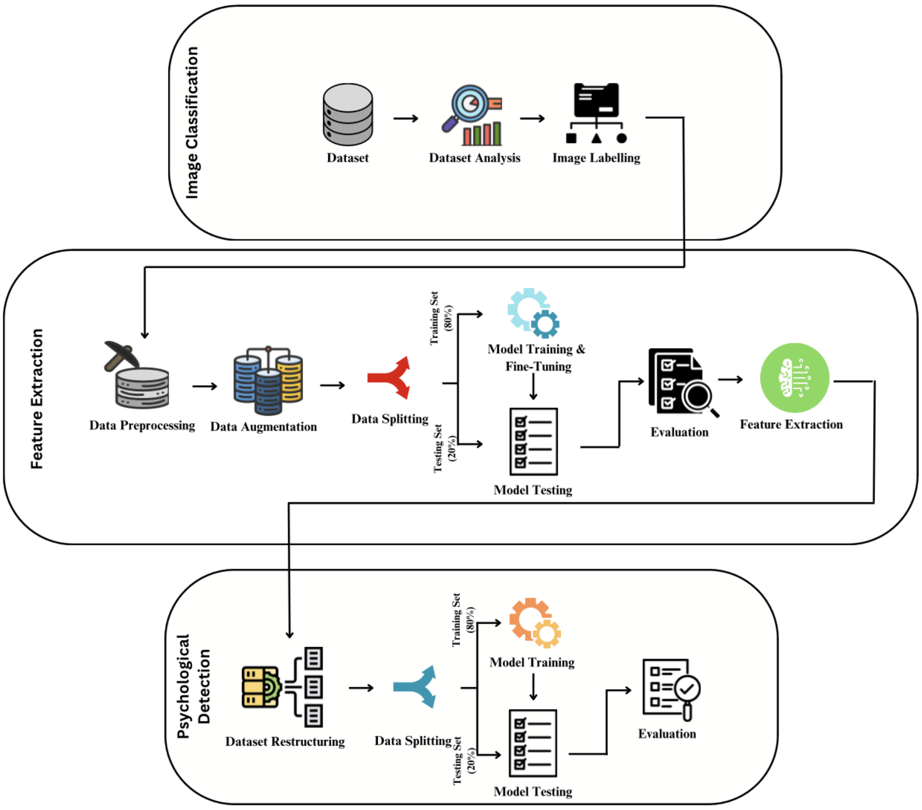


Figure 1: Proposed Project Structure

3.1 Dataset Analysis & Preprocessing

Based on the findings of Section 2's literature study, it is clear that the dataset selected for this project which aims to detect the psychological states of HTP test drawings using artificial intelligence models, should be created under the supervision of experts. It should include drawings from children suffering from psychological disorders, such as depression and anxiety, to ensure the dataset's reliability and relevance. As a consequence, the "HTP Images" dataset, created by psychologists from Istanbul Bilgi University's Faculty of Psychology, was chosen for this study, along with the appropriate ethical consents (Salar et al.; 2023). This data set was collected by obtaining the necessary permissions from the parents of the child clients of the psychologists of the Psychology Department. At the first stage, children were asked to draw houses, trees and people with black pencils on blank papers. At the end of the drawing step, questions were asked to the children about their drawings. Psychologists then evaluated the answers and assigned depression and anxiety scores based on the children's responses, using the Child Behaviour Checklist[3] as a reference. Figure 2 shows samples of house, tree, and person drawings from the HTP images dataset, which consists of drawings from 83 child patients.

In order to keep up with patient privacy regulations, the house, tree, and person drawings in the figure were randomly selected from the dataset and arranged in such a

[3]Child Behaviour Checklist: https://en.wikipedia.org/wiki/Child_Behavior_Checklist

Figure 2: HTP Images Dataset Samples

way that they are free of any details that could reveal the illustrations' owners.

For the purpose of easier data analysis, pre-processing, and model training, the drawings from the dataset's PDF files with different page numbers were converted to JPG format and named using the structure "patient number_patient name_page number." As a result, it was found that the image dataset consists of 450 HTP drawings which were unlabelled.

In addition to the HTP drawings, the dataset includes an Excel file created by professionals from Istanbul Bilgi University's Psychology Department, that provides detailed information and psychological reviews on the drawings. This file contains important data about the owners of the drawings, such as age, gender, IQ score, name and patient ID, as well as depression, anxiety and externalising scores given according to their answers to questions about the drawings. The data has been organised into a data frame to make the information easier to access and use effectively as can be seen in Figure 3.

	image_file	patient_name	page_number	patient_ID	sex	age	externalising_score	anxious_score	depressive_score	iq_score
0	81_1.jpg	OFB	1	81	1	6	10.0	7.0	4.0	96.0
1	61_3.jpg	ES	3	61	1	6	10.0	7.0	4.0	116.0
2	44_2.jpg	REG	2	44	1	7	27.0	17.0	6.0	49.0
3	8_1.jpg	AAA	1	8	1	6	9.0	5.0	3.0	121.0
4	45_1.jpg	NEŞ	1	45	0	8	3.0	6.0	2.0	96.0
...	...	...	...	...	...	...	...	...	...	...
445	35_3.jpg	BK	3	35	1	12	16.0	8.0	3.0	93.0
446	64_4.jpg	ARÖ	4	64	1	13	6.0	6.0	3.0	85.0
447	65_5.jpg	HP	5	65	0	7	10.0	7.0	1.0	115.0
448	2_5.jpg	YG	5	2	0	9	3.0	5.0	3.0	115.0
449	68_5.jpg	HÇ	5	68	1	10	7.0	4.0	4.0	81.0

450 rows × 10 columns

Figure 3: HTP Images Dataset Image Information Table

A detailed exploratory data analysis was performed by using all of these data to better understand the overall dataset structure. This analysis gave useful insights and guided the methods that were chosen for this study. Figure 4 illustrates the gender distribution of the children, where female is represented as 0 and male represented as 1.

This dataset, which consists of 53% male and 47% female, is considered balanced in terms of gender. Therefore, it is not expected to have problems caused by gender bias. Similarly, the ages of the children participating in the HTP test were examined.

As visualised in Figure 5, the dataset includes HTP test drawings from children aged 4 to 16, confirming its usefulness for analysing children's psychological states with artificial intelligence by using the HTP test data.

As part of the exploratory data analysis, the next feature examined was the anxiety, depression and externalising scores assigned by psychologists according to the answers given to the questions related to drawings, as Buck (1948) suggested. Figure 6 shows the histograms of externalising, anxiety and depression scores, respectively. The introversion scores in the dataset varied from 0 to 40, with the majority of images getting a score of 5. Similarly, anxiety scores varied between 0.0 to 17.5, and depression values from 0

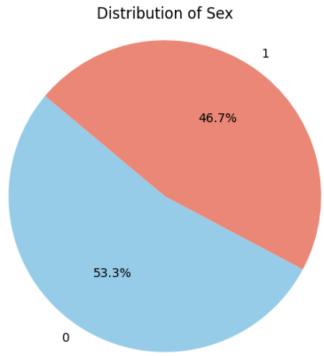


Figure 4: Distribution of Gender

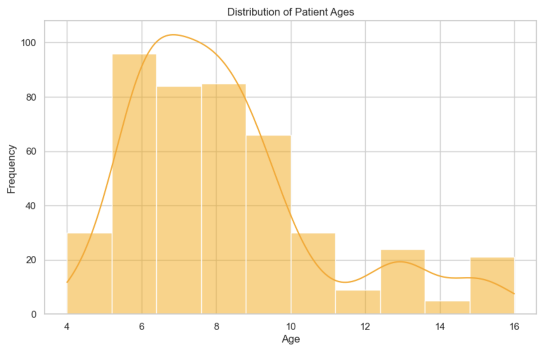


Figure 5: Age Distribution of Patients

to 14. These findings contributed to the project's next steps, which will be described in following sections.

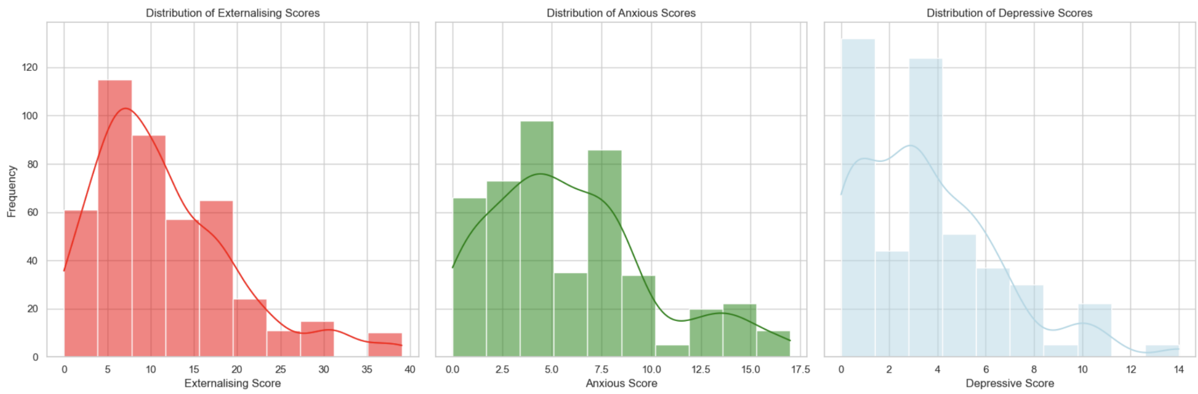


Figure 6: Psychological Scores' Distribution

Following the detailed analysis of the dataset, pre-processing methods such as resizing and normalising were applied to ensure that all images had the same size and structure of 224x224 pixels.

3.2 Image Classification Approach

The HTP image dataset, which is analysed in detail, consists of images of houses, trees and people that are not labelled, as mentioned earlier. This lack of labelling poses a significant challenge for artificial intelligence models, which are designed to infer psychological features. As a result of the use of unlabelled images, the models can not detect patterns between psychological characteristics and drawings. Therefore, the first step of the project is to label the drawings into three categories: house, tree and person.

A detailed literature review proved that automatic image classification or object detection algorithms applied to house, tree, and person drawings, which are used primarily because they are drawn by children and have not gone through any psychological analysis, outperform other projects in the literature (Lee et al.; 2024). However, in projects where drawings were carefully analysed by professionals and datasets were properly combined, object recognition or image classification algorithms showed limited accuracy (Salar et al.; 2023). Considering the ethical importance of providing reliability in the field of child psychology, it is critical to minimise errors. As a result, this project differs from others as it uses manual image labelling.

After careful analysis, it was decided that the use of various data labelling tools would be appropriate for the project due to the user-friendly interface, efficiency and various additional features such as annotation. For this reason, the Label Studio [4], which is an open source data labelling application, was used in this project. The image classification process starts by determining the labels that will be used in the project, which are house, tree and person in this case. Then, the images in the dataset imported to the project. Following these steps, each image had been classified one by one with the proper labels, as shown in Figure 7. Another advantage provided by this tool, which not only assigns labels to the images but also identifies and highlights the positions of objects within each image.

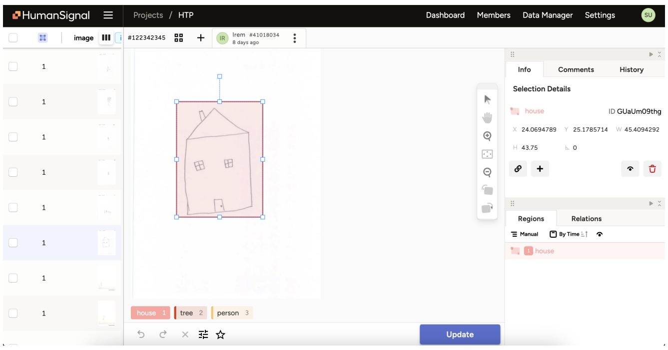


Figure 7: Label Studio Image Labelling

Once all of the images were successfully classified, they were exported in COCO format, which consists of a folder keeping the image files and a JSON file containing the labels and object coordinates for each image, which makes it easier to use in the project. In order to guarantee that the models only focus on the labelled objects in the drawings while ensuring that non-relevant drawings such as the house, tree, and person have no impact on model performance, the images have been trimmed using the object coordinates stored in the JSON file. Blank pages with no objects have been removed from the dataset as a consequence of the necessary modifications, leaving a total of 390 images in the dataset.

[4]Label Studio: `https://labelstud.io`

3.3 Feature Extraction Methods

The second part of the project involves extracting the features of the HTP test drawings. This phase is important since thorough analysis of these features allows the models implemented to accurately differentiate between house, tree, and person drawings.

3.3.1 Data Augmentation

Due to the limited dataset, a data augmentation approach was implemented to increase the number of data that is used to train the proposed models. As part of this method, the QuickDraw [5] dataset was used, which is a drawing game developed by Google that has millions of drawings in 345 categories, all drawn by real people from around the world (Salar et al.; 2023). Considering there was no person category in the QuickDraw dataset, 500 drawings of human faces from the "face" category were used as part of the data augmentation strategy, which also included 500 pictures from each of the house and tree categories by using "quickdraw" library of Python. Figure 8 shows the category distribution in the dataset before and after the data augmentation process, confirming the more balanced distribution achieved by this approach.

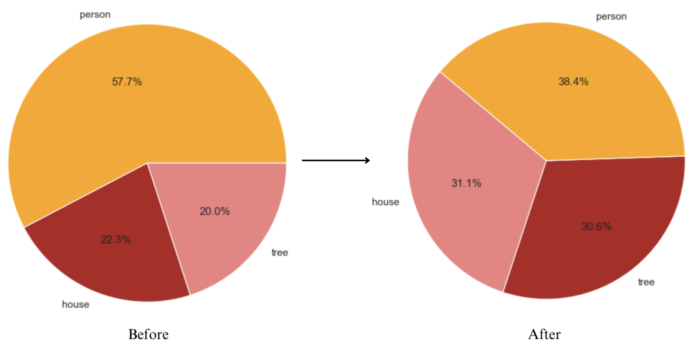


Figure 8: Distribution of Images by Category: Before and After

Table 2 provides information about the structure of the dataset used for the feature extraction part after all these data augmentation steps.

Data Source	House	Tree	Person
HTP Images Data	87	78	225
Quick Draw! Data	500	500	500

Table 2: Feature Extraction Dataset Distribution

3.3.2 ResNet50

ResNet50 (Residual Network) is a 50-layer convolutional neural network (CNN) architecture. The model, which has been pre-trained with the data in the ImageNet database consisting of millions of images, has been used in various disease detection projects due to its high performance in detecting complex patterns in images. As Kim et al. (2021) highlights, this approach performs well since it accurately translates inputs to outputs by learning residual values and using skip connections, helping it to successfully address the problems of deep learning.

[5] QuickDraw!: `https://quickdraw.withgoogle.com/data`

3.3.3 VGG16

Visual Geometry Group 16 (VGG16) model is a 16-layer convolutional neural network (CNN) approach that is pre-trained using ImageNet's database of millions of images. Although this model has not been used in other studies using HTP images in the literature, considering children's drawings usually include complex patterns and small details which can be difficult for others to understand, the application of this model is suggested due to its useful image processing capabilities.

3.3.4 EfficientNet

The EfficientNet model, which was previously trained in ImageNet's database of millions of images, was developed by Google Brain. The model, which has a highly advanced structure, balances the depth, width, and resolution parameters equally. Due to all these features, it was used in this project to be used on complex HTP images drawn by children.

3.4 Psychological Detection Methods

3.4.1 Random Forest

The random forests approach, which runs as a group of Decision Trees using a system in which each tree is trained individually using random subsets of the dataset's features, was used to assess the amount of anxiety and depression based on the features extracted in this study.

3.4.2 Decision Tree

The decision tree technique, which divides datasets into sections according to the most important attribute values and provides easier classification, was proposed as the second machine learning model for detecting psychological states by using extracted features from images in this project.

3.4.3 Gaussian Naive Bayes

The Gauss Naive Bayesian method is a model that calculates the posterior probability of each class given input properties using Bayes Theorem. This method is particularly used in this project due to its simplicity and effectiveness in handling large amounts of data.

3.4.4 Support Vector Machine

The Support Vector Machine (SVM) which was chosen for this project due to its simple structure and speed, is a machine learning technique used to identify the most suitable hyperplane with the largest distances between data points. Pan et al. (2022) have used this approach, which generated successful results in a variety of health-related artificial intelligence initiatives.

3.4.5 Logistic Regression

Logistic Regression is a machine learning model that is efficient and easy to use in binary classification projects. This model has showed significant performance in various disease detection projects. As a result of all these features, it has been applied in this project to analyse the performance it will show on the HTP dataset compared to other proposed machine learning models.

3.5 Evaluation Metrics

Within the scope of this project, various performance metrics have been taken into consideration in order to determine the most successful models in the feature extraction or psychological detection parts.

In the evaluation of the models' results implemented in the project, TP (true positive), TN (true negative), FP (false positive), and FN (false negative) were used.

3.5.1 Accuracy

The accuracy score is a performance metric that is primarily considered, which indicates how accurately the model predicted. In other words, the ratio of the total number of correct predictions to the total number of predictions gives the accuracy score. The accuracy score varies between 0 and 1, with 1 representing maximum accuracy and 0 representing minimum accuracy.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

3.5.2 Precision

Another performance metric is precision, which is the ratio of accurately predicted positive observations to the total number of positive predictions. The formula of the precision is as follows:

$$precision = \frac{TP}{TP + FP}$$

3.5.3 Recall

Recall also referred as the sensitivity, assesses a model's capability to recognise all positive data. The recall score ranges from 0 to 1, with 1 representing the highest recall score. The formula of this performance metric is the following:

$$recall = \frac{TP}{TP + FN}$$

3.5.4 F1-Score

F1-Score which is also known as F-Measure, evaluates how effectively recall and precision are balanced. A model maximum F1 score can be 1 and minimum 0. It is calculated using the following formula:

$$f1 - score = \frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

3.5.5 Grad-CAM

Grad-CAM (Gradient-weighted Class Activation Mapping) is an explainable AI method that helps people understand deep learning model decisions by visualising the areas and related inputs of data on a heat map. While the most significant characteristics are highlighted in red, the colour shifts to blue as the level of importance decreases.

4 Design Specification

In Section 3, the approaches proposed to be used in this project on the detection of children's psychological states by using HTP drawings with the help of artificial intelligence algorithms are explained briefly. Figure 9 shows an in-depth illustration of the techniques proposed as well as the project's architectural design.

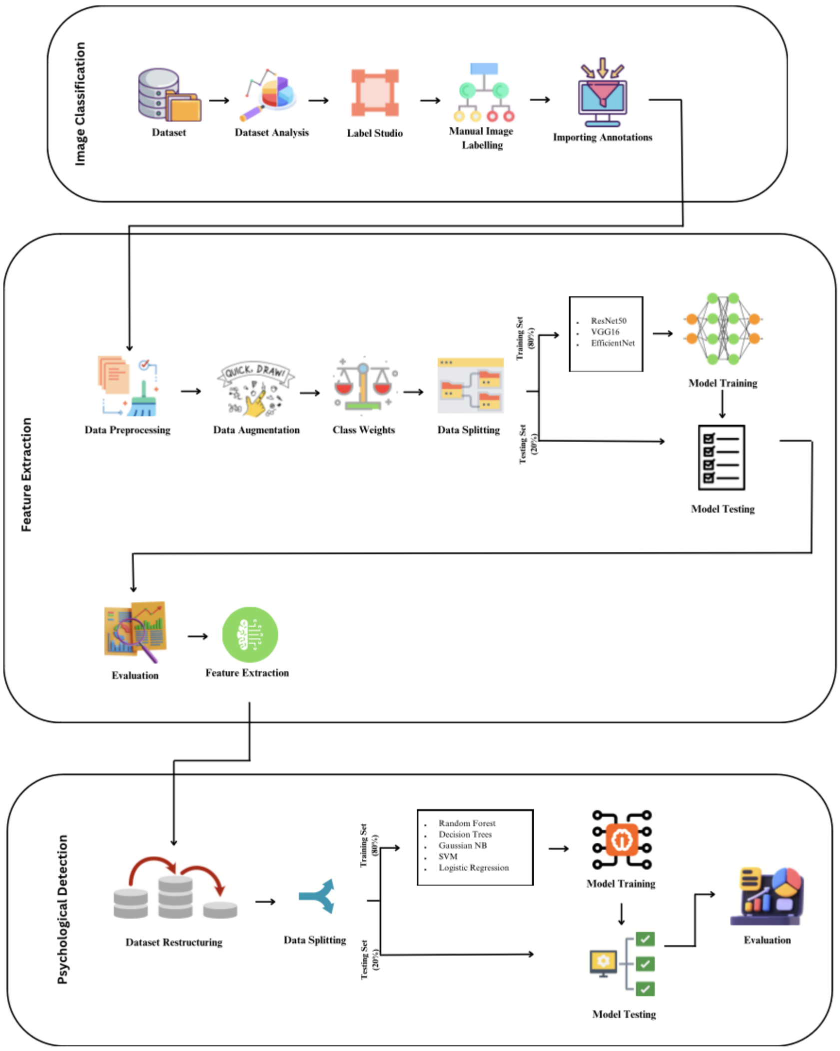


Figure 9: Project Design Architecture Diagram

As can be understood from the diagram, the project consists of harmonised three main steps: image classification using advanced tools, feature extraction with deep learning, and psychological detection utilising machine learning models.

4.1 Image Classification

In the image classification, which is the first step of the project, a manual approach has been adopted, as mentioned earlier. The main reason for this is that artificial intelligence algorithms fail to completely understand children's drawings, which are difficult to interpret even by human beings. For this strategy, the Label Studio application, which provides tools for several AI and data tasks, was chosen. The HTP image dataset, which had 450 data file, was reviewed one by one and labelled with the appropriate home, tree, or person label, as well as the object's positions. Following the labelling process, the results were exported in COCO format and then included in the project.

The manual image labelling method guaranteed that the limited dataset, which was first created by an authorised institution, was changed in order to avoid any errors in automatic image classification. The study applied this approach by following ethical standards in order to improve the field of child psychology by preventing the detection of incorrect results that may affect children who need the right treatment.

4.2 Feature Extraction

The second step of the project has been determined as feature extraction. After following the necessary pre-processing and data balancing steps by using QuickDraw dataset as described in the previous sections, various deep learning approaches have been used. It is expected that using models that have previously been trained with millions of images and have achieved great success in the field of image processing would be beneficial in this stage. Various performance metrics of the Resnet50, VGG16 and EfficientNet models, whose parameters were carefully selected, were analysed and the model with the best performance was saved for use in later steps. Unlike the projects reviewed in the related work section, the models apply a fine tuning technique with a very low learning rate. This approach guarantees that the pre-trained models continually evolve, improving their predictions based on new data while maintaining the generalised information they has previously learnt.

4.3 Psychological Detection

Since it shows the best performance in the feature extraction step, the saved model is used to extract the features of the images in the dataset. Following this, the final stage of the project started: Psychological Detection. The outputs of the feature extraction are stored in the data frame which is used in this final stage. Using these data frames, the performance of a number of machine learning algorithms was compared based on their ability to understand relationships within text data, which consists of extracted features of images, as well as their simplicity of implementation.

5 Implementation

5.1 Technologies & Tools

The project, which used artificial intelligence techniques to detect psychological disorders of children using HTP test images, was developed on a MacBook Pro with Apple M2 chip. This system has an 8-core design, with 4 performance cores and 4 efficiency cores. The project was developed with the open-source Python programming language and ran in the Jupyter Notebook environment via Anaconda Navigator. The main reason for this is Python and Jupyter Notebook's extensive libraries, which are quite useful for projects using artificial intelligence models. Figure 10 classifies the technologies and libraries used according to their specific purposes.

Figure 10: Technologies and Libraries Used in The Project

5.2 Feature Extraction

Following completing the data set pre-processing and data augmentation stages, which are detailed in Section 3, and the image classification procedure, which is one of the project's most critical tasks, the feature extraction part was started. Due to the fact that there is still a imbalance in the distribution of labels in the dataset after the data augmentation, the class weights method has been applied. Since the labels for the images in the "feature extraction" dataset that contains original dataset's images and QuickDraw images, are categorical, such as house, tree, and person. These categorical labels were converted to numerical values, 0, 1, and 2, respectively, using LabelEncoder.

As it was underlined before, the first model applied within the scope of the model application was ResNet50, which was selected due to its success in the field of image classification. Similarly, the VGG16 and EfficientNet models, which were previously trained with the images in the ImageNet database, were also used in the project due to their advanced structure and their success in the task of image classification. Keras and Tensorflow libraries were used to create and train the structure of these models. Figure 11 represents the general architecture of the all models which were built with the additional layers. The only variable in the model architecture is the "Base Model" layer. According to the applied deep learning method, the base model varies as ResNet50, VGG16 or EfficientNet.

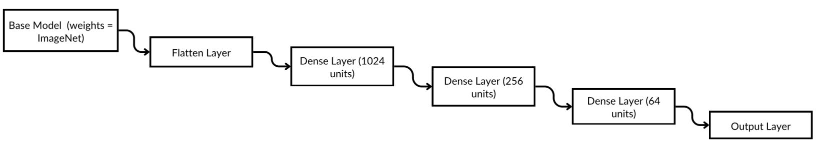


Figure 11: Base Model Architecture

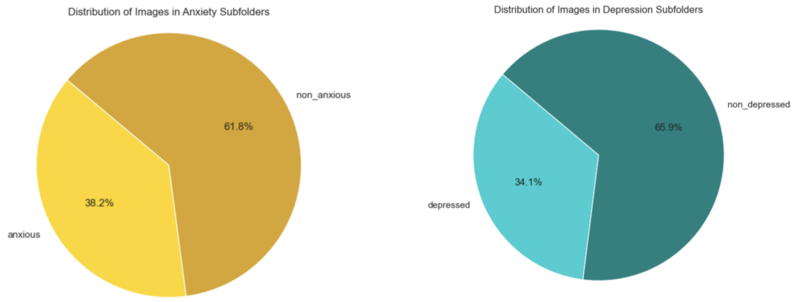
According to the general model architecture, the pre-trained model was saved as a basic model without classification layers in order to change its structure according to the project objectives by adding new layers. The first layer of the architecture of the base model used in the project was "Flatten", which allows the conversion of the multidimensional base model structure into a one-dimensional structure. This layer is followed by three "Dense" layers connected to each other. These layers consist of 1024, 256 and 64 units with ReLU activation, respectively. The parameters for these layers were carefully selected by testing to remove potential issues, such as over-fitting. The last layer, represented as the "Output Layer", consists of 3 units with softmax activation that create options for each of the three house, tree, and person classes. The last 10 layers are set to be trainable to preserve the learned features from the previously trained basic model, speed up training, and improve model performance. This fine-tuning keeps the learning method focuses on the latest layers. The model was developed with the Adam optimiser at a very low learning rate of 1e-5 to avoid problems such as over-fitting and to prevent the model from repeating the incorrect judgements made in earlier stages. "SparseCategoricalCrossentropy" was utilised as the loss function for this task. The models were trained for 10 epochs with the training data (80%), then tested with test data (20%). As mentioned earlier, class weights were taken into account during the training. Finally, the model with the best performance was saved for later use.

5.3 Psychological Detection

Using the saved model with the best performance in feature extraction, the data set that will be used in the Psychological detection task was reshaped before the features of the images in the dataset were extracted. In this task's dataset, there are only the data of the "HTP Images", which have been examined by psychologists and have undergone the necessary labelling processes. For easier data analysis based on anxiety and depression levels, the dataset was divided into folders named "Anxiety" and "Depression". In the "Anxiety" folder, two sub-folders were created: "non_anxious" and "anxious." Similarly, in the "Depression" folder, sub-folders named "non_depressed" and "depressed" were created. Images with depression and anxiety scores that had been analysed in detail were moved to the "non_anxious" sub-folder if the anxiety scores were less than 6, to the "anxious" sub-folder if they were higher. Similarly, they were moved to the "non_depressed" folder if the depression scores were less than 4, and to the "depressed" folder if they were higher. Following this reorganisation, the distribution of the "Anxiety" and "Depression" groups is illustrated in Figure 12.

The next step in the psychological detection section, where the imbalance between the data is resolved by class weight, is the feature extraction of the images using the saved model. The "get_features" function, which includes steps such as image resizing, converting to array and normalising, is then passed through the saved model to obtain feature vectors, resulting in feature extractions in the array structure.

The data obtained are divided into training and test classes, 80% and 20% respectively, using the "ml_models" function. This function defines the procedure for training, testing, and evaluating different machine learning models. Some of the models described in previous sections, which are activated by using this function, need unique parameter values. First of all, the Random Forest was set to use 300 estimators and a maximum depth of 300 since the dataset contained data that was complex. Second, a linear kernel was employed to determine the hyperplane in a best way, which separates the categor-



Anxiety Folder

Depression Folder

Figure 12: Anxiety and Depression Distributions

ies in the SVM model. Finally, the Logistic Regression model was optimised with 1000 iterations.

6 Evaluation

6.1 Feature Extraction

Feature extraction, which is detailed in depth in Section 5, is critical for the success of the next psychological detection step. The performances of the ResNet50, VGG16 and EfficientNet models used for this task were evaluated by taking into account the performance metrics specified in Section 3.

The ResNet50 model, which was first applied, has shown a remarkable success by achieving an accuracy rate of 89%, precision of 91%, recall of 89% and F1-score of 89% after 10 epochs of training. The details of this model's performance metrics and the confusion matrix are illustrated in Figure 13.

Classification Report:

	precision	recall	f1-score	support
0	0.97	0.86	0.91	115
1	0.79	0.96	0.87	135
2	0.96	0.85	0.90	128
accuracy			0.89	378
macro avg	0.91	0.89	0.89	378
weighted avg	0.90	0.89	0.89	378

Training Time: 682.3160371780396
Prediction Time: 10.558568000793457

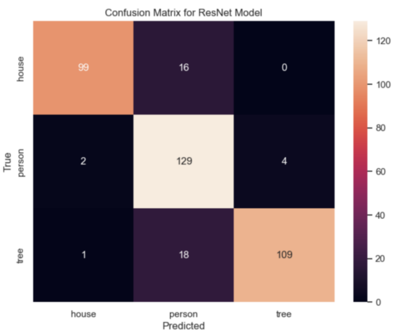


Figure 13: ResNet50 Model Performance Metrics

The second model applied, the pre-trained VGG16 model, achieved a significant accuracy rate of 97%. This model also performed well in other measures of performance, achieving 97% in precision, recall, and F1 scores. These results show that the proposed VGG16 model is very effective at correctly classifying drawings as house, tree, or person. The performance details of the model are shown in Figure 14.

EfficientNet, whose performance data on the classification of HTP images are shown in Figure 15, is the last model that has been trained and tested for the task of extracting

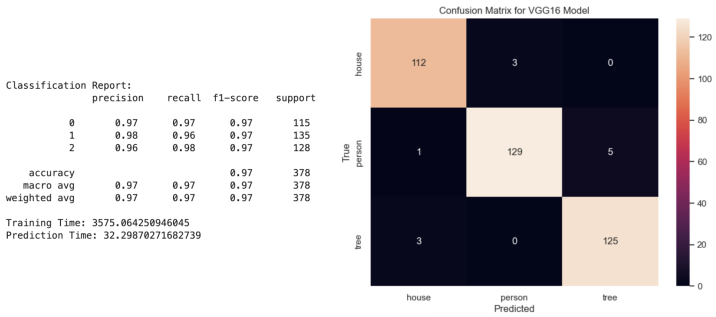


Figure 14: VGG16 Model Performance Metrics

the image features. The model, which showed a very low accuracy rate of 32%, was insufficient to understand the important features of house, tree and person drawings.

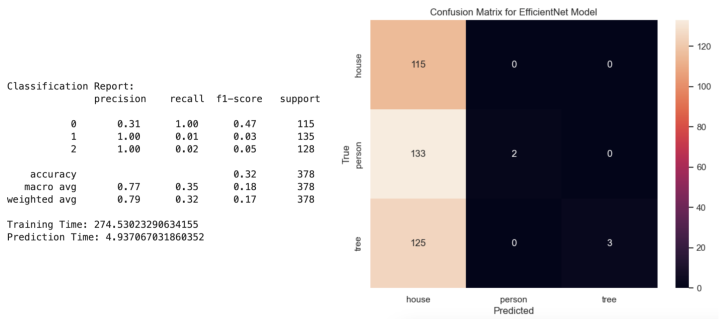


Figure 15: EfficientNet Model Performance Metrics

The graphs of the accuracy and loss rates achieved by the three models analysed in detail compared to the epochs in which they were trained are as shown in Figure 16. The results show that EfficientNet performed poorly in analysing the important characteristics of the images. It produced many errors while classifying house, tree, and person drawings, which caused a high rate of incorrect classifications based on the image characteristics.

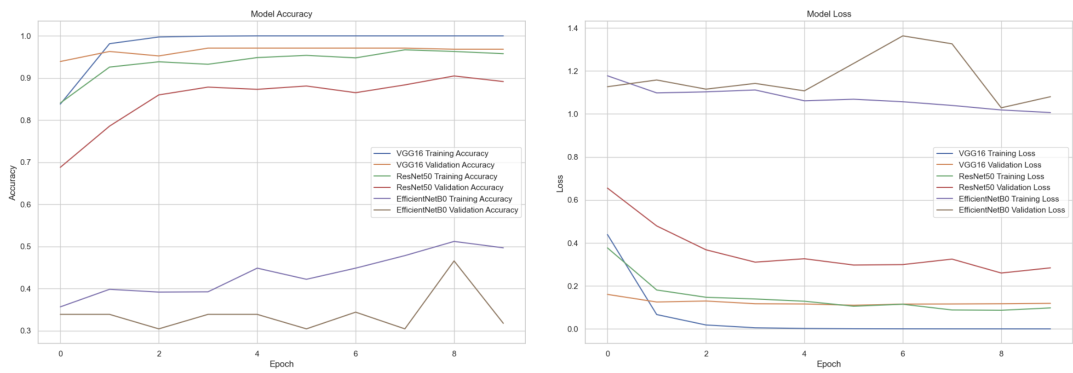


Figure 16: Models' Accuracy and Loss by Epochs

On the other hand, the ResNet50 and VGG16 models have proven their success in understanding the differences between images with their accuracy rates of 89% and 97%, respectively. The VGG16 model, in particular, showed a significant improvement in accuracy and a decrease in loss as the training period passed, proving that it is the best model for the task of image feature extraction among the methodologies tested. Figure 17 shows the time spent running the models, which highlights the relationship between accuracy and processing time.

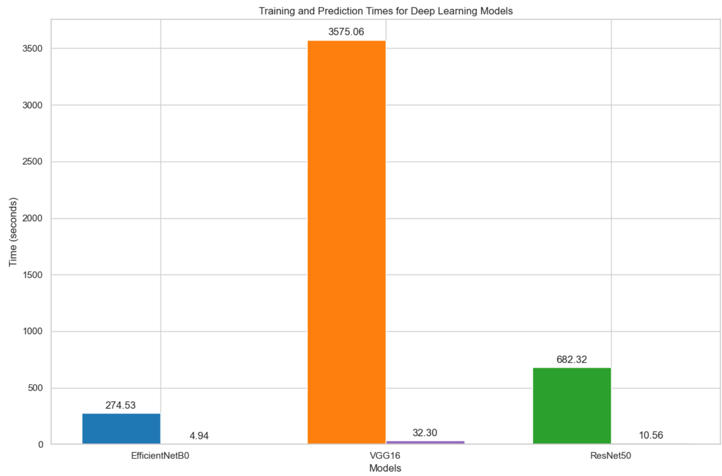


Figure 17: Training and Prediction Times of Deep Learning Models

Although the VGG16 model, which has the highest performance scores, has the longest processing time as can be seen from the visual, this time spent on complex calculations has enabled a more advanced and effective model. Figure 18 represents the logic of the ResNet50 model in understanding HTP images with a heat map, while Figure 19 shows the features that the VGG16 model gives attention to decision-making. As can be understood from these visuals, the VGG16 model has shown a better performance due to being more sensitive to the details of drawings and more aware of their important features.

Figure 18: ResNet50 Grad-CAM

Considering all of these, the VGG16 model was chosen to extract the features of the drawings in the HTP images dataset since it performed the best among the models tested.

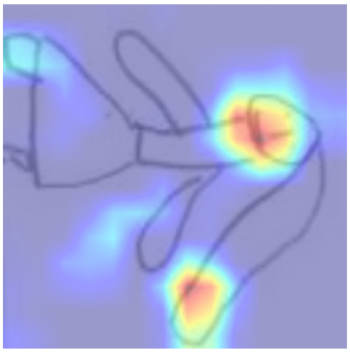


Figure 19: VGG16 Grad-CAM

6.2 Psychological Detection

The psychological detection procedure has begun once specific characteristics have been identified using the chosen model for analysing the differences between the house, tree, and person drawings. The machine learning techniques presented in Section 3 were first trained using the edited "Depression" dataset to detect symptoms of depression in HTP drawings. Table 3 displays the performance metrics of the models applied in the detection of depression.

Machine Learning Models	Precision	Recall	F1-Score	Accuracy
Random Forest	69%	72%	68%	72%
Decision Tree	60%	59%	59%	59%
Gaussian Naive Bayes	64%	69%	62%	69%
SVM	79%	79%	79%	79%
Logistic Regression	76%	77%	77%	77%

Table 3: Machine Learning Models' Performances for Depression Detection

As previously explained, the first model applied to determine whether the owners of the drawings in the dataset have psychological problems such as depression and anxiety using these extracted features is Random Forest. The system designed to detect depression in the drawings achieved an overall accuracy rate of 72%. This model performed well in detecting signs of depression, with a precision of 69%, recall of 72%, and an F1 score of 68%. The decision tree model, which is the second model studied, was first tested in the field of depression detection. The model achieved an accuracy rate of 59%, with a precision score of 60%, recall of 59%, and an F1 score of 59%. Third model which was Gaussian Naive Bayes, achieved a 69% accuracy, a precision of 64%, a recall of 69%, and an F1-score of 62%. The next model applied, which was the SVM model, achieved a rate of 79% in terms of accuracy, precision, recall and f1- scores. The final model evaluated for depression detection, Logistic Regression, achieved an accuracy rate of 77%. The precision score was 76%, while the recall and F1 scores were also 77%. These results showed that the Support Vector Machine (SVM) model and Logistic Regression model performed significantly better. Considering all of the performance metrics, the SVM model emerged as the most effective method for detecting depression, as can be seen in Figure 20.

All of the machine learning approaches that were previously evaluated for their depression detection performance, were also tested for their accuracy in detecting anxiety throughout the Psychological Detection section. Following feature extraction for the images in the "Anxiety" dataset, the models were trained and their performance, as shown in Table 4, was evaluated.

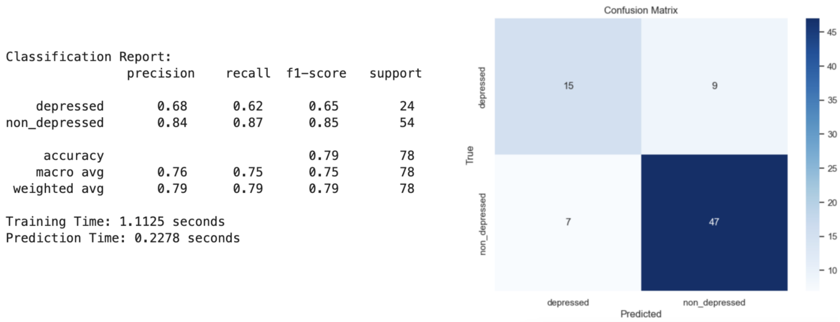


Figure 20: Support Vector Machine Model Performance Metrics for Depression

Machine Learning Models	Precision	Recall	F1-Score	Accuracy
Random Forest	74%	68%	61%	68%
Decision Tree	61%	62%	61%	62%
Gaussian Naive Bayes	49%	56%	49%	56%
SVM	66%	67%	66%	66%
Logistic Regression	62%	63%	62%	63%

Table 4: Machine Learning Models' Performances for Anxiety Detection

The Random Forest model which was implemented for anxiety detection, achieved a 68% accuracy rate as well as score of 74% precision, 68% recall and 61% f1-score. Secondly, Decision Tree model utilised for the same task. As a result, this model showed a good performance by achieving 62% accuracy, 61% precision, 62% recall, and a 61% F1 scores. Then, the Gaussian Naive Bayesian model was applied, showing 56% accuracy, 56% recall, 49% precision and F1 scores, which was less effective compared to the previous models. The SVM model was the next model applied in the detection of anxiety. It achieved an accuracy rate of 66%, demonstrating comparable performance to the Random Forest model. The SVM model showed 66% precision, 67% recall, and a 66% F1 score, showing its effectiveness in anxiety detection. Although the final model analysed, Logistic Regression, showed considerable effectiveness in detecting anxiety (accuracy and recall of 63%, precision and F1 score of 62%), it was not as effective as the Random Forest model. In conclusion, a thorough analysis of all models showed that the Random Forest and SVM models performed similarly in identifying anxiety via HTP test images. However, the Random Forest model demonstrated the best performance overall, as can be seen in Figure 21.

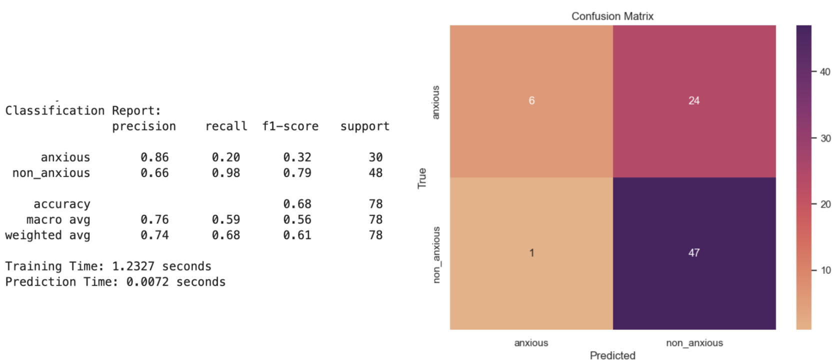


Figure 21: Random Forest Model Performance Metrics for Anxiety

7 Conclusion and Future Work

The HTP test is an useful art therapy technique, especially for children, where therapists assess images of house, tree, and person drawn by the participants. By assessing the responses to the questions related to these images, insight into the individual's family dynamics, personalities, and social relationships can be gained. This project aims to improve the accuracy of HTP test analyses by automating the psychological assessment process of the drawings by artificial intelligence approaches. In this study, advanced image classification tools are used, which differ from the approaches that recent projects utilised. Following the image classification, the project continues to feature extraction and psychological detection steps to analyse HTP test data.

In the feature extraction section, the performance of pre-trained and fine-tuned deep learning techniques in terms of recognise house, tree and person drawings was evaluated. The most successful model, VGG16 with 97% accuracy, was applied to extract significant features of images. These features were then used in model training to examine the effectiveness of multiple machine learning models for detecting depression and anxiety. The SVM model scored 79% accuracy in the depression diagnosis tests, while Random Forest achieved 68% accuracy in the anxiety detection tests.

In future studies, there is a significant need to expand the current research by working with larger and labelled HTP test datasets. This large dataset should be created using HTP test images collected in a professional environment, with all necessary ethical issues addressed and permissions secured. Furthermore, to increase automated evaluation of HTP test data via AI models, it is essential to improve the algorithms' performance by including new characteristics. For example, implementing features such as object size and amount, which psychologists give attention to in HTP drawings, could significantly boost the study's effectiveness. By addressing these issues, the analysis of house, tree, and person drawings could provide useful information into children's states of anxiety or depression, allowing for early detection and treatment.

References

Buck, J. (1948). The h-t-p test, *Journal of Clinical Psychology* **4**(2): 151–159.

Case, C. and Dalley, T. (2014). *The Handbook of Art Therapy*, Routledge.

Du, X., An, P., Leung, J., Li, A., Chapman, L. E. and Zhao, J. (2024). Deepthink: Designing and probing human-ai co-creation in digital art therapy, *International Journal of Human - Computer Studies* **181**(103139).

Eaton, L. G., Doherty, K. L. and Widrick, R. M. (2007). A review of research and methods used to establish art therapy as an effective treatment method for traumatized children, *The Arts in Psychotherapy* **34**(3): 256–262.

Kim, T., Yoon, Y., Lee, K., Kwahk, K. and Kim, N. (2021). Application of deep learning in art therapy, *International Journal of Machine Learning and Computing* **11**(6).

Lee, M., Kim, Y. and Kim, Y.-K. (2024). Generating psychological analysis tables for children's drawings using deep learning, *Data Knowledge Engineering* **149**(102266).

Malchiodi, C. A. (2020). *Trauma and expressive arts therapy: Brain, body, and imagination in the healing process*, Guilford Publications.

Pan, T., Zhao, X., Liu, B. and Liu, W. (2022). Automated drawing psychoanalysis via house-tree-person test, *2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI)* .

Salar, A. A., Faiyad, H., Sönmez, E. B. and Hafton, S. (2023). Artificial intelligence contribution to art-therapy using drawings of the house-person-tree test, *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)* pp. 1–6.