# Analysing the Impact of Images and Text for Predicting Human Creativity Through Encoders

Amaia Pikatza-Huerga[1][a], Pablo Matanzas de Luis[1][b], Miguel Fernandez-de-Retana Uribe[1][c], Javier Peña Lasa[2][d], Unai Zulaika[1][e] and Aitor Almeida[1][f]

[1]*Faculty of Engineering, University of Deusto, Unibertsitate Etorb., 24, Bilbao, Spain*

[2]*Faculty of* *ilbao, Spain*

{*a.pikatza, javier.pena, unai.zulaika, aitor.almeida*}*@deusto.es*

Keywords: Machine Learning, Creativity Assessment, Originality Evaluation, Artistic Expression, Text and Image Analysis, EEG.

Abstract: This study explores the application of multimodal machine learning techniques to evaluate the originality and complexity of drawings. Traditional approaches in creativity assessment have primarily focused on visual analysis, often neglecting the potential insights derived from accompanying textual descriptions. The research assesses four target features: drawings' originality, flexibility and elaboration level, and titles' creativity, all labelled by expert psychologists. The research compares different image encoding and text embeddings to examine the effectiveness and impact of individual and combined modalities. The results indicate that incorporating textual information enhances the predictive accuracy for all features, suggesting that text provides valuable contextual insights that images alone may overlook. This work demonstrates the importance of a multimodal approach in creativity assessment, paving the way for more comprehensive and nuanced evaluations of artistic expression.

## 1 INTRODUCTION

The assessment of creativity is a dynamic field where artificial intelligence (AI) opens possibilities to enhance the objectivity, scalability, and depth of evaluations across various tasks. Traditionally, creativity assessments, such as the Alternate Uses Task (AUT) in verbal creativity and drawing-based tasks in visual domains, have relied heavily on human judgment, facing challenges in consistency and efficiency. AI, however, introduces data-driven methods to quantify creativity aspects like originality and flexibility with objective precision. For instance, platforms like SemDis (Beaty and Johnson, 2020) use natural language processing to measure semantic distance, automating the scoring of verbal creativity and reducing the subjectivity and labor intensity of manual evaluations (Allen et al., 2015; Shaban-Nejad et al., 2022). Such advances lay a foundation for reliable, large-scale creativity assessments, making comprehensive analysis feasible (Stojnic et al., 2022).

In visual creativity assessment, a specialized area focuses on drawing completion tests commonly used in psychology to explore aspects of personality, emotional state, and cognitive style. These tests, like the Thematic Apperception Test (TAT) and Draw-A-Person Test (DAP), ask participants to complete drawings, revealing deeper psychological traits. While valuable, these assessments have traditionally depended on subjective evaluations, limiting consistency and scalability. AI enhances these assessments by analysing graphic features—such as line quality, shape complexity, and spatial arrangement—objectively, improving reliability and identifying subtle patterns that might be missed by human evaluators (Liu et al., 2020; Wang et al., 2023; Tan et al., 2023; Gado et al., 2021).

AI-based approaches in figural creativity have shown particular promise. Convolutional neural networks (CNNs) (O'Shea and Nash, 2015) have been applied to measure originality in drawing tasks, aligning closely with expert ratings and reducing both time and costs while ensuring consistency (Cropley

[a] https://orcid.org/0009-0003-9080-6242
[b] https://orcid.org/0009-0009-8897-5796
[c] https://orcid.org/0009-0002-0883-1303
[d] https://orcid.org/0000-0002-0041-7020
[e] https://orcid.org/0000-0002-7366-9579
[f] https://orcid.org/0000-0002-1585-4717



Pikatza-Huerga, A., Matanzas de Luis, P., Fernandez-de-Retana Uribe, M., Lasa, J. P., Zulaika, U. and Almeida, A.
Analysing the Impact of Images and Text for Predicting Human Creativity Through Encoders.
DOI: 10.5220/0013203600003938
In *Proceedings of the 11th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2025)*, pages 15-24
ISBN: 978-989-758-743-6; ISSN: 2184-4984

and Marrone, 2022; Kvam et al., 2023). Similarly, platforms like AuDrA (Patterson et al., 2023) utilize modified ResNet (He et al., 2015) architectures to score features like elaboration and divergent thinking, achieving correlations with human evaluations and highlighting AI's potential to standardize creativity assessments (Easton et al., 2019; Davis et al., 2022). This trend reflects the broader integration of AI in education, where it increasingly supports learning and assessment practices (Pezzulo et al., 2023).

Building on these advances, recent work has explored supervised learning techniques, such as Vision Transformers and Random Forest classifiers, to automate scoring in tasks like the Torrance Tests of Creative Thinking-Figural (TTCT-F) (Acar et al., 2024). This approach extends this by integrating textual data, including titles or descriptions accompanying drawings, to enable a multimodal assessment of creativity. This combined analysis of visual and textual data allows AI to capture nuanced aspects—particularly in originality and flexibility—that may be overlooked by image-only models (Weidinger et al., 2022). As human creativity often involves both visual and verbal expression, this multimodal approach is essential for more comprehensive evaluation (Bahcecik, 2023).

AI's applications in assessing emotional content within drawings also show promise. Using sentiment analysis, AI can detect emotional cues in human figure drawings, traditionally used to evaluate emotional well-being and intelligence. This ability enables faster, more precise emotional assessments, aligning with the increasing recognition of emotional intelligence in psychological evaluation (Imuta et al., 2013; Røed et al., 2023; Devedzic, 2020).

Beyond assessment, AI holds potential for therapeutic applications. By tracking changes in patient drawings over time, AI helps clinicians monitor emotional and cognitive progress throughout therapy, facilitating personalized interventions that leverage the creative process as a therapeutic tool (Zhang et al., 2024; Lee et al., 2015). AI's role in these settings not only enhances therapeutic effectiveness but also underscores the healing potential of creativity (Searle, 2018).

AI's capacity to compare drawings against normative data further enhances its diagnostic capabilities, helping detect psychological conditions early by identifying deviations from typical profiles (Sheng et al., 2019; Ferrara and Qunbar, 2022). Additionally, by assessing cognitive styles in drawing tasks, AI offers insights into thought processes and personality traits, advancing diagnostic accuracy and deepening our understanding of individual differences in creativity (Gigi, 2015; Creely and Blannin, 2023; Cetinic and She, 2021).

This research advances these developments by introducing a novel tool that combines visual and textual data for a more thorough creativity assessment in drawing tasks. The influence of visual and textual data has been studied separately, and by integrating titles or descriptions often accompanying drawings, this approach captures added layers of creativity, especially in originality and flexibility, where text can enrich insights beyond what image-based models offer. The model, trained on expert evaluations, aligns closely with human expertise while minimizing subjective inconsistencies, allowing for scalable, precise assessments across diverse psychological tests (Harré and El-Tarifi, 2023).

# 2 METHODS

## 2.1 Data

The drawings in this dataset were collected as part of a study aimed at investigating the effects of intracranial stimulation on creativity. In this original study, 53 participants were asked to create drawings in two phases: one before receiving intracranial stimulation (pre-stimulation phase) and another after the stimulation (post-stimulation phase). For each drawing, participants were also asked to provide a title or a brief description that reflected their interpretation or concept of the image.

The primary goal of that study was to explore whether intracranial stimulation could influence the originality of the participants' drawings. The dataset used in this research consists of these drawings along with their corresponding titles, both from the pre-stimulation and post-stimulation phases. These images were collected and made available for further analysis, and they serve as the foundation for the current study.

The dataset used consists of 486 samples, including numerical and textual data related to scanned images of drawings. Each image is assigned labels corresponding to various features, including the title given by the participant to their drawing, which is used for classification with a deep learning model. The study targets four numerical variables: O, FLE, E, and T, which represent different aspects of creativity that the model seeks to predict. Each entry in the dataset contains the following fields:

- IMAGE: A scanned image of one of the drawings completed by the participant.
- TEXT: Title assigned to the drawing by the par-

ticipant (in Spanish).

- O: Label given by an expert psychologist indicating whether the drawing demonstrates creativity, expressed as 0 (not original) or 1 (original).
- FLE: Label provided by an expert psychologist that assesses the participant's flexibility in drawing. Each drawing is assigned a numerical category related to its theme (people, landscapes, etc.), and flexibility is calculated based on the number of different categories represented.
- E: Label assigned by an expert psychologist that measures the level of elaboration of the drawing.
- T: Label provided by an expert psychologist indicating whether the title given to the drawing is creative or not, also expressed as O (0 or 1).

## 2.2 Participants

In total, 53 participants contributed to the creation of the dataset. Demographic and personal characteristics of the participants include age, gender, educational level, mother tongue and certain habits, such as stimulant and tobacco use, as well as number of hours of sleep.

The main features of the participant dataset are:

- Gender: Gender of the participant, recorded as 'M' (male) or 'F' (female). The distribution was balanced, with 49.1% men and 50.9% women.
- Age: Age in years of participants, ranging from 10 to 60 years.
- Mother tongue: The majority of participants have Spanish as their mother tongue.
- Education: Educational level ranges from compulsory secondary education to postgraduate studies.
- Sleeping hours: Sleeping hours were recorded the night before the experimental session.
- Stimulants and Tobacco: Participants reported on the consumption of stimulants (e.g. coffee) and tobacco before the sessions.
- Observations: Additional notes on participants, such as medical or behavioural observations during the study.

## 2.3 Model Building

In this study, we aim to predict four creativity-related variables — O (originality), E (elaboration), FLE (flexibility), and T (title originality) — using multimodal data that combines visual information (images of the drawings) with textual information (titles assigned to the drawings by the participants). To achieve this, deep learning models were employed to analyze both the visual features of the drawings and the semantic features of the titles. The objective is to evaluate the performance of these models in predicting the mentioned variables and to explore to what extent each data modality (image or text) contributes to the model's accuracy.

To obtain a more detailed understanding, all possible combinations of text and image models, which are described in the following subsections, were tested. Additionally, experiments were conducted using only images and only text to predict each of the four creativity variables separately, allowing us to assess how much information each modality contributes independently.

The visual features of the drawings were processed using convolutional neural networks (CNN) as encoders to create an image embedding, while the titles assigned to the drawings by the participants were processed using different text embedding models. For models that utilized both text and image data, the final layer before the output of each model was concatenated with the other modality, allowing both sources of information to be combined effectively.

### 2.3.1 Image-Based Models

- ResNet50: A deep network with 50 layers widely used in image classification tasks due to its ability to handle degradation problems in deep networks. (He et al., 2015)
- InceptionV3: A modular network design model that efficiently uses computational resources, improving image analysis accuracy. (Szegedy et al., 2015)
- EfficientNetB0: This model optimizes both network size and accuracy, offering a balance between computational performance and feature extraction capacity. (Tan and Le, 2020)
- Xception: Based on depthwise separable convolutions, this model excels in image classification tasks, improving accuracy without significantly increasing computational cost. (Chollet, 2016)

The ResNet50, InceptionV3, EfficientNetB0 and Xception architectures were chosen due to their relevance and diversity in CNN design strategies. These architectures have consistently demonstrated high performance in image classification tasks, such as those in ImageNet benchmarks, and represent key ap-

proaches in the evolution of CNNs. ResNet50 incorporates residual connections that enable the training of deep networks; InceptionV3 optimizes computational efficiency with convolutions of varying sizes; EfficientNetB0 introduces compound scaling to balance accuracy and efficiency; and Xception utilizes depthwise separable convolutions, achieving improved accuracy with reduced computational cost. This selection ensures a representative and diverse analysis of visual encoding capabilities in the context of the problem studied.

#### 2.3.2 Text Embedding Models

- BETO (Cañete et al., 2020): A model based on the Transformer architecture that provides contextualized representation of words in the titles, capturing both local and global meaning of the text.
- FastText (Joulin et al., 2016): This model generates word embeddings that include morphological information, which is particularly useful for short titles or unknown words.
- Keras Embedding layer: A simpler model that enables efficient text representation using dense layers, suitable for fast and efficient classification tasks.

The BETO, FastText, and Keras Embedding layer models were selected to encompass diverse strategies in semantic text representation. BETO is a Transformer-based model pre-trained specifically in Spanish, making it ideal for capturing linguistic nuances in the analyzed titles. FastText generates embeddings based on subword information, allowing it to handle out-of-vocabulary words and morphological features, which are particularly useful for short titles. Finally, the Keras Embedding layer offers an efficient and flexible approach for dense text representation in classification tasks. Combining these approaches provides a comprehensive and complementary analysis of the semantic characteristics of the titles within the context of creativity.

### 2.4 Model Evaluation

To assess the models' performance in predicting creativity-related variables, we employ distinct metrics tailored to classification (binary and multiclass) and regression tasks.

#### 2.4.1 Classification Metrics

For both binary and multiclass tasks, the following metrics provide a comprehensive view of classification performance:

- ROC AUC: Evaluates the model's ability to distinguish between classes, using:
  - Binary AUC: Direct comparison between two classes.
  - One-vs-Rest AUC (multiclass): Calculates AUC for each class, revealing overall discrimination ability.
- Accuracy: Measures the proportion of correctly predicted labels across all classes.
- Recall (Sensitivity): Proportion of actual positive instances correctly identified, highlighting the model's capability in capturing positive cases.
- Precision: Accuracy of predicted instances per class, showing how well each class is identified.
- Specificity: Proportion of true negatives correctly identified, useful for understanding false positive avoidance.
- F1 Score: Harmonic mean of Precision and Recall, balancing performance in cases of class imbalance.

#### 2.4.2 Regression Metrics

For predicting continuous variables (e.g., elaboration scores), we apply:

- Loss (Mean Squared Error (MSE)): Measures average squared error, penalizing larger deviations between predicted and actual values.
- Mean Absolute Error (MAE): Represents the average absolute difference between predicted and true values, offering intuitive error measurement.
- Root Mean Squared Error (RMSE): Square root of MSE, emphasizing larger errors and enhancing interpretability.
- $R^2$ Score: Proportion of variance explained by the model, indicating overall predictive strength for continuous outcomes.

## 3 RESULTS

The following section presents the results of the model evaluation for predicting each of the four target features. Metrics are presented for all combinations of models, including those using image data only, text data only and both combined. The performance of each model is evaluated using a comprehensive set of metrics that reflect the quality of predictions in both classification and regression tasks. These results provide an in-depth view of the performance of each combination on different prediction targets and allow

Table 1: Results Predicting Originality (O).

| Embedding | CNN | ROC AUC | Accuracy | Recall | Precision | Specificity | F1 score |
|---|---|---|---|---|---|---|---|
| - | ResNet50 | 0,76 | 0,77 | 0,71 | 0,40 | 0,84 | 0,51 |
| - | InceptionV3 | 0,81 | 0,78 | 0,66 | 0,42 | 0,88 | 0,51 |
| - | EfficientNetB0 | 0,72 | 0,68 | 0,55 | 0,30 | 0,81 | 0,39 |
| - | Xception | 0,82 | 0,81 | 0,76 | 0,44 | 0,88 | 0,56 |
| BETO | - | 0,78 | 0,73 | 0,71 | 0,37 | 0,81 | 0,49 |
| BETO | ResNet50 | 0,78 | 0,69 | 0,92 | 0,36 | 0,67 | 0,52 |
| BETO | **InceptionV3** | 0.84 | **0,83** | 0,76 | **0,46** | **0,90** | **0,57** |
| BETO | EfficientNetB0 | 0,68 | 0,67 | 0,87 | 0,30 | 0,53 | 0,45 |
| BETO | Xception | 0,80 | 0,79 | 0,82 | 0,43 | 0,82 | 0,56 |
| FastText | - | 0,80 | 0,78 | 0,66 | 0,40 | 0,89 | 0,50 |
| FastText | ResNet50 | 0,74 | 0,66 | **0,97** | 0,34 | 0,59 | 0,50 |
| FastText | **InceptionV3** | **0,85** | 0,80 | 0,71 | 0,42 | 0,89 | 0,53 |
| FastText | EfficientNetB0 | 0,77 | 0,74 | 0,55 | 0,33 | 0,86 | 0,54 |
| FastText | Xception | 0,80 | 0,73 | 0,82 | 0,38 | 0,75 | 0,52 |
| Keras | - | 0,81 | 0,76 | 0,87 | 0,38 | 0,76 | 0,53 |
| Keras | ResNet50 | 0,74 | 0,76 | 0,66 | 0,38 | 0,86 | 0,48 |
| Keras | InceptionV3 | 0,80 | 0,71 | 0,87 | 0,37 | 0,72 | 0,52 |
| Keras | EfficientNetB0 | 0,74 | 0,61 | 0,74 | 0,37 | 0,78 | 0,49 |
| Keras | Xception | 0,75 | 0,71 | 0,82 | 0,37 | 0,63 | 0,51 |

a better understanding of the contribution of text, image and their joint use in the prediction process.

## 3.1 Predicting Originality (O)

The table 1 presents the performance metrics for various model combinations used to predict the binary variable O. The models are evaluated across five metrics: ROC AUC, accuracy, recall (sensitivity), specificity, and F1 score.

The accuracy values range from 0.61 to 0.83, with the highest accuracy observed in the model using the combination of BETO and InceptionV3. The ROC AUC values vary between 0.68 and 0.85, with the best performance in this regard achieved by the combination of FastText and InceptionV3.

Recall (sensitivity) scores span from 0.55 to 0.97. The model with the highest recall is the one that uses FastText and ResNet50, whereas EfficientNetB0 without using text results in the lowest recall. Specificity values range between 0.53 and 0.90, where the combination of BETO and InceptionV3 achieves the highest specificity, while the FastText and InceptionV3 combination shows the lowest.

Finally, F1 scores in the table range from 0.50 to 0.64, with the highest score obtained by the FastText and InceptionV3 combination. The performance of each model varies across different metrics, indicating that no single combination of models consistently outperforms the others in all areas.

## 3.2 Predicting Elaboration (E)

Table 2 presents the performance metrics for predicting the continuous variable E. The evaluation metrics provided include loss, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R2 score.

The lowest loss values are observed for the models without embeddings that use either InceptionV3 (2.82) or Xception (2.81), with these models also showing the best overall performance in other metrics. Specifically, the InceptionV3 model achieves the lowest MAE (1.07) and RMSE (1.30), along with the highest R2 score (0.48). The Xception model follows closely, with an MAE of 1.14 and an RMSE of 1.37, and an R2 score of 0.44.

In contrast, models incorporating embeddings show substantially higher loss values. For instance, the BETO embedding combined with InceptionV3 results in a loss of 9.02, an MAE of 2.22, and an RMSE of 2.82, with a negative R2 score of -1.91. Similar trends are observed for combinations involving other embeddings, where the R2 scores are consistently negative, indicating poor model performance for predicting E.

Overall, the results suggest that models using only image data outperform those incorporating embeddings when predicting E, as evidenced by lower error metrics and higher R2 values.

Table 2: Results Predicting Elaboration (E).

| Embedding | CNN | loss | MAE | RMSE | R2 score |
|---|---|---|---|---|---|
| - | ResNet50 | 3,95 | 1,36 | 1,74 | 0,05 |
| **-** | InceptionV3 | 2,82 | **1,07** | **1,30** | **0,48** |
| - | EfficientNetB0 | 7,98 | 1,98 | 2,76 | -2,76 |
| **-** | Xception | **2,81** | 1,14 | 1,37 | 0,44 |
| BETO | - | 8,44 | 2,14 | 2,70 | -1,61 |
| BETO | ResNet50 | 8,77 | 2,18 | 2,77 | -1,81 |
| BETO | InceptionV3 | 9,02 | 2,22 | 2,82 | -1,91 |
| BETO | EfficientNetB0 | 7,65 | 2,02 | 2,56 | -1,32 |
| BETO | Xception | 8,58 | 2,13 | 2,77 | -1,86 |
| FastText | - | 8,75 | 2,21 | 2,77 | -1,82 |
| FastText | ResNet50 | 8,88 | 2,18 | 2,79 | -1,84 |
| FastText | InceptionV3 | 8,99 | 2,24 | 2,81 | -1,90 |
| FastText | EfficientNetB0 | 8,60 | 2,20 | 2,71 | -1,58 |
| FastText | Xception | 9,17 | 2,22 | 2,84 | -1,93 |
| Keras | - | 8,49 | 2,14 | 2,74 | -1,78 |
| Keras | ResNet50 | 8,97 | 2,18 | 2,81 | -1,89 |
| Keras | InceptionV3 | 8,80 | 2,15 | 2,79 | -1,87 |
| Keras | EfficientNetB0 | 7,72 | 2,04 | 2,64 | -1,66 |
| Keras | Xception | 8,67 | 2,17 | 2,77 | -1,84 |

### 3.3 Predicting Flexibility (FLE)

Table 3 provides the results for predicting the categorical variable FLE. Metrics such as ROC AUC, accuracy, recall, precision, accuracy, specificity, and F1 score are reported for each model configuration.

Based on ROC AUC, the best performing model is FastText without a convolutional neural network (CNN), with an AUC value of 0.91. This model also achieves the best accuracy (0.56) and F1 score (0.66), maintaining a reasonable balance between recall (0.37) and precision (0.80). In contrast, the highest recall (0.53) is observed in the Keras-InceptionV3 model, which achieves an F1 score of 0.65 but shows lower performance in terms of ROC AUC (0.79).

Regarding precision, BETO without a CNN yields the highest score (1.00), though it has low recall (0.10), suggesting that it correctly identifies positive instances when detected but misses many positive cases overall. On the other hand, combinations involving ResNet50 and EfficientNetB0 show lower recall and precision, indicating underperformance in comparison to other model combinations.

Models involving embeddings, particularly BETO and FastText, tend to exhibit more consistent performance across several metrics, although with varying degrees of success depending on the metric of focus.

### 3.4 Predicting Title Originality (T)

Table 4 presents the evaluation results for predicting the binary variable T. The metrics shown include ROC AUC, accuracy, recall, specificity, and F1 score.

The BETO embedding without a CNN stands out as the best-performing model across most metrics. It achieves the highest ROC AUC (0.91), the highest accuracy (0.90), and the highest F1 score (0.90). The recall of this model is also relatively high (0.80), with a balanced specificity of 0.74.

Any of the embeddings combined with CNNs perform worse when compared to the performance, in terms of ROC AUC or accuracy, achieved without using images. Despite this, the recall for many of these models remains relatively high, with some models, such as Keras with ResNet50 and EfficientNetB0, achieving perfect recall (1.00). However, these same models exhibit very low specificity, indicating a tendency to over-predict positive instances.

Results generally indicate that the combination of text-based embeddings such as BETO and FastText, especially without an image model, performs well in T prediction, showing a balanced trade-off between sensitivity and specificity.

Table 3: Results Predicting Flexibility (FLE).

| Embedding | CNN | ROC AUC | Accuracy | Recall | Precision | Specificity | F1 score |
|---|---|---|---|---|---|---|---|
| - | ResNet50 | 0,86 | 0,54 | 0,43 | 0,70 | 0,00 | 0,65 |
| - | InceptionV3 | 0,89 | 0,54 | 0,50 | 0,64 | 0,00 | 0,65 |
| - | EfficientNetB0 | 0,81 | 0,26 | 0,09 | 0,60 | 0,00 | 0,21 |
| - | Xception | 0,83 | 0,54 | 0,50 | 0,65 | 0,00 | 0,65 |
| BETO | - | 0,86 | 0,40 | 0,10 | **1,00** | 0,00 | 0,42 |
| BETO | ResNet50 | 0,84 | 0,50 | 0,39 | 0,79 | 0,00 | 0,62 |
| BETO | InceptionV3 | 0,83 | 0,52 | 0,50 | 0,63 | 0,00 | 0,63 |
| BETO | EfficientNetB0 | 0,78 | 0,35 | 0,28 | 0,51 | 0,00 | 0,50 |
| BETO | Xception | 0,89 | 0,52 | 0,48 | 0,81 | 0,00 | 0,63 |
| **FastText** | **-** | **0,91** | **0.56** | 0,37 | 0,80 | 0,00 | **0,66** |
| FastText | ResNet50 | 0,82 | 0,48 | 0,42 | 0,62 | 0,00 | 0,60 |
| FastText | InceptionV3 | 0,80 | 0,51 | 0,46 | 0,61 | 0,00 | 0,62 |
| FastText | EfficientNetB0 | 0,80 | 0,36 | 0,11 | 0,92 | 0,00 | 0,44 |
| FastText | Xception | 0,82 | 0,49 | 0,46 | 0,71 | 0,00 | 0,61 |
| Keras | - | 0,80 | 0,53 | 0,38 | 0,86 | 0,00 | 0,64 |
| Keras | ResNet50 | 0,82 | 0.50 | 0,48 | 0,60 | 0,00 | 0,62 |
| Keras | InceptionV3 | 0,79 | 0,54 | **0,53** | 0,64 | 0,00 | 0,65 |
| Keras | EfficientNetB0 | 0,86 | 0,30 | 0,07 | **1,00** | 0,00 | 0,20 |
| Keras | Xception | 0,81 | 0.51 | 0,50 | 0,61 | 0,00 | 0,62 |

# 4 DISCUSSION

The primary goal of this research was to explore whether using textual descriptions of drawings, in addition to images, can improve the prediction of creativity-related characteristics such as originality (O), thematic flexibility (FLE), elaboration (E), and title creativity (T). Unlike previous studies that have focused exclusively on image analysis to assess these aspects, our research introduces text as an additional (or even primary) source of information. The results obtained allow us to reflect on whether text alone is sufficient and whether combining it with images provides significant added value.

A key finding is that models based solely on text were surprisingly competitive in predicting the originality of the drawing (O). For instance, the FastText without CNN model achieved a recall of 0.97, indicating that textual descriptions have great potential in capturing whether a drawing is original or not. This is a significant result, given that previous studies have relied solely on images, which may have limited the detection of more abstract aspects of originality.

However, when observing other metrics such as precision and specificity, models that combine both text and images (such as FastText + InceptionV3) showed improvements by reducing false positives. This suggests that while text alone provides valuable information, combining both data types allows for a more balanced and accurate prediction.

The prediction of thematic flexibility (FLE) showed that text alone is not only sufficient but, in many cases, the most effective data source. Text-only models, such as FastText without CNN, achieved the highest scores in ROC AUC and F1 score, outperforming models based solely on images. This indicates that textual descriptions of drawings capture the variety of themes represented well, an aspect that appears to be more abstract and conceptual and may escape purely visual evaluation.

The fact that images do not significantly improve FLE prediction suggests that thematic categories are more easily expressed and understood through language than by observing the visual details of the drawing.

In the prediction of elaboration (E), images proved to be clearly superior to text. Models relying exclusively on visual data (such as InceptionV3) achieved better results in terms of MAE, RMSE, and R2 score, indicating that the visual details of the drawing are essential for assessing its level of elaboration. Text-based models, or combinations of text and images, were unable to effectively capture the visual nuances related to the complexity of the drawing.

This finding suggests that for characteristics like elaboration, which depend on the direct perception of visual details, the text does not provide sufficient information and may introduce noise into the analysis. Consequently, image analysis becomes crucial to ac-

Table 4: Results Predicting Title Originality (T).

| Embedding | CNN | ROC AUC | Accuracy | Recall | Precision | Specificity | F1 score |
|---|---|---|---|---|---|---|---|
| - | ResNet50 | 0,57 | 0,45 | **1,00** | 0,54 | 0,00 | 0,70 |
| - | InceptionV3 | 0,50 | 0,41 | 0,86 | 0,54 | 0,02 | 0,66 |
| - | EfficientNetB0 | 0,54 | 0,50 | 0,95 | 0,57 | 0,10 | 0,71 |
| - | Xception | 0,62 | 0,61 | 0,59 | 0,78 | 0,47 | 0,67 |
| **BETO** | **-** | **0,91** | **0,90** | 0,80 | **1,00** | **0,74** | **0,90** |
| BETO | ResNet50 | 0,83 | 0,83 | 0,73 | 0,94 | 0,68 | 0,82 |
| BETO | InceptionV3 | 0,78 | 0,83 | 0,93 | 0,63 | 0,26 | 0,75 |
| BETO | EfficientNetB0 | 0,79 | 0,64 | 0,93 | 0,64 | 0,31 | 0,76 |
| BETO | Xception | 0,78 | 0,76 | 0,61 | 0,82 | 0,64 | 0,70 |
| FastText | - | 0,87 | 0,80 | 0,73 | 0,73 | 0,65 | 0,73 |
| FastText | ResNet50 | 0,55 | 0,50 | 0,98 | 0,56 | 0,08 | 0,71 |
| FastText | InceptionV3 | 0,69 | 0,65 | 0,57 | 0,68 | 0,53 | 0,62 |
| FastText | EfficientNetB0 | 0,51 | 0,49 | 0,98 | 0,56 | 0,08 | 0,71 |
| FastText | Xception | 0,73 | 0,74 | 0,66 | 0,92 | 0,61 | 0,77 |
| Keras | - | 0,87 | 0,78 | 0,82 | 0,80 | 0,57 | 0,81 |
| Keras | ResNet50 | 0,48 | 0,45 | **1,00** | 0,54 | 0,00 | 0,70 |
| Keras | InceptionV3 | 0,65 | 0,60 | 0,66 | 0,70 | 0,42 | 0,68 |
| Keras | EfficientNetB0 | 0,52 | 0,47 | **1,00** | 0,55 | 0,02 | 0,71 |
| Keras | Xception | 0,59 | 0,57 | 0,66 | 0,66 | 0,37 | 0,66 |

curately assess these characteristics.

The analysis of title creativity (T) showed that models based on text are the most effective tool for this task. Since title creativity is expressed exclusively through language, text-based models like BETO without CNN performed exceptionally well, achieving a ROC AUC of 0.91 and an F1 score of 0.90. In contrast, models based solely on images were ineffective, highlighting the irrelevance of visual data for this prediction.

## 5 CONCLUSION

A key contribution of this research is the demonstration that text not only provides relevant information but can, in some cases, be more informative than images in predicting certain aspects of creativity. In previous research, the focus has mainly been on images, overlooking the informative potential of textual descriptions. Our results reveal that:

- For features such as originality (O) and thematic flexibility (FLE), text alone is a very valuable source, and it may be even more suitable than pictures for capturing abstract concepts.
- For title creativity (T), text is the only relevant data source, as title creativity cannot be evaluated through images.
- For elaboration (E), images remain the best option, as this aspect depends more on the direct perception of visual details.

The combination of text and images only proved advantageous in some cases, particularly for reducing false positives in the prediction of O. However, in most cases, text alone was sufficient or even more effective than images.

This study demonstrates that incorporating text as a data source in the evaluation of creativity in drawings provides significant value, especially in predicting abstract characteristics such as originality and thematic flexibility. While images remain crucial for visual traits like elaboration, researchers should seriously consider using text in future studies, as it offers a complementary, and in some cases, more powerful perspective for capturing creativity.

## ACKNOWLEDGEMENTS

We would like to thank the Deustek5 group of the University of Deusto who have made this research possible.

## REFERENCES

Acar, S., Organisciak, P., and Dumas, D. (2024). Automated scoring of figural tests of creativity with computer vision. *Journal of Creative Behavior*. Cited by: 2.

Allen, T. E., Chen, M., Goldsmith, J., Mattei, N., Popova, A., Regenwetter, M., Rossi, F., and Zwilling, C.

(2015). *Beyond Theory and Data in Preference Modeling: Bringing Humans into the Loop*, page 3–18. Springer International Publishing.

Bahcecik, S. O. (2023). I trends security politics and artificial intelligence: Key trends and debates. *International Political Science Abstracts*, 73(3):329–338.

Beaty, R. E. and Johnson, D. R. (2020). Automating creativity assessment with semdis: An open platform for computing semantic distance. *Behavior Research Methods*, 53(2):757–780.

Cañete, J., Chaperon, G., Fuentes, R., Ho, J.-H., Kang, H., and Pérez, J. (2020). Spanish pre-trained bert model and evaluation data. In *PML4DC at ICLR 2020*.

Cetinic, E. and She, J. (2021). Understanding and creating art with ai: Review and outlook.

Chollet, F. (2016). Xception: Deep learning with depthwise separable convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1800–1807.

Creely, E. and Blannin, J. (2023). The implications of generative ai for creative composition in higher education and initial teacher education. *ASCILITE Publications*, page 357–361.

Cropley, D. H. and Marrone, R. L. (2022). Automated scoring of figural creativity using a convolutional neural network. *Psychology of Aesthetics, Creativity, and the Arts*.

Davis, J. L., Shank, D. B., Love, T. P., Stefanik, C., and Wilson, A. (2022). *Gender Dynamics in Human-AI Role-Taking*, page 1–22. Emerald Publishing Limited.

Devedzic, V. (2020). Is this artificial intelligence? *Facta universitatis - series: Electronics and Energetics*, 33(4):499–529.

Easton, K., Potter, S., Bec, R., Bennion, M., Christensen, H., Grindell, C., Mirheidari, B., Weich, S., de Witte, L., Wolstenholme, D., and Hawley, M. S. (2019). A virtual agent to support individuals living with physical and mental comorbidities: Co-design and acceptability testing. *Journal of Medical Internet Research*, 21(5):e12996.

Ferrara, S. and Qunbar, S. (2022). Validity arguments for ai-based automated scores: Essay scoring as an illustration. *Journal of Educational Measurement*, 59(3):288–313.

Gado, S., Kempen, R., Lingelbach, K., and Bipp, T. (2021). Artificial intelligence in psychology: How can we enable psychology students to accept and use artificial intelligence? *Psychology Learning & Teaching*, 21(1):37–56.

Gigi, A. (2015). Human figure drawing (hfd) test is affected by cognitive style. *Clinical and Experimental Psychology*, 02.

Harré, M. S. and El-Tarifi, H. (2023). Testing game theory of mind models for artificial intelligence. *Games*, 15(1):1.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.

Imuta, K., Scarf, D., Pharo, H., and Hayne, H. (2013). Drawing a close to the use of human figure drawings as a projective measure of intelligence. *PLoS ONE*, 8.

Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H., and Mikolov, T. (2016). Fasttext.zip: Compressing text classification models. *arXiv preprint arXiv:1612.03651*.

Kvam, P. D., Sokratous, K., Fitch, A., and Hintze, A. (2023). Using artificial intelligence to fit, compare, evaluate, and discover computational models of decision behavior.

Lee, S. W., Kwak, D. S., Jung, I. S., Kwak, J. H., Park, J. H., Hong, S. M., Lee, C. B., Park, Y. S., Kim, D. S., Choi, W. H., and Ahn, Y. H. (2015). Partial androgen insensitivity syndrome presenting with gynecomastia. *Endocrinology and Metabolism*, 30(2):226.

Liu, J., Xue, Z., Vann, K. R., Shi, X., and Kutateladze, T. G. (2020). Protocol for biochemical analysis and structure determination of the zz domain of the e3 ubiquitin ligase herc2. *STAR Protocols*, 1.

O'Shea, K. and Nash, R. (2015). An introduction to convolutional neural networks.

Patterson, J. D., Barbot, B., Lloyd-Cox, J., and Beaty, R. E. (2023). Audra: An automated drawing assessment platform for evaluating creativity. *Behavior Research Methods*, 56(4):3619–3636.

Pezzulo, G., Parr, T., Cisek, P., Clark, A., and Friston, K. (2023). Generating meaning: Active inference and the scope and limits of passive ai.

Røed, R. K., Baugerud, G. A., Hassan, S. Z., Sabet, S. S., Salehi, P., Powell, M. B., Riegler, M. A., Halvorsen, P., and Johnson, M. S. (2023). Enhancing questioning skills through child avatar chatbot training with feedback. *Frontiers in Psychology*, 14.

Searle, J. R. (2018). *Minds, Brains and Programs*, page 18–40. Routledge.

Shaban-Nejad, A., Michalowski, M., Bianco, S., Brownstein, J. S., Buckeridge, D. L., and Davis, R. L. (2022). Applied artificial intelligence in healthcare: Listening to the winds of change in a post-covid-19 world. *Experimental Biology and Medicine*, 247(22):1969–1971.

Sheng, L., Yang, G., Pan, Q., Xia, C., and Zhao, L. (2019). Synthetic house-tree-person drawing test: A new method for screening anxiety in cancer patients. *Journal of Oncology*, 2019.

Stojnic, G., Gandhi, K., Yasuda, S., Lake, B. M., and Dillon, M. R. (2022). Commonsense psychology in human infants and machines.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision.

Tan, M. and Le, Q. V. (2020). Efficientnet: Rethinking model scaling for convolutional neural networks.

Tan, T., Rodriguez-Ruiz, A., Zhang, T., Xu, L., Beets-Tan, R. G. H., Shen, Y., Karssemeijer, N., Xu, J., Mann, R. M., and Bao, L. (2023). Multi-modal artificial intelligence for the combination of automated 3d breast ultrasound and mammograms in a population of women with predominantly dense breasts. *Insights into Imaging*, 14(1).

Wang, W., Kofler, L., Lindgren, C., Lobel, M., Murphy, A., Tong, Q., and Pickering, K. (2023). Ai for psycho-

metrics: Validating machine learning models in measuring emotional intelligence with eye-tracking techniques. *Journal of Intelligence*, 11(9):170.

Weidinger, L., Reinecke, M. G., and Haas, J. (2022). Artificial moral cognition: Learning from developmental psychology.

Zhang, R., Zeng, B., Yi, W., and Fan, Z. (2024). *Artificial Intelligence Painting: A New Efficient Tool and Skill for Art Therapy*.