CLOCK DRAWING TEST: FORM VS NO FORM CLASSIFICATION

Maria Jose Viveros Riquelme

ITAI 1378

Huston City College

December, 2025

INTRODUCTION

- The Clock Drawing Test is widely used to evaluate planning, memory, and visuospatial reasoning. Its interpretation, however, can be subjective. This project investigates whether a deep learning model can reliably classify CDT drawings by identifying whether they present a recognizable clock form, offering a more consistent and automated approach to assessment.

DATA SET

- Clock-drawing images were obtained from the **NHATS (National Health and Aging Trends Study)** public-use cognitive assessment dataset. CDT drawings were originally labeled clinically; for this project, labels were reframed into **form** and **no_form** based on geometric structure.

INITIAL CLINICAL APPROACH

The first version of the model tried to predict five different clinical categories from the Clock Drawing Test, moving from severely impaired drawings to completely normal clocks. In practice, this approach did not work well: the accuracy stayed low because many of the clinical levels look almost identical on the page, and our dataset was too small for the network to learn such fine-grained differences. After reviewing the errors and the literature, I shifted from a purely clinical perspective to a more practical, visual one: instead of asking the model to infer cognitive status, I asked it to decide whether the drawing has a recognizable clock form or no form at all. **This shape–based** criterion fits much better with how convolutional networks actually learn, because it relies on clear visual cues like edges, symmetry, and overall structure rather than subtle clinical nuances.

WHY CLINICAL CATEGORIES FAILED

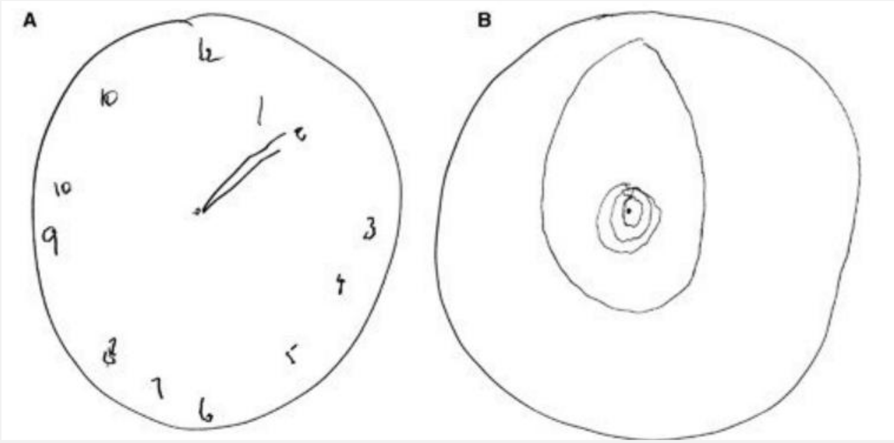
- Clinical scoring depends on cognitive interpretation, not purely visual patterns.
- Small datasets cannot support fine-grained distinctions.

BINARY CLINICAL PASS/FAIL

A simplified pass/fail model still produced low accuracy.
Visual similarity between categories limited model performance.

FINAL APPROACH: FORM VS NO FORM

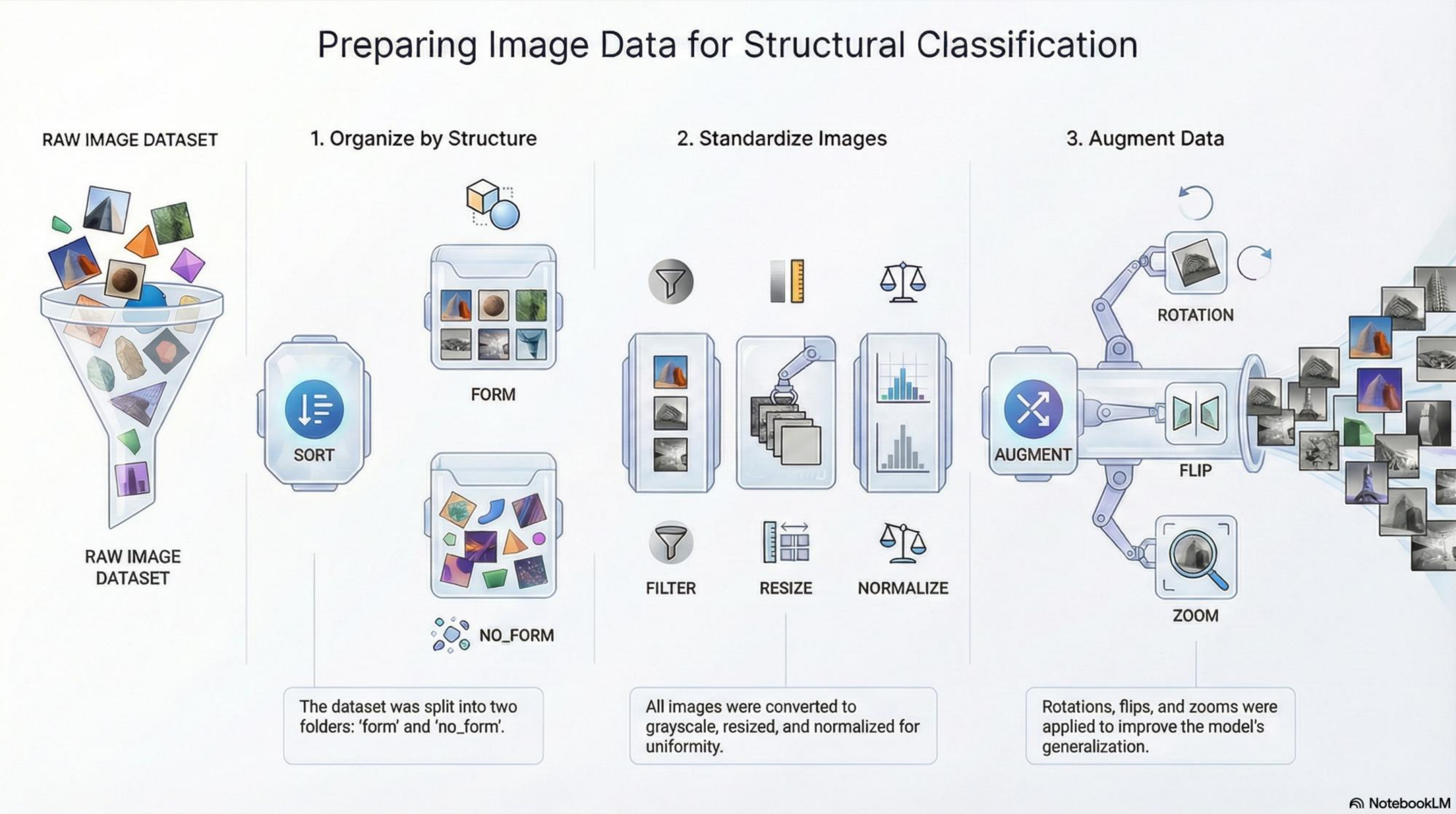
- A review of prior studies suggests that clock-drawing assessment is more reliable when analyzed through geometric and structural features rather than subjective clinical labels. CNN architectures are optimized for extracting hierarchical spatial features, making them ideal for tasks that rely on contour, symmetry, and shape-based pattern recognition.



FORM NO FORM

DATASET AND PREPROCESSING

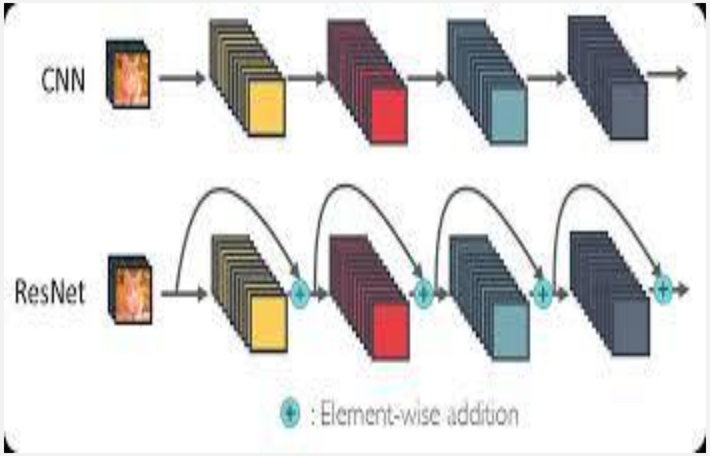
- The dataset was reorganized into two folders, **form and no form**, to reflect the structural criterion used in the final model.
- **This separation allowed the network to learn from visually consistent groups rather than relying on subtle clinical distinctions.**
- Before training, all images were converted to grayscale, resized to a standard resolution, and normalized to ensure uniformity across samples.
- Additional preprocessing steps included controlled data augmentation—such as slight rotations, flips, and zoom—to improve the model's ability to generalize and to compensate for the small size of the dataset.



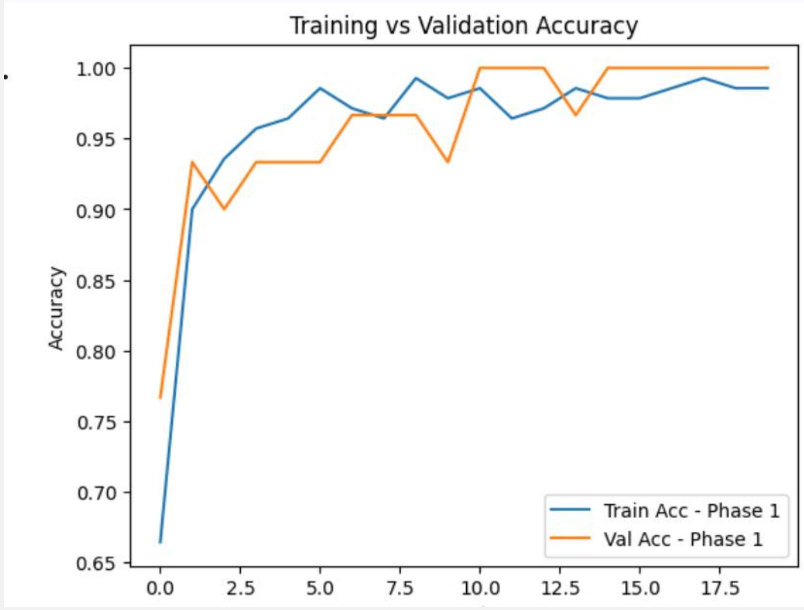
Preparing Image Data for Structural Classification
RAW IMAGE DATASET
1. Organize by Structure
2. Standardize Images
3. Augment Data
RAW IMAGE DATASET
SORT
FORM
NO_FORM
FILTER
RESIZE
NORMALIZE
AUGMENT
ROTATION
FLIP
ZOOM
The dataset was split into two folders: 'form' and 'no_form'.
All images were converted to grayscale, resized, and normalized for uniformity.
Rotations, flips, and zooms were applied to improve the model's generalization.
NotebookLM

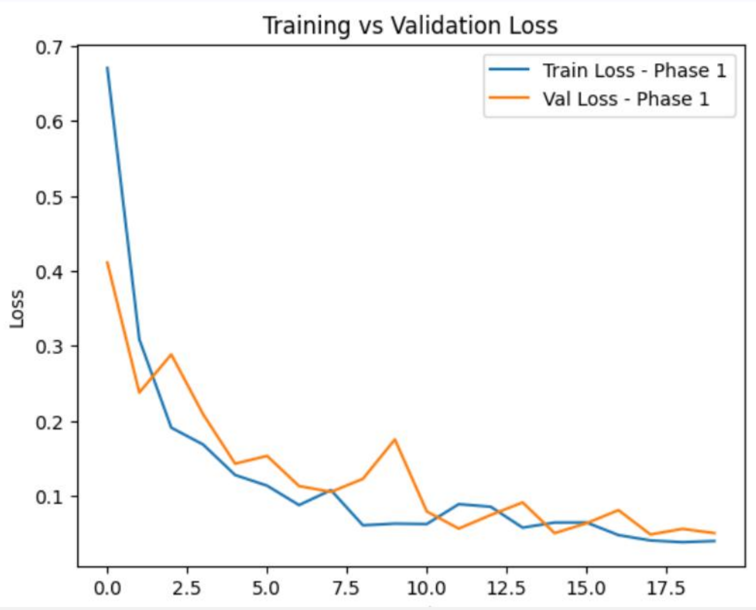
MODEL ARCHITECTURE

- The model was built on top of **a ResNet50 backbone** pretrained on ImageNet, but the backbone was kept frozen during training. This allowed the network to reuse strong geometric and edge-detection features without overfitting to the small CDT dataset. On top of the backbone, a custom dense classifier was added to learn the final decision boundary for the form versus no_form task. This architecture was intentionally chosen to balance expressive power with stability, ensuring that the model could generalize well even with limited data.



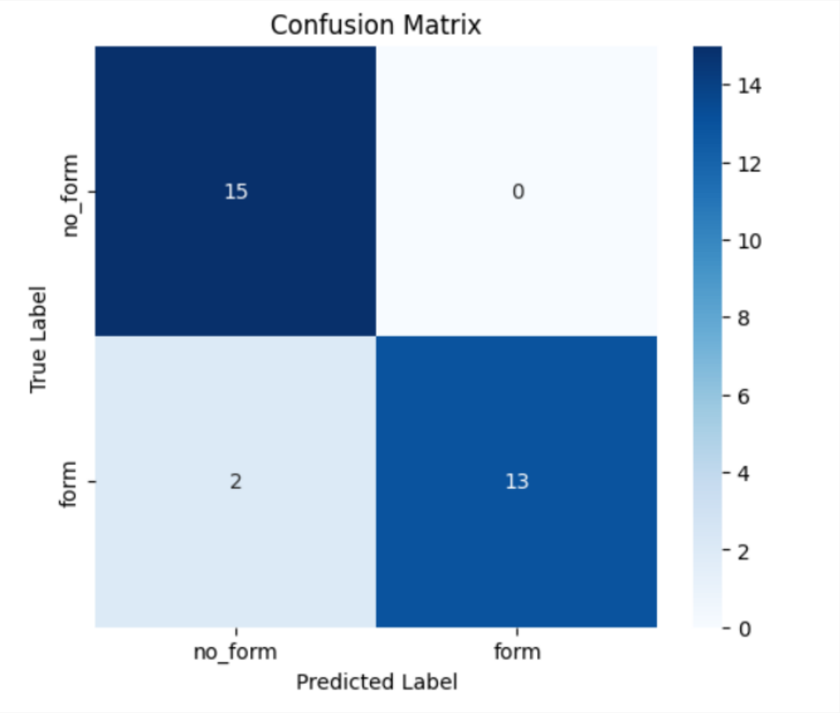
TRAINING PERFORMANCE





- The training and validation curves show nearly parallel trends, with both accuracy lines increasing steadily and both loss lines decreasing consistently. This indicates that the model is learning meaningful patterns without memorizing the data. The validation accuracy reaches approximately 0.98, demonstrating strong generalization even with a small dataset.
- No signs of overfitting were detected, as the validation loss does not rise and remains close to the training loss throughout training. These results confirm that the ResNet50-based architecture, combined with regularization and callbacks, was effective and stable for this classification task.

TEST RESULTS



- The model generalized well to new CDT images.
- Structural differences were consistently detected
- Accuracy 0.93

MODEL INTERPRETATION

- **The model learned to recognize edges, symmetry, circularity, and spatial cues**, which means it started focusing on the same structural features that clinicians examine in the Clock Drawing Test. For instance, detecting continuous edges and balanced shapes allows the network to differentiate between organized drawings and those where the visual structure breaks down. Likewise, identifying circularity helps the model determine whether the figure preserves the basic geometry of a clock or shifts into irregular, fragmented forms often associated with cognitive decline.

CONCLUSION AND FUTURE WORK

- **The form vs no_form approach is stable and effective** because it focuses on clear visual structure, letting the model learn consistent cues like shape, symmetry, and spatial organization. This reduces ambiguity and keeps performance reliable even with a small dataset.
- **Future work may add stroke-sequence data or expand the dataset**, allowing the model to capture planning behavior and improve generalization across more drawing styles.