Can we read the "age" of the brain from how someone writes and draws?

Lixia Chen [1] **Hazel Lim** [1]

Abstract

Brain age prediction from neuroimaging has emerged as a biomarker for neurological health, but MRI-based approaches remain expensive and inaccessible for widespread screening. We investigate whether age classification from handwriting and drawing can serve as a behavioral proxy for brain age assessment. We compiled a dataset of 3,672 handwriting and drawing samples across six age classes spanning preschool (3-6 years) to seniors (65+ years) from neurotypical individuals. We evaluated multiple deep learning architectures including ResNet-18, ResNet-50, Vision Transformer, and a feature extraction-based MLP operating on 63 handcrafted motor control features. The feature-based MLP achieved 86.3% accuracy, outperforming image-based models, demonstrating that stroke-level characteristics such as pressure dynamics, curvature smoothness, and structural topology provide age-discriminative information. Applying our model to clinical populations revealed interesting motor age gaps: young adults with Parkinson's disease were predicted as 20-40 years older, while children with autism spectrum disorder and conduct disorder were predicted as 4-6 years younger than their chronological ages. These bidirectional deviations validate motor age gaps as quantitative biomarkers of neurological dysfunction, offering a low-cost, accessible complement to neuroimaging for early detection and disease monitoring.

1. Introduction

Neurological disorders often progress silently for years before clinical diagnosis. Motor and cognitive changes accumulate over years, yet many patients only receive a diagnosis once symptoms are disruptive to daily life. By that point, however, structural and functional alterations in the brain are usually well established and difficult to reverse (McDonald, 2017). Early detection is therefore a central challenge in neurology, particularly for conditions such as Parkinson's disease, where subtle changes in movement and fine motor control can precede a formal diagnosis by several years.

Recent studies show that brain aging can provide one way of rethinking about this problem. Aging of the brain is a complex, multifaceted process involving loss of neural circuits and plasticity, along with gray and white matter changes, and altered metabolism in key cortical and subcortical regions (Kim et al., 2025; Thal et al., 2004). These age-related changes increase vulnerability to neurodegenerative diseases, including Alzheimer's disease, Parkinson's disease, and Huntington's disease (Mungas et al., 2002; Yan et al., 2024; Horvath et al., 2016). In response, more work has been done to analyze structural magnetic resonance images (MRI) using machine learning models to estimate "brain age" (Peng et al., 2021; Franke & Gaser, 2019). The difference between predicted brain age and chronological age, often called the brain age gap, has been proposed as a biomarker of brain health. A positive brain age gap, where the brain appears older than expected, has been associated with higher mortality and increased risk of neurodegenerative disease in several studies (Franke & Gaser, 2012; Gaser et al., 2013). Convolutional neural networks (CNNs) have demonstrated strong performance in predicting brain age gap, and therefore corresponding neurological disorders, from MRI scans (Lee et al., 2022; Kim et al., 2025).

Despite this promise, MRI-based biomarkers have practical and methodological limitations. MRI is expensive, time-consuming, and unevenly available across clinical settings, which makes routine screening difficult at scale. Deep learning models for 3D neuroimaging also require large, carefully curated datasets and high computational resources, which are not always feasible in everyday clinical environments. As a result, even though brain age models based on MRI are informative, they are not yet positioned for widespread early screening in the general population or in community clinics.

Parkinson's disease (PD) suggests a complementary path. It is characterized by motor symptoms that often involve handwriting and drawing: micrographia, tremor, and changes in speed and fluency of pen strokes (Thomas et al., 2017). Existing work has used handwriting and drawing tasks to classify Parkinson's versus healthy controls (Pereira et al., 2016; Taleb et al., 2017; Ngo et al., 2024; Pradeep & Kamalakannan, 2025), and there are also studies that predict

age or gender from handwriting and sketches (Alaei & Alaei, 2023; Singla et al., 2024). However, prior age-prediction work on handwriting typically focuses on limited developmental ranges (no comprehensive framework exists for predicting age across the full developmental spectrum—from preschool to seniors—using handwriting and drawing data) and achieves modest accuracy (averaging 75%), and most Parkinson's handwriting studies aim at diagnosis, not at estimating a continuous measure such as brain age or brain age gap. This limits the range of observations possibly made from looking at detail-rich data like handwriting and drawings. Moreover, studies that use MRI scans to make brain age gap estimations focus on positive delta (i.e., a brain that looks older than the actual age of the participant) and overlook the possible effect of negative delta.

In this project, we ask whether the brain age framework can be extended beyond neuroimaging to behavioral data that are closely tied to motor function. Specifically, we investigate whether deep networks trained on handwriting and drawing can predict age and, in turn, provide a proxy for brain age gap. Age is reflected in the maturation and decline of fine motor skills, so age-related patterns may be encoded in stroke dynamics, shape complexity, and spatial organization in drawings and written text. We divide age into six main developmental stages, from preschool (3-6 years) to seniors (65+ years). We first train deep learning models on handwriting and children's drawings from healthy participants, using chronological age as the ground truth. We then use these models to estimate brain age gaps in participants with Parkinson's disease and examine whether this behavioral brain age gap differs from that of age-matched healthy controls. We extend our observations and test our models on children with severe Autism spectrum disorder (ASD) and conduct disorder (CD) to analyze how our models predict disorders commonly seen in the younger population.

Our long-term aim is to explore whether such a behavioral brain age gap can serve as a low-cost, accessible early marker of neurological risk. If a simple drawing or handwriting task completed on a tablet can flag individuals whose "brain age" appears older than expected, clinicians could prioritize those individuals for detailed neurological evaluation, rather than relying solely on MRI or waiting for overt symptoms to emerge. Beyond Parkinson's disease, we are also interested in extending this approach to developmental and mental health conditions that affect motor and visuospatial function in childhood, such as autism spectrum disorder co-occurring with conduct disorder, where subtle differences in drawing and writing may appear early but are difficult to quantify by eye.

2. Related Work

2.1. Brain age prediction from neuro-imaging data

The study of brain age prediction initially focused on structural MRI and PET, where early deep learning models showed that a network trained on large neuroimaging cohorts can estimate chronological age and that the gap between predicted and actual age relates to cognitive decline and dementia risk (Mungas et al., 2002; Franke & Gaser, 2012; Gaser et al., 2013; Franke & Gaser, 2019). Later work improved this framework in two main ways.

First, some studies focused on clinical realism: they trained on research-grade 3D MRI and then adapted to routine 2D clinical scans, showing that brain age gap on these scans increases in Alzheimer's disease and Parkinson's disease and tracks disease progression (Taleb et al., 2017; Thomas et al., 2017; Gonzalez-Ortiz et al., 2023; Yan et al., 2024). Second, lightweight 3D convolutional networks were introduced to reduce parameter count and memory usage while preserving accuracy, through architectural simplification and training tricks such as data augmentation and model ensembling (Peng et al., 2021; Lee et al., 2022; Soumya Kumari & Sundarrajan, 2024; Kim et al., 2025).

Together, these studies established the brain age gap as a meaningful imaging biomarker. However, they remain tied to MRI acquisition, large curated datasets, and substantial computing resources, which limits their use as a frequent, low-cost screening tool.

2.2. Handwriting and drawing as clinical signals

In parallel, handwriting and drawing have been studied as behavioral markers of neurological and psychiatric conditions. For Parkinson's disease, early systems used digital tablets to record timing, position, and pressure during standardized writing or drawing tasks, extracted kinematic and spatio-temporal features, and trained classical classifiers such as SVMs. These models achieved high sensitivity and specificity for distinguishing early PD from age-matched controls (Pereira et al., 2016; Thomas et al., 2017; Taleb et al., 2017). More recent work applies deep convolutional networks and hybrid architectures directly to handwriting images, again with strong performance for PD detection (Ngo et al., 2024; Pradeep & Kamalakannan, 2025).

Other studies use similar ideas for different outcomes. Age estimation from handwriting has been framed as a classification problem, with deep residual networks and attention mechanisms achieving around 80% accuracy on standard datasets (Alaei & Alaei, 2023; Singla et al., 2024; Zhao et al., 2024). Clock drawing tests for dementia have been modeled with semi-supervised generative networks that learn latent factors describing clock shape and construction, which then separate dementia from non-dementia (Bandyopad-

hyay et al., 2023). Handwriting has also been explored as a marker of mood state in bipolar disorder and as a diagnostic signal for schizophrenia using transfer learning with CNNs (AYAZ et al., 2022; Castro et al., 2024).

These results show that fine motor behavior, captured through writing and drawing, carries robust information about brain function. Most of this work, however, focuses on disease-versus-control classification or discrete age categories, rather than on a continuous notion of brain age or brain age gap.

2.3. Gaps in current research and the positioning of our work

Current MRI-based brain age models give a principled way to quantify deviation from healthy aging, but they depend on expensive imaging and high-resource infrastructure. Handwriting and drawing models, in contrast, are cheap and accessible, yet they typically ignore the brain age framework and do not attempt to estimate an age gap that could generalize across conditions or developmental stages.

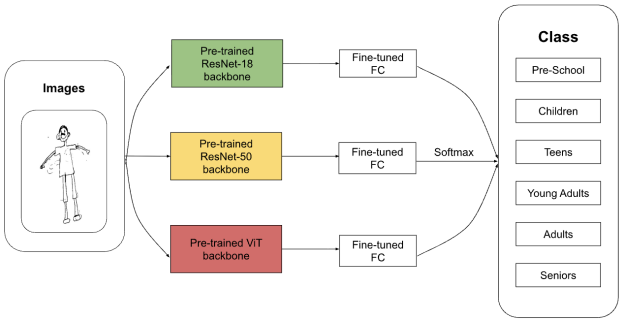
Our study is designed to sit between these two lines of work. We extend the brain age gap concept from neuroimaging to behavioral data by training models on handwriting and drawings from healthy participants and asking whether the resulting "behavioral brain age gap" differs in Parkinson's disease. Methodologically, we explore both a Feature MLP that integrates features from handwriting and drawing, and image-based architectures such as ResNet and Vision Transformers. Our dataset is more heterogeneous than many prior handwriting studies, spanning multiple languages and drawing types, which allows us to probe brain health across a broader range of motor behaviors. We also use our models to explore negative delta brain age ap estimations. We believe that a negative delta might indicate possible developmental deficiencies in the younger developmental stages (e.g., a teen or child having a predicted brain age of a pre-school).

While our experiments focus on a limited number of neurological disorders (mainly PD and some ASD), we view this as a first step toward applying a similar behavioral brain age framework to common developmental and mental health conditions, such as attention-deficit/hyperactivity disorder (ADHD), where subtle motor and visuospatial differences may emerge early and are difficult to capture with imaging alone.

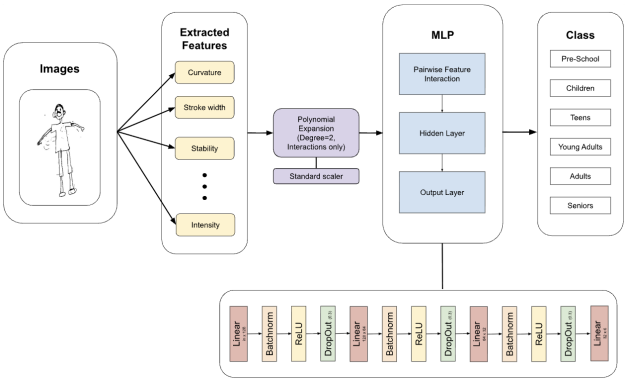
3. Method

3.1. Model Architectures

We evaluate two complementary model families for age prediction from handwriting and drawings: (i) image-based architectures that operate directly on rasterized images (ResNet-18, ResNet-50, Vision Transformer), and (ii) a feature-extraction–based multilayer perceptron (MLP) that operates on handcrafted stroke descriptors. An overview of these pipelines is shown in Fig. 1.



(A) Image-based architectures (ResNet-18/50, ViT)



(B) Feature-extraction–based MLP

Figure 1. Overview of the model architectures used in this study. Panel **A** illustrates the image-based pipelines (ResNet-18, ResNet-50, and ViT), where handwriting and drawing images are processed by deep convolutional or Transformer backbones followed by a classification head. Panel **B** shows the feature-extraction–based MLP, which operates on 63 handcrafted stroke features instead of raw pixel inputs.

Panel **A** shows the image-based architectures (ResNet-18, ResNet-50, and Vision Transformer), where raw handwriting or drawing images are first passed through a deep feature-extraction backbone. These backbones consist of stacked convolutional blocks (ResNets) or multi-head self-attention layers (ViT), which progressively transform low-level pixel information into hierarchical representations encoding stroke textures, shapes, spacing, and global structural patterns. After the backbone processing, each model feeds its final embedding into a task-specific classification layer that outputs the predicted age group.

Panel **B** presents the feature-extraction–based MLP, which focuses purely on stroke-level information. Instead of operating on raw pixel values, the model receives a hand-

crafted feature vector describing pressure intensity, curvature smoothness, stroke width variability, skeleton geometry, and histogram-based intensity distributions. These 63 descriptors are expanded through a second-order interaction-only polynomial transformation to model nonlinear relationships between stroke characteristics. The resulting feature vector is then passed through a three-layer multilayer perceptron with batch normalization, ReLU activation, and dropout, ultimately producing age-group predictions.

3.2. ResNet-18

ResNet-18 is a convolutional neural network architecture introduced by (He et al., 2015) as part of the Residual Network family. The architecture consists of 18 layers organized into four residual blocks, each containing two convolutional layers with batch normalization and ReLU activation functions. The key innovation of ResNet architectures lies in the introduction of skip connections, or identity mappings, that allow gradient information to flow directly through the network by bypassing one or more layers. These residual connections address the degradation problem observed in deep networks, where accuracy saturates and then degrades rapidly as network depth increases. Mathematically, instead of learning a direct mapping H(x), residual blocks learn the residual function F(x) = H(x) - x, which has been empirically shown to be easier to optimize.

The ResNet-18 architecture begins with a 7×7 convolutional layer followed by max pooling, then proceeds through four stages with [2, 2, 2, 2] residual blocks containing 64, 128, 256, and 512 filters respectively. Spatial dimensions are reduced through strided convolutions rather than pooling operations, maintaining computational efficiency while preserving important spatial information. The network concludes with global average pooling and a fully connected layer for classification. For our age classification task, we replaced the final fully connected layer to output predictions for six age groups. The relatively shallow depth of ResNet-18 compared to its deeper counterparts (ResNet-50, ResNet-101) makes it computationally efficient while maintaining strong representational capacity, particularly suitable for datasets of moderate size where overfitting concerns exist.

3.3. ResNet-50

ResNet-50 extends the residual learning framework to a deeper architecture consisting of 50 layers organized into four stages with [3, 4, 6, 3] residual blocks. Unlike ResNet-18, which uses basic residual blocks containing two convolutional layers, ResNet-50 employs bottleneck blocks to manage computational complexity while increasing network depth. Each bottleneck block consists of three convolutional layers: a 1×1 convolution that reduces dimensionality, a 3×3 convolution that performs spatial processing, and a 1×1 convolution that restores dimensionality. This bottleneck design reduces the number of parameters and computational cost compared to using only 3×3 convolutions throughout, enabling the construction of significantly deeper networks without prohibitive resource requirements.

The architecture progresses through four stages with filter sizes of 256, 512, 1024, and 2048, applying skip connections around each bottleneck block to facilitate gradient flow. Downsampling occurs through strided convolutions at the beginning of each stage except the first. The increased depth of ResNet-50 provides greater representational capacity compared to ResNet-18, potentially capturing more subtle and hierarchical features in the input data. This deeper architecture has demonstrated superior performance on complex image classification tasks, particularly when sufficient training data is available. However, the increased number of parameters (approximately 25 million compared to ResNet-18's 11 million) requires more computational resources and may be more prone to overfitting on smaller datasets. For our study, we employed ResNet-50 to investigate whether the additional depth and representational capacity would improve age classification accuracy from handwriting and drawing samples.

3.4. Vision Transformer (ViT)

The Vision Transformer represents a departure from convolutional architectures by adapting the Transformer architecture, originally designed for natural language processing, to computer vision tasks. Introduced by (Dosovitskiy et al., 2021), ViT divides an input image into fixed-size patches (typically 16×16 pixels), which are then linearly embedded and treated as sequences analogous to tokens in text processing. Each patch embedding is concatenated with a learnable position embedding that encodes spatial information, as the Transformer architecture itself is permutation-invariant and requires explicit positional encoding. A special learnable [CLS] token is prepended to the sequence, and its final hidden state serves as the image representation for classification tasks.

The core of the ViT architecture consists of multiple Transformer encoder blocks, each containing multi-head self-attention mechanisms and feed-forward neural networks with layer normalization and residual connections. The self-attention mechanism allows the model to weigh the importance of different image patches when processing each patch, enabling the network to capture long-range dependencies across the entire image without the inductive biases inherent in convolutional operations such as locality and translation equivariance. While this flexibility allows ViT to learn more general representations, it typically requires larger datasets for effective training compared to CNNs, as the model must learn spatial relationships from scratch

rather than leveraging built-in convolutional priors. For our age classification task, we employed a pre-trained ViT model fine-tuned on our handwriting and drawing dataset to investigate whether the global attention mechanisms would capture age-discriminative features that differ from those learned by convolutional architectures. The patch-based processing of ViT may be particularly relevant for handwriting analysis, where relationships between different regions of the image (such as letter spacing, line consistency, and overall composition) could provide important age-related cues.

3.5. Feature Extraction-based MLP

Deep convolutional architectures such as ResNet-18, ResNet-50, and Vision Transformer variants are effective at learning hierarchical visual representations, but they introduce a limitation for this task. These models tend to capture both object-level visual semantics (e.g., recognizable shapes or figures in drawings) and stroke-level graphical properties (e.g., line smoothness, curvature, or density). Because the goal of our study is to predict age-related patterns reflected in the construction of lines and strokes, rather than to classify the semantic content of the drawings, reliance on architectures that jointly encode object and stroke information may lead to unintended biases.

To ensure the model attends exclusively to the manner in which lines were drawn, we developed a feature-extraction–based multilayer perceptron (MLP). Instead of learning directly from raw images, the model operates on a set of 63 handcrafted stroke descriptors capturing pressure, curvature, stability, skeleton geometry, spacing, and histogram-based intensity patterns, encoded as a feature vector $\mathbf{x} \in \mathbb{R}^{63}$. This design constrains the classifier to learn purely from stroke behavior rather than from visual semantics of the drawings, thereby aligning the model's inductive bias with the goal of age prediction.

To model nonlinear relationships between these stroke descriptors without relying on deep convolutional layers, we apply a degree-2 polynomial interaction expansion, generating pairwise feature interactions of the form

$$\phi(\mathbf{x}) = \{\, x_i,\ x_i x_j \ \mid\ 1 \le i < j \le 63 \,\},$$

that capture higher-order dependencies (e.g., how curvature variability interacts with local contrast). The expanded feature dimensionality is given by $63 + \binom{63}{2} = 2016$. The expanded feature set is standardized using z-score normalization and then passed to a three-layer MLP with 128, 64, and 32 hidden units. Each hidden layer includes batch normalization, ReLU activation, and dropout (p = 0.3), which collectively improve training stability and reduce overfitting. A final linear layer maps the 32-dimensional representation to logits over the six age categories.

We evaluated multiple depths and hidden unit configurations; deeper networks tended to overfit, while smaller ones underfit, and the chosen configuration provided the most reliable performance. The model is trained using cross-entropy loss

$$\mathcal{L} = -\sum_{c=1}^{6} y_c \log p(y = c \mid \mathbf{x})$$

and the Adam optimizer with weight decay, and early stopping is applied based on validation accuracy. As demonstrated in Section 5, the feature-based MLP achieves strong performance (test accuracy ≈ 0.86) and outperforms all image-based architectures. These results demonstrate that a representation constrained to stroke-level graphical characteristics is better aligned with the age-classification objective and provides a more robust basis for subsequent analyses.

4. Experiments

4.1. Dataset Collection

We compiled a comprehensive multi-source dataset spanning the complete developmental and aging spectrum from pre-school to seniors. Our dataset aggregates handwriting and drawing samples from seven primary sources: the ASD Handwriting & Drawing dataset (Liaqat, 2024), Ecuador Children Drawings (GTS.AI, 2024), Draw-a-Person Sketches (Rakhmanov et al., 2020), Hebrew Handwriting Dataset (HHD) (Rabaev et al., 2020), HandPD (Pereira et al., 2016), BanglaWriting (Mridha et al., 2020), and additional samples collected from various online sources.

Table 1 provides an overview of the total images collected from each source and the age classes they cover. In total, we gathered 3,672 images across six age classes: Pre-school (3-6 years), Children (7-12 years), Teens (13-17 years), Young Adults (18-40 years), Adults (40-65 years), and Seniors (65+ years).

Table 1. Summary Statistics by Dataset Source

Dataset Source	Total Images	Age Classes Covered
ASD Handwriting & Drawing	291	Pre-school, Children
Ecuador Children Drawings	359	Pre-school
Draw-a-Person Sketches	914	Pre-school, Children
Hebrew Handwriting (HHD)	976	Teens, Young Adults, Adults
HandPD (control)	912	Teens, Young Adults, Adults, Seniors
BanglaWriting	64	Young Adults
Other Sources	156	Seniors
GRAND TOTAL	**3,672**	**All 6 age classes**

The detailed distribution of images across age classes for each dataset source is presented in Table 2. The distribution varies significantly across age groups, with the Adults class containing the most samples (972 images) and Seniors containing the fewest (212 images). Pre-school and Children classes contain 881 and 683 images respectively, while Teens and Young Adults classes include 447 and 477

images.

Table 2. Images Collected per Dataset Source by Age Class

Dataset Source	Pre-school (3-6)	Children (7-12)	Teens (13-17)	Young Adults (18-40)	Adults (40-65)	Seniors (65+)
ASD Handwriting & Drawing	118	173	0	0	0	0
Ecuador Children Drawings	359	0	0	0	0	0
Draw-a-Person Sketches	404	510	0	0	0	0
Hebrew Handwriting (HHD)	0	0	439	333	204	0
HandPD (control)	0	0	8	80	768	56
BanglaWriting	0	0	0	64	0	0
Other Sources	0	0	0	0	0	156
TOTAL	**881**	**683**	**447**	**477**	**972**	**212**

For model training and evaluation, we employed a standard split strategy: approximately 70% of data for training, 15% for validation, and 15% for testing. As shown in Table 3, this resulted in 2,579 training images, 531 validation images, and 562 test images across all age classes. The split was performed while maintaining class balance to ensure each age group was adequately represented in all three sets. Here we only train on healthy control's handwritings and drawings for dataset that included patients drawing, such as HandPD (Parkinson's dataset) and ASD Handwriting & Drawing (Autism spectrum and conduct disorder patients).

Table 3. Train/Validation/Test Split by Age Class

Split	Pre-school (3-6)	Children (7-12)	Teens (13-17)	Young Adults (18-40)	Adults (40-65)	Seniors (65+)
Training	615	478	312	325	676	173
Validation	130	101	66	75	140	19
Testing	136	104	69	77	156	20
TOTAL	**881**	**683**	**447**	**477**	**972**	**212**

4.2. Data Preprocessing

4.2.1. Line Segmentation

The HandPD dataset (Pereira et al., 2016) presented unique preprocessing challenges due to its data collection methodology. In this dataset, participants (both Parkinson's disease patients and healthy controls) were instructed to trace preprinted black reference lines on paper using blue ink pens. The resulting images contained both the original black template lines and the blue traced lines, which could confound age-related motor pattern analysis. To isolate the motor execution characteristics from the template, we developed a color-based extraction pipeline. First, we converted the RGB images to HSV color space to enable robust color-based segmentation. We then applied HSV thresholding to detect and isolate blue ink pixels (H: 90-140, S: 50-255, V: 50-255), effectively creating a binary mask that selectively captured the participant's traced strokes while excluding the black template lines. Morphological operations (opening followed by dilation with a 2×2 kernel) were applied to the mask to remove noise and ensure stroke continuity. The isolated blue ink strokes were then placed on a white background and converted to grayscale, where darker gray pixels indicate higher pen pressure (stronger ink deposition) and lighter gray pixels indicate lower pressure. This grayscale representation preserves pressure dynamics as intensity variations, which serve as critical features for characterizing motor control patterns. Figure 2 illustrates the complete preprocessing pipeline, showing the original image with both black template and blue tracing (left), the extracted blue mask (center), and the final grayscale result on white background (right). This preprocessing approach ensures that our models learn exclusively from the participant's motor execution patterns rather than from the printed template, thereby capturing age-discriminative features such as stroke stability, pressure consistency, and trajectory smoothness.

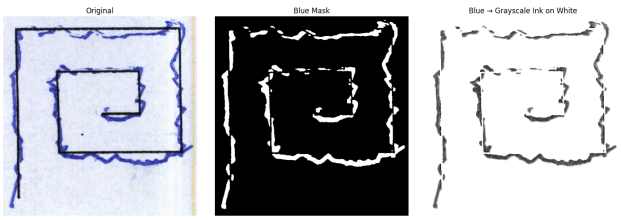


Figure 2. HandPD dataset preprocessing pipeline. **Left:** Original image showing black template lines and blue traced strokes. **Center:** Binary mask isolating blue ink pixels through HSV color space thresholding. **Right:** Final preprocessed output with blue strokes converted to grayscale on white background, where intensity variations encode pen pressure dynamics.

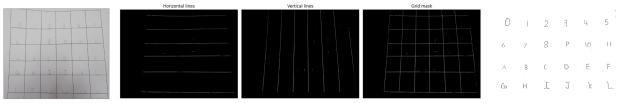


Figure 3. ASD Handwriting dataset preprocessing for grid removal. **Left:** Original handwriting sample with horizontal and vertical grid lines. **Center-left:** Horizontal line mask detected using morphological operations. **Center:** Vertical line mask. **Center-right:** Combined grid mask. **Right:** Final preprocessed image with grid lines removed and handwriting strokes preserved on white background.

Similarly, the ASD Handwriting & Drawing dataset (Liaqat, 2024) required preprocessing to remove interfering visual elements. The original handwriting samples were collected on lined paper with horizontal and vertical grid lines intended to guide writing. While these grids served a pedagogical purpose during data collection, they introduced systematic artifacts that could bias the model toward learning grid patterns rather than handwriting characteristics. To address this, we developed a grid removal pipeline using directional line detection and masking. As illustrated in Figure 3, we first detected horizontal lines by applying morphological operations with an elongated horizontal structuring element, creating a horizontal line mask. We repeated this process with a vertical structuring element to generate a vertical line mask. These two masks were then combined to produce a

comprehensive grid mask that captured all ruled lines in the image. Finally, we applied inpainting techniques to remove the detected grid lines while preserving the handwritten strokes, resulting in clean handwriting on a white background. This preprocessing ensures that the model focuses exclusively on stroke morphology, pressure variation, and spatial layout of the handwriting itself. For other datasets such as Ecuador Children Drawings (GTS.AI, 2024), Draw-a-Person Sketches (Rakhmanov et al., 2020), Hebrew Handwriting (Rabaev et al., 2020), and BanglaWriting (Mridha et al., 2020), we performed similar preprocessing operations tailored to each dataset's specific characteristics. These generally involved extracting black or gray strokes, placing them on uniform white backgrounds, and converting to grayscale format to ensure consistency across all datasets while preserving pressure-related intensity variations that are critical for motor pattern analysis.

4.2.2. Data Augmentation

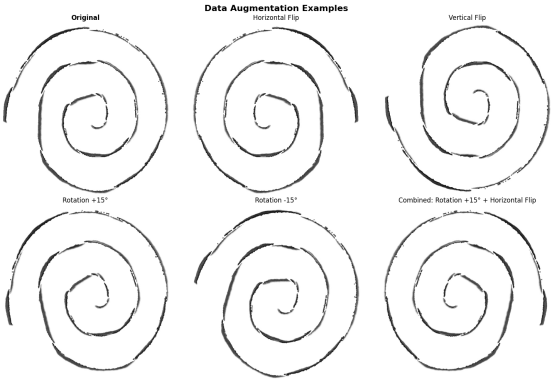


Figure 4. Data augmentation pipeline applied to Seniors, Adults, and Young Adults age classes. **Top row:** Original image (left), horizontal flip (center), and vertical flip (right). **Bottom row:** Rotation by +15° (left), rotation by -15° (center), and combined augmentation with rotation +15° followed by horizontal flip (right). All transformations preserve stroke characteristics and pressure dynamics while introducing geometric variation.

To address class imbalance and improve model robustness, we applied data augmentation techniques to age classes with limited sample sizes. Specifically, the Seniors (65+ years), Adults (40-65 years), and Young Adults (18-40 years) classes received augmentation to expand their training sets and enhance model generalization. Our augmentation pipeline consisted of three geometric transformations that preserve the fundamental motor characteristics of handwriting and drawing while introducing controlled variation. As illustrated in Figure 4, we applied horizontal flipping (mirroring the image along the vertical axis), vertical flipping (mirroring along the horizontal axis), and rotations within a range of ±15 degrees. These transformations simulate natural variations in paper orientation and writing angle without distorting the underlying stroke patterns, pressure dynamics, or spatial relationships that encode age-related motor control features. Importantly, we avoided photometric augmentations such as brightness adjustment, contrast modification, or noise injection, as these could alter the grayscale intensity values that encode pen pressure information, a critical feature for pattern analysis.

4.3. Feature Extraction

Our Features MLP operates on 63 handcrafted features extracted from preprocessed handwriting and drawing images. These features were designed to capture motor control characteristics across multiple dimensions: intensity patterns (encoding pen pressure), local texture (reflecting stroke quality), edge properties (measuring precision), geometric shape (describing trajectory planning), and skeletal topology (quantifying structural complexity). Table 4 provides a complete enumeration of all extracted features organized by category.

Table 4. Complete List of Handcrafted Features for Age Classification

Category	Feature Name	Count
2*Intensity Statistics	Mean stroke intensity, Std stroke intensity	2
2*Local Contrast	Mean local contrast, Std local contrast	2
2*Edge Energy	Mean gradient magnitude, Std gradient magnitude	2
Laplacian	Laplacian variance	1
Gabor Filters	4 orientations (0°, 45°, 90°, 135°), mean & std each	8
Intensity Histogram	32 bins (grayscale distribution)	32
2*Curvature	Mean curvature, Curvature variance	2
2*Stroke Width	Mean stroke width, Stroke width variance	2
Orientation Histogram	8 directional bins	8
Motion Stability	Shakiness (high-frequency angle changes)	1
Motion Dynamics	Jerk (rate of curvature change)	1
Structural Complexity	Number of connected components (segments)	1
Orientation Entropy	Directional pattern entropy	1
TOTAL		**63**

While all features contribute to the model's discriminative capacity through their polynomial interactions, several feature categories warrant detailed discussion due to their particularly strong biological interpretability and predictive importance. First, **intensity statistics and histogram features** (34 features total) capture pen pressure dynamics by measuring the distribution of grayscale values across the image. Darker pixels indicate higher pen pressure resulting from stronger grip force and greater motor control, while lighter pixels suggest reduced pressure application. Across the developmental spectrum, these features encode systematic patterns: young children typically produce lighter strokes due to developing fine motor skills and weaker grip strength; adults demonstrate consistent moderate-to-dark strokes reflecting mature motor control and stable pressure application; and seniors may exhibit either reduced pressure (due to decreased muscle strength) or highly variable pressure (due to tremor-induced fluctuations). The intensity histogram specifically captures the global pressure distribution, for

instance, adults typically show a concentration of pixels in the darker intensity ranges, whereas children show broader distributions skewed toward lighter values, and seniors may display bimodal distributions reflecting baseline lightness punctuated by tremor-induced pressure spikes.

Second, **Laplacian variance** quantifies the variance of the image's second spatial derivative, effectively measuring the density of fine-scale intensity fluctuations. This feature is exceptionally sensitive to micro-tremor and stroke stability. High Laplacian variance indicates rapid local intensity changes, which manifest as jagged or fuzzy stroke edges characteristic of motor instability. Young children exhibit moderate-to-high Laplacian variance due to unsteady hand movements and inconsistent pressure control during motor skill acquisition. Adults typically show low Laplacian variance, reflecting smooth, controlled strokes with minimal fluctuation. Seniors, particularly those with age-related tremor or early neurodegenerative changes, demonstrate elevated Laplacian variance as involuntary hand oscillations introduce high-frequency intensity variations along stroke edges.

Third, **jerk features** measure the rate of change of curvature along strokes, effectively quantifying the smoothness of motion trajectories. In biomechanics, jerk (the third derivative of position) is a fundamental measure of movement fluidity and motor planning efficiency. Low jerk indicates smooth, well-planned movements executed with minimal corrections, characteristic of mature motor control in adults. High jerk reflects abrupt directional changes and frequent trajectory adjustments, which occur during childhood motor development when planning and execution are still maturing, and in older adults when motor control deteriorates.

Fourth, **skeleton topology features** (5 features) analyze the structural properties of the thinned stroke skeleton, including the total number of skeleton points, endpoints (where strokes terminate), and junctions (where strokes branch or intersect). These features capture the complexity and connectivity of the drawing or writing.

Together, these four feature categories, intensity patterns, fine-scale stability (Laplacian variance), motion fluidity (jerk), and structural topology, span the full range of motor control aspects relevant to age discrimination: from the biomechanical (pressure and tremor) to the kinematic (smoothness) to the cognitive (planning complexity).

5. Results and Analysis

5.1. Model Performance

We evaluated four model architectures on the age classification task: ResNet-50, ResNet-18, Vision Transformer (ViT), and our proposed Features MLP. Performance was assessed using three metrics: overall accuracy, macro-averaged F1 score, and macro-averaged AUC (Area Under the ROC Curve). As illustrated in Figure 5, the Features MLP achieved the highest performance across all metrics, with an accuracy of 86.3%, macro-F1 of 87.3%, and macro-AUC of 98.2%. This superior performance demonstrates that features capturing stroke-level characteristics, such as pressure dynamics, curvature smoothness, and spatial geometry, provide more discriminative information for age classification than end-to-end learned representations from raw images.

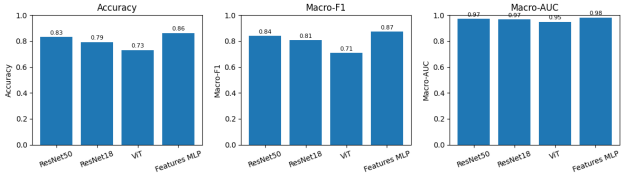


Figure 5. Model performance comparison across four architectures. **Left:** Overall accuracy on the test set. **Center:** Macro-averaged F1 score. **Right:** Macro-averaged AUC. The Features MLP outperforms all image-based models (ResNet-50, ResNet-18, ViT) across all metrics, achieving 86.3% accuracy, 87.3% macro-F1, and 98.2% macro-AUC.

Among the image-based architectures, ResNet-50 demonstrated the strongest performance with 83.1% accuracy, 83.9% macro-F1, and 97.1% macro-AUC, followed by ResNet-18 (79.4% accuracy, 80.7% macro-F1, 96.9% macro-AUC). The superior performance of ResNet-50 over ResNet-18 suggests that deeper architectures with greater representational capacity can capture more nuanced hierarchical patterns in handwriting and drawing samples, including subtle variations in stroke texture, spatial organization, and local features that correlate with aging. The Vision Transformer achieved the lowest performance among all models, with 73.1% accuracy, 71.0% macro-F1, and 94.8% macro-AUC. The model typically requires substantially larger training datasets to effectively learn spatial relationships through self-attention mechanisms, as it lacks the inductive biases (locality and translation equivariance) inherent in convolutional architectures. Given our dataset size of 3,672 images, ViT likely suffered from insufficient training data to fully leverage its global attention capabilities.

The confusion matrix for the Features MLP, shown in Figure 6, reveals detailed classification patterns across the six age groups. The model achieved perfect classification for Young Adults (18-40 years) with 100% accuracy, and strong performance for Adults (40-65 years) at 92.4% and Pre-school children (3-6 years) at 90.7%. These high accuracies suggest that these age classes exhibit distinctive signatures that captured by stroke-level features. The model also performed well on Seniors (65+ years) with 81.3% accuracy, though with some confusion with Pre-school chil-

dren (17.3% misclassification). This confusion pattern is interpretable: both age class have relatively lower control stability, irregular stroke trajectories, and variable pressure application, leading to overlapping feature representations. Children's developing motor skills and seniors' age-related motor decline may produce superficially similar handwriting characteristics despite fundamentally different underlying causes.

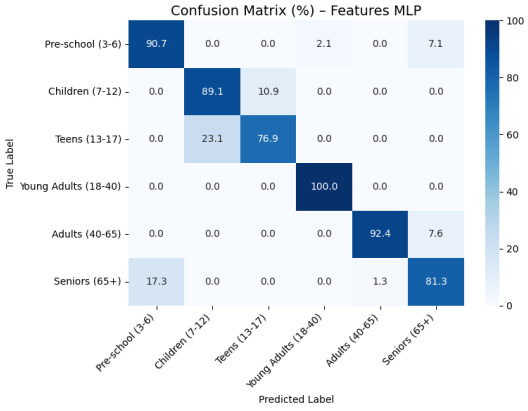


Figure 6. Confusion matrix for Features MLP showing classification accuracy (in percentages) across six age groups. Rows represent true age labels and columns represent predicted labels.

The Teens class (13-17 years) showed the most distributed confusion pattern, achieving 76.9% accuracy with notable misclassifications into Children (23.1%). This reflects the transitional nature of adolescence, where handwriting characteristics gradually shift from childhood patterns toward adult stability but retain substantial individual variation in motor maturity. The Children class (7-12 years) achieved 89.1% accuracy with some confusion with Teens (10.9%), which is consistent with the overlapping developmental trajectory of motor skill refinement during these years. Overall, the confusion matrix demonstrates that age-discriminative motor patterns exist across the lifespan, with the strongest distinctions in stable adult phases and expected overlap during transitional developmental periods and between the extremes of motor development and decline. These results validate our hypothesis that stroke-level features encode meaningful age-related motor control signatures and provide a foundation for extending this framework to detect deviations associated with neurological disorders.

5.2. Permutation Feature Importance

To interpret which handcrafted features most strongly influence age classification in our Features MLP model, we conducted permutation feature importance analysis. This technique measures the decrease in model accuracy when individual features are randomly shuffled, thereby breaking their relationship with the target labels. Features that cause substantial accuracy drops when permuted are considered more important for prediction. We performed this analysis on the polynomial-expanded feature space (2,346 dimensions derived from 68 original features through second-order interaction terms) using the test set, with 5 repetitions per feature to ensure stable importance estimates. The baseline model accuracy before any permutation was 86.3%.

Table 5 presents the top 20 most important features identified through this analysis. The most critical feature is the interaction between `orient_bin_5` (stroke orientation in the fifth angular bin) and `skel_points` (number of skeleton branch points), with an importance score of 0.0105, indicating that permuting this feature reduces accuracy by approximately 1.05 percentage points. This interaction captures the relationship between stroke directionality and structural complexity, which evolves systematically across age groups as motor planning and execution mature and decline. The second most important feature is the interaction between `hist_bin_1` and `hist_bin_30` (intensity histogram bins at opposite ends of the grayscale spectrum), with an importance of 0.0034. This feature encodes the relationship between very light and very dark pixels, effectively capturing pen pressure dynamics and stroke contrast patterns that reflect age-related changes in grip strength and pressure control.

Table 5. Top 20 Most Important Features from Permutation Importance Analysis

Rank	Feature Name	Importance
1	orient_bin_5 × skel_points	0.010546
2	hist_bin_1 × hist_bin_30	0.003390
3	gabor_5 × skel_junctions_ratio	0.002637
4	hist_bin_1 × hist_bin_29	0.002637
5	hist_bin_2 × hist_bin_27	0.002637
6	hist_bin_5 × hist_bin_26	0.002637
7	hist_bin_4 × skel_points	0.002260
8	hist_bin_3 × hist_bin_26	0.002260
9	hist_bin_2 × hist_bin_30	0.002260
10	lap_var × segments	0.002260
11	inten_std × hist_bin_28	0.002260
12	inten_std × hist_bin_29	0.002260
13	hist_bin_6 × hist_bin_25	0.002260
14	lap_var × orient_bin_6	0.002260
15	stroke_width_var × segments	0.001883
16	stroke_width_mean × segments	0.001883
17	segments × gabor_5	0.001883
18	jerk × segments	0.001883
19	stroke_width_mean × skel_points	0.001883
20	hist_bin_4 × hist_bin_26	0.001883

Several patterns emerge from the top-ranked features. First, histogram bin interactions dominate the list (appearing

in 12 of the top 20 features), highlighting that intensity distribution patterns, which encode pressure variation, stroke darkness, and background contrast—are fundamental discriminators of age-related handwriting characteristics. Second, skeleton-based features (`skel_points`, `skel_junctions_ratio`) appear in multiple top interactions, indicating that the topological complexity and branching structure of strokes provide age-discriminative information about planning and execution precision. Third, stroke geometry features including orientation bins (`orient_bin_5`, `orient_bin_6`), stroke width statistics (`stroke_width_mean`, `stroke_width_var`), and texture descriptors (Gabor filter responses) contribute substantially to classification performance. Fourth, dynamic features such as `jerk` (rate of change of acceleration) and `lap_var` (Laplacian variance, measuring edge sharpness) capture temporal and spatial smoothness properties that distinguish fluid adult writing from the more variable patterns of children and seniors.

Notably, the polynomial expansion strategy proved crucial, many of the most important features are second-order interaction terms rather than single features, suggesting that age-related patterns are encoded in relationships between multiple stroke characteristics rather than in individual measurements alone. For instance, the interaction between Laplacian variance and segment count (`lap_var` × `segments`) may capture how edge sharpness relates to stroke complexity, with younger children producing simpler but less sharp strokes, adults producing complex and sharp strokes, and seniors showing degraded sharpness despite maintained complexity. These interpretable feature importance patterns validate our feature engineering approach and provide insights into the biomechanical and neuromotor signatures that distinguish handwriting across developmental stages and aging.

6. Discussions

6.1. Predicting Motor Age Gap

To test whether motor-based age prediction reveals dysfunction in clinical populations, we applied our Features MLP (trained exclusively on neurotypical individuals) to handwriting from young adults with Parkinson's disease and children with Autism Spectrum Disorder co-occurring with Conduct Disorder.

6.1.1. Parkinson's Disease: Accelerated Motor Aging

Parkinson's disease (PD) produces characteristic handwriting abnormalities including micrographia, reduced stroke amplitude, and tremor-induced irregularities (Pereira et al., 2016; Taleb et al., 2017). We analyzed 412 handwriting samples from PD patients aged 18 to 40 years. As shown in Figure 7, the model predicted 61% as Adults (40-65) and 39% as Seniors (65+), with zero predictions in the correct Young Adults category. This systematic upward shift indicates a motor age gap of approximately 20 to 40 years.

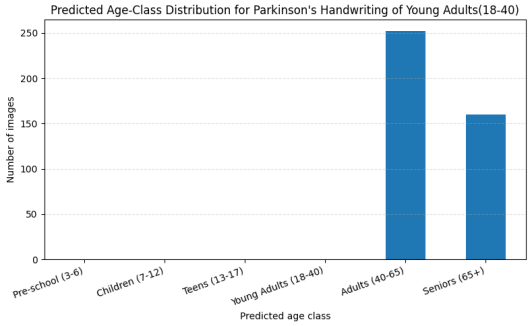


Figure 7. Predicted age-class distribution for handwriting samples from individuals with Parkinson's disease (chronological age 18 to 40 years). The model predominantly predicts older age classes: 61% as Adults (40-65) and 39% as Seniors (65+), indicating a motor age gap of approximately 20 to 40 years.

This upward prediction reflects the fundamental neuropathology of Parkinson's disease: progressive degeneration of dopaminergic neurons in the substantia nigra disrupts basal ganglia circuits responsible for motor control, resulting in motor dysfunction that resembles accelerated aging. The loss of dopamine impairs the basal ganglia's ability to regulate movement amplitude, speed, and smoothness, producing bradykinesia, rigidity, and tremor that mirror motor decline in natural aging but occur decades prematurely. Neuroimaging studies demonstrate that PD patients exhibit accelerated structural brain aging, with brain age gaps correlating with disease severity and progression (Yan et al., 2024; Kim et al., 2025). Our behavioral motor age gap provides a functional manifestation of this accelerated neurodegeneration: just as dopamine depletion causes premature aging of motor circuits, the resulting handwriting abnormalities produce signatures indistinguishable from those of individuals 20 to 40 years older. This convergence of structural and functional aging biomarkers validates the motor age gap as a meaningful measure of disease-related neural decline.

The upward prediction reflects specific PD motor features: tremor increases Laplacian variance and shakiness; bradykinesia reduces pen pressure mimicking age-related muscle weakness; micrographia creates abnormal stroke width patterns; and rigidity increases jerk values. These features converge to produce signatures the model associates with older individuals.

6.1.2. Autism Spectrum Disorder with Conduct Disorder: Delayed Motor Development

Autism Spectrum Disorder (ASD) is frequently accompanied by motor skill deficits, with fine motor control particularly affected due to disruptions in motor planning and sensorimotor integration (Liaqat, 2024). We analyzed 83 handwriting and drawing samples from children with severe ASD and CD aged 7 to 12 years. As shown in Figure 8, 94% were classified as Pre-school (3 to 6 years) and only 6% as the correct Children category, revealing a negative motor age gap of approximately 4 to 6 years.

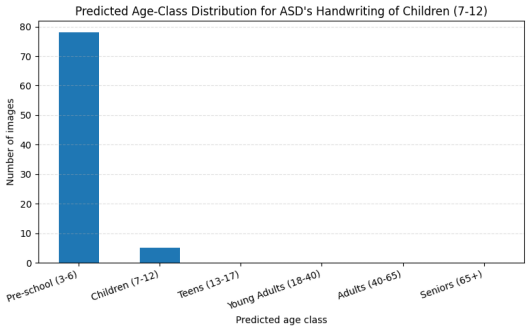


Figure 8. Predicted age-class distribution for handwriting and drawing samples from children with Autism Spectrum Disorder and Conduct Disorder (chronological age 7 to 12 years). The model overwhelmingly predicts Pre-school (3 to 6 years) for 94% of samples, indicating a motor age gap of approximately 4 to 6 years.

This downward prediction reflects the neurobiological underpinnings of ASD, where atypical development of motor-related brain regions, including the cerebellum, basal ganglia, and motor cortex, leads to persistent deficits in motor skill acquisition and refinement. Children with ASD often exhibit delayed maturation of cortico-cerebellar circuits responsible for motor learning and coordination, resulting in motor patterns that resemble those of younger, typically developing children. The comorbidity of conduct disorder, associated with prefrontal cortex dysfunction affecting impulse control and sustained attention, further compounds these motor deficits by impairing the child's ability to plan, execute, and complete drawing tasks with the deliberate control expected of their chronological age. Thus, the 4 to 6 year motor age gap quantified by our model reflects genuine neurodevelopmental delays rather than mere behavioral differences, providing an objective measure of functional motor impairment.

This downward prediction reflects multiple features: reduced pen pressure from weaker grip strength or force modulation deficits; increased shakiness and jerk indicating fragmented movements; simpler skeleton topology suggesting reduced motor planning sophistication; higher Laplacian variance reflecting stroke instabilities; and increased segmentation showing stop-start patterns.

6.2. Clinical Implications, Limitations and Future Directions

The contrasting patterns in Parkinson's disease (accelerated aging) and ASD with CD (delayed development) demonstrate bidirectional motor age gaps detectable from simple handwriting tasks. This approach offers key advantages: dramatically greater accessibility than neuroimaging (requiring only paper and smartphone), non-invasiveness enabling repeated assessments, functional behavioral measurement complementing structural imaging, and rapid administration across diverse settings including community clinics and telemedicine platforms.

However, several limitations must be acknowledged. First, we lack ground truth brain age gap measurements (e.g., from MRI) to validate whether behavioral motor age gaps accurately reflect neural aging or developmental delays. Second, our training dataset of 3,672 images, while spanning six age classes, remains modest compared to large-scale neuroimaging studies, potentially limiting the model's ability to capture subtle variations within age groups. Third, motor age gaps reflect motor dysfunction specifically and may not generalize to cognitive or other neurological domains. Fourth, categorical age predictions limit sensitivity to small longitudinal changes compared to continuous age estimation.

Future work should establish ground truth correlations by collecting paired handwriting and neuroimaging data to validate motor age gaps against brain age gaps. Expanding training datasets with larger sample sizes across more granular age bins would enable regression-based continuous age prediction. Longitudinal studies tracking motor age gaps alongside clinical progression and treatment response are essential for establishing prognostic utility. Extending this framework to other conditions (multiple sclerosis, Huntington's disease, stroke, ADHD, Developmental Coordination Disorder) would establish generalizability and diagnostic specificity. Finally, integrating motor age gaps with other accessible biomarkers (voice, gait, cognition) could create multimodal behavioral age estimation systems with enhanced clinical utility for early detection and disease monitoring.

7. Conclusion

Our work has demonstrated the feasibility of using handwriting and drawing samples as accessible biomarkers for age classification and potential neurological disorder detection. We developed and evaluated multiple deep learning

architectures—including ResNet-18, ResNet-50, and Vision Transformer models, as well as a feature extraction-based multilayer perceptron—for predicting age across six distinct groups spanning preschool children (3-6 years) to seniors (65+ years). Our feature extraction-based MLP achieved the highest classification accuracy, followed by ResNet-50, suggesting that carefully engineered features capturing stroke patterns, pressure variations, and spatial characteristics may provide more discriminative information than end-to-end learned representations for this task. The strong performance of ResNet-50 among the deep learning models indicates that deeper architectures can effectively capture hierarchical motor control patterns when sufficient data is available.

A key contribution of this work lies in the comprehensive data collection and preprocessing pipeline we established. We aggregated handwriting and drawing samples from multiple diverse datasets, including healthy children's drawings from Ecuador, BanglaWriting samples, and clinical handwriting specimens. Our preprocessing approach addressed the inherent variability in data collection methods by standardizing images to consistent formats: removing grid lines while preserving pressure variations in handwriting samples, converting all images to PNG format with black strokes on white backgrounds, and ensuring uniform resolution. This preprocessing pipeline proved critical for enabling effective transfer learning, as it reduced domain-specific noise while retaining age-discriminative motor characteristics. The resulting dataset spans the full developmental and aging spectrum, filling a significant gap in existing literature that has largely focused on narrow age ranges.

Our application of the trained model to Parkinson's disease patients revealed deviations in predicted age classifications, providing preliminary evidence that motor-based age prediction could serve as a screening tool for neurological conditions. The concept of a "motor age gap"—analogous to the brain age gap derived from MRI scans—offers a low-cost, accessible alternative for preliminary neurological assessment. This approach addresses the practical limitations of neuroimaging-based methods, requiring only paper, pencil, and a smartphone camera or scanner for image capture. Such accessibility makes the method particularly valuable for resource-limited settings, rural communities, and large-scale screening programs where MRI is prohibitively expensive or unavailable.

The implications of this work extend beyond Parkinson's disease detection. The framework we developed could be adapted for identifying other neurological and developmental disorders that manifest through motor control changes. In our study, we observed a negative delta in the younger participants. For instance, we noticed that children with Autism spectrum disorder and conduct disorder presented distinct motor signatures in handwriting and drawing that was detected through a younger looking brain than the child's actual age. Furthermore, extending this methodology to understudied conditions such as schizophrenia could provide novel insights into motor manifestations of psychiatric disorders. Schizophrenia has been associated with subtle motor abnormalities that precede psychotic symptoms, yet research on motor biomarkers for this condition remains limited. Analyzing whether schizophrenia patients exhibit characteristic motor age gaps could contribute to earlier detection and better understanding of the disorder's progression.

Future work should pursue several directions to strengthen and expand this approach. First, longitudinal studies tracking individuals over time would validate whether motor age gaps predict subsequent neurological decline or disease onset. Second, collecting larger datasets of clinical populations would enable training disorder-specific classifiers rather than relying solely on deviations from healthy aging patterns. Third, developing mobile applications that guide users through standardized drawing and handwriting tasks would facilitate widespread deployment and data collection. Such applications could provide immediate preliminary assessments while flagging cases requiring professional evaluation. Fourth, investigating which specific motor features drive age predictions—such as stroke smoothness, pressure consistency, spatial planning, or tremor characteristics—would provide mechanistic insights into how neurological conditions alter motor output across the lifespan.

Integration into clinical workflows represents another important avenue for future development. The method could serve as a first-line screening tool in primary care settings, with positive results triggering referrals for comprehensive neurological evaluation. Combining motor-based assessments with other accessible biomarkers, such as voice analysis or gait patterns, could further improve diagnostic accuracy. Additionally, expanding the age classification framework to include finer-grained age bins or regression-based age prediction would increase sensitivity to subtle deviations from expected motor development and aging trajectories.

In conclusion, this work establishes motor output analysis as a viable approach for age classification and preliminary neurological screening. Our multi-dataset integration, comprehensive preprocessing pipeline, and comparative evaluation of multiple architectures provide a foundation for accessible, scalable neurological assessment. The feature extraction-based approach's superior performance highlights the value of domain knowledge in medical applications, while the strong results from ResNet-50 demonstrate that deep learning can effectively capture age-discriminative motor patterns when properly trained. By extending the brain age gap concept beyond expensive neuroimaging to simple hand-

writing and drawing tasks, we offer a practical tool that could facilitate earlier detection of neurological disorders and ultimately improve patient outcomes through timely intervention. The framework's adaptability to various neurological and psychiatric conditions positions it as a versatile platform for future research into motor manifestations of brain health across the lifespan.

8. Code Availability

To access our code, follow the linked GitHub.

9. Data Availability

To access our data, follow the linked directory.

10. Author Contributions

10.1. Data Collection & Processing

Both of us equally contributed to gathering data and processing it. Lixia was in charge of processing three age groups (Young Adults, Adults, and Seniors) and Hazel was in charge of processing the other half of the age groups (Preschool, Children, and Teens). Lixia wrote the processing code for the HandPD dataset and Hazel wrote the processing algorithm for the ASD with CD dataset.

10.2. Model Training & Testing

Both of us equally contributed to model training, validation, and testing on the same datasets. Both of us planned on what our feature-extraction model would look like and partook in research of relevant pre-existing models. Lixia was in charge of training the feature-extraction-based model as well as the Vision Transformer model. Hazel was in charge of training the ResNet-18 and ResNet-50 models. For testing, we both ran the highest accuracy models to test the brain age gap with Parkinson's dataset and ASD dataset for discussion.

10.3. Report & Presentation

Both of us equally contributed to writing the report. Lixia was in charge of writing the abstract, half of the methods (Feature based MLP section), experiments, and results. Hazel was in charge of writing the introduction, related work, half of the methods (ResNet-18, ResNet-50, and ViT), conclusion, and references + in-text citations. Both of us created the diagrams; Lixia created the tables and graphs, and Hazel modified the model architecture diagrams. While we initially divided the work in half, at the end, both of us edited each other's work and mutually agreed on the final draft. Both of us also equally divided the presentation slides among ourselves.

References

Alaei, F. and Alaei, A. Review of age and gender detection methods based on handwriting analysis. *Neural Computing and Applications*, 35(33):23909–23925, November 2023. ISSN 1433-3058. doi: 10.1007/s00521-023-08996-x. URL https://doi.org/10.1007/s00521-023-08996-x.

AYAZ, N., CELBİS, O., ZAYMAN, E. P., KARLIDAĞ, R., and ÖNAR, B. S. The use of handwriting changes for the follow-up of patients with bipolar disorder. *Archives of Neuropsychiatry*, 59(1):3–9, January 2022. ISSN 1300-0667. doi: 10.29399/npa.27666. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8895811/.

Bandyopadhyay, S., Wittmayer, J., Libon, D. J., Tighe, P., Price, C., and Rashidi, P. Explainable semi-supervised deep learning shows that dementia is associated with small, avocado-shaped clocks with irregularly placed hands. *Scientific Reports*, 13(1):7384, May 2023. ISSN 2045-2322. doi: 10.1038/s41598-023-34518-9. URL https://www.nature.com/articles/s41598-023-34518-9.

Castro, R., Patel, I., Patanjali, T., and Iyer, P. Automated schizophrenia detection from handwriting samples via transfer learning convolutional neural networks, July 2024. URL http://arxiv.org/abs/2408.06347. arXiv:2408.06347.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., and Houlsby, N. An image is worth 16x16 words: transformers for image recognition at scale, June 2021. URL http://arxiv.org/abs/2010.11929. arXiv:2010.11929.

Franke, K. and Gaser, C. Longitudinal changes in individual *BrainAGE* in healthy aging, mild cognitive impairment, and alzheimer's disease. *GeroPsych*, 25(4):235–245, January 2012. ISSN 1662-9647, 1662-971X. doi: 10.1024/1662-9647/a000074. URL https://econtent.hogrefe.com/doi/10.1024/1662-9647/a000074.

Franke, K. and Gaser, C. Ten years of BrainAGE as a neuroimaging biomarker of brain aging: what insights have we gained? *Frontiers in Neurology*, 10, August 2019. ISSN 1664-2295. doi: 10.3389/fneur.2019.00789. URL https://www.frontiersin.org/journals/neurology/articles/10.3389/fneur.2019.00789/full.

Gaser, C., Franke, K., Klöppel, S., Koutsouleris, N., Sauer, H., and Alzheimer's Disease Neuroimaging Initiative.

Brainage in mild cognitive impaired patients: predicting the conversion to alzheimer's disease. *PloS One*, 8(6): e67346, 2013. ISSN 1932-6203. doi: 10.1371/journal.pone.0067346.

Gonzalez-Ortiz, F., Kac, P. R., Brum, W. S., Zetterberg, H., Blennow, K., and Karikari, T. K. Plasma phospho-tau in Alzheimer's disease: towards diagnostic and therapeutic trial applications. *Molecular Neurodegeneration*, 18(1):18, March 2023. ISSN 1750-1326. doi: 10.1186/s13024-023-00605-8.

GTS.AI. Ups writing skills dataset: Ecuador children's drawings. GTS.AI Dataset Repository, 2024. URL https://gts.ai/dataset-download/ups-writing-skills/.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition, December 2015. URL http://arxiv.org/abs/1512.03385. arXiv:1512.03385.

Horvath, S., Langfelder, P., Kwak, S., Aaronson, J., Rosinski, J., Vogt, T. F., Eszes, M., Faull, R. L. M., Curtis, M. A., Waldvogel, H. J., Choi, O.-W., Tung, S., Vinters, H. V., Coppola, G., and Yang, X. W. Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging*, 8(7):1485–1512, July 2016. ISSN 1945-4589. doi: 10.18632/aging.101005.

Kim, H., Park, S., Seo, S. W., Na, D. L., Jang, H., Kim, J. P., Kim, H. J., Kang, S. H., and Kwak, K. A novel deep learning-based brain age prediction framework for routine clinical MRI scans. *npj Aging*, 11 (1):70, July 2025. ISSN 2731-6068. doi: 10.1038/s41514-025-00260-x. URL https://www.nature.com/articles/s41514-025-00260-x.

Lee, J., Burkett, B. J., Min, H.-K., Senjem, M. L., Lundt, E. S., Botha, H., Graff-Radford, J., Barnard, L. R., Gunter, J. L., Schwarz, C. G., Kantarci, K., Knopman, D. S., Boeve, B. F., Lowe, V. J., Petersen, R. C., Jack, C. R., and Jones, D. T. Deep learning-based brain age prediction in normal aging and dementia. *Nature Aging*, 2(5): 412–424, May 2022. ISSN 2662-8465. doi: 10.1038/s43587-022-00219-7. URL https://www.nature.com/articles/s43587-022-00219-7.

Liaqat, I. Handwriting and drawing based early detection of autism. Kaggle Dataset, 2024. URL https://www.kaggle.com/code/imranliaqat32/handwriting-and-drawing-based-early-detection-of-a.

McDonald, W. M. Overview of neurocognitive disorders. *Focus: Journal of Life Long Learning in Psychiatry*, 15 (1):4–12, January 2017. ISSN 1541-4094. doi: 10.1176/appi.focus.20160030. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6519631/.

Mridha, D. M. F., Quwsar Ohi, A., Ali, M. A., Emon, M. I., and Kabir, M. M. BanglaWriting: A multi-purpose offline Bangla handwriting dataset, October 2020. URL https://data.mendeley.com/datasets/r43wkvdk4w/1.

Mungas, D., Reed, B. R., Jagust, W. J., DeCarli, C., Mack, W. J., Kramer, J. H., Weiner, M. W., Schuff, N., and Chui, H. C. Volumetric MRI predicts rate of cognitive decline related to AD and cerebrovascular disease. *Neurology*, 59(6):867–873, September 2002. ISSN 0028-3878. doi: 10.1212/wnl.59.6.867.

Ngo, Q. C., McConnell, N., Motin, M. A., Polus, B., Bhattacharya, A., Raghav, S., and Kumar, D. K. Neurodiag: software for automated diagnosis of parkinson's disease using handwriting. *IEEE Journal of Translational Engineering in Health and Medicine*, 12:291–297, January 2024. ISSN 2168-2372. doi: 10.1109/JTEHM.2024.3355432. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10896420/.

Peng, H., Gong, W., Beckmann, C. F., Vedaldi, A., and Smith, S. M. Accurate brain age prediction with lightweight deep neural networks. *Medical Image Analysis*, 68:101871, February 2021. ISSN 1361-8415. doi: 10.1016/j.media.2020.101871. URL https://www.sciencedirect.com/science/article/pii/S1361841520302358.

Pereira, C. R., Weber, S. A. T., Hook, C., Rosa, G. H., and Papa, J. P. Deep learning-aided parkinson's disease diagnosis from handwritten dynamics. In *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 340–346, October 2016. doi: 10.1109/SIBGRAPI.2016.054. URL https://ieeexplore.ieee.org/document/7813053. ISSN: 2377-5416.

Pradeep, P. and Kamalakannan, J. Automated detection of Parkinson's disease using improved linknet-ghostnet model based on handwriting images. *Scientific Reports*, 15(1):30731, August 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-12636-w. URL https://www.nature.com/articles/s41598-025-12636-w.

Rabaev, I., Kurar Barakat, B., Churkin, A., and El-Sana, J. The hhd dataset. In *2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, pp. 228–233, September 2020. doi: 10.1109/ICFHR2020.2020.00050. URL https://ieeexplore.ieee.org/document/9257619.

Rakhmanov, O., Nnanna, N. A., and Adeshina, S. *Experimentation on Hand Drawn Sketches by Children to Classify Draw-a-Person Test Images in Psy-*

chology. PhD thesis, Nile University of Nigeria, May 2020. URL https://cdn.aaai.org/ocs/18457/18457-79397-1-PB.pdf.

Singla, C., Maini, R., and Kumar, M. Age, gender and handedness prediction using handwritten text: A comprehensive survey. *Engineering Applications of Artificial Intelligence*, 128:107432, February 2024. ISSN 0952-1976. doi: 10.1016/j.engappai.2023.107432. URL https://www.sciencedirect.com/science/article/pii/S0952197623016160.

Soumya Kumari, L. K. and Sundarrajan, R. A review on brain age prediction models. *Brain Research*, 1823: 148668, January 2024. ISSN 1872-6240. doi: 10.1016/j.brainres.2023.148668.

Taleb, C., Khachab, M., Mokbel, C., and Likforman-Sulem, L. Feature selection for an improved Parkinson's disease identification based on handwriting. In *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, pp. 52–56, April 2017. doi: 10.1109/ASAR.2017.8067759. URL https://ieeexplore.ieee.org/document/8067759.

Thal, D. R., Del Tredici, K., and Braak, H. Neurodegeneration in normal brain aging and disease. *Science of Aging Knowledge Environment*, 2004(23), June 2004. ISSN 1539-6150, 1539-6150. doi: 10.1126/sageke.2004.23.pe26. URL https://www.science.org/doi/10.1126/sageke.2004.23.pe26.

Thomas, M., Lenka, A., and Kumar Pal, P. Handwriting analysis in parkinson's disease: current status and future directions. *Movement Disorders Clinical Practice*, 4(6): 806–818, November 2017. ISSN 2330-1619. doi: 10.1002/mdc3.12552. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6174397/.

Yan, S., Lu, J., Zhu, H., Tian, T., Qin, Y., Li, Y., and Zhu, W. The influence of accelerated brain aging on coactivation pattern dynamics in Parkinson's disease. *Journal of Neuroscience Research*, 102(5):e25357, May 2024. ISSN 1097-4547. doi: 10.1002/jnr.25357.

Zhao, L., Wu, X., and Chen, X. Enhancing age estimation from handwriting: a deep learning approach with attention mechanisms. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 15(5), May 2024. ISSN 2156-5570. doi: 10.14569/IJACSA.2024.0150574. URL https://thesai.org/Publications/ViewPaper?Volume=15&Issue=5&Code=IJACSA&SerialNo=74.