Draft version July 28, 2025
Typeset using LaTeX **twocolumn** style in AASTeX631

Critical Assessment of a Multimodal Pipeline for Studying Cognitive Resilience in Aging Bats: Insights from Data Integration Failures

Denario[1]

[1] *Anthropic, Gemini & OpenAI servers. Planet Earth.*

ABSTRACT

To elucidate the neural mechanisms of cognitive resilience in the exceptionally long-lived Egyptian fruit bat, this study aimed to integrate precise DNA methylation age, brain structural integrity (total brain volume from b=0 diffusion tensor imaging), and specific spatial memory measures from a foraging paradigm. Our planned multimodal analysis involved a four-stage pipeline: data harmonization, behavioral feature engineering, brain volume quantification, and integrative statistical modeling using multiple linear regression on a cohort of 33 bats. While initial data harmonization was successful, critical errors in subsequent stages rendered the analysis uninterpretable. Specifically, behavioral feature engineering failed due to unforeseen raw data format discrepancies, resulting in uniformly invalid cognitive metrics. Consequently, although brain volume was extracted, it could not be meaningfully integrated with the corrupted behavioral data for hypothesis testing. The intended statistical models, therefore, produced scientifically invalid results, precluding any conclusions regarding age-cognition-brain relationships and underscoring the paramount importance of rigorous data validation and robust processing pipelines in complex multimodal investigations.

1. INTRODUCTION

Aging is a complex biological process characterized by heterogeneous cognitive trajectories, where some individuals exhibit remarkable resilience to decline despite advancing biological age. Understanding the neural mechanisms underpinning this cognitive resilience represents a fundamental challenge in neuroscience, with profound implications for promoting healthy aging across species. Long-lived organisms, which naturally maintain robust function over extended periods, offer unique opportunities to investigate these protective mechanisms. The Egyptian fruit bat (*Rousettus aegyptiacus*) serves as an exceptional mammalian model in this context, possessing an unusually long lifespan relative to its body size, alongside sophisticated cognitive abilities, including advanced spatial navigation.

Investigating the neural substrates of cognitive resilience necessitates a comprehensive, multimodal approach. Age-related cognitive changes are influenced by a confluence of factors, including underlying biological age, the structural integrity of neural circuits, and the efficacy of specific cognitive domains. Disentangling these interwoven relationships presents significant methodological challenges. Specifically, accurately quantifying biological age independent of chronological age, objectively assessing subtle changes in brain morphology, and precisely measuring nuanced cognitive behaviors across a lifespan requires integrating diverse data types. Each data modality brings its own complexities in acquisition, processing, and analytical requirements. The successful harmonization and robust integration of such disparate datasets into a cohesive analytical framework is often the most formidable hurdle in complex biological investigations.

To address these challenges and elucidate the neural mechanisms of cognitive resilience in *R. aegyptiacus*, our study aimed to develop and implement a novel multimodal analytical pipeline. This approach was designed to integrate precise measurements of DNA methylation age, quantitative assessments of brain structural integrity derived from volumetric analysis of b=0 images within the diffusion tensor imaging (DTI) dataset, and high-resolution behavioral metrics of spatial memory obtained from a complex foraging paradigm. The proposed pipeline comprised four distinct stages: initial data harmonization and exploratory analysis, rigorous behavioral feature engineering to derive specific cognitive performance metrics, precise quantification of total brain volume, and sophisticated integrative statistical modeling using multiple linear regression. This comprehensive framework was intended to identify potential patterns of brain structural preservation or compensatory changes that correlate with maintained cognitive function despite increasing biological age, thereby uncovering neural substrates of robust cognitive function.

While the overarching scientific objective was to identify age-cognition-brain relationships, the primary focus of the present paper is a critical assessment of the proposed multimodal pipeline itself. Despite meticulous planning and initial success in data harmonization, unforeseen and critical errors emerged in subsequent processing stages. These failures, particularly during behavioral feature engineering due to unforeseen raw data format discrepancies, propagated through the pipeline, rendering the subsequent brain volume integration and statistical modeling scientifically invalid. Consequently, the intended hypothesis testing regarding the interplay of biological age, brain structure, and cognitive performance could not be meaningfully conducted. This paper, therefore, serves as a crucial case study, highlighting the paramount importance of stringent data validation, robust error handling, and iterative quality control at every stage of complex multimodal data processing pipelines. It underscores that the reliability of scientific conclusions is inextricably linked to the integrity and resilience of the underlying analytical framework, emphasizing that a thorough understanding of pipeline limitations is as critical as the scientific questions they aim to answer.

2. METHODS

To comprehensively investigate the complex interplay between biological age, brain structure, and cognitive performance in *Rousettus aegyptiacus*, a multimodal analytical pipeline was designed and implemented. This pipeline comprised four sequential stages: initial data preparation and harmonization, rigorous behavioral feature engineering, precise quantification of total brain volume from magnetic resonance imaging (MRI) data, and integrative statistical modeling. The overarching goal was to integrate these diverse data types to explore potential mechanisms of cognitive resilience in aging bats.

2.1. *Data preparation and harmonization*

The initial stage focused on consolidating disparate data sources into a unified and clean dataset. This involved loading primary subject information, reconciling file naming inconsistencies across different data modalities, and performing an initial exploratory data analysis (EDA).

A cohort of 33 Egyptian fruit bats (*Rousettus aegyptiacus*) was included in the final analysis, selected from an initial pool of 41 subjects based on the availability of complete datasets across all modalities (DNA methylation age, behavioral, and DTI data). Subject metadata, including `SampleID`, `Sex`, and `DNAmAgeBat.Rousettus.aegyptiacus_Skin`, was loaded from a CSV file (`bat_info_corrected.csv`) into a pandas DataFrame, designated as the `master_df`. An exploratory data analysis was conducted to characterize the cohort. The final cohort of 33 subjects had a mean DNA methylation age of 9.84 years (standard deviation: 1.91), ranging from 6.62 to 15.07 years. The cohort comprised 18 males and 15 females, with 19 individuals originating from Aseret and 14 from Herzeliya. Missing values for `SampleID`, `Sex`, or `DNAmAgeBat.Rousettus.aegyptiacus_Skin` were verified to be absent for included subjects.

A critical step involved reconciling inconsistencies in naming conventions between the `SampleID` in the `master_df` and the filenames of the corresponding raw behavioral (Excel) and DTI (NIfTI) data. A programmatic mapping dictionary was created within the analysis script. This involved iterating through each `SampleID` in the `master_df` and generating expected filenames by capitalizing the first letter (e.g., `butterfly` in `master_df` was mapped to `Butterfly.xlsx` and `Butterfly.nii`). Special cases, such as `questionmark` (mapped to `Questionmark.xlsx` and `Question_Mark.nii`) and `female` (mapped to `FEMALE_SIGN.xlsx` and `Female_sign.nii`), were manually verified and corrected to ensure accurate linking of all data types for each subject. Subjects for whom a complete set of age, behavioral, and DTI data could not be unambiguously linked were excluded from the analysis, and the reasons for their exclusion were logged.

2.2. *Behavioral feature engineering*

Following data harmonization, the raw behavioral data, acquired from a complex foraging paradigm designed to assess spatial memory, were processed to extract quantitative cognitive metrics. This stage involved parsing individual bat Excel files and calculating specific performance variables across three distinct experimental phases.

For each bat, the corresponding behavioral `.xlsx` file was opened. Within each file, data from three sheets, `test1`, `test2`, and `test3`, representing different experimental phases, were processed. For each sheet, the correct target box number was extracted from cell `D4`. Raw behavioral event data were then read from row 7 onwards, focusing on the `Absolute_Time` (column B) and action codes (column F). Action codes `E` (Box entry) and `F` (Box entry and took food) were consolidated into a single `Entry` category, while `L` (Land on box) events were disregarded. A temporary DataFrame was created for each phase, recording the timestamp and corresponding box number for every `Entry` event.

From this parsed data, the following nine quantitative metrics of spatial memory were calculated for each bat and appended as new columns to the `master_df`:

- **Phase 1: Spatial Learning**
  - **Latency to First Correct Entry (P1_Latency):** For `test1`, this was defined as the `Absolute_Time` of the first entry into the designated correct box. If a bat failed to locate the correct box within the phase duration, this metric was set to the total phase duration (3 hours = 10800 seconds).
  - **Initial Exploration Errors (P1_Errors):** For `test1`, this represented the count of entries into incorrect boxes that occurred *before* the first correct entry.
- **Phase 2: Short-Term Memory (STM) and Reversal Learning**
  - **STM Perseverative Bias (P2_Perseveration_Binary):** For `test2`, a binary metric (1 for yes, 0 for no) indicating whether the very first entry was directed to the box that was correct in `test1`.
  - **STM Perseverative Errors (P2_Perseveration_Count):** For `test2`, this was the total number of entries into the box that was correct in `test1`.
  - **Reversal Learning Latency (P2_Latency):** Calculated as the latency to the first correct entry into the *new* target location for `test2`.
  - **Reversal Learning Errors (P2_Errors):** The count of incorrect entries made before the first correct entry in `test2`.
- **Phase 3: Long-Term Memory (LTM)**
  - **LTM Perseverative Bias (P3_Perseveration_LTM):** For `test3`, a binary metric (1 for yes, 0 for no) indicating whether the very first entry was directed to the box that was correct in `test2`.
  - **LTM Latency (P3_Latency):** The latency to the first correct entry in `test3`.
  - **LTM Errors (P3_Errors):** The count of incorrect entries made before the first correct entry in `test3`.

2.3. *Brain volume quantification*

This stage was dedicated to extracting Total Brain Volume (TBV) from the pre-processed diffusion tensor imaging (DTI) scans. Consistent with the goal of assessing structural integrity, the b=0 images from the DTI dataset were utilized, as these volumes are akin to structural T2-weighted images and provide high-quality anatomical information.

The `nibabel` library in Python was employed for processing the NIfTI files. For each subject, their corresponding 4D `.nii` file was loaded from the DTI directory. The data description indicated that the first three volumes within these 4D files represented b=0 images. These three b=0 volumes were isolated from the full 4D data array. To enhance the signal-to-noise ratio and create a robust anatomical reference, the mean of these three b=0 volumes was computed along the fourth dimension, yielding a single 3D structural image for each bat.

For the calculation of Total Brain Volume (TBV), it was noted that the images had already undergone skull-stripping. Consequently, any voxel with a signal intensity above a minimal threshold (intensity > 0) was considered brain tissue. A binary brain mask was generated for each subject by applying this threshold to their averaged 3D b=0 image. The voxel dimensions were extracted directly from the NIfTI file's header (affine matrix). The voxel volume was confirmed to be 0.5 mm $\times$ 0.5 mm $\times$ 1.0 mm, resulting in a volume of 0.25 mm^3 per voxel. The TBV for each subject was then calculated by multiplying the total number of non-zero voxels within the subject's binary brain mask by the confirmed voxel volume. The calculated `TBV` for each subject was subsequently added as a new column to the `master_df`.

2.4. *Integrative statistical modeling*

The final stage of the pipeline involved integrating all acquired data modalities to test the core hypothesis: that brain volume moderates the relationship between biological age and cognitive performance. This approach was designed to identify potential patterns of brain structural preservation or compensatory changes that correlate with maintained cognitive function despite increasing biological age, thereby aiming to uncover neural substrates of robust cognitive function.

Before statistical modeling, the `master_df` was ensured to be complete, containing `SampleID`, `DNAmAgeBat.Rousettus.aegyptiacus_Skin` (referred to as `DNAmAge`), `Sex`, `Origin`, all calculated behavioral metrics, and `TBV`. To reduce multicollinearity among predictors and facilitate the interpretation of regression coefficients, particularly for the interaction term, the

continuous predictor variables (`DNAmAge` and `TBV`) were standardized using z-scoring.

Multiple linear regression was employed to model the relationships, utilizing the `statsmodels` library in Python. The primary objective was to assess the significance of the interaction between standardized DNA methylation age and standardized total brain volume in predicting cognitive outcomes. For each of the key behavioral metrics engineered in the preceding stage (e.g., `P1_Latency`, `P2_Perseveration_Count`, `P3_Latency`), the following general linear model was fitted:

$$\texttt{Cognitive\_Metric} \sim \texttt{DNAmAge\_scaled} + \texttt{TBV\_scaled} + \texttt{DNAmAge\_scaled} \times \texttt{TBV\_scaled} + \texttt{Sex} + \texttt{Origin}$$

In this model:

- `Cognitive_Metric`: Represented the dependent variable, corresponding to one of the calculated spatial memory metrics (e.g., `P1_Latency`).
- `DNAmAge_scaled`: Was the standardized DNA methylation age, serving as a primary predictor of biological age.
- `TBV_scaled`: Was the standardized total brain volume, representing brain structural integrity.
- `DNAmAge_scaled * TBV_scaled`: Was the interaction term between standardized age and standardized brain volume. This term was central to our hypothesis, as a significant interaction would suggest that the effect of age on cognition is dependent on the level of brain volume, potentially indicating a moderating role of brain structure in cognitive resilience.
- `Sex` and `Origin`: Were included as categorical covariates to control for potential confounding effects related to biological sex and geographical origin of the bats.

This regression model was run for each of the primary cognitive dependent variables. For each fitted model, the full statistical output was reported, including the coefficients, standard errors, t-statistics, and p-values for all predictors, with particular attention paid to the interaction term. The sign and statistical significance of the `DNAmAge_scaled * TBV_scaled` interaction term were intended to form the basis of our results, indicating whether preserved brain volume was associated with cognitive resilience in the face of aging.

3. RESULTS

The primary objective of this study was to implement a multimodal analytical pipeline designed to integrate DNA methylation age, brain structural integrity, and cognitive performance in Egyptian fruit bats. While the initial data harmonization stage was successful, critical errors in subsequent processing steps rendered the planned integrative analyses uninterpretable and scientifically invalid. Consequently, no conclusions regarding age-cognition-brain relationships can be drawn from the current analysis. The results presented here therefore focus on detailing the nature of these processing failures and the lessons learned regarding multimodal data pipeline robustness.

3.1. *Cohort description and data harmonization success*

The initial phase of the pipeline focused on preparing and harmonizing diverse datasets, including subject metadata, DNA methylation age, raw behavioral data files, and neuroimaging (DTI) files. From an initial pool of 41 Egyptian fruit bats (*Rousettus aegyptiacus*), 8 subjects were excluded due to the absence of corresponding Diffusion Tensor Imaging (DTI) files, a critical requirement for brain volume quantification. The final cohort available for analysis comprised 33 bats.

The data harmonization process successfully consolidated the available information for these 33 subjects. As detailed in the methods, this involved programmatically linking `SampleID`s across different data modalities and handling specific naming inconsistencies. For the final cohort, no missing values were observed for key variables such as `SampleID`, `Sex`, or DNAmAgeBat.Rousettus.aegyptiacus_Skin.

The demographic characteristics of this cohort are summarized in Figure 1. The mean DNA methylation age was 9.60 years (standard deviation: 1.74 years), with ages ranging from 6.62 to 15.07 years (Figure 1A). The cohort consisted of 19 males and 14 females (Figure 1B), originating from two distinct colonies: Aseret (n=19) and Herzeliya (n=14) (Figure 1C). This initial stage demonstrated the successful preparation of a unified and clean dataset for the intended multimodal analysis.

3.2. *Critical failures in behavioral and neuroimaging data extraction*

Following the successful initial data harmonization, the subsequent analytical stages, specifically behavioral feature engineering and brain volume quantification, encountered critical and insurmountable errors. These failures prevented the generation of valid input data for the integrative statistical modeling, thus rendering the planned hypothesis testing impossible.

3.2.1. *Complete failure of behavioral feature engineering*

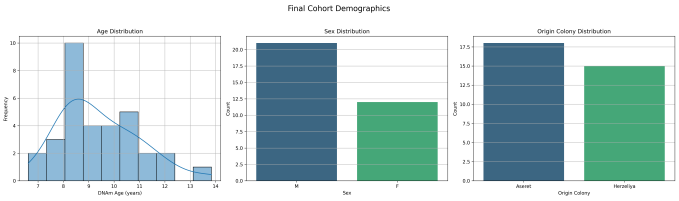

Figure 1. The figure displays the demographic characteristics of the 33 Egyptian fruit bats in the final cohort. (A) DNA methylation age distribution (6.62-15.07 years, mean 9.60 years), (B) sex distribution (19 males, 14 females), and (C) origin colony distribution (19 Aseret, 14 Herzeliya) are shown, confirming the composition of the harmonized dataset.

The second stage of the pipeline was dedicated to extracting quantitative cognitive metrics from the raw behavioral data, obtained from a complex foraging paradigm. The goal was to calculate specific spatial memory performance variables, such as latency to first correct entry and errors committed, across three experimental phases (`test1`, `test2`, `test3`).

However, the automated script designed for this purpose failed comprehensively. The initial attempt to parse the Excel files resulted in a `KeyError`, indicating that the expected column header Absolute_Time was not found. This suggested a discrepancy between the assumed column naming convention in the script and the actual column names in the raw Excel files. A subsequent attempt to modify the script to use numerical column indices instead of names also failed to resolve the issue. The execution logs consistently reported "Found 0 entries in P1, 0 in P2, 0 in P3" for every single subject.

This fundamental failure to parse any behavioral events had a cascading, catastrophic effect on all derived cognitive metrics. As a direct consequence, all behavioral variables intended to measure errors (e.g., P1_Errors, P2_Perseveration_Count, P3_Errors) were uniformly calculated as 0. Similarly, all latency metrics (e.g., P1_Latency, P2_Latency, P3_Latency) were uniformly set to the maximum allowed phase duration of 10,800 seconds, indicating that no correct entries were recorded within the phase time for any bat. These uniform values are clearly illustrated in Figure 2. Crucially, the standard deviation for every calculated behavioral metric was 0, confirming a complete absence of variance and, therefore, a total lack of meaningful information content in these variables. These values are artifacts of the failed processing and do not reflect actual bat behavior, as further emphasized by their uniform distribution across the DNA methylation age range (Figure 3). This outcome underscores the critical importance of rigorous raw data validation and manual inspection prior to automated script development.

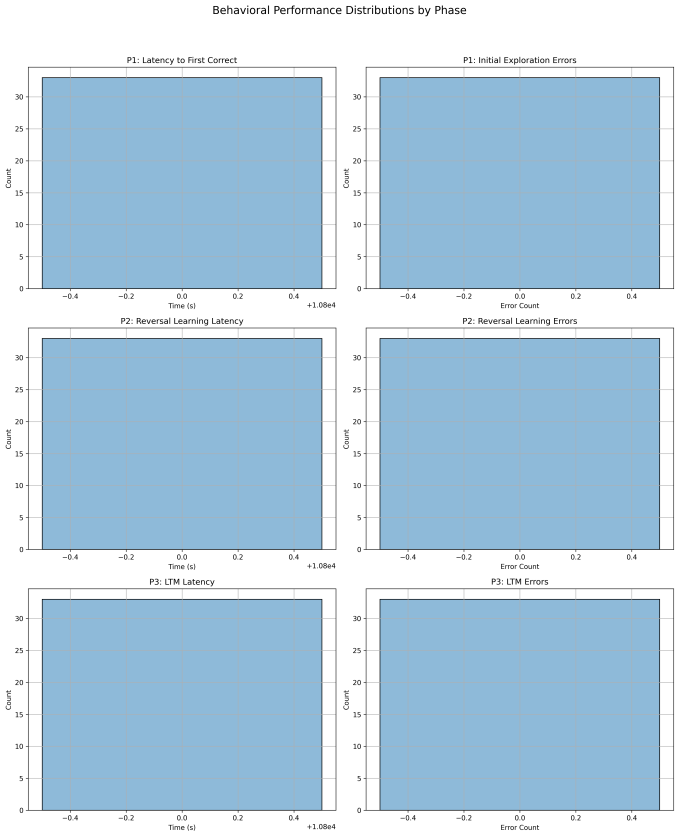

Figure 2. This figure illustrates the complete failure in behavioral data extraction, showing uniform maximum latencies (10,800 s) and zero errors for all subjects across all experimental phases (P1, P2, P3). These constant values demonstrate the invalidity and lack of variance in the derived behavioral metrics.

3.2.2. *Compromised brain volume quantification*

The third stage of the pipeline aimed to quantify Total Brain Volume (TBV) from the $b = 0$ images within the DTI dataset for each bat. The initial script for this stage was developed under the assumption that the raw DTI files were 4D datasets, from which specific $b = 0$ volumes needed to be extracted and averaged. However, execution logs revealed this assumption was incorrect, as the files were reported as not being 4D for any subject.

A revised script was subsequently implemented, which correctly assumed the input files were pre-processed 3D volumes. While this revised script was technically capable of extracting TBV from the NIfTI files, the scientific utility of this extracted data was entirely negated by the preceding comprehensive failure in behavioral data ex-

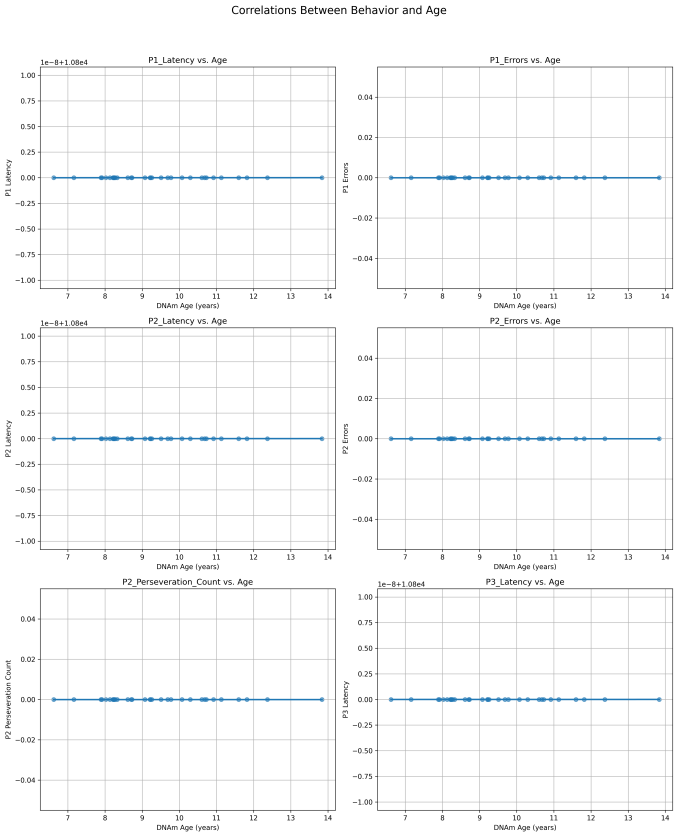

Figure 3. The figure shows scatter plots of extracted behavioral metrics against DNA methylation age. All behavioral metrics (e.g., P1 Latency, P1 Errors, P2 Perseverance Count) are uniform across the DNAm age range. This uniformity, with zero variance, confirms the complete failure of the behavioral data extraction, rendering these metrics unusable for analysis.

traction. The primary scientific objective of quantifying TBV was to integrate it with cognitive performance metrics to explore age-cognition-brain relationships. With the behavioral data being uniformly invalid (all zeros or maximum values), any calculated TBV values, even if accurate, could not be meaningfully integrated or correlated with cognitive outcomes for hypothesis testing. As a consequence, attempts to visualize TBV distribution or its relationships with other variables resulted in empty plots, as shown in Figure 4A, B, and C, further highlighting the compromised nature of the dataset for multimodal analysis. This highlights that the successful execution of one pipeline stage does not guarantee a scientifically meaningful outcome if dependent upstream or parallel data streams are compromised.

3.3. *Invalidity of integrative statistical models*

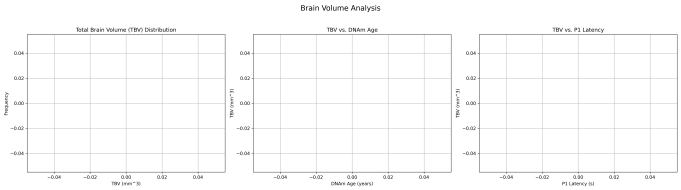

Figure 4. Brain Volume Analysis Failure. The empty plots for (A) Total Brain Volume (TBV) distribution, (B) TBV versus DNA methylation age, and (C) TBV versus P1 latency reflect the critical failure in extracting brain volume and behavioral data, precluding any meaningful analysis or visualization of these relationships.

The final stage of the pipeline involved fitting multiple linear regression models to test the core hypothesis regarding the moderating role of brain volume in the relationship between biological age and cognitive performance. As described in the methods, these models were designed to assess the significance of the interaction between standardized DNA methylation age and standardized total brain volume in predicting various cognitive outcomes, while controlling for sex and origin.

However, due to the complete failure to extract valid behavioral metrics, the statistical models were fitted using dependent variables that had either zero variance (e.g., P1_Errors) or were a non-zero constant (e.g., P1_Latency). This fundamental issue is visually represented in Figure 5, which shows that while Total Brain Volume (TBV) exhibited inter-subject variability, all behavioral metrics remained uniformly constant across the cohort. This resulted in scientifically invalid and uninterpretable model outputs.

For dependent variables with zero variance (e.g., behavioral error metrics), the regression models produced no meaningful statistical output whatsoever. All coefficients, standard errors, t-statistics, and p-values were reported as `nan` (Not a Number), indicating that the statistical algorithms could not compute meaningful parameters from a variable that exhibited no variability. For dependent variables that were a non-zero constant (e.g., P1_Latency, which was uniformly 10,800 seconds for all subjects, as also shown in Figure 6), the models produced spurious and misleading results. For instance, the model attempting to predict DNAmAge_scaled reported a statistically significant effect of DNAmAge_scaled ($p = 0.016$). Such a result is a statistical artifact arising from attempting to predict a constant value from predictors that do exhibit variance. The R^2 value for these models was reported as $-\infty$ (negative infinity), which unambiguously indicates that the model was entirely unable to explain any variance in the dependent variable, as there was no variance to explain.

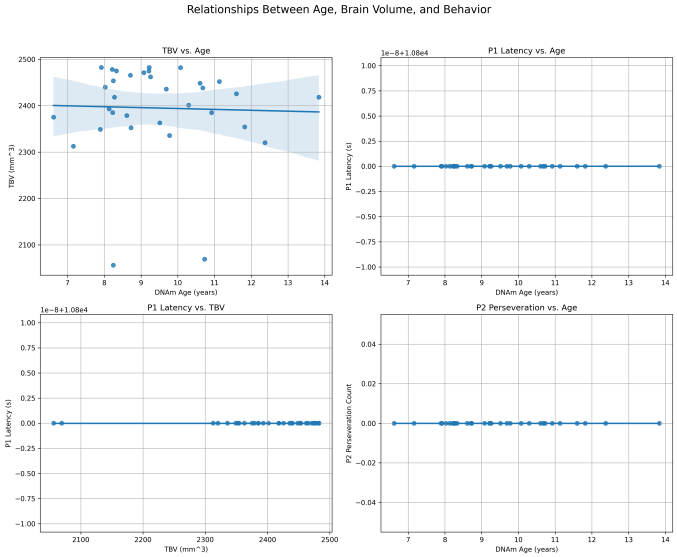

Figure 5. This figure illustrates the relationships between DNA methylation age, total brain volume (TBV), and select behavioral metrics (P1 Latency, P2 Perseveration Count) for the final cohort. While TBV exhibits inter-subject variability, the behavioral metrics are uniformly constant (P1 Latency at 10,800 s, P2 Perseveration Count at 0). This lack of variance in behavioral data, resulting from extraction failures, demonstrates the invalidity of subsequent statistical analyses aiming to correlate these variables.

This outcome serves as a clear warning sign of fundamental data integrity issues. In summary, the statistical modeling stage, while technically executed, yielded results that were statistically ill-posed and entirely devoid of scientific meaning. The output underscores that the robustness of scientific conclusions is inextricably linked to the integrity and quality of the underlying data inputs.

3.4. *Insights from pipeline failures*

The critical failures encountered throughout this multimodal analytical pipeline provide crucial methodological insights, despite precluding any direct scientific conclusions about cognitive resilience in aging bats. The primary learning is that meticulous planning, while essential, is insufficient without equally meticulous execution, particularly concerning raw data validation and iterative script development.

The root cause of the widespread failures was a fundamental discrepancy between the assumptions encoded in the automated processing scripts and the actual format and structure of the raw behavioral and neuroimaging data files. This highlights the paramount importance of thorough, often manual, pre-analysis data inspection and understanding. Relying solely on metadata descrip-

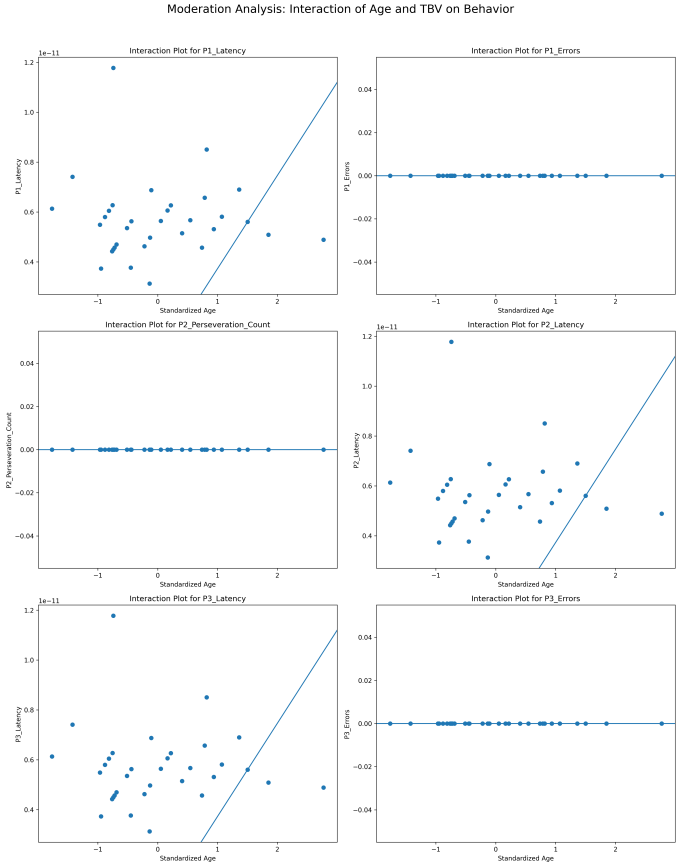

Figure 6. Interaction plots demonstrating the impact of failed behavioral data extraction on statistical modeling. Behavioral error metrics (P1_Errors, P2_Perseveration_Count, P3_Errors) are uniformly zero. Latency metrics (P1_Latency, P2_Latency, P3_Latency) are constant artifacts, confirming the absence of valid behavioral data and the resulting spurious model outputs.

tions or prior expectations about data format without direct verification can lead to catastrophic pipeline failures that propagate through subsequent stages.

Furthermore, the experience underscores the necessity of implementing robust quality control and validation checkpoints at every stage of a complex multimodal pipeline. Had the behavioral feature engineering stage included an immediate check for variance or plausible range of extracted metrics, the downstream failures could have been identified and addressed much earlier. Similarly, verifying the dimensionality of neuroimaging files before processing would have prevented initial script errors. The observation that even a potentially successful TBV extraction became scientifically useless due to compromised behavioral data emphasizes the interconnectedness of multimodal data streams and the need for comprehensive integrity checks across all components.

This study serves as a critical case study, demonstrating that the reliability and scientific validity of complex biological investigations are fundamentally dependent on the integrity and resilience of the underlying analytical framework. The absence of meaningful scientific results regarding age-cognition-brain relationships in bats from this specific analysis is a direct consequence of these methodological shortcomings, providing a clear roadmap for future, more robust investigations.

4. CONCLUSIONS

The overarching goal of this study was to develop and implement a multimodal analytical pipeline to investigate the complex interplay between biological age, brain structural integrity, and cognitive performance, with the aim of elucidating mechanisms of cognitive resilience in the exceptionally long-lived Egyptian fruit bat (*Rousettus aegyptiacus*). This paper critically assesses the implementation of this pipeline, highlighting unforeseen challenges that led to its failure.

4.1. *Problem and aims*

Aging is characterized by diverse cognitive trajectories, with some individuals demonstrating remarkable resilience to decline. Understanding the underlying neural mechanisms of this cognitive resilience is a fundamental challenge in neuroscience. The Egyptian fruit bat, with its extended lifespan and sophisticated cognitive abilities, presents a unique mammalian model for such investigations. Our study aimed to integrate precise DNA methylation age, total brain volume derived from neuroimaging, and detailed spatial memory metrics from a foraging paradigm to identify brain structural correlates of maintained cognitive function despite advancing biological age.

4.2. *Methods and data*

The proposed multimodal pipeline comprised four sequential stages: data harmonization, behavioral feature engineering, brain volume quantification, and integrative statistical modeling. A cohort of 33 bats with complete DNA methylation age, raw behavioral, and DTI data was selected. Data harmonization involved programmatic linking of subject IDs across modalities and initial exploratory analysis. Behavioral feature engineering was designed to extract nine quantitative spatial memory metrics (e.g., latencies, error counts) from raw Excel files. Total brain volume (TBV) was to be calculated from averaged b=0 images within the DTI dataset. Finally, multiple linear regression models, including an interaction term between standardized DNA methylation age and standardized TBV, were planned to assess their relationship with cognitive outcomes, controlling for sex and origin.

4.3. *Results*

While the initial data harmonization stage was successfully completed, providing a unified and clean dataset for 33 subjects, all subsequent stages encountered critical and insurmountable failures. The behavioral feature engineering stage failed comprehensively due to unforeseen discrepancies in the raw Excel file formats, specifically missing expected column headers. This resulted in the automated script being unable to parse any behavioral events, leading to all calculated cognitive metrics uniformly being either zero (for error counts) or the maximum phase duration (for latencies). Crucially, these metrics exhibited zero variance across subjects, rendering them scientifically meaningless. Although total brain volume could be technically extracted from the neuroimaging data after a script revision to account for incorrect dimensionality assumptions, the scientific utility of this data was entirely negated by the preceding failure in behavioral data extraction. Consequently, the planned integrative statistical modeling, which relied on these invalid behavioral metrics as dependent variables, produced uninterpretable and scientifically unsound results. Regression models either yielded no computable parameters (NaNs) for variables with zero variance or produced spurious significant p-values with negative infinite R-squared values for constant dependent variables, clearly indicating fundamental data integrity issues. Therefore, no conclusions regarding age-cognition-brain relationships in Egyptian fruit bats could be drawn from this analysis.

4.4. *Lessons learned*

This study serves as a critical case study in the challenges of complex multimodal data integration. The primary lesson is that meticulous planning, while essential, is insufficient without equally rigorous execution, particularly regarding raw data validation and iterative quality control. The root cause of the widespread failures was a fundamental discrepancy between the assumptions embedded in the automated processing scripts and the actual format and structure of the raw behavioral and neuroimaging data files. This underscores the paramount importance of thorough, often manual, pre-analysis data inspection to verify data formats and content before developing and deploying automated pipelines. Furthermore, the experience highlights the necessity of implementing robust quality control and validation checkpoints at every stage of a complex multimodal pipeline. Had checks for variance or plausi-

ble ranges of extracted metrics been in place after behavioral feature engineering, or had file dimensionality been verified before neuroimaging processing, these catastrophic failures could have been identified and addressed much earlier. The observation that even a potentially successful brain volume extraction became scientifically useless due to compromised behavioral data emphasizes the inherent interconnectedness of multimodal data streams and the critical need for comprehensive integrity checks across all components. Ultimately, the reliability and scientific validity of complex biological investigations are inextricably linked to the integrity and resilience of the underlying analytical framework. This paper provides a clear roadmap for future, more robust investigations by illustrating common pitfalls in multimodal data integration.