Thesis Proposal

Multivariate Normative Models Using Variational Auto-Encoders: A Study on Covariate Embedding and Robustness to Site-Variance using Gen R Data

Remy Duijsens
Delft University of Technology
Delft, The Netherlands
r.w.j.p.duijsens@student.tudelft.nl

***Abstract*—Normative modeling is a promising framework that is an alternative for case-control studies in medical research and other fields. Current efforts have studied these models for the univariate case and applied traditional machine learning for behavioral predictive modeling. This thesis will explore the design of a multivariate normative model by exploiting the latent space properties of Variational Auto-Encoders. The model is used for brain age estimation, an inherently multivariate task that can benefit directly from multivariate normative modeling. The effectiveness of different methods of embedding important covariates (e.g., sex and age) into the model is studied. Furthermore, the robustness of the model to confounding factors that introduce noise (e.g., site) is studied and compared to different harmonization methods applied in univariate normative modeling. Lastly, this work looks at the advantages and disadvantages of using multivariate normative models over traditional univariate ones. Expected findings include a better understanding of the properties of Variational Auto-Encoders in normative modeling, improved model performance due to effective covariate embedding, improved robustness against noise, and improved individual normative estimations. Both clinicians and academics can benefit from and apply these findings in relevant work where normative modeling can be used.**

I. Preamble

This TU Delft Computer Science master's thesis is based on an external collaboration with ErasmusMC under the supervision of Hugo Schnack and Ryan Muetzel at the Department of Child and Adolescent Psychiatry/Psychology. The project builds upon earlier work involving predictive modeling of various psychological disorders using brain imaging data and previous research on normative modeling techniques. This thesis project can demonstrate how the intersection of computer science and applied fields, such as behavioral predictive modeling in the medical domain, can result in useful theoretical and practical findings. Using recent machine learning techniques, particularly in developing multivariate normative models, this work will try to provide new insights that can improve current clinical practices related to normative modeling. In addition to the possible impact of the findings of this work, this collaboration is also a very valuable personal learning experience.

II. Introduction

Normative modeling is an alternative to case-control studies where each individual is compared to a reference population group based on clinically relevant covariates, enabling statistical inferences at the individual's level [1] [2]. The most common example of normative modeling in everyday clinical practice is the use of growth charts, such as those for head circumference, height, and weight, during pediatric consultations. These charts have been used for years to compare an individual's physical development to a population reference [3]. Normative modeling is particularly compelling in the study of brain images, as it allows researchers and clinicians to compare an individual's brain measures to a reference population, thereby identifying deviations that may be clinically relevant. This particular application of normative modeling only recently gained traction. This delay is likely due to the complexity of brain data and the challenges involved in modeling the many factors that influence brain structure and function.

Normative modeling has been predominantly studied on single variables (e.g., weight, whole brain volume) [4]. In cases where multiple coherent variables are present, each is typically modeled independently. If we model these variables in a single multivariate normative model, we can use the coherence between the variables. This approach can be very valuable if the number of variables is large since (clinically) working with them becomes impractical. Furthermore, the model can base its prediction on a complete representation of the input data and could, therefore, be more accurate and robust to noise than univariate models. Different types of brain images, such as structural and functional MRI data, can be used to train these normative models. Additionally, interesting nuisance factors such as age, sex, and IQ can be used to improve normative models by accounting for natural variations in brain structures. Noise factors and other sources of variability, such as the site, also play an important role and must be accounted for to ensure the robustness and reliability of these models.

Recent works has begun to explore multivariate normative models, with several studies demonstrating the potential of deep learning models, such as various types of Auto-Encoders (AE), in this context [5] [6] [7]. These models, when used for normative modeling, must meet specific requirements. Primarily, they must reduce the dimensionality of the data, producing a significantly smaller latent space representation that can be easily analyzed and interpreted in a clinical setting. This latent representation must also align with or be transformable to a normative multivariate distribution. Some initial works in this field using deep learning models have proposed different variants of AEs. The Variational Auto-Encoder (VAE) is a promising model since its latent representation is expressed in distribution estimates, which forms a good starting point for a multivariate normative model [5]. Recent literature has shown new types of VAE regularizers that show great potential in shaping the distribution estimates to a desired distribution [10]. Another promising property of the usage of VAEs is the large body of literature showing the increased performance in dimensionality reduction over linear dimensionality reduction algorithms. Furthermore, the robustness of VAEs is a promising property and could replace existing efforts in data harmonization techniques to accommodate for site variation.

While the initial work on VAEs in normative modeling is promising, current studies have primarily focused on model performance based on accuracy. However, it is necessary to explore additional properties to understand the effectiveness of VAEs in multivariate normative models compared to traditional univariate normative models. This work will examine the model's robustness to noise factors, where site is the confounding variable of interest. Furthermore, the importance of nuisance factors in accounting for natural variations in brain structures will be addressed. Here, the effectiveness of different covariate embedding strategies in VAEs is studied. At last, this work will compare the multivariate approach with traditional univariate models to determine the relative advantages and disadvantages.

Through external collaboration with ErasmusMC, this study is tied to a valuable case study focused on predicting the biological brain age of children and estimating their brain age gap relative to their peers. Furthermore, the implementation of a multivariate normative model supports the study of normal variation, the heterogeneity shown in biological systems, in brain measures. This research is made possible through access to large datasets, including the Generation R study [8] and the Healthy Brain Network (HBN) [9]. The Generation R study, based in Rotterdam, provides a large, population-based birth cohort dataset, which includes MRI images (specifically T1-weighted images) and health-related covariates from children and adolescents. The HBN dataset presents a comparable dataset. The brain images in these datasets have been converted into high-dimensional representations of various brain measurements using a validated preprocessing pipeline.

It is important to clarify that normative models and biological age estimation are distinct yet related concepts, especially in the context of brain research. Biological brain age is often conceptualized as a multivariate measure, which considers the interaction of various brain properties. In contrast, normative models have traditionally been used to model single variables. This thesis seeks to bridge the gap between these approaches by exploring how multivariate normative models can inform biological age estimation.

The remainder of this proposal is organized as follows. Section III presents the aim and objectives of this work and presents and motivates the research questions. Section IV contains the research design and methodology. At last, the project timeline is discussed in Section V.

III. Research Objectives

Current efforts in normative modeling have primarily focused on the univariate case and applied traditional machine learning for behavioral predictive modeling. However, these univariate models do not fully exploit the coherence between multiple input variables. Moreover, they can become clinically impractical when dealing with many input features, as this necessitates the interpretation of multiple normative scales. The field is currently exploring multivariate normative modeling, with deep learning emerging as a promising technique for improving normative models. While some initial work has been done to verify the accuracy of AE-based normative models, the full potential of deep learning in this context, including its ability to address other important properties and behavior of the multivariate normative model, is yet to be fully explored.

A. Research Aim

The aim is to create a normative model on a low-dimensional latent representation of the high-dimensional properties of the child/adolescent brain. The key input variables are thus the brain properties (X) that we want to reconstruct as good as possible from a latent representation (Z). In this process, we need to take into account other variables too. These variables are of a different nature. The first type of variables are characterised as nuisance factors, which we also call covariates, and are of direct interest and include variables such as age and sex. These variables provide information about the underlying biology and can help to improve the model. The quality of the models will improve if these factors are embedded efficiently. The second type of variable is usually referred to as noise. These variables include parameters describing the random noise and variation due to unknown biological factors and parameters describing the fixed effects of measurement design, such as the use of different measurement hardware and configurations (systematic noise). Currently, several approaches have been used to incorporate these variables into the modeling process, including separation of handling them and the actual normative modeling (e.g., first harmonizing the data from different scanners, making separate models for males and females), and compromising the nature of these variables (e.g., using one-hot encoding for the continuous variable age). This study aims to follow a more structured framework to incorporate these variables into the normative modeling process.

B. Proposed Solution

The contribution of this thesis project is focused on creating a multivariate normative model using Variational Auto-Encoders and studying different techniques to embed covariates effectively into the model. Furthermore, the robustness to systematic noise due to site variance is studied. This requires implementing and verifying a multivariate normative model in the applied domain of brain age prediction using brain images and other covariates (e.g. age and sex). The covariates' effects on the model's quality and different methods of embedding the covariates in the multivariate normative model are studied. Likewise, the impact of confounding factors (e.g. site) is investigated. The robustness of the multivariate normative model to systematic noise of confounding factors is explored. Here, the intrinsic robustness of the model is compared to noise reduction and data harmonization techniques currently used in normative modeling practices. This naturally leads to distinct research questions which will be explained in the next sections.

C. Research Topic 1: Systematic Noise Harmonization

RQ1: How does systematic site variance influence the robustness and accuracy of multivariate normative models, and can deep learning models improve on traditional data harmonization techniques?

To motivate this question, it is essential to understand that site variance, which arises when data is collected from different locations or scanners, can introduce systematic noise that degrades model performance. Site variance results from a combination of two main factors. First, due to the sample differences between sites (e.g., population A at site A vs. population B at site B, along with sampling and inclusion effects). Second, due to measurement differences, such as the use of different scanner models and protocols across sites. These factors often correlate and separating these confounding factors is complex. Hence, this work will simply refer to these factors as site variance.

Existing techniques, such as ComBat [13], are commonly used to harmonize data across sites. However, their effectiveness in preserving the integrity of covariates while improving model performance has yet to be fully understood. For univariate normative models, different techniques have been proposed [14]. This research question explores whether multivariate normative models based on Variational Auto-Encoders can account for site variance, potentially eliminating the need for traditional preprocessing methods. Additionally, it seeks to identify the best metrics for evaluating model robustness against site variance and to explore strategies for enhancing robustness without compromising the model's normative properties or the quality of the covariates.

D. Research Topic 2: Covariate Embedding Techniques

RQ2: What are the most effective methods for embedding covariates in VAE-based multivariate normative models, and how do these methods impact model performance?

This question is motivated by the need to understand the role of covariates, such as age and sex, in the performance of multivariate normative models. Covariates can significantly influence the accuracy and generalizability of these models, and their proper integration is crucial for making meaningful predictions due to underlying natural variations in the data. This work will explore various methods for embedding covariates within deep learning-based normative models, aiming to identify the approaches that most effectively improve model performance. A particular focus will be on whether specific covariates, like age, can be disentangled from the latent space in VAE-based models, providing deeper insights into the underlying mechanisms of these models.

E. Research Topic 3: Comparative Study of multivariate and univariate normative models

RQ3: What are the potential advantages and disadvantages of using multivariate normative models over traditional univariate models, and which methods can be used to compare these approaches?

This research question stems from the need to critically evaluate the benefits and limitations of multivariate normative models compared to their univariate counterparts. Multivariate models have the potential to capture more complex relationships between input variables, which could lead to more accurate and clinically useful predictions. However, the design of these models also raises questions about their interpretability and practical application. This study will compare the performance of multivariate models against traditional univariate models using both existing datasets, such as the Generation R dataset, and multi-site data, like the Healthy Brain Network (HBN). In addition, this work also compares the model's age prediction performance with other existing biological age models. By presenting the findings in an accessible and reproducible manner, the research aims to facilitate future studies and establish benchmarks for comparing normative models across different settings.

F. Theoretical Contribution

The primary theoretical contribution of this thesis lies in advancing the understanding and properties of multivariate normative modeling. Previous research primarily addressed the accuracy of such models. Here, we look at the application of Variational Auto-Encoders, compare different techniques for embedding important covariates (e.g. age and sex), and examine the impact of noise factors on these models. Furthermore, comparing the performance of multivariate models to traditional univariate approaches will give a better understanding of the potential of these multivariate normative models. The expected findings include improved model performance since many brain measures can now be integrated in a single model. This results in a better reflection of individuals' positions with respect to the norm. Furthermore, the model design is improved by estimating age and accounting for natural variation in the same architecture. Finally, the model might be more robust in terms of systematic noise. A comparison to existing data harmonization techniques that counter these noise factors is made, which can provide valuable insights into the robustness properties of VAE-based multivariate normative models.

G. Practical Implications

From a practical standpoint, this thesis aims to contribute to the existing framework for normative modeling. The presented use case particularly addresses normative modelling in the context of brain age prediction using brain imaging data. These findings will naturally extend to other clinical use cases of normative modelling that try associating biological measures with clinical covariates. Exploring deep learning techniques to address site variance and effectively embed covariates will provide actionable insights for practitioners seeking to implement multivariate normative models in clinical settings. Reducing the number of variables in the resulting models and unifying multiple independent variables into a single model could be closer to underlying biology and more practical to work with. Additionally, the comparative analysis will help guide the selection of modeling approaches based on specific research or clinical needs, improving the usability of normative models in clinical settings.

IV. Research Design

A. Datasets

Through collaboration with ErasmusMC, this project benefits from access to a large, high-quality dataset from the Generation R study. This dataset includes structural MRI images (specifically T1-weighted images) of the brain and health-related covariates (e.g., age and sex) from children and adolescents in the Rotterdam area. In addition, the Healthy Brain Network (HBN) dataset provides another comparable dataset from 10,000 children and adolescents (ages 5-21) in the New York City area. The HBN dataset is important for testing the quality of the model and assessing its robustness to multi-site variance. This dataset includes brain imaging collected using multiple types of scanners.

The brain images in these datasets are converted into a high-dimensional representation of various brain measurements using a validated preprocessing pipeline. The Generation R anatomical scans have been processed using FreeSurfer 6.0.0. The processed results include images and tabular data containing measurements from different brain regions.

The Generation R dataset is based on a population-based birth cohort (9,778 pregnant mothers) in the Rotterdam area. It includes three waves of data collection:

- **Wave 1:** Includes data from 1,070 children aged 6-9 years.
- **Wave 2:** Includes data from 4,087 children aged 9-11 years, with 3,992 obtaining parental consent for research. Of these, 3,959 children completed a full T1-weighted MRI sequence, 3,687 received a T2 scan, 3,777 underwent a DTI scan, and 3,439 had a resting-state fMRI scan. A total of 3,937 scans were successfully reconstructed using FreeSurfer.
- **Wave 3:** Includes data from 3,725 children aged 13-17 years.

Both datasets include brain images and a set of core data containing important covariates such as age and sex. For this study, these covariates are merged with the corresponding individual's tabular brain measurements to simplify the training process.

B. Model Architecture

The architecture of the VAE must meet several requirements for our use case. Firstly, the model should be capable of capturing and accurately representing the normative variation within the brain imaging data, which involves learning a latent space that reflects typical developmental trajectories. To achieve this, the architecture must handle the inherent variability in brain measurements while distinguishing between normative and pathological deviations. Robustness is an important property to account for variability introduced by multiple imaging sites and scanner types. Additionally, the embedding of covariates, such as age, is important for improving the model's representation of normal variations.

Covariate Embedding: In the context of normative modeling covariates such as age need to be embedded effectively. Different strategies for embedding these covariates will be explored:

- **No Covariate Modeling:** The simplest approach is to ignore the covariate, thereby not leveraging the information about the individual's age in the model. This method, shown in Figure 1, serves as a baseline to compare against more sophisticated approaches.

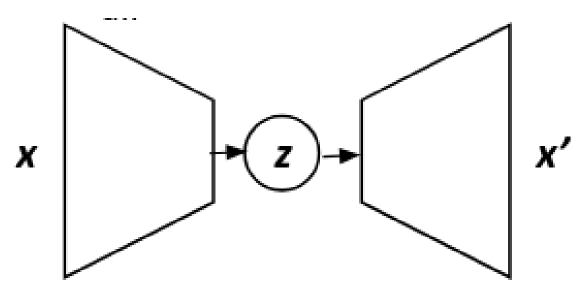

Fig. 1. VAE without age embedding.

- **Age Embedding in Decoding Phase:** In this method, age is included during the decoding phase. The model architecture, shown in Figure 2, is as follows: $X \rightarrow$ Encoder $\rightarrow Z$, and $\{Z, \text{age}\} \rightarrow$ Decoder $\rightarrow X'$. The aim here is for Z to become age-independent, meaning that the latent representation Z captures features of the brain imaging data that are invariant to age.

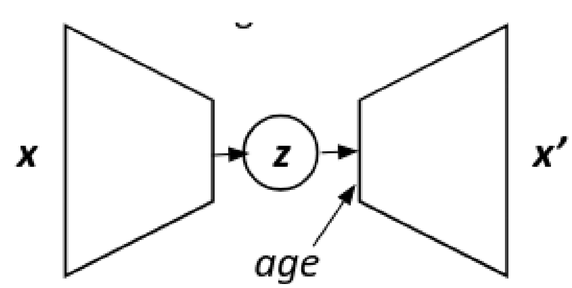

Fig. 2. VAE with age embedded in decode phase.

- **Latent Dimension for Biological Age:** We can introduce an additional latent dimension Z_{BA} representing biological age. The model architecture, shown in Figure 3, is: $\{X, \text{age}\} \rightarrow \{Z, Z_{BA}\} \rightarrow \{X', \text{age}' = BA\}$, where BA stands for brain age. The loss function will be a combination of the reconstruction loss for X and a weighted loss that penalizes deviations between predicted age and Z_{BA}. This approach helps Z to remain age-independent while explicitly modeling the developmental state of the instances through Z_{BA}.

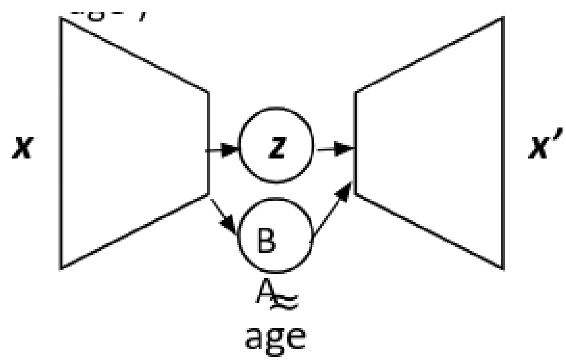

Fig. 3. VAE with separate latent dimension for age.

- **Brain-Age Gap as Latent Dimension:** Another strategy involves modeling the brain-age gap as a separate latent dimension, which is learned through backpropagation without conditioning on age. This gap represents the difference between the predicted brain age and chronological age, capturing deviations in brain development. The architecture is shown in Figure 4.

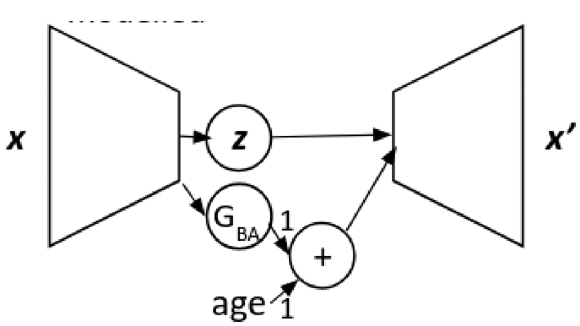

Fig. 4. VAE with brain age gap latent dimension.

- **Disentangled Age Representation:** A more advanced approach involves disentangling age from all dimensions of the latent space, based on recent work in the field [15]. This method aims to separate age-related variance from other latent factors, although it requires specific assumptions about the structure of the latent space.

Each of these approaches will be evaluated based on their effectiveness in capturing age-related variability and improving the model's performance on the given datasets.

C. Evaluation

The following methods will be employed to evaluate the model:

a) Reconstruction Error: Reconstruction error is a primary metric for evaluating the performance of our models. This metric measures how well the model can reconstruct the original brain features from the latent space representation. In the context of this project, it is anticipated that reconstruction error might be higher compared to univariate normative models due to the complexity of reconstructing multiple brain features (approximately tens to hundreds) from a reduced number of latent factors. This comparison will help us understand the trade-offs of using multivariate models versus univariate approaches.

b) Handling Multi-Site Variability: One key consideration is the model's ability to handle multi-site variations in the data. We will evaluate whether the multivariate normative models offer improved performance in managing these site-specific differences compared to univariate models. By analyzing the robustness of our model across multiple sites, we can assess its generalizability and reliability. r-robustness [16] is a metric which can be used to evaluate the robustness of VAEs. This paper used adversarial attacks to evaluate the metric but should also generalize it to test multi-site robustness.

c) Comparison to Univariate Baseline Models: The created multivariate normative models will be compared to a univariate baseline model to assess the added value of incorporating covariates and using a more complex latent space. This comparison will include metrics such as reconstruction error and deviance detection to determine whether the multivariate normative model gives significant improvements over univariate normative models.

d) Correlation with Behavioral and Clinical Measures: Another useful evaluation method is correlating the detected deviance from the norm (z-scores) with behavioral and clinical measures, such as symptom scores, IQ, and other relevant metrics. This correlation will help validate whether the model's representations are meaningful and aligned with clinically relevant outcomes.

D. Research Approach

A structured workplan is presented in this section. This workplan is designed to support a systematic approach to model development, evaluation, and comparison, as outlined in the objectives of this thesis. The approximate timeline for this thesis is outlined in Section V.

1) Development of the Multivariate Normative Model (Single-site Data):

a) VAE Design and Implementation:

- Begin with designing and implementing a Variational Auto-Encoder (VAE) using the Generation R (GenR) dataset.

- Input data will consist of brain imaging measurements (e.g., T1-weighted MRI scans) and clinically relevant covariates such as age and sex.
- Experiment with different model configurations (e.g., number of layers, nodes per layer, latent space dimensionality) to determine which architecture best balances complexity and performance.
- Use the training dataset to train an initial VAE model by minimizing reconstruction error, ensuring the latent representation Z captures the normative patterns of the brain data.

b) Creation of the Multivariate Normative Model:

- Use the VAE model to create a multivariate normative model.
- This involves using the VAE's latent space representation as a basis for generating individual-level normative predictions, allowing for the identification of outliers.
- Inspect the properties of the latent space. Interesting aspects are the shape of the latent space and the mapping of healthy data points and outliers to the space.

c) Model Evaluation and Fine-Tuning:

- Evaluate the initial model based on key performance metrics, focusing on reconstruction error to assess how well the model reconstructs the original data.
- Adjust the model hyperparameters (e.g., latent space dimensionality, learning rate) iteratively to minimize reconstruction error and improve model robustness.
- If necessary, use regularization techniques (e.g., dropout, weight decay, distribution regularizers) to prevent overfitting and improve the model's generalizability.

2) Development of a Univariate Normative Benchmark Model:

a) Model Implementation:

- Develop a univariate normative model based on traditional approaches (e.g., Gaussian Process Regression) for brain imaging features. Fine-tuning an existing model from one of the involved research groups is also a viable option.
- Train the model using the same GenR dataset, modeling each brain feature independently to serve as a benchmark for evaluating the multivariate model.

3) Evaluation of Model Robustness to Systematic Noise (Multi-site Data):

a) Multi-Site Data Preparation:

- After the initial multivariate normative model is created using single-site data (GenR), evaluate the Healthy Brain Network (HBN) dataset for testing the model's robustness to site variance.
- Process the HBN dataset to match the GenR dataset in terms of data format, brain feature extraction, and covariate selection.

b) Analyse Robustness and Harmonization Techniques:

- Test the multivariate normative model on the multi-site data, particularly focusing on its robustness to site variability introduced by different scanner types, locations, and protocols.
- Investigate the impact of systematic noise by analyzing model performance across different sites.
- Compare the model's performance with and without using traditional harmonization techniques (e.g., ComBat) to control for site-related variance. Analyze whether the VAE-based model inherently handles site variance better than traditional univariate models.
- Assess different strategies for embedding site-related information into the VAE to improve the model's robustness to systematic noise.

4) Covariate Embedding Techniques and Impact on Model Performance:

a) Exploring Covariate Embedding Approaches:

- Experiment with various methods for embedding covariates (e.g., age, sex) into the VAE, including:
  - No covariate modeling (baseline),
  - Covariate embedding during the decoding phase,
  - Separate latent dimensions for biological age,
  - Brain-age gap as a latent dimension,
  - Disentangled age representation.

b) Performance Evaluation:

- Evaluate the impact of each embedding technique on the model's performance and robustness.
- Compare the effectiveness of covariate embedding approaches by analyzing their influence on the latent space. Furthermore, evaluate whether age is correctly captured and disentangled from the latent dimensions.

E. Available Resources

Several computational resources are available for this research project to facilitate data processing, model development, and evaluation. The ResearchSuite environment at ErasmusMC serves as the primary local computing environment. It hosts the Generation R datasets and provides sufficient computational power for initial data exploration and analysis. ResourceSuite is a Linux environment with the software and tools required for handling the Generation R data. Additionally, access to the Snellius supercomputer [17] in the Netherlands has been provided through the Child and Adolescent Psychiatry/Psychology NeuroImaging (CAPPNI) research group. Snellius is the most powerful supercomputer in the country. It has many GPU nodes that can be used to train the models. Documentation is available for both computing resources and the Gen R dataset. To comply with the data policies of ErasmusMC, a mandatory data security session was held.

V. Timeline

The following timeline outlines the planning and key milestones throughout this thesis project. While the dates are approximate and may be adjusted as the project progresses, this timeline serves as a structured guide to ensure that each phase can be completed on schedule.

1) **Weeks 1-3: Research Preparation**
   - Finalize literature review.
   - Set up the working environment (e.g. ResearchSuite and Snellius).
   - Get acquainted with the systems and project methodology by doing a full loop from raw data to evaluation.
2) **Weeks 4-10: Model Development**
   - Design and implement the Variational Auto-Encoder (VAE) architecture.
   - Train the VAE models with Generation R data (also important to get good time estimates for this process).
   - Experiment with different methods for embedding covariates (e.g., age, sex).
   - Start preliminary analysis on model robustness to site variance.
   - Create a simple univariate normative baseline model.
3) **Week 10: First Stage Review**
   - Prepare a detailed report on the current results.
   - Review progress and refine objectives based on initial findings.
4) **Weeks 11-16: Model Refinement**
   - Apply more advanced covariate embedding techniques (e.g. covariate disentanglement).
   - Compare the robustness of the baseline model using different data harmonization techniques to our model.
   - Refine model architecture based on found results.
   - Hyperparameter tuning for the model.
5) **Weeks 17-20: Model Evaluation**
   - Evaluate the VAE-based multivariate normative model on Generation R and HBN datasets.
   - Conduct a comparative study between multivariate and univariate normative models.
   - Test the robustness to systematic noise and site variance.
   - Analyze the effectiveness of covariate embedding techniques using statistical tests.
6) **Weeks 21-24: Processing Results**
   - Create detailed documentation of all experiments and results.
   - Check the significance of the found results.
7) **Week 24: Greenlight Review**
   - Present results and analyses.
   - Obtain feedback and approval to proceed to the final phase.
8) **Weeks 25-30: Thesis Writing and Finalization**
   - Finalize writing the final thesis document.
   - Incorporate feedback from supervisors.
   - Conduct final proofreading and formatting.
   - Submit the final thesis to supervisors.

References

[1] Marquand, A. F., Rezek, I., Buitelaar, J., & Beckmann, C. F. (2016). Understanding Heterogeneity in Clinical Cohorts Using Normative Models: Beyond Case-Control Studies. Biological psychiatry, 80(7), 552–561. https://doi.org/10.1016/j.biopsych.2015.12.023

[2] Rutherford, S., Kia, S.M., Wolfers, T. et al. (2022). The normative modeling framework for computational psychiatry. Nat Protoc 17, 1711–1734. https://doi.org/10.1038/s41596-022-00696-5

[3] A health professional's guide for using the new WHO growth charts. (2010). Paediatrics child health, 15(2), 84–98. https://doi.org/10.1093/pch/15.2.84

[4] Rutherford, Saige and Barkema, Pieter and Tso, Ivy F and Sripada, Chandra and Beckmann, Christian F and Ruhe, Henricus G and Marquand, Andre F. (2023) Evidence for embracing normative modeling eLife 12:e85082.

[5] Lawry Aguila, A., Chapman, J., Altmann, A. (2023). Multi-modal Variational Autoencoders for Normative Modelling Across Multiple Imaging Modalities. In: Greenspan, H., et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. MICCAI 2023. Lecture Notes in Computer Science, vol 14220. Springer, Cham. https://doi.org/10.1007/978-3-031-43907-0_41

[6] Kumar S, Payne PRO, Sotiras A. Normative Modeling using Multimodal Variational Autoencoders to Identify Abnormal Brain Volume Deviations in Alzheimer's Disease. Proc SPIE Int Soc Opt Eng. 2023 Feb;12465:1246503. doi: 10.1117/12.2654369. Epub 2023 Apr 7. PMID: 38130873; PMCID: PMC10731988.

[7] Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. Hum Brain Mapp. 2019 Feb 15;40(3):944-954. doi: 10.1002/hbm.24423. Epub 2018 Oct 11. PMID: 30311316; PMCID: PMC6492107.

[8] White T, Muetzel RL, El Marroun H, Blanken LME, Jansen P, Bolhuis K, Kocevska D, Mous SE, Mulder R, Jaddoe VWV, van der Lugt A, Verhulst FC, Tiemeier H. Paediatric population neuroimaging and the Generation R Study: the second wave. Eur J Epidemiol. 2018 Jan;33(1):99-125. doi: 10.1007/s10654-017-0319-y. Epub 2017 Oct 24. PMID: 29064008; PMCID: PMC5803295.

[9] Alexander, L., Escalera, J., Ai, L. et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci Data 4, 170181 (2017). https://doi-org.tudelft.idm.oclc.org/10.1038/sdata.2017.181

[10] Duda, Jarek. (2018). Gaussian Auto-Encoder.

[11] Hrdlička, A. (1934). Normal Variation. Proceedings of the American Philosophical Society, 74(3), 253–261. http://www.jstor.org/stable/984718

[12] Wrigglesworth, J., Ward, P., Harding, I.H. et al. Factors associated with brain ageing - a systematic review. BMC Neurol 21, 312 (2021). https://doi.org/10.1186/s12883-021-02331-4

[13] Maxwell Reynolds, Tigmanshu Chaudhary, Mahbaneh Eshaghzadeh Torbati, Dana L. Tudorascu, Kayhan Batmanghelich, ComBat Harmonization: Empirical Bayes versus fully Bayes approaches, NeuroImage: Clinical, Volume 39, 2023, 103472, ISSN 2213-1582, https://doi.org/10.1016/j.nicl.2023.103472.

[14] Johanna M.M. Bayer, Richard Dinga, Seyed Mostafa Kia, Akhil R. Kottaram, Thomas Wolfers, Jinglei Lv, Andrew Zalesky, Lianne Schmaal, Andre Marquand, Accommodating site variation in neuroimaging data using normative and hierarchical Bayesian models, NeuroImage, Volume 264, 2022, 119699, ISSN 1053-8119, https://doi.org/10.1016/j.neuroimage.2022.119699.

[15] Leeb, Felix Bauer, Stefan Schölkopf, Bernhard. (2021). Interventional Assays for the Latent Space of Autoencoders.

[16] Camuto, Alexander Willetts, Matthew Roberts, Stephen Holmes, Chris Rainforth, Tom. (2020). Towards a Theoretical Understanding of the Robustness of Variational Autoencoders.

[17] Snellius: The National Supercomputer. SURF.nl. (n.d.). https://www.surf.nl/en/services/snellius-the-national-supercomputer