GEMCONT: Genetics-based Multimodal Contrastive Learning Enhances Phenotypic embeddings and Boosts Genetic Discovery

Daniel Sens[1,2,3] DANIEL.SENS@HELMHOLTZ-MUNICH.DE
Liubov Shilova[1,2,3,4] LIUBOV.SHILOVA@HELMHOLTZ-MUNICH.DE
Adrian V. Dalca[5,6] ADALCA@MIT.EDU
Julia A. Schnabel[3,7,8] JULIA.SCHNABEL@HELMHOLTZ-MUNICH.DE
Francesco Paolo Casale[1,2,3] PAOLO.CASALE@HELMHOLTZ-MUNICH.DE

[1] *Institute of AI for Health, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany*
[2] *Helmholtz Pioneer Campus, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany*
[3] *School of Computation, Information and Technology, Technical University of Munich, Garching, Germany*
[4] *Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*
[5] *A.A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*
[6] *Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA*
[7] *Institute of Machine Learning in Biomedical Imaging, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany*
[8] *School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK*

Editors: Under Review for MIDL 2026

Abstract

Genetic variation provides stable, time-invariant markers of disease risk and can therefore reveal upstream mechanisms underlying complex traits. Genome-wide association studies (GWAS) have identified thousands of loci associated with disease, yet most remain difficult to interpret because the intermediate phenotypes linking genotype to disease are unknown. Here, we address the question whether disease-associated genetic loci can be directly used to extract such risk-related features from quantitative phenotypes, including functional tests and medical imaging. We introduce GEMCONT, a multimodal contrastive learning framework that aligns genotype and phenotype representations in a shared latent space. By using known disease-associated variants as supervision, GEMCONT learns phenotypic embeddings guided by genetic information. To reflect the weak, additive nature of genetic effects, it employs a linear genetic encoder alongside a deep phenotypic encoder. We validate GEMCONT in controlled simulations and apply it to two real-world settings: spirometry curves for asthma and retinal fundus images for glaucoma. In both, GEMCONT improves disease risk prediction and enhances recovery of genetic associations compared with standard unsupervised or polygenic risk–based models. Altogether, our results demonstrate that incorporating stable genetic supervision into multimodal representation learning enables the extraction of genetically informed risk traits, refining disease phenotypes and improving the interpretability of association studies.

Keywords: Multimodal Contrastive Learning, Imaging Genetics, Genome-Wide Association Studies, Machine Learning–Derived Phenotypes, Medical Imaging

1. Introduction

Genome-wide association studies (GWAS) have identified thousands of genetic loci associated with human diseases and complex traits (Visscher et al., 2017; Manolio et al., 2009). Because germline variation is fixed and precedes disease onset, genetic associations provide upstream information about biological mechanisms. Yet for most loci, the functional link between genotype and phenotype remains unknown (Tam and Patel, 2019). Recent work in imaging genetics has begun to address this challenge by using high-dimensional biomedical data—such as medical images or physiological recordings—to derive quantitative phenotypes that better reflect underlying biology (Wright and Herzberg, 2021; Tracy, 2008; Robinson, 2012). Early approaches relied on manually defined regions or handcrafted measurements, while more recent studies leverage machine learning to learn compact phenotypic representations directly from raw data (Zech et al., 2018). These representations have proven valuable for association studies: for instance, supervised networks trained on clinical outcomes can uncover novel genetic loci (Rakowski et al., 2024; Kirchler et al., 2022), and unsupervised embeddings can reveal heritable structure (Yun et al., 2024b; Xie et al., 2024). However, unsupervised representation learning tends to capture the dominant axes of variation in the data and may overlook disease-related effects when these correspond to more subtler or less frequent patterns (Shilova et al., 2025).

To address these challenges, we introduce GEMCONT, a multimodal contrastive learning framework for imaging-genetics analysis (Fig. 1). GEMCONT leverages known disease-associated variants as supervision to jointly embed genetic and phenotypic data within a shared latent space, extracting structure enriched for disease relevance. This differs from prior multimodal approaches such as ContIG (Taleb et al., 2022) and MRM (Yang et al., 2023), which use molecular or genetic modalities for broad, task-agnostic pretraining. In contrast, GEMCONT focuses on disease-specific representation learning guided by GWAS knowledge, using genetics not merely as an auxiliary modality but as an explicit supervisory signal to extract phenotypic traits that are proximal to genetic risk and predictive of early or future disease states.

The contributions of this work are threefold:

1. **Genetics-informed contrastive learning.** We adapt multimodal contrastive learning to disease-focused imaging-genetics analysis through GEMCONT, which (i) selects disease-associated variants from external GWAS summary statistics to define targeted genetic supervision, and (ii) employs a linear genetic projector for efficient and interpretable variant contributions (Fig. 1).

2. **Benchmarking across genetic architectures.** We benchmark GEMCONT using controlled simulations with known disease-associated latent traits and causal variants, evaluating performance across sample sizes and genetic architectures to determine when contrastive alignment improves latent trait recovery.

3. **Applications to population imaging.** We apply GEMCONT to two use cases in the UK Biobank (Sudlow et al., 2015). First, we recover asthma-related spirometry embeddings by integrating asthma-associated variants with flow–volume curve data. Second, we recover a glaucoma-related latent trait from retinal fundus images using glaucoma-associated variants and imaging data. In both settings, genetics-guided

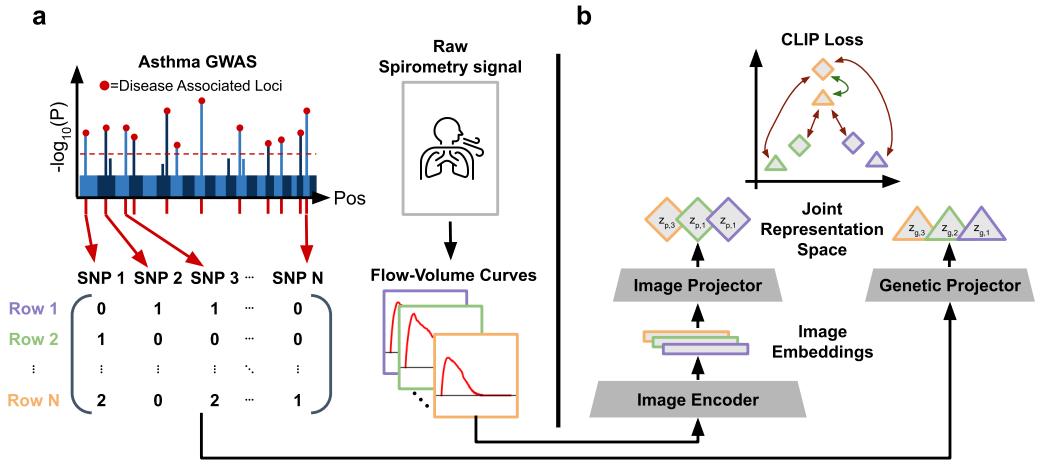

Figure 1: **Application of GEMCONT to asthma variants and spirometry images.** **(a)** Genetic variants significantly associated with asthma are extracted from the imputed UK Biobank data. **(b)** Corresponding raw spirometry signals are converted into Flow-Volume Curve images, and an asymmetric dual-encoder is trained to align genotype and image embeddings in a joint representation space through the CLIP Loss.

contrastive learning improves disease risk prediction and strengthens genetic association analyses compared to standard self-supervised approaches.

Together, these results establish contrastive learning with genetic supervision as a principled approach for constructing disease-specific, GWAS-aware phenotypes from high-dimensional medical imaging data.

2. Methods

The approach implemented in GEMCONT can be formulated as two-step process: (i) learning a joint representation space where genetic and imaging-derived features are co-embedded, and (ii) validating the extracted imaging embeddings in statistical association analyses and disease classification. Below, we describe the co-embedding pipeline and the statistical validation procedures.

2.1. Contrastive Learning for Genetics-Image Alignment

Contrastive learning for genetics-imaging alignment builds on the CLIP (Radford et al., 2021) framework, introducing key adaptations to address the unique challenges of genetic data, which are characterized by sparse and weak additive effects on phenotypes. Formally, given a dataset of paired genotype and phenotype samples $\mathcal{D} = \{(x_{g,i}, x_{p,i})\}_{i=1}^{N}$, each genotype sample $x_g \in \{0, 1, 2\}^S$ represents an allele count vector of S disease-associated variants, while x_p denotes a high-content phenotype (e.g., medical images). Following the principle implemented in the CLIP model, GEMCONT learns joint embeddings of genetic and imaging data by maximizing agreement between modalities from the same individual

while encouraging separation between individuals. This objective enables the co-embedding of genetic and phenotypic features into a shared latent space, facilitating the discovery of biologically meaningful genotype-phenotype relationships.

Multimodal Contrastive Learning Objective. GEMCONT processes genotype and phenotype data through modality-specific encoders (Fig. 1). The phenotype encoder f_{θ_p} maps x_p to an intermediate embedding e_p, which is then projected onto a unit-norm latent space as z_p. The genotype data are processed through a linear genotype projector, mapping x_g to a unit-norm latent embedding z_g. During training, we sample a mini-batch $\mathcal{B} \subset \mathcal{D}$ of paired genotype–phenotype samples and optimize the phenotype encoder, phenotype projector, and genotype projector to align genetic and phenotypic projections by maximizing similarity within individuals while minimizing it across individuals. This is achieved using the multimodal contrastive loss (Radford et al., 2021):

$$\mathcal{L} = \frac{1}{2}\left(\mathcal{L}_{g\to p} + \mathcal{L}_{g\leftarrow p}\right), \tag{1}$$

where

$$\mathcal{L}_{g\to p} = -\sum_{j\in\mathcal{B}} \log \frac{\exp(z_{g,j}^T z_{p,j}/\tau)}{\sum_{k\in\mathcal{B},\, k\neq j} \exp(z_{g,j}^T z_{p,k}/\tau)}, \tag{2}$$

and $\mathcal{L}_{g\leftarrow p}$ is defined analogously, swapping g and p. Similar to CLIP, $\tau > 0$ is a learnable temperature parameter.

Adaptations for genetic data. To address the sparsity and additive nature of genetic effects, GEMCONT introduces two key adaptations:

1. **Selection of informative variants.** We extract relevant genetic features from genome-wide association study (GWAS) summary statistics, which quantify associations (e.g., p-values, effect sizes) between millions of variants and a disease of interest. To ensure independence, we apply standard clumping (Purcell et al., 2007), iteratively retaining the most significant variant while removing correlated neighbors within a 5 Megabase (Mb) window.

2. **Linear projection of genotypes.** Genetic effects are predominantly additive with limited evidence for interactions between variants (Hill et al., 2005), making a linear projection sufficient for mapping selected variants into the latent space. This reduces model complexity while preserving key genetic signals.

2.2. Genetic Association Analysis of Learned Embeddings

Multi-trait GWAS for embedding analysis. To assess whether GEMCONT-derived embeddings capture meaningful genetic signals, we perform a multi-trait genome-wide association study (GWAS) on a held-out set of samples not used for training. We adapt the single-variant model from (Lippert et al., 2014; Casale et al., 2015), modeling each embedding dimension as a quantitative trait influenced by genetic variation. Let $\boldsymbol{E} \in \mathbb{R}^{N\times D}$ be the embedding matrix, $\boldsymbol{g} \in \{0,1,2\}^{N\times 1}$ a genotype vector, and $\boldsymbol{F} \in \mathbb{R}^{N\times K}$ a matrix of covariates. The model is:

$$\boldsymbol{E} = \boldsymbol{F}\boldsymbol{A} + \boldsymbol{g}\boldsymbol{b}^T + \boldsymbol{\Psi}, \tag{3}$$

where $\boldsymbol{A} \in \mathbb{R}^{D\times K}$ and $\boldsymbol{b} \in \mathbb{R}^{D\times 1}$ capture covariate and genetic effects, respectively, and $\boldsymbol{\Psi} \sim \mathcal{N}(0, \boldsymbol{C})$ models residual noise with a learnable covariance matrix $\boldsymbol{C} \in \mathbb{R}^{D\times D}$. Following (Lippert et al., 2014; Casale et al., 2015), $\boldsymbol{C}$ is estimated under the null model, while a single scaling factor per variant is optimized under the alternative model to control false positives efficiently (Korte et al., 2012). We test whether $\boldsymbol{b} \neq 0$, obtaining p-values via a likelihood ratio test with D degrees of freedom. For simulations (Sec. 3.1), we do not adjust for covariates. In our real-world application (Sec. 3.2), we control for genotyping array, assessment center, sex, age, age^2, sex-by-age, sex-by-age^2, height, height^2, BMI, and the top 20 genetic principal components to account for confounders. To address feature correlation and non-Gaussian distributions, embeddings are projected onto their top D principal components and rank-normalized before association testing. Independent genome-wide significant loci ($p < 5 \times 10^{-8}$) are identified using PLINK's clumping procedure (Purcell et al., 2007), which retains only approximately independent genetic associations by removing variants in linkage disequilibrium ($r^2 < 0.05$) within a 5 Mb window.

Assessing overlap with disease GWAS. To evaluate whether GEMCONT-derived embeddings capture known disease-associated genetic signals, we compare the genomic loci identified in our embedding-based GWAS to those from a standard disease GWAS. Specifically, we measure the fraction of independent genome-wide significant loci ($p < 5 \times 10^{-8}$) identified in the disease GWAS that are also recovered at genome-wide significance in the embedding GWAS. This assessment is performed on a held-out test set, distinct from the training data used for learning embeddings.

External disease GWAS and meta-analysis. To define the disease-specific variant panels used by GEMCONT, we rely on large external genome-wide association studies (GWAS) for asthma and glaucoma from the Million Veteran Program (Verma and Huffman, 2023) and FinnGen (Kurki and Karjalainen, 2023). For each disease, we harmonize summary statistics across cohorts and combine them using a fixed-effect inverse-variance meta-analysis in METAL (Willer et al., 2010). From the resulting meta-analytic GWAS, we selected variants with association $p < 10^{-5}$ and applied LD clumping in PLINK (5 Mb window, $r^2 < 0.05$) (Purcell et al., 2007), yielding an approximately independent set of disease-enriched SNPs. The 10^{-5} threshold is an intermediate cut-off that has been used when selecting variants from GWAS loci for Mendelian randomization analyses (Davey Smith and Hemani, 2014; Jin et al., 2024) and lies within the range of p-value thresholds typically explored in clumping-and-thresholding polygenic risk score methods (Choi et al., 2020). This choice provides a panel of variants that is strongly enriched for disease-associated signal while remaining sufficiently large to supervise the phenotype encoder.

3. Experiments and Results

We evaluate GEMCONT's ability to (i) enhance genetic signal for phenotype- or disease-associated variants and (ii) recover the underlying latent phenotype in simulations, and via disease risk prediction as a proxy in real data applications. We conduct three experiments: a controlled simulation study to assess performance under varying genetic architectures and two real-world applications to UK Biobank (Sudlow et al., 2015) data. In the first application we analyze flow-volume curves - used in asthma diagnosis (Jayasooriya and Stolbrink, 2023) - and integrate genetic variants associated with asthma. In the second we

apply our framework to retina fundus images, which are used in glaucoma diagnosis (Saha et al., 2023), and co-embed them with variants associated with glaucoma.

Compared methods. In the simulation and spirometry experiments, we compare GEMCONT against two established self-supervised embedding methods: a variational autoencoder (VAE), which learns latent representations by optimizing a reconstruction objective under a latent prior (Kingma and Welling, 2014), and SimCLR (Chen et al., 2020), a contrastive learning approach that maximizes agreement between augmented views of the same input. In the two real-data applications, we additionally include baselines tailored to disease prediction. First, we use a simple multimodal model in which the genetic branch is reduced to a single polygenic risk score (PRS) for the target disease, computed as the sum of GEMCONT's input variants weighted by their GWAS effect sizes and fed as a univariate input to the genetic projector. Second, to assess whether genetics-driven phenotype embeddings provide added value over conventional clinical markers, we benchmark all spirometry models against the FEV_1/FVC ratio and all fundus models against the cup-to-disc ratio, both widely used functional (Lambert et al., 2015) and imaging-derived (Gordon and et al., 2002; Foster et al., 2002) biomarkers in their respective diagnostic domains. Finally, in the fundus experiment we leverage a strong retinal foundation model (a DINOv2-pretrained ViT backbone) and consider three configurations on top of it: a frozen RetFound (Zhou and Wang, 2025) baseline, GEMCONT, and a supervised upper-bound model (Sec. 3.3).

3.1. Benchmarking in Simulated Data

Setup. To evaluate GEMCONT in a controlled setting, we design a simulation framework where a latent genetic trait influences imaging features, mimicking real-world genotype-phenotype relationships. We use EMNIST (Cohen et al., 2017), a dataset of 814,255 grayscale handwritten characters across 62 classes, and define the latent phenotype as the rotation angle of each character. We systematically vary key factors: training set size (N_{train}), genetic variance explained (h_g), and phenotype transformation strength ($\alpha_{\max}$). A fixed 100K test set is used across all experiments, with five random splits for training, where N_{train} is varied (default: 100K). After training, we extract image embeddings and evaluate:

1. **Phenotype recovery:** Predicting z from embeddings using ridge regression, measured by R^2 on the test set.

2. **Genetic recovery:** Identifying genome-wide significant variants ($p < 5 \times 10^{-8}$) via multi-trait GWAS (Sec. 2.2).

Simulation strategy. We first subsample N images stratified across character labels. Next, we simulate a genotype matrix $\boldsymbol{G} \in \{0, 1, 2\}^{N \times S}$ for S variants, where each entry represents allele counts drawn from a binomial distribution: $\boldsymbol{G}_{i,j} \sim \text{Binomial}(2, f_j)$ with minor allele frequency f_j. The latent phenotype $\boldsymbol{z}$ is generated as a weighted combination of genetic effects and environmental noise, controlling the proportion of variance explained by genetics (h_g):

$$\boldsymbol{z} = \sqrt{h_g} \cdot \text{std}(\boldsymbol{G}\boldsymbol{\beta}) + \sqrt{1 - h_g} \cdot \text{std}(\boldsymbol{z}_n), \tag{4}$$

where $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_S)$ represents variant effect sizes and $\boldsymbol{z}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$ models environmental noise. We then define rotation angles as $\boldsymbol{\alpha} = \alpha_{\max} \cdot \tanh(c \cdot \boldsymbol{z})$, where c is chosen to prevent

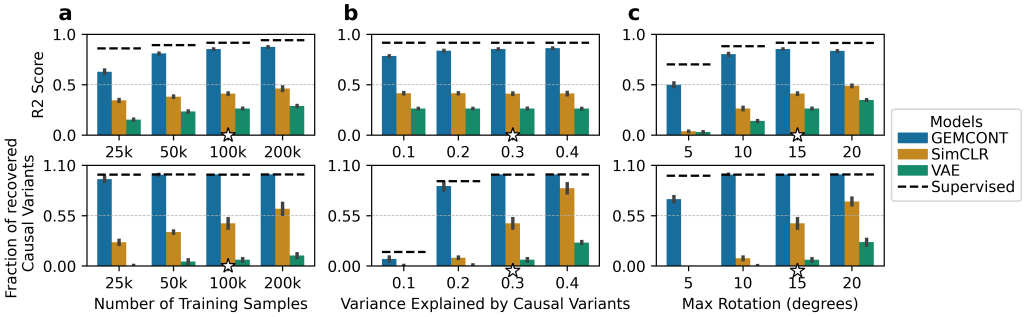

Figure 2: **Evaluation of phenotype and genetic signal recovery.** Performance of GEMCONT, SimCLR, VAE, and a supervised model is assessed under varying training set sizes (a), genetic variance explained (b), and maximum rotation (c). The first row shows the R^2 score for predicting the latent phenotype z from the learned embeddings, while the second row presents the fraction of causal variants identified at genome-wide significance ($p < 5 \times 10^{-8}$). Each bar represents the mean $\pm$ standard deviation across five random splits. Stars denote standard values held constant while other parameters were varied.

saturation of tanh across samples. Each image is rotated by its corresponding α_j, creating a dataset of genetic-image pairs directly linked through rotation.

Results. Figure 2 summarizes the results. GEMCONT outperforms SimCLR and VAE across all settings. Performance saturates beyond 100K training samples, though GEMCONT maintains a significant advantage (Fig. 2a). Genetic variance (h_g) has minimal impact on phenotype recovery but strongly affects variant detection, with GEMCONT consistently identifying more causal variants (Fig. 2b). Finally, lower rotation angles ($\alpha_{\max}$) degrade baseline performance more than GEMCONT, which remains robust across conditions (Fig. 2c). As expected, a supervised model serves as an upper bound for both phenotype and variant recovery.

3.2. Application to Spirometry and Asthma

Experimental setup. We generate flow–volume curves following (Yun et al., 2024a) and compute the FEV_1/FVC ratio, a key biomarker for asthma diagnosis (Lambert et al., 2015). Similar to recent work that applies CNNs directly to images of spirometry flow–volume curves for quality control (Martins et al., 2025; Wang and Li, 2022), we rasterize each trajectory into a standardized 256×256 grayscale image at 200 dpi and use these images as input to the encoder. Genetic variants associated with asthma are selected from external GWAS summary statistics using clumping (Sec. 2.2), yielding 551 approximately independent variants. To ensure that spirometry curves reflect baseline lung function, we exclude participants who reported using a chest inhaler or smoking a cigarette within the last hour before testing, in line with clinical spirometry preparation guidelines (Paraskeva et al., 2011). After matching imaging and genetic data and restricting to individuals of European ancestry, we retain 227,332 participants. To obtain stable estimates and quantify variability, we perform five random 50/50 train/validation splits and evaluate (i) disease recovery and (ii) genetic signal enrichment in the embeddings. Disease recovery is assessed using L2-regularized logistic

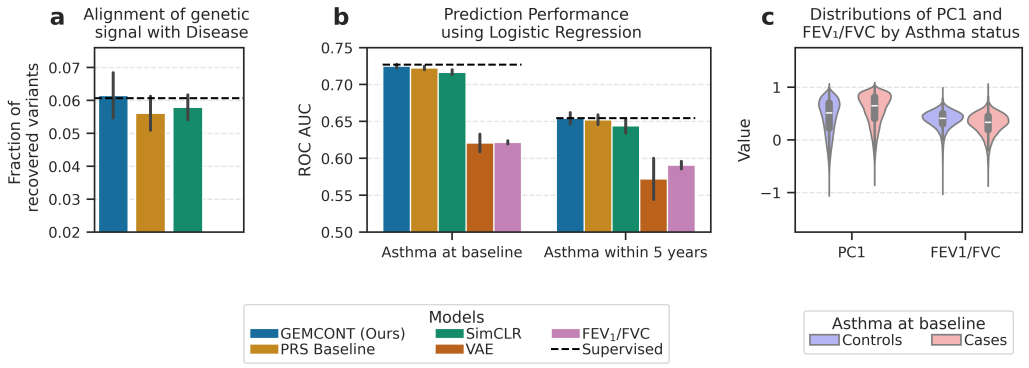

Figure 3: **Asthma prediction and genetic signal enrichment.** Comparison of GEMCONT, PRS baseline, SimCLR, VAE, and a supervised model trained on the latent phenotype. (a) Fraction of independent genome-wide significant asthma-associated variants recovered in method-specific GWAS after multiple testing correction. (b) ROC AUC for asthma classification at baseline and within 5 years post-assessment. (c) Distributions of the first principal component (PC1) of GEMCONT image embeddings and of FEV_1/FVC, stratified by asthma status. Violin plots show normalized values for controls (blue) and cases (red), with black bars indicating median and inter-quartile range. Results are mean ± standard deviation across 5 random 50/50 splits.

regression to predict asthma based on (i) pre-spirometry diagnosis and (ii) diagnosis within five years post-assessment, reporting ROC AUC. Genetic signal enrichment is quantified by performing multi-trait GWAS on the embeddings and computing the fraction of independent asthma-associated variants that remain significant after Bonferroni correction (Sec. 2.2). Figure 3 summarizes the results.

Results. GEMCONT achieves the highest recall of asthma-associated loci, though all models recover only a small fraction (Fig. 3a), consistent with expectations given our smaller sample size relative to the effective sample size of the GWAS meta-analysis. For asthma prediction, GEMCONT significantly ($p < 0.05$) outperforms the compared models at baseline and approaches the supervised model's upper bound for future diagnoses (Fig. 3b). Finally, Fig. 3c displays violin plots of the first principal component of the image embeddings (PC1) and the FEV_1/FVC ratio, stratified by asthma status; both measures show modest distributional shifts between cases and controls, indicating that PC1 captures asthma-related variation that is comparable in magnitude to the classical biomarker but derived directly from the flow–volume curves.

3.3. Application to Fundus Images and Glaucoma

Experimental setup. We analyzed color fundus images from the first imaging visit of UK Biobank participants (Sudlow et al., 2015) and filtered images using the MCF-Net model (Fu and Wang, 2019), excluding images with a rejection probability above 80%. We further excluded fundus images from participants who reported prior surgery or laser treatment for glaucoma, as this affects biomarkers for glaucoma risk and can impact fundus morphology (Lesk and Spaeth, 1999; Raghu and Pandav, 2012; Pillunat and Kretz, 2023).

Genetic variants associated with glaucoma were selected from external GWAS summary statistics using clumping (Sec. 2.2), yielding 1,535 approximately independent variants. After merging with genetic data for individuals of European ancestry, we retained 36,349 participants with at least one usable fundus image. For individuals with two images, we randomly sampled left or right eye with equal probability during training whenever the individual was drawn into a batch. During validation, if both eyes were available, we extracted image embeddings for each eye and used their mean as the final embedding. We build on a Vision Transformer (ViT) base encoder pretrained using DINOv2 on retinal images (Zhou and Wang, 2025), which we keep frozen and use as an online feature extractor during training (Kolesnikov and Beyer, 2020; Vo and Wang, 2025). Standard image augmentations are applied before feeding inputs through the frozen encoder (Sec. 3.4). On top of this backbone, we consider three image-based configurations. First, RetFound uses the frozen ViT features with simple mean pooling over patch tokens; no additional representation learning is performed, and the resulting embeddings are used directly in downstream GWAS and logistic regression. Second, GEMCONT adds an attention-pooling layer (Ilse and Tomczak, 2018) and a lightweight two-layer MLP to map pooled features to image embeddings, which are then aligned with genetic embeddings using the multimodal contrastive objective. Third, a supervised model shares the same architecture as GEMCONT but is optimized directly for glaucoma classification, providing an approximate upper bound. Given the availability of this strong fundus-specific backbone, we focus on these configurations rather than training additional generic self-supervised image models such as SimCLR or VAE on fundus images. As in the spirometry experiment (Sec. 3.2), we perform five random 50/50 train/validation splits and evaluate (i) disease recovery and (ii) genetic signal enrichment in the embeddings. Disease recovery is assessed using L2-regularized logistic regression to predict glaucoma based on (i) diagnosis at image acquisition and (ii) diagnosis within five years post-assessment, reporting ROC AUC. Genetic signal enrichment is quantified via multi-trait GWAS on the embeddings, measuring the fraction of independent glaucoma-associated variants that remain significant after Bonferroni correction (Sec. 2.2). Figure 4 summarizes the results.

Results. GEMCONT recovers the largest fraction of independent glaucoma-associated variants, outperforming the PRS and RetFound baselines as well as the supervised upper bound in terms of alignment between embeddings and disease loci (Fig. 4a). For disease prediction, GEMCONT is competitive with the supervised model and consistently outperforms both the PRS baseline and the cup-to-disc ratio for glaucoma at image acquisition and for incident glaucoma within five years (Fig. 4b). Finally, both the first principal component of the GEMCONT embeddings and the cup-to-disc ratio show case–control shifts, with PC1 exhibiting slightly stronger separation, indicating that the learned representation captures glaucoma-related variation that is at least comparable to this established imaging biomarker (Fig. 4c).

3.4. Implementation Details

All models were implemented in PyTorch (Paszke et al., 2019) and trained for 150 epochs with a batch size of 1024 using AdamW (Loshchilov and Hutter, 2017) (base learning rate 3×10^{-4}, weight decay 1×10^{-4}) and a cosine-annealing schedule with a 10-epoch warm-up. For the supervised baselines we additionally applied early stopping on the validation loss

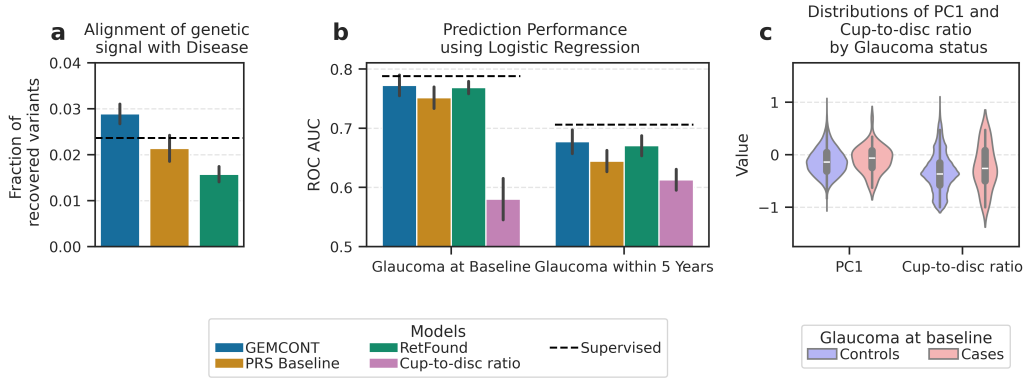

Figure 4: **Glaucoma prediction and genetic signal enrichment.** Comparison of GEMCONT, a PRS-only baseline, RetFound embeddings, a supervised upper-bound model (dashed line), and the cup-to-disc ratio clinical biomarker. (a) Fraction of independent genome-wide significant glaucoma-associated variants recovered after multiple testing correction. (b) ROC AUC for glaucoma classification at baseline and within 5 years post-assessment. (c) Distributions of the first principal component (PC1) of GEMCONT image embeddings and of cup-to-disc ratio, stratified by glaucoma status. Violin plots show normalized values for controls (blue) and cases (red), with black bars indicating median and inter-quartile range. Results are mean ± standard deviation across 5 random 50/50 splits.

with a patience of 50 epochs. For the genetic branch, weights connected to each input variant were initialized to the corresponding effect size from the external meta-analytic GWAS, and genetic inputs were augmented using SCARF (Bahri and Jiang, 2021) with corruption probability $p = 0.1$. Image augmentations were adapted to each experiment: random erasing for the EMNIST simulation to preserve the rotation signal, random resized crops with Gaussian blur for spirometry, and random resized crops, color jitter, and Gaussian blur for fundus images.

4. Conclusion and Future Work

We introduced GEMCONT, a genetics-based multimodal contrastive learning framework that aligns genotype and imaging embeddings to emphasize disease-relevant variation. In simulations, GEMCONT improved phenotype and variant recovery compared to self-supervised baselines, and in the asthma application it outperformed unsupervised methods in both disease prediction and genetic signal enrichment. Future work includes extending GEMCONT to richer imaging phenotypes such as MRI for neurodegenerative disorders and integrating it with genetic causal inference frameworks such as PRiMeR (Sens et al., 2024) to test whether genetically guided embeddings capture putative causal intermediates of disease. A key challenge remains the limited availability of large, harmonized imaging–genetics datasets; as such resources expand, genetics-informed contrastive learning may further refine genotype–phenotype mapping and support more mechanistic and predictive models for precision medicine.

References

Dara Bahri and et al. Jiang. SCARF: Self-supervised contrastive learning using random feature corruption. arXiv [cs.LG], 2021. URL https://openreview.net/pdf?id=CuV_qYkmKb3.

Francesco Paolo Casale, Barbara Rakitsch, Christoph Lippert, and Oliver Stegle. Efficient set tests for the genetic analysis of correlated traits. Nature methods, 2015. doi: 10.1038/nmeth.3439.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. arXiv [cs.LG], 2020. URL http://proceedings.mlr.press/v119/chen20j/chen20j.pdf.

Shing Wan Choi, Timothy Shin-Heng Mak, and Paul F O'Reilly. Tutorial: a guide to performing polygenic risk score analyses. Nature protocols, 2020. URL https://www.nature.com/articles/s41596-020-0353-1.

Gregory Cohen, Saeed Afshar, Jonathan Tapson, and André van Schaik. EMNIST: an extension of MNIST to handwritten letters. arXiv [cs.CV], 2017. URL http://arxiv.org/abs/1702.05373.

George Davey Smith and Gibran Hemani. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. Human molecular genetics, 2014. URL http://dx.doi.org/10.1093/hmg/ddu328.

Paul J Foster, Ralf Buhrmann, Harry A Quigley, and Gordon J Johnson. The definition and classification of glaucoma in prevalence surveys. The British journal of ophthalmology, 2002. URL https://bjo.bmj.com/content/86/2/238.short.

Huazhu Fu and et al. Wang. Evaluation of retinal image quality assessment networks in different color-spaces. In Lecture Notes in Computer Science. Springer International Publishing, 2019. URL http://dx.doi.org/10.1007/978-3-030-32239-7_6.

Mae O Gordon and Beiser et al. The ocular hypertension treatment study: baseline factors that predict the onset of primary open-angle glaucoma. Archives of ophthalmology, 2002. URL http://dx.doi.org/10.1001/archopht.120.6.714.

William (bill) W G Hill, Michael E Goddard, and Peter M Visscher. Data and theory point to mainly additive genetic variance for complex traits. PLoS genetics, 2005. doi: 10.1371/journal.pgen.1000008.eor.

Maximilian Ilse and et al. Tomczak. Attention-based deep multiple instance learning. In Proceedings of the 35th International Conference on Machine Learning. PMLR, 2018.

S Jayasooriya and et al. Stolbrink. Clinical standards for the diagnosis and management of asthma in low- and middle-income countries. The international journal of tuberculosis and lung disease: the official journal of the International Union against Tuberculosis and Lung Disease, 2023. URL http://dx.doi.org/10.5588/ijtld.23.0203.

Yongxiu Jin, Chenxi Han, Dongliang Yang, and Shanlin Gao. Association between gut microbiota and diabetic nephropathy: a mendelian randomization study. Frontiers in microbiology, 2024. URL http://dx.doi.org/10.3389/fmicb.2024.1309871.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. 2014. URL http://arxiv.org/abs/1312.6114.

Matthias Kirchler, Stefan Konigorski, Matthias Norden, Christian Meltendorf, Marius Kloft, Claudia Schurmann, and Christoph Lippert. transferGWAS: GWAS of images using deep transfer learning. Bioinformatics (Oxford, England), 2022. doi: 10.1093/bioinformatics/btac369.

Alexander Kolesnikov and et al. Beyer. Big transfer (BiT): General visual representation learning. In Computer Vision – ECCV 2020. Springer International Publishing, 2020. URL http://dx.doi.org/10.1007/978-3-030-58558-7_29.

Arthur Korte, Bjarni J Vilhjálmsson, Vincent Segura, Alexander Platt, Quan Long, and Magnus Nordborg. A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nature genetics, 2012. doi: 10.1038/ng.2376.

Mitja I Kurki and et al. Karjalainen. FinnGen provides genetic insights from a well-phenotyped isolated population. Nature, 2023. URL http://dx.doi.org/10.1038/s41586-022-05473-8.

Allison Lambert, M Bradley Drummond, Christine Wei, Charles Irvin, David Kaminsky, Meredith McCormack, and Robert Wise. Diagnostic accuracy of FEV1/forced vital capacity ratio z scores in asthmatic patients. The journal of allergy and clinical immunology, 2015. doi: 10.1016/j.jaci.2015.02.027.

Mark R Lesk and et al. Spaeth. Reversal of optic disc cupping after glaucoma surgery, analysed with a scanning laser tomograph. Journal of glaucoma, 1999. URL http://dx.doi.org/10.1097/00061198-199902001-00022.

Christoph Lippert, Franceso Paolo Casale, Barbara Rakitsch, and Oliver Stegle. LIMIX: genetic analysis of multiple traits. bioRxiv, 2014. doi: 10.1101/003905.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. arXiv [cs.LG], 2017. URL http://arxiv.org/abs/1711.05101.

Teri A Manolio, Francis S Collins, and et al. Cox. Finding the missing heritability of complex diseases. Nature, 2009. doi: 10.1038/nature08494.

Carla Martins, Henrique Barros, and André Moreira. Transfer learning in spirometry: CNN models for automated flow-volume curve quality control in paediatric populations. Computers in biology and medicine, 2025. URL http://dx.doi.org/10.1016/j.compbiomed.2024.109341.

Miranda A Paraskeva, Brigitte M Borg, and Matthew T Naughton. Spirometry. Australian family physician, 2011. URL https://www.racgp.org.au/getattachment/b2aef6c3-a6fb-46bf-9acf-f0215484f04c/Spirometry.aspx.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html.

Karin R Pillunat and et al. Kretz. Effectiveness and safety of VISULAS green selective laser trabeculoplasty: a prospective, interventional multicenter clinical investigation. International ophthalmology, 2023. URL http://dx.doi.org/10.1007/s10792-022-02617-7.

Shaun Purcell, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A R Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I W de Bakker, Mark J Daly, and Pak C Sham. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics, 2007. doi: 10.1086/519795.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. arXiv [cs.CV], 2021. URL https://proceedings.mlr.press/v139/radford21a/radford21a.pdf.

N Raghu and et al. Pandav. Effect of trabeculectomy on RNFL thickness and optic disc parameters using optical coherence tomography. Eye, 2012. URL http://dx.doi.org/10.1038/eye.2012.115.

Alexander Rakowski, Remo Monti, and Christoph Lippert. TransferGWAS of T1-weighted brain MRI data from UK biobank. PLoS genetics, 2024. doi: 10.1371/journal.pgen.1011332.

Peter N Robinson. Deep phenotyping for precision medicine. Human mutation, 2012. doi: 10.1002/humu.22080.

Sajib Saha, Janardhan Vignarajan, and Shaun Frost. A fast and fully automated system for glaucoma detection using color fundus photographs. Scientific reports, 2023. URL http://dx.doi.org/10.1038/s41598-023-44473-0.

Daniel Sens, Liubov Shilova, Ludwig Gräf, Maria Grebenshchikova, Bjoern M Eskofier, and Francesco Paolo Casale. Genetics-driven risk predictions leveraging the Mendelian randomization framework. Genome Research, 34(9):1276–1285, 2024.

Liubov Shilova, Daniel Sens, and Ayshan et al. Aliyeva. REECAP: Contrastive learning of retinal aging reveals genetic loci linking morphology to eye disease. medRxiv: the preprint server for health sciences, 2025. URL http://dx.doi.org/10.1101/2025.11.19.25340555.

Cathie Sudlow et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine, 12(3):e1001779, 2015.

Aiham Taleb, Matthias Kirchler, Remo Monti, and Christoph Lippert. ContIG: Self-supervised multimodal contrastive learning for medical imaging with genetics. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022. doi: 10.1109/cvpr52688.2022.02024.

Vivian Tam and et al. Patel. Benefits and limitations of genome-wide association studies. Nature reviews. Genetics, 2019. URL http://dx.doi.org/10.1038/s41576-019-0127-1.

Russell P Tracy. 'deep phenotyping': characterizing populations in the era of genomics and systems biology. Current opinion in lipidology, 2008. doi: 10.1097/MOL.0b013e3282f73893.

Anurag Verma and et al. Huffman. Diversity and scale: Genetic architecture of 2,068 traits in the VA million veteran program. medRxiv: the preprint server for health sciences, 2023. URL http://dx.doi.org/10.1101/2023.06.28.23291975.

Peter M Visscher, Naomi R Wray, Qian Zhang, Pamela Sklar, Mark I McCarthy, Matthew A Brown, and Jian Yang. 10 years of GWAS discovery: Biology, function, and translation. The American Journal of Human Genetics, 2017. doi: 10.1016/j.ajhg.2017.06.005.

Hung Q Vo and et al. Wang. Frozen large-scale pretrained vision-language models are the effective foundational backbone for multimodal breast cancer prediction. IEEE journal of biomedical and health informatics, 2025. URL http://dx.doi.org/10.1109/JBHI.2024.3507638.

Yimin Wang and et al. Li. Deep learning for spirometry quality assurance with spirometric indices and curves. Respiratory research, 2022. URL http://dx.doi.org/10.1186/s12931-022-02014-9.

Cristen J Willer, Yun Li, and Gonçalo R Abecasis. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics (Oxford, England), 2010. URL https://dx.doi.org/10.1093/bioinformatics/btq340.

J T Wright and M C Herzberg. Science for the next century: Deep phenotyping. Journal of dental research, 2021. doi: 10.1177/00220345211001850. URL http://dx.doi.org/10.1177/00220345211001850.

Ziqian Xie, Tao Zhang, Sangbae Kim, Jiaxiong Lu, Wanheng Zhang, Cheng-Hui Lin, Man-Ru Wu, Alexander Davis, Roomasa Channa, Luca Giancardo, Han Chen, Sui Wang, Rui Chen, and Degui Zhi. iGWAS: Image-based genome-wide association of self-supervised deep phenotyping of retina fundus images. PLoS genetics, 2024. doi: 10.1371/journal.pgen.1011273.

Qiushi Yang, Wuyang Li, Baopu Li, and Yixuan Yuan. MRM: Masked relation modeling for medical image pre-training with genetics. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2023. URL http://dx.doi.org/10.1109/iccv51070.2023.01961.

Taedong Yun, Justin Cosentino, Babak Behsaz, Zachary R McCaw, Davin Hill, Robert Luben, Dongbing Lai, John Bates, Howard Yang, Tae-Hwi Schwantes-An, Yuchen Zhou, Anthony P Khawaja, Andrew Carroll, Brian D Hobbs, Michael H Cho, Cory Y McLean, and Farhad Hormozdiari. Unsupervised representation learning on high-dimensional clinical data improves genomic discovery and prediction. Nature genetics, 2024a. doi: 10.1038/s41588-024-01831-6.

Taedong Yun, Justin Cosentino, and et al. Behsaz. Unsupervised representation learning on high-dimensional clinical data improves genomic discovery and prediction. Nature genetics, 2024b. doi: 10.1038/s41588-024-01831-6.

John R Zech, Marcus A Badgeley, Manway Liu, Anthony B Costa, Joseph J Titano, and Eric K Oermann. Confounding variables can degrade generalization performance of radiological deep learning models. arXiv [cs.CV], 2018. URL http://arxiv.org/abs/1807.00431.

Yukun Zhou and et al. Wang. Revealing the impact of pre-training data on medical foundation models. Research Square, 2025. URL http://dx.doi.org/10.21203/rs.3.rs-6080254/v1.