

Can Clustering of Deep Image Representations Reveal Latent Breast Cancer Risk Subtypes in a Multiethnic Population?

Lydia Sollis, Thomas Wolfgruber¹, Brandon Quon¹, Cade Kane², Dustin Valdez¹, Brenda Y. Hernandez¹, Karla Kerlikowske³, Peter Sadowski², and John A. Shepherd¹

¹University of Hawai‘i Cancer Center, ²University of Hawai‘i at Mānoa, ³University of California, San Francisco



Introduction

Deep learning (DL) has the potential to uncover hidden features predictive of breast cancer (BC) risk from mammographic images. Models like Mirai have outperformed traditional risk models, but their “black-box” nature limits explainability and clinical trust [1, 2]. One avenue for improving explainability is unsupervised clustering of learned image features. Clustering is widely used in healthcare for subgroup discovery in high-dimensional data [3]. It has been applied for risk stratification of patients using electronic health record (EHR) data with joint optimization of representation learning and subgroup discovery [4], as well as to analyze radiomic features of mammographic parenchymal complexity and their associations with BC risk factors such as age, BMI, and breast density [5]. While Gu et al. applied k-means clustering to risk factors identified through Cox regression models, followed by Kaplan–Meier survival curves to characterize breast cancer risk subtypes [6], to our knowledge, no study has applied unsupervised clustering directly to the latent features extracted from a DL-based survival model such as Mirai. This represents a novel opportunity to uncover clinically meaningful imaging-derived subgroups that reflect patterns learned by state-of-the-art risk models.

Methods

Study Design: We conducted a prospective analysis of women receiving 2D screening mammograms from 2009–2021 across sites in the Hawai‘i and Pacific Islands Mammography Registry (NIH R01CA263491, U54CA143728). Inclusion required diagnostic BI-RADS 0/1/2 mammographic exams with the standard four views, known race/ethnicity and breast density, and age 40–79. Cases were matched to controls (3:1) on birth year, race/ethnicity, first visit, and MG model. Cases also required ≥6 months between imaging and diagnosis, with known stage.

AI Risk Model & Clustering: Mammograms were preprocessed to PNG for ingestion into Mirai [1]. Clinical risk scores and 512-dimensional latent representations (mean-pooled across views) were extracted from Mirai’s final hidden image representation layer. K-means clustering was applied to these embeddings. The optimal 6-cluster solution was selected using the elbow and silhouette methods, and visualized with t-SNE as shown in Fig. 1. A Decision Tree classified clusters using clinical variables and explained clinical variable importance. **Statistical Analysis:** Cluster-level differences were tested via Kruskal–Wallis and chi-square with Bonferroni correction. **Model Evaluation:** Decision tree performance was assessed using accuracy, AUC, and Mean Decrease in Impurity interpretation of key predictors per cluster.

Results

Figure 3: Decision splits for Decision Tree trained on Mirai deep feature cluster assignments

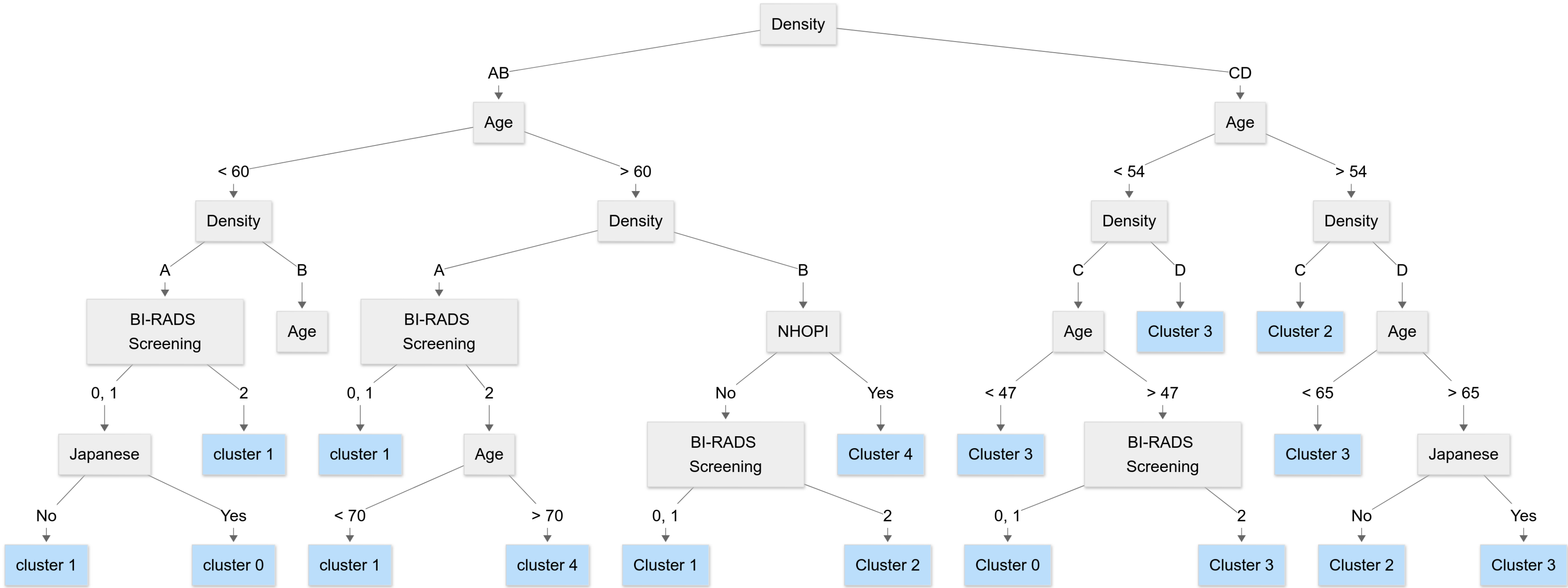


Table 2: Mean values for clinical variables per cluster, with clusters ordered from low to high Mirai risk values

Mean ± SD of Clinical Variables per Cluster						
Clinical Variable	Cluster 0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Breast Density	2.794 ± 0.657	1.892 ± 0.531	2.596 ± 0.567	3.117 ± 0.665	2.04 ± 0.509	2.706 ± 0.577
Menopause Status	0.678 ± 0.467	0.855 ± 0.352	0.868 ± 0.339	0.526 ± 0.499	0.949 ± 0.221	0.846 ± 0.362
BI-RADS Assessment	1.234 ± 0.555	1.367 ± 0.577	1.405 ± 0.648	1.338 ± 0.661	1.458 ± 0.663	1.483 ± 0.733
Prior Breast Cancer Diagnosis	0.003 ± 0.057	0.004 ± 0.062	0.017 ± 0.131	0.012 ± 0.107	0.008 ± 0.091	0.024 ± 0.153
Age at Mammogram	56.087 ± 9.205	61.925 ± 8.557	61.798 ± 8.174	52.612 ± 8.888	66.634 ± 7.128	61.429 ± 8.303
Mirai 5-Year Risk Assessment	0.012 ± 0.002	0.014 ± 0.003	0.019 ± 0.006	0.021 ± 0.009	0.031 ± 0.021	0.058 ± 0.038

Table 1: Patient-level counts for demographic and clinical risk data.

	Cases	Controls
Women, N	1,283	3,159
Age at First Exam		
40-49	324	720
50-59	418	975
60-69	395	1,062
70-79	146	402
Race/Ethnicity		
NHPI	272	508
White	182	138
Chinese	99	178
Filipino	161	476
Japanese	444	174
Other/NOS Asian	62	869
Other	16	43
Breast Density		
Fatty (A)	35	189
Scattered (B)	620	1,438
Heterogeneous (C)	524	1,272
Extremely dense (D)	128	260
Menopause Status		
Menopausal	904	2,314
Pre-menopausal	379	845

A total of 1,283 cases and 3,159 matched controls met our inclusion criteria. Table 1 shows the clinical and demographic characteristics of the study population. Across the six clusters derived from deep learning representations, significant differences were observed in nearly all key clinical variables (Bonferroni-adjusted $p < 0.008$), with variation in across groups visualized as the distribution in percent of each category in Figure 2 and mean and SD values in Table 2. These included breast density, menopausal status, BI-RADS screening scores, ethnicity, prior breast cancer history, age at mammogram, and Mirai 5-year breast cancer risk score. A Decision Tree model with a maximum depth of 5 was trained to predict cluster membership from clinical metadata, achieving an AUC of 0.74. The most predictive features were breast density, age at mammogram, and BI-RADS assessment, consistent with their Mean Decrease in Impurity values of 0.56, 0.32, and 0.10, respectively. Notably, the DT was unable to successfully categorize the highest risk cluster, based on clinical variables, as shown in Figure 3.

Conclusion

Clustering of deep learning-derived mammographic features reveals phenotypes that are statistically and clinically distinct, suggesting potential utility for personalized screening strategies. Future work will explore end-to-end deep clustering approaches, incorporate additional clinical variables, and apply advanced explainability techniques across external cohorts to further validate and interpret subgroup differences.

References

- [1] A. Yala et al., "Toward robust mammography-based models for breast cancer risk," Science Translational Medicine, vol. 13, no. 578, p. eaba4373, 2021, doi: doi:10.1126/scitranslmed.aba4373.
- [2] Z. X et al., "Deep Learning Predicts Interval and Screening-detected Cancer from Screening Mammograms: A Case-Case-Control Study in 6369 Women - PubMed," Radiology, vol. 301, no. 3, 2021, doi: 10.1148/radiol.2021203758.
- [3] M. Karim, M. Beyan, M. Zappa, and H. Martinetz, "Deep learning-based clustering approaches for bioinformatics," Briefings in Bioinformatics, vol. 22, no. 1, pp. 393–415, Jan 2021, doi: 10.1093/bib/bbz170.
- [4] H. Huang et al., "Deep significance clustering: A novel approach for identifying risk-stratified and predictive patient subgroups," Journal of the American Medical Informatics Association, vol. 28, no. 12, pp. 2641–2652, Dec 2021, doi: 10.1093/jamia/ocab192.
- [5] E. Kontos, R. Bakic, S. Kontos, and A. Maidment, "Radiomic Phenotypes of Mammographic Parenchymal Complexity: Toward Augmenting Breast Density in Breast Cancer Risk Assessment," Radiology, vol. 290, no. 1, pp. 41–49, Jan 2019, doi: 10.1148/radiol.2018180551.
- [6] Y. Gu et al., "Unveiling Breast Cancer Risk Profiles: A Survival Clustering Analysis Empowered by an Online Web Application," Future Oncology, vol. 19, no. 40, pp. 2651–2667, Oct 2023, doi: 10.2217/fon-2023-0736.
- [7] J. Tice, S. Cummings, R. Smith-Bindman, L. Ichikawa, W. Barlow, and K. Kerlikowske, "Using clinical factors and mammographic breast density to estimate breast cancer risk: development and validation of a new predictive model - PubMed," Annals of Internal Medicine, vol. 148, no. 5, 2008, doi: 10.7326/0003-4819-148-5-200803040-00004.

Figure 2: Clinical and demographic characteristics of significant clinical features by cluster

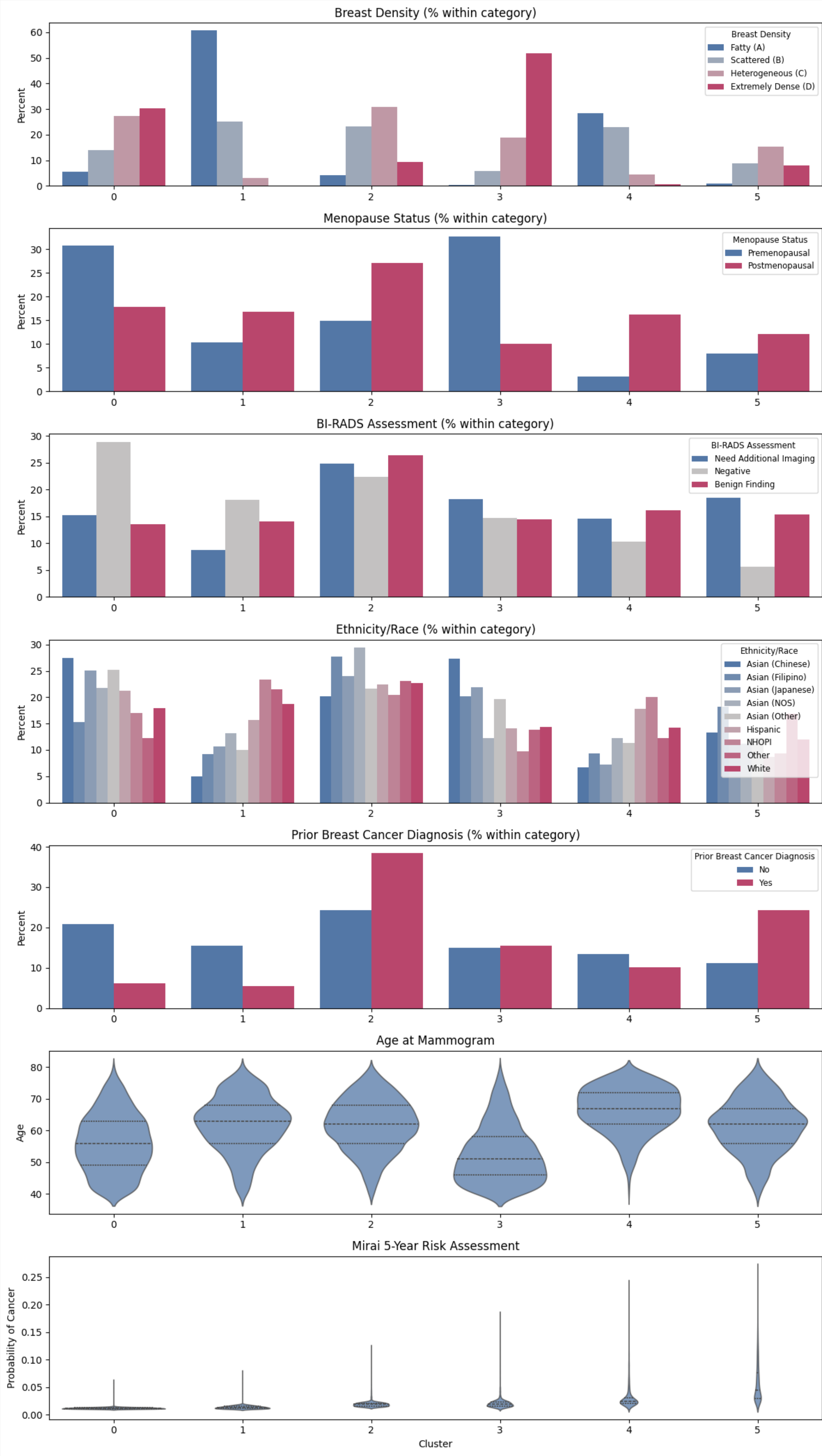


Figure 1: UMAP visualization of deep risk features grouped into k=6 clusters

