Figure 1: Clustering visualizations (All Features)

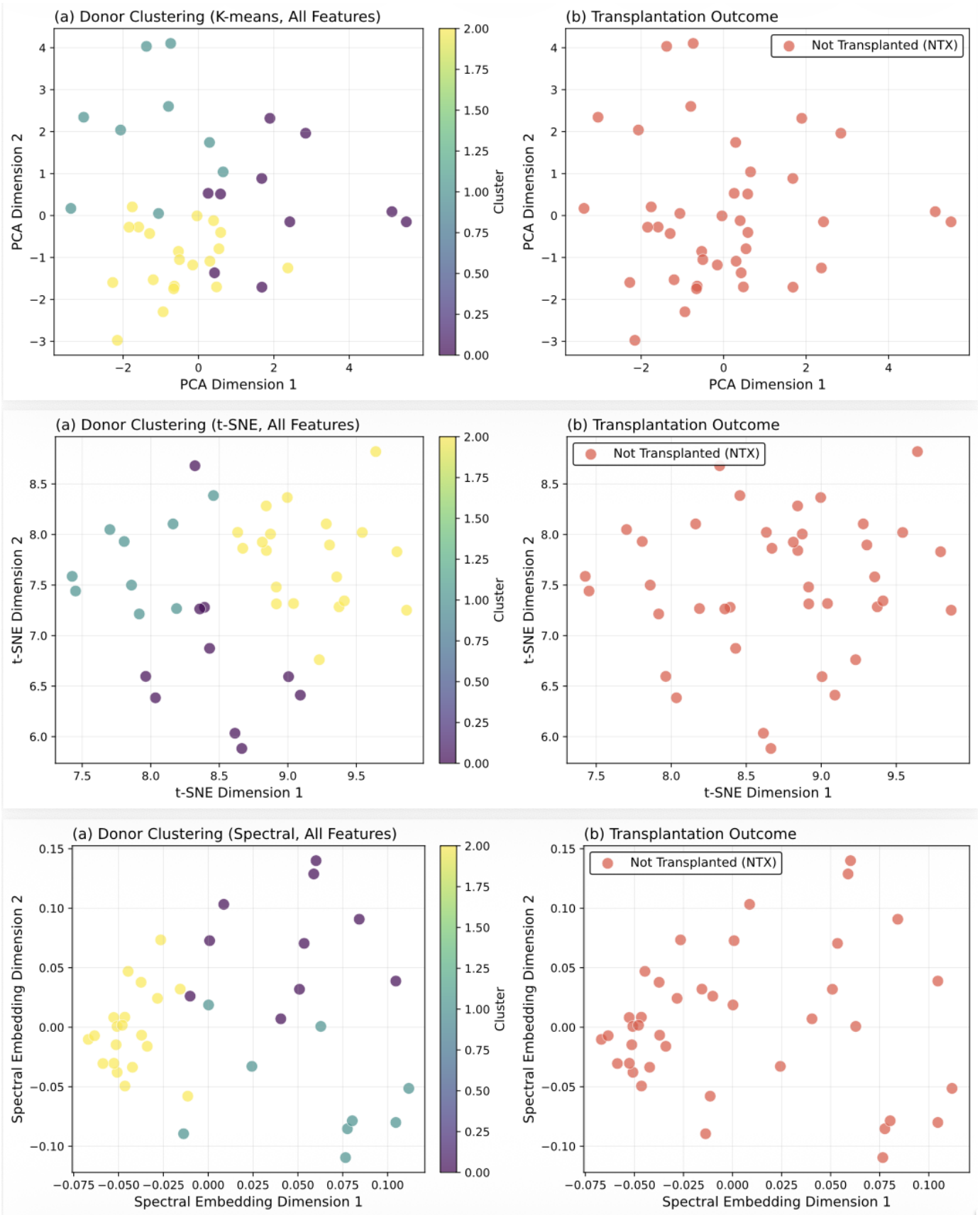


Figure 2: Clustering visualizations (Key Features from JSON)

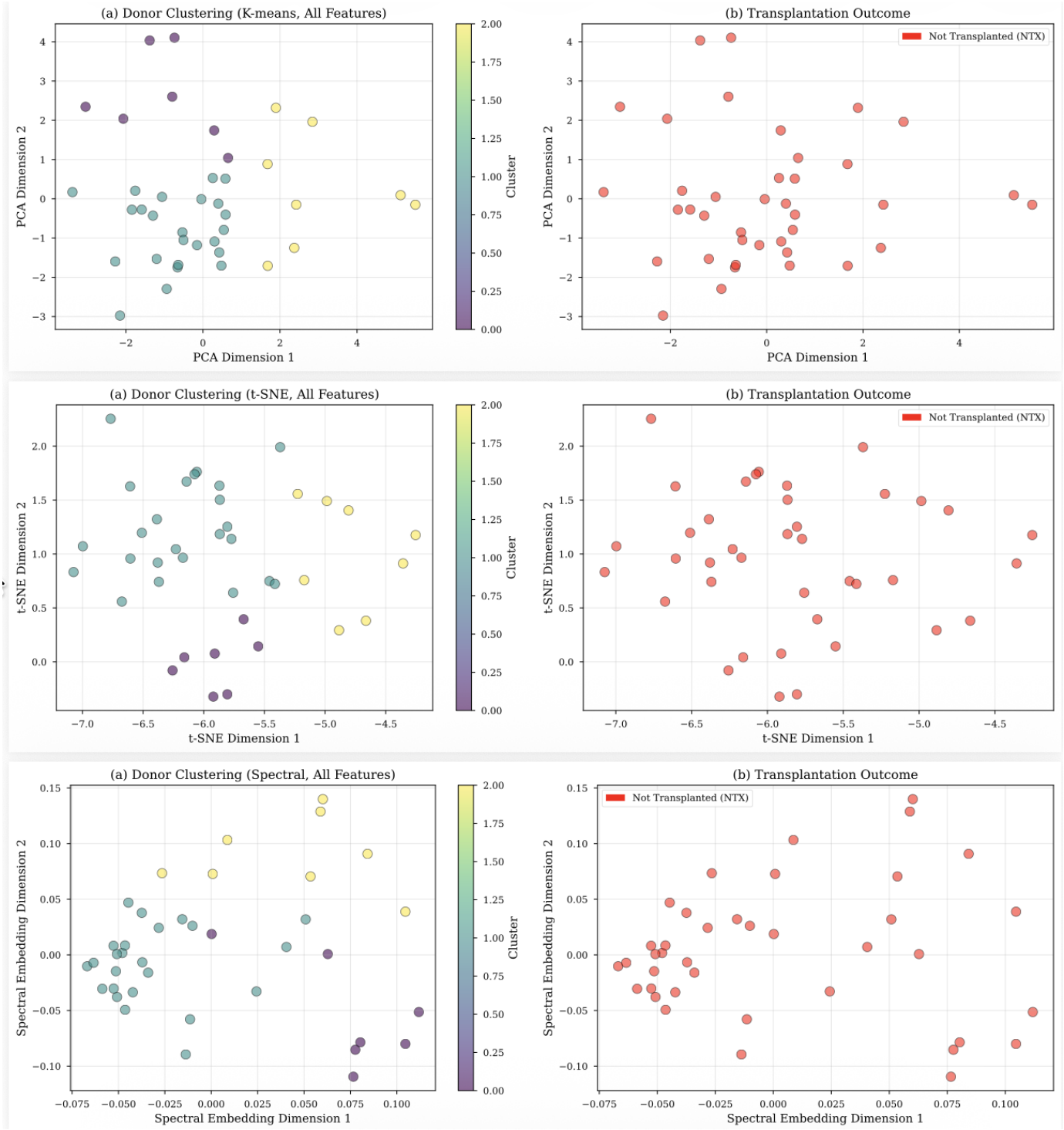
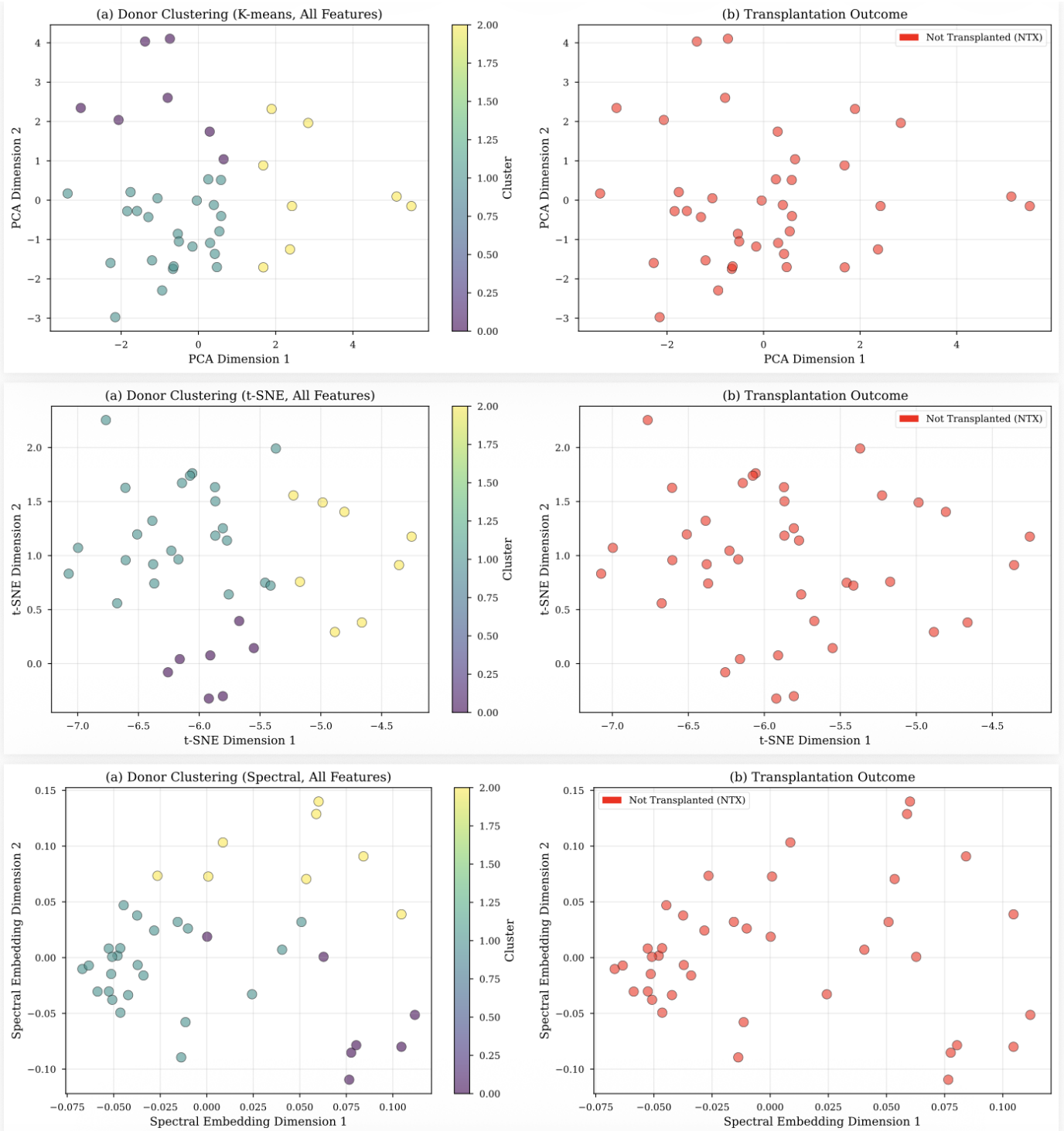


Figure 3: Clustering visualizations (Key Features from PDF highlights)



Interpretation Summary (Thesis-aligned)

Dataset: N=39 donor DIF-derived JSON records (hospital-internal).
Label (TX vs NTX): derived from presence of liver reperfusion timestamp fields in Transplantation sec

What the plots show:
• Left panels: unsupervised donor clusters (K-means) in a 2D embedding space.
• Right panels: the same embedding colored by the clinical outcome (TX/NTX).

Key patterns (consistent with your thesis narrative):
1) Heterogeneous donor subpopulations: clusters suggest multiple donor 'profiles' rather than a sing
2) Weak alignment between clusters and TX/NTX in raw feature space: outcome points are not clean
indicating that transplantability is driven by nuanced multi-factor interactions and missingness pat
3) Key-feature embeddings show slightly more structured geometry: when restricting to clinically me
the embedding often becomes more compact and interpretable, reflecting reduced noise and less

Clinical interpretation hook:
• Clusters can be treated as 'donor phenotypes' that a downstream SSL encoder should preserve and
• The mismatch between clusters and outcomes supports your motivation for representation learnin
the model must learn latent structure beyond naive proximity in raw tabular space.

Top feature-attribution signals (proxy) used for clinical discussion:

Feature (JSON field)	|weight|	Clinician key?
Main.WAIST_CIRCUMFERENCE	0.229	No
Therapy.THERAPY_MEDICATION___OTHER__	0.226	No
Therapy.THERAPY_MEDICATION_< ****** >	0.226	No
Main.STATUS_EFF	0.225	No
Main.STATUS_UTI	0.225	No
Main.BMI	0.198	Yes
UrineResults.URINERES_PROT_RESULT_+	0.196	No
Main.BLOOD_GROUP_B	0.178	Yes
Main.WEIGHT	0.178	Yes
Heart.ECG_RHYTHM_SR	0.177	No
Chest.NODULE_MASS_LR	0.174	No
AdmCircumstances.ADMISSION_CAUSE_OHCA	0.169	Yes

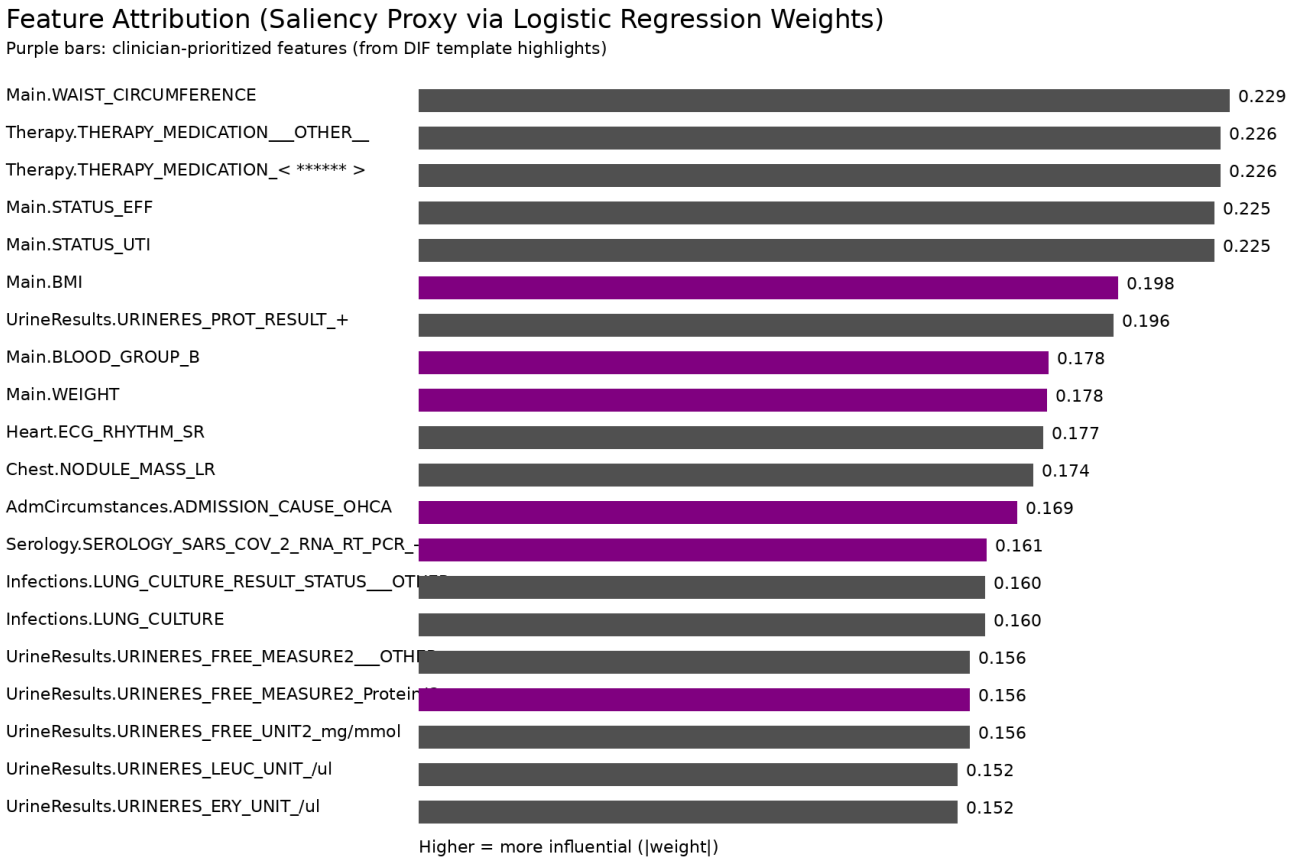



Figure 4: Feature attribution (Gradient-based Saliency Proxy)