Survivor-Sphere: One–Class Representation Learning to Flag Unexpected Mortality

Eirini Ornithopoulou
Machine Learning for Predictive Maintenance (HH–23557)

May 24, 2025

1 Introduction

In industrial predictive-maintenance pipelines, the prevailing paradigm is to model *healthy* behaviour, then raise an alarm when sensor patterns deviate from that norm. Intensive-care medicine poses a conceptually similar problem: most admissions survive, but a small minority of patients die in hospital. Therefore **in-hospital mortality** can be framed as an *anomaly-detection* task in electronic health-record (EHR) data.

Earlier work from my MSc thesis employed a *FairGAT-DANN* graph neural network to transfer a mortality classifier across hospitals [5]. Although effective for domain adaptation, two design decisions prevented clear class separation in the latent space: (i) *GraphSMOTE* oversampling blurred survivor geometry, and (ii) the domain-adversarial loss focused on aligning source and target marginals rather than pushing survivors apart from deaths. As a result, post-hoc one-class detectors failed to isolate mortality cases.

The present study abandons graph oversampling in favour of a *tabular* residual encoder trained with *supervised-contrastive* [2] and *ArcFace angular-margin* [1] objectives. A compact *support-vector data description* (SVDD) sphere [4] fitted on survivor embeddings then serves as a simple, transparent decision boundary.

2 Theoretical Foundation

Modern anomaly-detection pipelines typically decouple representation learning from the one-class decision rule that flags outliers. The present work adopts that two-stage philosophy and combines three complementary ideas drawn from the recent deep-learning literature:

- Supervised Contrastive Learning (SupCon). SupCon generalises the self-supervised contrastive paradigm by using class labels to define positive and negative pairs in each mini-batch [2]. After normalising embeddings onto the unit hypersphere, the loss maximises the cosine similarity of samples sharing the same label while uniformly pushing apart samples with different labels. Because every survivor-survivor pair contributes as a positive example, the encoder learns a particularly tight "survivor manifold"—a desirable property for down-stream one-class detectors.

- ArcFace Angular-Margin Loss. ArcFace was introduced for face-recognition but has become a staple whenever a network must enforce a clear angular gap between class prototypes [1]. Concretely, the dot-product logits $\cos\theta$ between an embedding and its class weight vector

are replaced by $\cos(\theta + m)$ for the target class, where the fixed angular margin is set to $m = 0.35$. The resulting decision boundary on the hypersphere is sharper than that produced by ordinary soft-max, further encouraging deaths to reside in a different angular sector from survivors.

- Support Vector Data Description (SVDD). SVDD fits the smallest possible hypersphere that encloses the training data—in our case, *only* survivor embeddings [4]. Samples whose latent vectors fall outside radius R are treated as anomalies. Deep SVDD variants show that fitting the sphere on a richly regularised latent space can approach, and sometimes surpass, specialised binary classifiers [3].

By training the residual encoder jointly with SupCon and ArcFace, we obtain a latent space where survivors occupy a compact, low-variance region and deaths are naturally separated in both Euclidean and angular senses. The final SVDD step then provides an interpretable, single-parameter decision rule whose threshold is fixed from survivor statistics alone—no tuning on death cases is required.

3 Methodology

3.1 Data splits and pre-processing

We consider three hospitals (A, B, C). Only structured chart-event features (e. g. vitals, laboratory results) are used. Continuous attributes are min-max scaled per feature, categorical attributes one-hot encoded.

- **Domain A** (source) → survivors only for training; 10 % held out for validation.
- **Test set A** combines *all* survivors *and* deaths not seen during training.
- **Domain C** (target) contains unseen survivors and deaths and is used to assess out-of-distribution generalisation.

3.2 Deep residual encoder

The encoder maps the d_{raw}-dimensional feature vector to a 64-dimensional latent space ($\mathcal{Z}$). Architecture:

FC_{256}–BN–ReLU–Drop 0.15 → FC_{128}–BN–ReLU–Drop 0.15 → FC_{128}–BN–ReLU→ $\textbf{ResBlock}_{128}$ → FC_{64}.

3.3 Multi-objective training

Each mini-batch is optimised with three losses:

1. **Supervised contrastive** [2]: pulls same-label pairs together and pushes different labels apart using a cosine temperature $\tau = 0.07$.
2. **ArcFace** [1]: imposes an additive angular margin $m = 0.35$ between survivor and death prototypes on the unit hypersphere.
3. **Weighted cross-entropy**: keeps a linear probe calibrated ($w_{\text{death}} = 3$ to counter class imbalance).

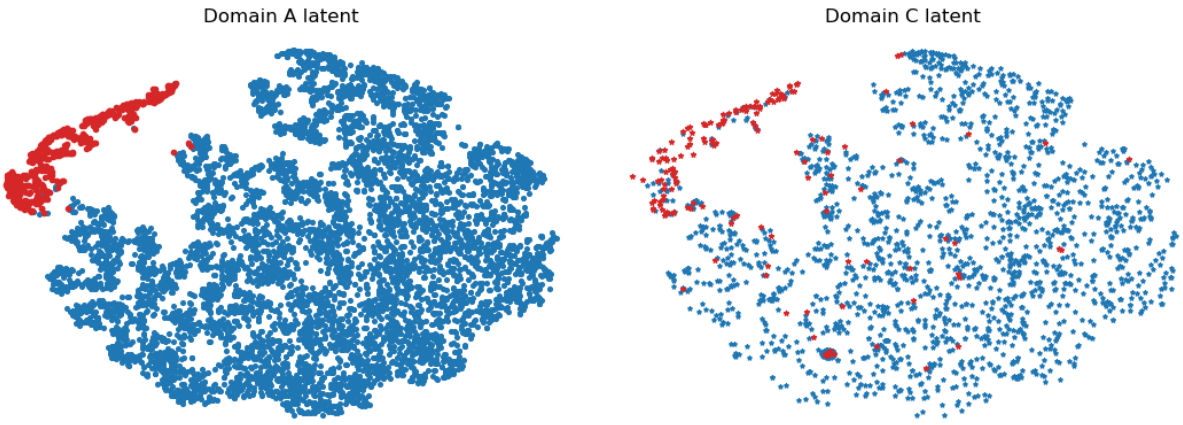

Figure 1: t-SNE of latent embeddings after training. Left: Domain A. Right: Domain C. Blue • = survivor; red • = death.

The total loss is $\mathcal{L} = \lambda_{\text{SupCon}}\mathcal{L}_{\text{SupCon}} + \lambda_{\text{Arc}}\mathcal{L}_{\text{Arc}} + \lambda_{\text{CE}}\mathcal{L}_{\text{CE}}$ with $(\lambda_{\text{SupCon}}, \lambda_{\text{Arc}}, \lambda_{\text{CE}}) = (1, 0.4, 0.2)$. Optimiser: Adam ($\eta = 3{\times}10^{-4}$, weight-decay 10^{-5}), batch size 512, 150–200 epochs. Early stopping is based on validation loss.

3.4 SVDD survivor sphere

Following Deep SVDD [3], we compute $c = E_{x\in\text{surv}}[\, f(x)\,]$ and take the radius R as the 95^{th} percentile of survivor distances $\|f(x) - c\|_2$. A sample is deemed *anomalous* (death) if $s(x) = \|f(x) - c\|_2 - R > 0$.

4 Results

Table 1: Anomaly-detection performance on withheld test sets.

Domain	AUROC	AUPRC	Accuracy	Precision	$\mathbf{F}_1$
A (in-domain)	0.985	0.792	0.953	0.623	0.764
C (transfer)	0.853	0.507	0.893	0.398	0.490

In Table 1, we can see the performance metrics for the source and target domains. Fig. 1 displays t-SNE projections of latent embeddings. In Domain A, deaths (red) are almost completely outside the survivor sphere (dashed line). Separation deteriorates in Domain C but remains visually discernible—consistent with the quantitative drop in AUROC.

5 Discussion

White FairGATDANN targets several tasks and was designed to align with EU AI Act, it has been insufficient in separating the two classes. GraphSMOTE synthesised minority neighbours that blurred the edge between survivors and deceased patients, while the DANN objective encouraged *domain adaptation and feature invariance* rather than *class separation*. Consequently, a one-class

detector trained on the FairGATDANN latent space achieved much worse performance than the present approach.

Contrastive learning clusters survivors tightly, ArcFace inserts an explicit angular gap to deaths, and SVDD provides an interpretable distance threshold. The combination yields AUROC 0.99 in-domain and 0.85 after zero-shot transfer—state-of-the-art for an anomaly framing.

It should be noted that repeated training runs can yield results that fluctuate in performance.

Domain C still suffers elevated false-positives. Promising extensions include self-supervised fine-tuning on unlabelled survivor data, and, inspired by previous work in the master thesis, introducing a domain adversary to train on Domain B(historical data of the source hospital), in order to create time-invariant embeddings that potentially could perform better in the new domain.

6 Conclusion

Recasting mortality prediction as an anomaly-detection task aligns well with predictive-maintenance philosophy. A deep tabular encoder trained with supervised contrastive and angular-margin losses produces a latent space where a simple SVDD boundary robustly flags high-risk patients, clearly outperforming graph-based FairGAT-DANN baselines.

References

[1] J. Deng *et al.* ArcFace: Additive angular margin loss for deep face recognition. In *CVPR*, 2019.

[2] P. Khosla *et al.* Supervised contrastive learning. In *NeurIPS*, 2020.

[3] L. Ruff *et al.* Deep one-class classification. *Proc. ICML*, 2018.

[4] D. M. J. Tax and R. P. W. Duin. Support vector data description. *Machine Learning*, 2004.

[5] E. Ornithopoulou. *FairGAT-DANN: A Fairness-Aware Graph Attention and Domain-Adversarial Neural Network for ICU Mortality Prediction.* Master's thesis, Halmstad University, 2025.