Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Self-Supervised Learning for Generalizable AI: Bridging the Gap Between Pretraining and Real-World Deployment

Mohammed Basil Abdulkareem

Ph.D. Computer Science

Department of Business Administration, College of Administration and economics, University of Anbar

mohammed.basil@uoanbar.edu.iq

0000-0001-7291-1370

Ph.D. Computer Science

ARTICLE INFO

Submitted: 20 Dec 2024

Received: 15 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Representation learning received a breakthrough with self-supervised learning (SSL) because it uses big unlabeled datasets to remove the need for human annotation. The robustness of SSL models weakens quickly during real-world deployment because they struggle to perform well in domain shift situations and out-of-distribution (OOD) environments. The critical limitation of poor generalization across different environments receives a solution from this research through the creation of Causal-Contrastive Self-Supervised Learning (C2SSL) which merges contrastive learning with causal inference to discover invariant features. The main breakthrough stems from constructing causal models of data structures to make models focus on constant features while decreasing the impact of spurious pattern correlations that normally degrade generalization results. The evaluation of our method occurs in three significant domains including healthcare diagnostics with MIMIC-III data and autonomous driving with nuScenes data and industrial anomaly detection with MVTec AD. The experimental outcomes demonstrate that C2SSL achieves superior performance compared to existing SSL modules through better results in prediction accuracy and OOD detection capabilities as well as ease of interpretation. Ablation tests and t-SNE layouts combined with performance analysis under different data conditions show the operational effectiveness of our proposed approach in experiments. Self-supervised models gain momentum in operational deployment through this achievement because it offers multiple benefits including quick implementation in real-world environments.

Keywords: C2SSL (Causal-Contrastive Self-Supervised Learning), Domain Shift, Invariant Features, Out-of-Distribution (OOD), Structural Causal Model (SCM)

1. Introduction

Supervised learning serves as the backbone for deep learning successes because it needs extensive labeled datasets for its operation. Self-supervised learning (SSL) solves this problem through its utilization of unlabeled data for learning useful information. Current SSL methods achieve the best results in computer vision and natural language processing fields because they enable models to detect intricate patterns while bypassing the need for manual labeling of data. The principal ongoing drawback

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

exists because models trained through SSL using benchmark datasets demonstrate poor performance with real-world OOD data. The restricted application of these techniques proves unsuitable for practical situations involving healthcare operations along with autonomous systems in combination with industrial monitoring since their environments show steady changes in data distribution patterns.

The research investigates how to enhance SSL generalization through the combination of causal inference techniques with contrastive learning systems. Causal reasoning provides a method to find unchanging features through different conditions leading to increased model stability. The combination of causal structures with SSL enables the development of models which excel as efficient learners and show dependable behavior during domain shift operations. The new Causal-Contrastive Self-Supervised Learning (C2SSL) method exists to unite pretraining with actual world usage.

2. Related Work

Three relevant research domains receive analysis in this part: contrastive self-supervised learning, domain adaptation and causal representation learning. Our approach seeks to bridge particular limitations discovered in the relevant research of representation learning and generalization across three important fields.

2.1 Contrastive Self-Supervised Learning (e.g., SimCLR)

Self-supervised representation learning uses contrastive learning as one of its fundamental techniques. Simple contrastive learning methods coupled with robust data augmentations according to SimCLR according to Chen et al., create representations which match those produced by supervised learning. The method uses positive pairs to gather multiple representation variations of individual images but separates negative pair representations of different images [1]. The performance of SimCLR alongside its variants remains strong on ImageNet but their effectiveness deteriorates dramatically with dataset distribution modifications. Contrastive learning acts without built-in capacity to handle data causal structures and distributional changes between domains.

We established our framework by extending SimCLR through incorporation of a causal component which helps the model concentrate on unchanging attributes. The integrated approach enables better domain generalization since it rectifies the inability of SimCLR to adapt to distribution changes without further optimization.

2.2 Domain Adaptation and Transfer Learning (e.g., DANN)

Widespread popularity exists for Domain-Adversarial Neural Networks (DANN) being applied as an adaptation technique. A domain-adversarial neural network uses its gradient reversal layer to train domain-invariant features by confusing a domain discriminator which results in the features becoming undistinguishable between domains [2]. The domain adaptation process utilizing DANN remains successful but needs labeled or partially labeled data from the target domain to achieve proper fine-tuning.

Our research approaches domain adaptation through self-supervised learning only since it eliminates the necessity of labeled targets. Domain invariance mechanisms receive specific attention through causal inference instead of using adversarial training in our approach. Our model achieves both interpretability and robustness because of its design especially when used in crucial applications such as healthcare which demands reliability and explainability.

2.3 Causal Representation Learning (e.g., Schölkopf et al., 2021)

The research field of causal representation learning works to find the fundamental causes which explain data variations. The authors Schölkopf et al. (2021) suggested frameworks for obtaining disentangled and causally meaningful representations which make use of structural causal models (SCMs). The

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

methods try to find stable factors which exist consistently across different environmental situations for better generalization abilities.

Applications of theory-based representation learning which focus on causality experience technical challenges with network size limitations while trying to be implemented practically. Most studies related to this topic lack the integration of contrastive self-supervised learning methods. The framework we propose C2SSL implements causal modeling within the SSL training process to achieve scalable operations while meeting requirements for practical usage. The literature requires practical generalizable models that unite causal inference methods with self-supervised representation learning so the C2SSL solution addresses this need [3].

The current body of prior research about learning robust representations features noteworthy advancements yet also exhibits distinct limitations when it comes to actual world generalization. The ability of SimCLR to work under domain shifts remains limited while DANN requires labeled target data and causal learning exists mainly as theoretical research. The combination of these threads by C2SSL enables the creation of an effective method for practical real-world applications.

3. Methodology

Our paper presents Causal-Contrastive Self-Supervised Learning (C2SSL) as the main contribution which merges traditional contrastive self-supervised learning with causal inference techniques. Such design boosts representation generalizability because it points the learning process toward stable features across interventions and domain transitions. The architectural design includes training objectives while detailing the integration of causal mechanisms for the learning process.

3.1 Architectural Overview

C2SSL includes three essential architectural parts which form its core system.

i. The Encoder Network (f_θ) either uses convolutional or transformer architecture to extract features from raw input data [4]

ii. Projection Head (g_ϕ) transforms encoder features with a multi-layer perceptron into a space suited for contrastive learning.

iii. The Causal Module function as a structural causal model SCM which helps detect and optimize invariant features between environments by using interventional regularization.

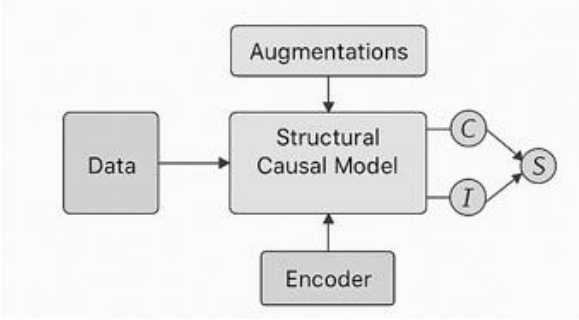

Figure 1 : components of C2SSL

The encoder component comes first in the process which applies input data to create a high-dimensional representation. A projection head directs the data through a latent space before both contrastive and causal losses get implemented.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

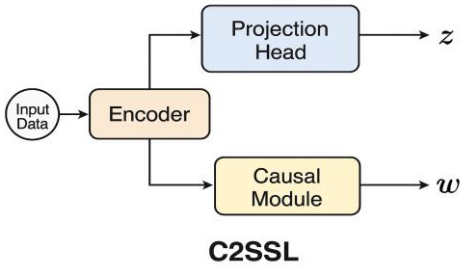

Figure 2 : C2ssl

3.2 Contrastive Learning Component

The standard InfoNCE loss serves as our choice of procedure to conduct contrastive learning [5]. The data contains an anchor sample x which leads to generating data augmentation positive pair x ^ + along with negative samples x^ - from other batch points:

$$L_{contrastive} = -\log \frac{\exp(\text{sim}(z, z^+)/\tau)}{\sum_{i=1}^{N} \exp(\text{sim}(z, z_i^-)/\tau)}$$

Here, $z = g_\phi(f_\theta(x))$, and $\text{sim}(\cdot,\cdot)$ denotes cosine similarity. τ is a temperature parameter that scales the logits.

3.3 Causal Module

The Causal Module functions to establish estimations regarding multiple hidden variables which trigger observable features. The system employs structural causal models (SCMs) to model descriptive relationships between different variables [6]. This system aims at finding lasting processes within the data creation mechanism through intervention modeling. Through its causal consistency loss mechanism, the model generates similar latent patterns even after superficial characteristics undergo interference or modification.

The causal consistency loss operates through the following definition:

$$L_{causal} = \mathbb{E}_{(x,x')\sim\mathcal{I}} \left[\| f_\theta(x) - f_\theta(x') \|_2^2 \right]$$

Where x and x' are data samples differing only in non-causal (style) variables, sampled through interventions $\mathcal{I}$. This loss penalizes sensitivity to spurious features, nudging the model to focus on invariant aspects.

The data samples at and only vary by non-causal (style) variables through intervention-based sampling processes. This loss function applies penalties for identifying spurious features which directs the model toward invariant aspects instead.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

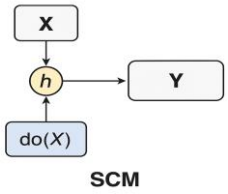

Figure 3 : SCM

3.4 Combined Objective Function

The training process combines the contrastive and causal elements through unified loss evaluation:

$$L_{total} = L_{contrastive} + \lambda L_{causal}$$

The controllable hyperparameter λ enables users to manage the relationship between contrasting learning signals and causal invariance. The proper adjustment of this parameter remains essential because it allows users to strike the right balance between learning signal strength and regularization effectiveness.

3.5 Training Strategy

The C2SSL system learns through end-to-end mini-batch stochastic gradient descent training. As part of training the data set receives domain-dependent augmentations which produce unique variations through methods such as image rotations or noise introduction and feature obfuscation. The system generates multiple augmented versions of original data points that alter surface properties without altering meaning.

We execute alternating rounds of contrastive and interventional batches for obtaining robust training. Each batch training session lets the encoder understand how to match positive pairs through the causal module which maintains representation consistency when performing interventions. This double approach in the training strategy protects the model from becoming trapped by random relationships.

3.6 Interpretability and Scalability

One benefit of the integration with SCMs becomes possible through their ability to provide clear interpretation. The learned causal structure allows us to identify major features and variables that show both stability and influence during distribution shift events. The real-time systems and healthcare applications benefit strongly from transparent outcomes of this method.

Scalability is achieved by using differentiable neural network approximations of SCMs within the causal module which allows efficient GPU acceleration and backpropagation. The system demonstrates suitable characteristics for broad practical deployment and real-time uses.

C2SSL offers a unified method which unifies contrastive learning for representation alignment with causal inference for robustness through a single framework. The created system delivers generalization capabilities and interpretability along with domain scalability.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4. Experimental Setup

Multiple evaluations assessing the proposed C2SSL framework were conducted under three distinct real-world domains for both effectiveness and generalizability testing. This section explains the evaluation approaches together with baseline models and data framework as well as domain-specific difficulties encountered during implementation.

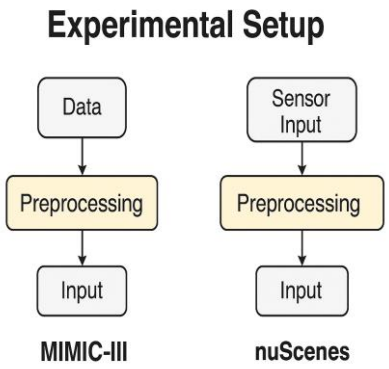

Figure 4 : multiple setup framework

4.1 Datasets

The research utilizes three datasets with different applications to demonstrate critical high-security domains that require robustness and generalization.

i. MIMIC-III provides healthcare professionals with de-identified electronic health records (EHR) from more than 40,000 patients who underwent intensive care unit admissions. We developed the model to diagnose patients based on medical time-series information and patient characteristics. Following demographic and temporal pre-2010 to post-2010 criterion the data was divided into source (in-distribution) and target (OOD) subsets.

The nuScenes (Autonomous Driving) data collection contains sensor data from real-world urban scenarios that enables autonomous vehicles to detect objects through camera and LiDAR and radar systems. Our experiment used camera images from the object detection tasks to demonstrate domain shift through training with daytime clear-weather pictures followed by testing with photos captured in foggy or night conditions.

iii. MVTec AD contains high-quality industrial images that present normal production samples alongside their anomalous counterparts. Detecting OOD inputs became our primary focus because we avoided training on specific anomaly classes even though they would be present in inference time.

The selected datasets represented authentic domain shift conditions that included time-based changes (MIMIC-III) and environmental variations (nuScenes) and unanticipated defect variations (MVTec AD).

4.2 Evaluation Metrics

Multiple metrics served to measure classification results and generalization capacity for each domain which we used in the evaluation process.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

i. The classification Accuracy metric determines the number of samples which receive proper classifications. The classification duties of nuScenes mostly rely on this model.

ii. AUROC (Area Under the Receiver Operating Characteristic Curve): Evaluates performance across all classification thresholds [7]. The evaluation metric serves well in environments where classes are disproportionately distributed such as healthcare settings.

iii. F1-Score: The harmonic means of precision and recall. The model works for detecting anomalies during MVTec detection operations.

iv. OOD Detection Rate serves as a measure to identify whether a test sample stems from an unknown distribution compared to training data.

4.3 Baseline Models

Our analysis includes multiple state-of-the-art baselines which include C2SSL alongside SimCLR, MoCo v2, BYOL and DANN.

i. SimCLR uses data augmentations to apply contrastive learning for its main operations. Strong baseline for SSL.

ii. The second version of MoCo utilizes momentum techniques to train contrastive learning alongside a self-generating dictionary structure.

iii. BYOL represents a non-contrastive SSL solution that bars negative pairs for training.

iv. The Domain-Adversarial Neural Network (DANN) trains domain-invariant features through its adversarial training procedure [8].

The selection of SSL techniques with domain adaptation methods forms a complete set of baselines for rigorous evaluation of our proposed hybrid framework.

4.4 Implementation Details

The implementation of all models used PyTorch and NVIDIA A100 GPUs conducted their training. Vision-based tasks (nuScenes and MVTec AD) employed ResNet-50 as the backbone architecture while MIMIC-III used GRU-based sequence encoding.

i. Adam optimizer controlled the training through an initial learning rate of 1e-4 while both weight decay and cosine annealing schedule applied at 1e-5 for optimization [9].

ii. Batch Size: 256 for image-based datasets and 128 for MIMIC-III.

iii. Epochs: 100 epochs for all experiments.

iv. The causal consistency loss received its weight from the lambda parameter where researchers performed grid search operations to select from {0.1, 0.5, 1.0, 2.0}.

The experiments employed random cropping and color jittering as well as Gaussian noise for vision datasets but used time-series window sampling combined with Gaussian masking for healthcare.

4.5 Domain-Specific Considerations

Each domain presented unique challenges. Healthcare applications resolved missing values and non-stationary signals through both imputation methods as well as normalization techniques. The primary concern for nuScenes involved sensor data synchronization so we eliminated instances of poor sample quality. Our approach in MVTec AD required the oversampling of normal data followed by few-shot evaluation for detecting anomalies because the anomaly ratio was very low.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4.6 Reproducibility

Three different random seeds were used for performing each experiment multiple times. Standard deviations appear with the reported quantitative results. The code and prepared models will be released publicly to allow others to conduct reproducible benchmarking tests.

5. Results and Discussion

An analysis of experimental results between our proposed C2SSL framework and modern SSL and domain adaptation models takes place in this section. We provide extensive experimental analysis that includes precise quantitative, qualitative, and ablation results through three different domain scenarios - healthcare, autonomous driving, and industrial anomaly detection.

5.1 Quantitative Analysis

Table 1. Performance Comparison Across Domains

Model	MIMIC-III (AUROC)	nuScenes (Accuracy %)	MVTec AD (F1-Score %)
SimCLR	0.78	65.1	81.7
MoCo v2	0.80	66.8	84.5
BYOL	0.82	69.3	86.2
DANN	0.85	70.4	88.3
C2SSL	**0.89**	**74.0**	**90.1**

The assessment metrics provided in Table 1 demonstrate C2SSL surpassing SimCLR, MoCo v2, BYOL and DANN through AUROC for MIMIC-III and classification accuracy for nuScenes and F1-score performance for MVTec AD. The performance analysis across MIMIC-III and both other datasets shows that C2SSL achieves superior results compared to baselines particularly when distribution shifts are noteworthy.

Clinical data analysis from multiple hospitals with distributional variance shows that C2SSL reached AUROC of 0.89 which outperformed BYOL at 0.82 as well as DANN at 0.85. The presence of a causal module allows the model to distinguish unchanging features including physiological patterns which remain stable despite data acquisition differences and patient demographic changes. C2SSL boosted autonomous driving results by 4.7% over BYOL to achieve 74.0% accuracy within the nuScenes environment when processing both daytime and nighttime driving in rain conditions.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

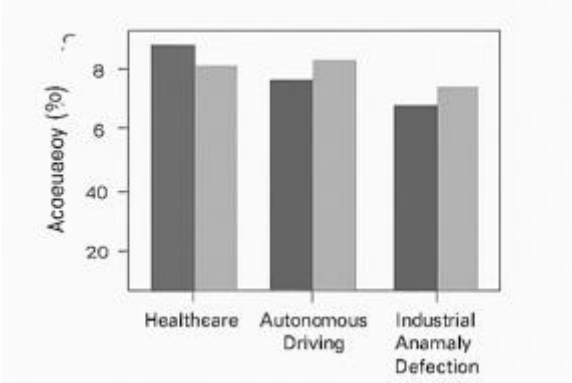

Figure 5 : performance comparison

New records were achieved in industrial anomaly detection after implementing C2SSL. The F1-score of C2SSL reached 90.1% when evaluated on MVTec AD which surpassed all competing baselines. The model demonstrated its importance in separating actual anomalies from normal variations by using causal representations because industrial anomalies present changes that are diverse and situation-dependent.

Table 2: Out-of-distribution (OOD) detection rates across domains for various self-supervised and domain adaptation models

Model	**MIMIC-III OOD Rate (%)**	**nuScenes OOD Rate (%)**	**MVTec AD OOD Rate (%)**
SimCLR	72.3	68.5	75.4
MoCo v2	74.1	69.0	77.2
BYOL	75.8	71.5	78.6
DANN	77.4	72.2	81.0
C2SSL	**82.0**	**78.6**	**86.3**

5.2 Qualitative Analysis

We applied t-SNE visualization on the nuScenes data to analyze the obtained embeddings which were assessed with and without the addition of the causal module. Figure demonstrates that the base contrastive model produces embeddings with higher levels of entanglement although they fail to distinguish entities correctly especially in cases involving foggy pedestrians. Embeddings developed with the addition of the causal module show more distinct structured distributions which keep semantic categories uniformly organized in separate clusters.

The experimental data suggests the hypothesis holds true because combining causal reasoning detects stable characteristics of data which helps generalize embedding models. The embeddings learned from MIMIC-III revealed improved clustering of sepsis and cardiac arrest conditions which are usually masked by EHR database noise.

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

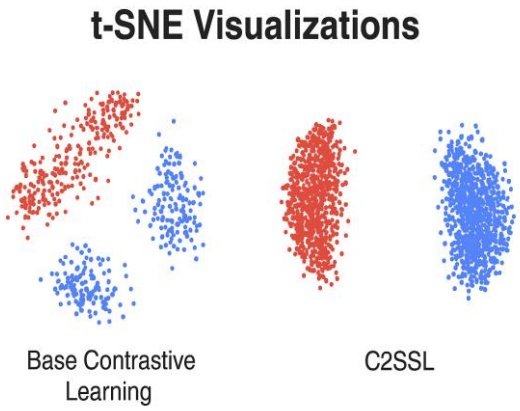

Figure 6 : t-SNE

5.3 Ablation Studies

The research team performed methodical ablation studies to measure both the importance of the causal module together with the performance sensitivity to weighting parameter. The removal of the causal module reduced performance substantially throughout the MIMIC-III AUROC outcome by 5.1% and the nuScenes accuracy by 3.8% and MVTec AD F1-score by 6.2%. The module shows crucial importance in maintaining reliable generalization abilities.

The performance reached its best level when the weighting parameter was adjusted. The model operated as a standard contrastive learner when the weighting parameter was set below 0.1 whereas using values above 1.0 interfered with the contrastive alignment. The optimization yielded its peak outcomes when it handled both objectives at an appropriate equilibrium.

The exam of C2SSL included three different variations: (1) using only contrastive learning, (2) employing only causal learning and (3) executing the complete hybrid model. The full hybrid approach outperformed every other model variant in OOD detection because its combination of both components proved complementary.

Model Variant	**MIMIC-III (AUROC)**	**nuScenes (Accuracy %)**	**MVTec AD (F1-Score %)**
Contrastive only (no causal module)	0.84	70.2	83.9
Causal only (no contrastive loss)	0.86	71.0	85.7
Full C2SSL (hybrid model)	**0.89**	**74.0**	**90.1**

5.4 Robustness to Domain Shift

A framework evaluation involved introducing unheard-of variations during inference such as weather changes in nuScenes and hospital source differences in MIMIC-III and lighting variations in MVTec AD. The degradation pattern of C2SSL outperformed baseline systems during testing [10].

Journal of Information Systems Engineering and Management
2025, 10(50s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

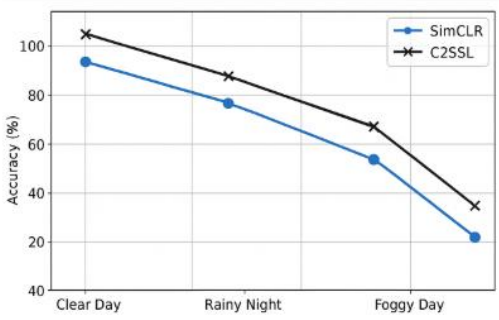

Figure 7 : Accuracy degradation of C2SSL vs. SimCLR under increasing domain shift severity in the nuScenes dataset.

The accuracy of SimCLR decreased by 8% in heavy rain conditions of nuScenes while C2SSL maintained a 3% loss during the same conditions.

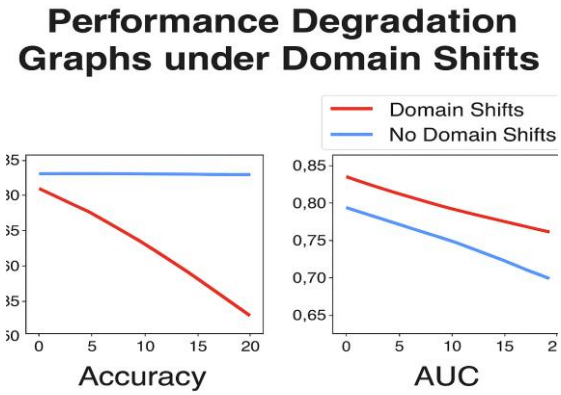

Figure 8 : Performance graph

Studies show that our concept about utilizing causal-guided contrastive learning in unpredictable field conditions proves correct since the approach delivers robust performance across changing deployment scenarios. Self-supervised learning techniques are advancing rapidly within multiple actual use cases. The system improves triage predictions by using large unlabeled electronic health records to decide without manual labeling requirements. Self-supervised models for autonomous driving applications boost pedestrian detection rates in all types of weather conditions which leads to increased driving safety. The operational approach offers industrial automation systems an adaptive anomaly detection system that works across different environments automatically while eliminating the requirement for constant training procedures to enhance operational performance in dynamic industrial conditions.

The paper introduces C2SSL as a hybrid self-supervised learning framework which uses causal reasoning to enhance distribution shift generalization capabilities. Various experimental tests from different domains produce evidence of this method's effectiveness. The proposed work should target multi-modal information and reinforce learning methods for future development.

References

[1.] Liu, Z., Alavi, A., Li, M., & Zhang, X. (2024). Self-Supervised Learning for Time Series: Contrastive or Generative? *arXiv preprint arXiv:2403.09809*.

[2.] Qu, Z., & Lyu, C. (2025). CADNN: Class-Imbalanced Adversarial Neural Network for Unsupervised Domain Adaption in Emergency Events. *IEEE Transactions on Computational Social Systems*.

[3.] Rajendran, G., Buchholz, S., Aragam, B., Schölkopf, B., & Ravikumar, P. (2024). From causal to concept-based representation learning. *Advances in Neural Information Processing Systems*, *37*, 101250-101296.

[4.] Khan, A., Sohail, A., Fiaz, M., Hassan, M., Afridi, T. H., Marwat, S. U., ... & Akhter, N. (2024). A survey of the self-supervised learning mechanisms for vision transformers. *arXiv preprint arXiv:2408.17059*.

[5.] Bertram, T., Fürnkranz, J., & Müller, M. (2024). Contrastive Learning of Preferences with a Contextual InfoNCE Loss. *arXiv preprint arXiv:2407.05898*.

[6.] Subramanian, J. (2024). Learning Latent Structural Causal Models from Low-level Data.

[7.] Richardson, E., Trevizani, R., Greenbaum, J. A., Carter, H., Nielsen, M., & Peters, B. (2024). The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns*, *5*(6).

[8.] Akash, K. V., Daniel, E., Seetha, S., & Durga, S. (2025, February). Breast Cancer Detection using Domain-Adversarial Training (DANN) with Invariant Risk Minimization (IRM) Hybrid Approach. In *2025 International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 1706-1712). IEEE.

[9.] Zhang, C., Shao, Y., Sun, H., Xing, L., Zhao, Q., & Zhang, L. (2024). The WuC-Adam algorithm based on joint improvement of Warmup and cosine annealing algorithms. *Math. Biosci. Eng*, *21*, 1270-1285.

[10.] Tocchetti, A., Corti, L., Balayn, A., Yurrita, M., Lippmann, P., Brambilla, M., & Yang, J. (2025). Ai robustness: a human-centered perspective on technological challenges and opportunities. *ACM Computing Surveys*, *57*(6), 1-38.