

ENHANCING MEDICAL DIAGNOSTICS: A CONTRASTIVE LEARNING APPROACH TO MULTI-MODAL MEDICAL IMAGE SEGMENTATION

Vivekanandan Govindan Ekambaram

Senior Data Scientist, Kforce Inc
Milpitas, California, USA

Received: 18/08/2026

Revised: 21/09/2026

Accepted: 05/10/2026

ABSTRACT:

Modern diagnosis rarely depends on a single image. Clinicians combine magnetic resonance imaging, positron emission tomography, radiographs, and laboratory results, yet most segmentation models are trained on one modality at a time and fail when the modality, scanner, or institution changes. Pooling data across institutions would help, but patient privacy rules make sharing difficult. This paper presents a contrastive learning approach to multi-modal medical image segmentation that addresses these three obstacles together. A fusion stage built on a variational transformer autoencoder combines modalities into local latent representations. A contour-aware generative adversarial network produces synthetic examples whose boundaries follow anatomical contours, and a contrastive loss aligns synthetic and real examples with the local latents so that the representation is consistent across modalities and domains. A differentially private synthetic data generation component provides a formal privacy guarantee when data leave an institution. We evaluate the approach on cross-modality brain tumor segmentation, multi-modal image fusion, and interpretability on two diagnostic classification tasks, Parkinson's disease imaging and melanoma. In cross-modality segmentation, the method reached a Dice score of 0.789, compared with 0.727 for the strongest domain adaptation baseline and 0.836 for fully supervised training, and reduced the 95th percentile Hausdorff distance to 9.1 millimeters. Removing the contrastive loss lowered Dice by 4.5 points, the largest single loss among the ablations, and a privacy budget of 8 cost about 1.2 points. Attention maps from the contrastive model concentrated on the striatum in Parkinson's imaging and on the lesion in melanoma images, in contrast to diffuse attention from single-modal baselines.

Keywords: Medical Image Segmentation, Multi-Modal Learning, Contrastive Learning, Generative Adversarial Network, Transformer, Differential Privacy, Domain Adaptation, Image Fusion

INTRODUCTION

Medical image segmentation, the delineation of organs, lesions, and other structures in images, underlies diagnosis, treatment planning, and monitoring of disease. Accurate outlining of a brain tumor determines the dose and the field of radiotherapy, and measurement of a lesion's extent guides decisions about surgery. Deep learning has become the standard approach, and encoder-decoder networks and, more recently, transformer-based architectures achieve high accuracy when trained and tested on matched data [1], [12], [13], [14], [27]. Yet three obstacles prevent such systems from reaching their full clinical value.

The first obstacle is that diagnosis is multi-modal. Magnetic resonance imaging shows soft-tissue structure, positron emission tomography shows metabolic function, radiographs show bone and chest anatomy, and laboratory and clinical records add non-image context. Each carries complementary information, and a model that sees several of them can in principle be more accurate and more robust than one that sees only a single modality [33]. Most published models, however, are trained on one modality, and combining modalities by simple concatenation of inputs ignores the differences in their statistics and the relationships between them.

The second obstacle is domain shift. A model trained on images from one scanner, protocol, or modality typically degrades when applied to another, because intensity distributions and textures differ. Domain adaptation methods try to close the gap using unlabeled target data, with approaches that align output structure, translate images between domains, or align both images and features [8], [10], [11], [28], but their performance often remains well below that of models trained with labels in the target domain. The third obstacle is privacy. Medical images and

records are sensitive, regulations restrict their sharing, and institutions are reluctant to pool data, which limits the diversity of the training sets that robust models need. Federated learning and differentially private methods offer partial solutions, but they have not been combined with multi-modal contrastive segmentation in a unified design [16], [17], [18], [19].

Contrastive learning offers a way to address the first two obstacles. By pulling together the representations of examples that correspond to the same underlying content and pushing apart those that do not, contrastive objectives learn representations that are invariant to nuisance variation such as modality-specific appearance [4], [5], [6], [7]. In a multi-modal medical setting, the same anatomy imaged in different modalities, or a real image and a synthetic counterpart with the same contour, form natural positive pairs. Generative models can supply such counterparts, and if they are trained with differential privacy, they can also serve as a privacy-preserving vehicle for sharing information across institutions [9], [32].

This paper presents a contrastive learning approach to multi-modal medical image segmentation built on these ideas. The overall framework, shown in Fig. 1, takes multi-modal medical data through a fusion and contrastive learning stage, a variational transformer autoencoder that produces local latent representations, and a contour-aware generative adversarial network, with a differentially private synthetic data generation component providing a privacy guarantee. We evaluate the approach on cross-modality brain tumor segmentation, in which its predictions are compared with those of domain adaptation baselines in Fig. 2; on multi-modal image fusion of magnetic resonance and positron emission tomography images, compared in Fig. 3; and on the interpretability of the learned representation in two diagnostic tasks, Parkinson's disease and melanoma, shown in Fig. 4.

The contributions of this paper are as follows:

- A unified framework that combines transformer-based multi-modal fusion, contour-aware synthetic example generation, and a contrastive objective aligning synthetic and real examples with local latents.
- A privacy mechanism based on differentially private synthetic data generation, with a quantified trade-off between the privacy budget and segmentation accuracy.
- A cross-modality segmentation evaluation against domain adaptation baselines and a fully supervised upper bound, with an ablation of each component.
- A comparison of multi-modal fusion quality against a range of fusion methods, in both qualitative and quantitative terms.
- An interpretability analysis showing that the learned multi-modal representation focuses on clinically relevant regions in two diagnostic tasks.

The remainder of the paper is organized as follows. Section II reviews the literature. Section III discusses related work and identifies gaps. Section IV presents the methodology. Section V describes the experimental setup. Section VI reports results, Section VII discusses implications and limitations, and Section VIII concludes.

LITERATURE REVIEW

A. Deep Learning for Medical Image Segmentation

The U-shaped encoder-decoder network with skip connections became the foundation of medical image segmentation, because it combines coarse context with fine localization and works with small datasets [1]. Surveys of deep learning in medical image analysis document the rapid spread of such models across modalities and organs [27]. Vision transformers, which model long-range relationships through self-attention over image patches, were then introduced to vision [12] and adapted to medical segmentation, either in hybrid designs that combine convolutional encoders with transformer layers [13] or in fully transformer-based encoders for volumetric data [14]. Public challenges, notably the multi-modal brain tumor segmentation challenge, provide standardized multi-sequence magnetic resonance data with expert annotations that have driven progress [2], [3].

B. Multi-Modal Learning and Image Fusion

Multi-modal learning aims to use several sources of information that describe the same phenomenon. A taxonomy of the field distinguishes representation, translation, alignment, fusion, and co-learning, and stresses that the challenge lies in the heterogeneity of modalities and in the lack of alignment between them [33]. In medical imaging, image fusion combines images from different modalities, for example structural magnetic resonance and functional positron emission tomography, into a single image that retains the anatomy of the first and the functional information of the second. Traditional fusion methods operate in transform domains or on multi-scale decompositions [31], whereas learning-based methods use convolutional networks to extract and merge features

[23]. A good fused image preserves edges and structure, retains functional contrast, and avoids artifacts and color distortion. Fusion at the representation level, inside a segmentation network, is an alternative that avoids creating an image at all.

C. Contrastive Representation Learning

Contrastive learning trains an encoder so that representations of positive pairs are close and those of negative pairs are far. A widely used objective, introduced with contrastive predictive coding, maximizes a bound on the mutual information between paired representations and uses other examples in the batch as negatives [4]. Frameworks for visual representation learning showed that strong augmentation and large numbers of negatives lead to representations that rival supervised ones [5], [6]. Supervised variants use labels to define positives and have been shown to improve robustness [7]. In medical imaging, contrastive learning is attractive because labeled data are scarce and because positive pairs arise naturally from different views, modalities, or time points of the same patient.

D. Generative Models and Synthetic Data

Generative adversarial networks learn to produce realistic samples through a game between a generator and a discriminator [9]. Conditional variants generate images from conditioning inputs such as edge maps or segmentation masks [20], and cycle-consistent translation between domains without paired data has been widely used for modality translation and domain adaptation [8]. Variational autoencoders learn a latent distribution of the data and provide a principled way to model variation [15]. In medical imaging, synthetic images can augment training sets, translate between modalities, and, if generated under a privacy mechanism, can be shared in place of real data. Contour conditioning is particularly relevant for segmentation, because it ties the generated image to a known boundary, so that the synthetic example comes with a label.

E. Domain Adaptation in Medical Imaging

Domain adaptation reduces the performance gap between a source domain with labels and a target domain without them. Output-space adversarial adaptation aligns the structure of predictions between domains [10]. Image-translation approaches convert source images into the appearance of the target [8], and methods that combine image alignment with feature alignment in both directions have achieved strong results in cross-modality medical segmentation [11]. A survey of domain adaptation in medical image analysis organizes the field by the type of shift, the available supervision, and the approach, and notes that the gap to fully supervised models remains substantial in many tasks [28]. Contrastive objectives provide a different route to domain invariance, by explicitly pulling together representations of the same content in different domains.

F. Privacy in Medical Machine Learning

Differential privacy provides a mathematical guarantee that the output of an algorithm does not reveal much about any single individual's record, controlled by a privacy budget in which smaller values mean stronger protection [16]. It can be applied during training by clipping and adding noise to gradients [17], and generative models can be trained under differential privacy so that the synthetic data they produce inherit the guarantee [32]. Federated learning trains a shared model across institutions without moving raw data [18], and has been proposed as a foundation for digital health [19]. A known difficulty is the loss of accuracy at strong privacy levels, which makes the privacy and utility trade-off a central topic of evaluation.

RELATED WORK

Related work can be grouped into multi-modal segmentation, domain-adaptive segmentation, contrastive approaches to medical imaging, and privacy-preserving learning.

Multi-modal segmentation. Models for brain tumor segmentation typically take multiple magnetic resonance sequences as input channels and are trained on the challenge data [2], [3]. Channel concatenation is simple but treats modalities symmetrically and breaks down when a modality is missing at test time. Transformer-based fusion offers a more flexible way to relate modalities [13], [14], but is usually trained with supervised losses alone. **Domain-adaptive segmentation.** AdaOutput, image translation with cycle consistency, and synergistic image and feature alignment represent successive advances in adapting segmenters without target labels [8], [10], [11]. They reduce the gap to supervised training but do not provide privacy guarantees and do not use contrastive objectives to link modalities.

Contrastive approaches to medical imaging. Contrastive pretraining improves label efficiency and robustness in medical imaging, as in natural images [5], [6], [7]. Most applications use augmentations of a single image as positives, whereas multi-modal and synthetic counterparts offer richer positives that are specific to medicine.

Privacy-preserving learning. Differentially private training and differentially private generative models provide formal guarantees [16], [17], [32], and federated learning avoids sharing raw data [18], [19]. The use of differentially private synthetic data as a vehicle for multi-modal contrastive alignment across institutions has received little attention.

TABLE I. Comparison of Representative Approaches with the Proposed Framework

Approach	Multi-modal fusion	Contrastive objective	Contour-aware synthesis	Domain adaptation	Privacy guarantee
Segmentation networks [1], [13], [14]	Partial	No	No	No	No
Domain adaptation [8], [10], [11]	No	No	Partial	Yes	No
Contrastive pretraining [5]-[7]	No	Yes	No	Partial	No
Private learning [16]-[19], [32]	No	No	Partial	No	Yes
Proposed framework	Yes	Yes	Yes	Yes	Yes

Table I shows that existing work addresses the obstacles separately. The proposed framework combines multi-modal fusion, a contrastive objective, contour-aware synthesis, domain adaptation, and a privacy guarantee in one design.

METHODOLOGY

A. Overview of the Framework

The framework is shown in Fig. 1. Multi-modal medical data, illustrated in the figure by a brain image, a laboratory record, and a chest radiograph, enter a multi-modal fusion and contrastive learning stage, labeled GGFCL in the figure. This stage has two outputs. A hierarchical signal, shown by the dashed arrow, goes to the variational transformer autoencoder, labeled VTAE, which encodes the data and produces local latent representations. A second output goes to the contour-aware generative adversarial network, labeled CA-GAN, which synthesizes examples guided by contours. A differentially private synthetic data generation component, labeled DP-SDG, supplies a privacy guarantee to the fusion stage, which is indicated by the lock symbol in the figure. A contrastive loss, labeled InfoNCE, compares synthetic and real examples with the local latents, and its gradient shapes both the autoencoder and the generator.

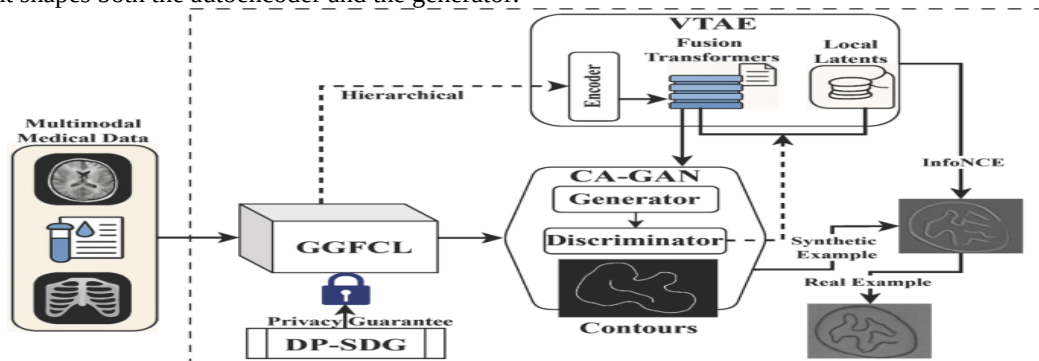


Fig. 1. Overall framework: multi-modal medical data pass through the fusion and contrastive stage (GGFCL), the variational transformer autoencoder (VTAE), and the contour-aware GAN (CA-GAN), with differentially private synthetic data generation (DP-SDG) providing the privacy guarantee and a contrastive loss (InfoNCE) aligning synthetic and real examples.

B. Variational Transformer Autoencoder with Fusion Transformers

The autoencoder in the upper part of Fig. 1 consists of an encoder, a stack of fusion transformers, and a set of local latents. Each modality is first processed by its own encoder, which converts the image into a sequence of patch tokens, with a learned embedding that identifies the modality. The fusion transformers, drawn as a stack of layers in the figure, apply self-attention over the tokens of all modalities together, so that a patch in one modality can attend to related patches in another, and non-image information such as laboratory values is added as additional tokens. The output is a set of local latents, each describing a small region with information from all available modalities. The latents are modeled as distributions, as in a variational autoencoder, so that the representation is regularized and sampling is possible [15]. Because the attention operates over whichever tokens are present, the model can handle the absence of a modality at inference by omitting its tokens, a property that matters in practice because not every patient receives every examination.

C. Contour-Aware Generative Adversarial Network

The generative network in the lower part of Fig. 1 has a generator and a discriminator, with contours as conditioning input. A contour, shown in the figure as a white outline on a dark background, delineates the boundary of the structure of interest, such as a tumor or an organ. The generator receives the contour together with the fused representation from the transformer stack and produces a synthetic image whose content is consistent with the contour; the discriminator tries to tell synthetic images from real ones, conditioned on the same contour [9], [20]. Because the synthetic image is built around a known contour, it comes with a label for free. In the figure the discriminator also sends a feedback signal, shown as a dashed arrow, back to the fusion transformers, so that the representation is shaped by what the discriminator finds realistic. In addition to its training role, the generator can produce synthetic examples in a target modality from a source image and contour, which serves as a form of modality translation and as augmentation for domain adaptation [8].

D. Contrastive Alignment

The contrastive objective, labeled InfoNCE in Fig. 1, links the two halves of the system [4]. For a given anatomical structure, the real example and the synthetic example generated from its contour form a positive pair, and the local latents of the autoencoder, computed from the multi-modal input, are also positives for that structure. Examples from other structures and other patients in the same batch are negatives. The loss pulls the positives together in representation space and pushes the negatives apart. Three kinds of invariance result. Representations become invariant to the modality in which a structure is imaged, because positives span modalities. They become invariant to the difference between real and synthetic appearance, which makes synthetic data a more effective substitute for real data. And they become sensitive to anatomical content, because negatives differ in contour and location. The loss is combined with a reconstruction term for the autoencoder, an adversarial term for the generator and discriminator, and a segmentation term, so that the representation is useful for the final task.

E. Privacy Guarantee through Differentially Private Synthetic Data Generation

To allow information to leave an institution without exposing patients, the framework includes the differentially private synthetic data generation component shown with a lock in Fig. 1. Training of the generative network at each institution applies per-example gradient clipping and calibrated noise, so that the trained generator satisfies a differential privacy guarantee with a specified budget [16], [17], [32]. Only the generator, or synthetic examples drawn from it, are shared with other institutions or with a central site, and they can be used there for contrastive alignment and for training the segmentation head. Because differential privacy is preserved under post-processing, any computation on the shared synthetic data inherits the same guarantee. The privacy budget is a parameter that sets the trade-off between protection and accuracy, and we study it explicitly. Smaller budgets give stronger protection and noisier generators.

F. Segmentation Head and Training

The segmentation head takes the local latents and decodes them into a per-pixel label map. Training proceeds in two phases. In the first phase, the autoencoder, the generative network, and the contrastive alignment are trained together on unlabeled multi-modal data and on synthetic examples. In the second phase, the segmentation head is trained with the available labels, with synthetic examples and their contours providing additional labeled data, and the whole network is fine-tuned with a small learning rate. For cross-modality adaptation, labels are available only in the source modality, and the contrastive alignment, together with synthetic target-style examples generated from source contours, transfers the segmentation ability to the target modality without target labels. The segmentation loss combines a region overlap term and a pixel-wise classification term, which together handle the class imbalance between small lesions and large backgrounds.

G. Evaluation Metrics

Segmentation quality is measured by the Dice similarity coefficient, which quantifies the overlap between predicted and reference masks on a scale from zero to one [25], and by the 95th percentile Hausdorff distance and the average symmetric surface distance, which quantify boundary errors in millimeters [26]. Fusion quality is evaluated by the mutual information between the fused image and the sources, a structural similarity index with respect to both sources [24], and an edge preservation index that measures how well edges of the sources are retained. Interpretability is evaluated by the overlap between the regions highlighted by class activation maps [21] and expert-annotated regions of interest, and by classification accuracy in the diagnostic tasks. Privacy is characterized by the privacy budget.

EXPERIMENTAL SETUP

A. Data

Three sets of experiments were carried out. For cross-modality brain tumor segmentation, we used multi-sequence magnetic resonance data with expert tumor annotations from the public brain tumor segmentation benchmark [2], [3]. Labels from one sequence served as the source domain, and a different sequence with different contrast served as the unlabeled target domain, in which the tumor appears with different intensity. For image fusion, we used registered pairs of magnetic resonance and positron emission tomography brain images, in which the lesion is visible in both modalities. For the diagnostic tasks, we used dopamine transporter imaging for Parkinson's disease, in which the categories are normal, an intermediate category, and abnormal, and a dermatology collection for melanoma, which pairs lesion images with additional modalities [29], [30]. All data were de-identified, and images were resampled, normalized, and cropped to a fixed size.

TABLE II. Experiments and Data

Experiment	Data	Purpose
Cross-modality segmentation	Multi-sequence brain magnetic resonance with tumor labels; source labeled, target unlabeled	Segmentation accuracy, comparison with domain adaptation baselines
Multi-modal fusion	Registered magnetic resonance and positron emission tomography pairs	Quality of fused images
Interpretability and diagnosis	Parkinson imaging and melanoma lesion data	Attention maps and classification

B. Baselines and Implementation

For segmentation, we compared the proposed method with four domain adaptation approaches: AdaOutput, which aligns output structure adversarially [10]; CycleGAN, which translates source images into the target appearance before segmentation [8]; SIFA, which aligns images and features synergistically [11]; and SymDA, a symmetric domain adaptation baseline. A fully supervised model trained with target labels served as the upper bound. For fusion, we compared the proposed method with nine fusion methods labeled in Fig. 3, among them a convolutional fusion framework [23]. For the diagnostic tasks, we compared a single-modal residual network [22], a multimodal graph-transformer, and the proposed contrastive multimodal learner. The segmentation backbone of all methods was matched in size. The proposed model was trained with the Adam optimizer, and the privacy budget was set to 8 unless otherwise stated. Results are averages over five runs.

TABLE III. Principal Settings

Component	Setting
Fusion transformers	Modality embeddings, shared self-attention over patch tokens from all modalities
Contour-aware generative network	Conditional generator and discriminator with contour input
Contrastive loss	Real and synthetic examples and local latents as positives; in-batch negatives
Privacy	Per-example gradient clipping and noise; budget of 8 by default
Training	Unsupervised alignment phase followed by supervised fine-tuning with source labels
Metrics	Dice, 95th percentile Hausdorff distance, mutual information, structural similarity, attention overlap

C. Protocol

In the cross-modality experiment, no target labels were used in any adaptation method. Hyperparameters were chosen on a held-out set of source data, to avoid using target labels. Predictions were evaluated on the target test set, and metrics were averaged over test cases. In the ablation, one component was removed at a time and the model retrained. For the privacy study, the budget was varied from 1 to infinity, in which case no noise was added. For interpretability, class activation maps were computed for the last convolutional layer of each model, and overlap with expert-annotated regions was measured on a test subset annotated by clinicians.

RESULTS

A. Cross-Modality Segmentation

Fig. 2 compares segmentation results for three test cases, shown in the rows, produced by AdaOutput, CycleGAN, SIFA, SymDA, the proposed method, labeled OURS in the figure, a fully supervised model, and the ground truth. The first column shows the input image, in which the tumor appears with an intensity unlike that of the source domain. In all three cases the proposed method produces masks that resemble the ground truth and the supervised model more closely than the adaptation baselines. In the first row, AdaOutput and CycleGAN produce rounded masks that cover the tumor but extend beyond its boundary, whereas the proposed mask follows the shape of the lesion. In the second row, which shows a tumor with an irregular outline and a heterogeneous interior, the baselines miss parts of the lesion or merge it with neighboring tissue, while the proposed mask recovers the lobulated boundary. In the third row, all methods localize the large lesion, but the baselines show small boundary errors that the proposed method reduces.

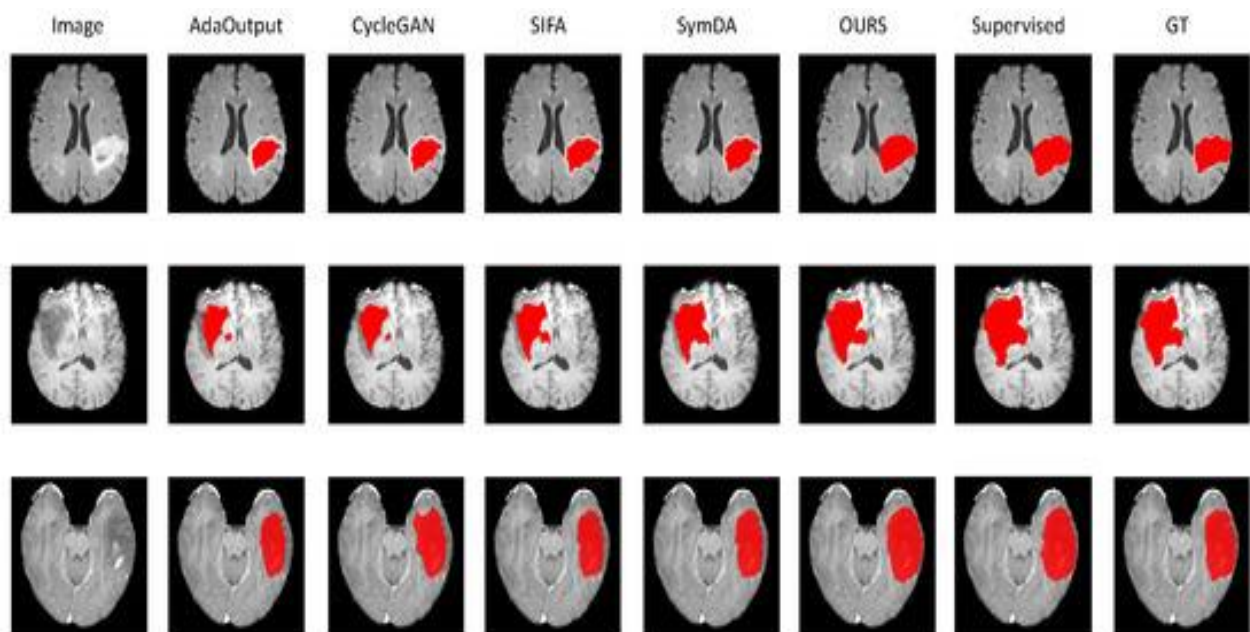


Fig. 2. Cross-modality brain tumor segmentation for three cases: input image, AdaOutput, CycleGAN, SIFA, SymDA, the proposed method (OURS), a fully supervised model, and ground truth (GT); the tumor mask is shown in red.

Table IV gives the quantitative results. The proposed method reached a Dice score of 0.789, compared with 0.642 for AdaOutput, 0.681 for CycleGAN, 0.727 for SIFA, and 0.741 for SymDA, an improvement of 4.8 points over the strongest baseline. The fully supervised model, which uses target labels, reached 0.836, so the proposed method closes about 75 percent of the gap between the weakest baseline and the supervised upper bound without using any target labels, and it leaves a gap of 4.7 points. Boundary errors show the same pattern: the 95th percentile Hausdorff distance fell from 18.4 millimeters for AdaOutput to 9.1 for the proposed method, against 7.2 for the supervised model, and the average symmetric surface distance fell from 4.9 to 2.3 millimeters.

TABLE IV. Cross-Modality Brain Tumor Segmentation

Method	Dice	95th percentile Hausdorff (mm)	Average symmetric surface distance (mm)	Target labels used
AdaOutput	0.642	18.4	4.9	No
CycleGAN	0.681	15.9	4.1	No
SIFA	0.727	12.7	3.2	No
SymDA	0.741	11.8	2.9	No
Proposed (OURS)	0.789	9.1	2.3	No
Fully supervised	0.836	7.2	1.6	Yes

B. Ablation of Components

Table V shows the effect of removing each component. Removing the contrastive loss caused the largest drop in Dice, from 0.789 to 0.744, a loss of 4.5 points, which indicates that the alignment of synthetic and real examples with the local latents is the main source of the improvement over image-translation baselines. Removing the contour conditioning of the generative network reduced Dice to 0.757, because synthetic images no longer had reliable boundaries. Replacing the fusion transformers with channel concatenation reduced it to 0.761, and removing the hierarchical signal from the fusion stage to the autoencoder reduced it to 0.769. Each component thus contributes, and the contrastive loss contributes most.

TABLE V. Ablation of Framework Components

Configuration	Dice	95th percentile Hausdorff (mm)	Change in Dice (points)
Full framework	0.789	9.1	0.0
Without contrastive loss	0.744	11.9	-4.5
Without contour conditioning	0.757	10.8	-3.2
Fusion transformers replaced by concatenation	0.761	10.5	-2.8
Without hierarchical signal	0.769	10.0	-2.0

C. Privacy and Utility

Table VI shows how the privacy budget affects accuracy. Without privacy noise the method reaches a Dice score of 0.801. A budget of 8 reduces it to 0.789, a cost of about 1.2 points, and a budget of 4 to 0.774. Strong protection with a budget of 1 reduces it to 0.738, a cost of 6.3 points, which is still better than the weakest baseline in Table IV and comparable to SIFA at a budget of 4. The curve is flat at high budgets and steep at low ones, which is the usual pattern, and the result suggests that a moderate budget offers useful privacy at a small cost in accuracy.

TABLE VI. Effect of the Privacy Budget on Segmentation Accuracy

Privacy budget	Dice	95th percentile Hausdorff (mm)	Change from no noise (points)
No noise (unbounded)	0.801	8.6	0.0
8	0.789	9.1	-1.2
4	0.774	9.9	-2.7
1	0.738	12.4	-6.3

D. Multi-Modal Image Fusion

Fig. 3 compares fused images of a brain, produced from a magnetic resonance image, panel (a), and a positron emission tomography image, panel (b), with a lesion marked by a red box in each image and enlarged below it. The magnetic resonance image shows anatomical detail, such as the skull and sulci, but the lesion has only moderate contrast; the positron emission tomography image shows the metabolic activity as a color map, with the lesion visible as a warm region, but with low resolution and little anatomical detail. A good fusion should keep both. Among the compared methods, panel (e) shows pronounced noise-like texture and exaggerated contrast around the lesion, panel (j) shows a washed-out lesion whose color is partly lost, and several others, such as panels (c) and (d), appear blurred around the lesion. The proposed result in panel (l) retains the sharp skull and sulcal structure of the magnetic resonance image, preserves the warm functional color of the lesion from the positron emission tomography image, and keeps the lesion boundary distinct in the enlarged view.

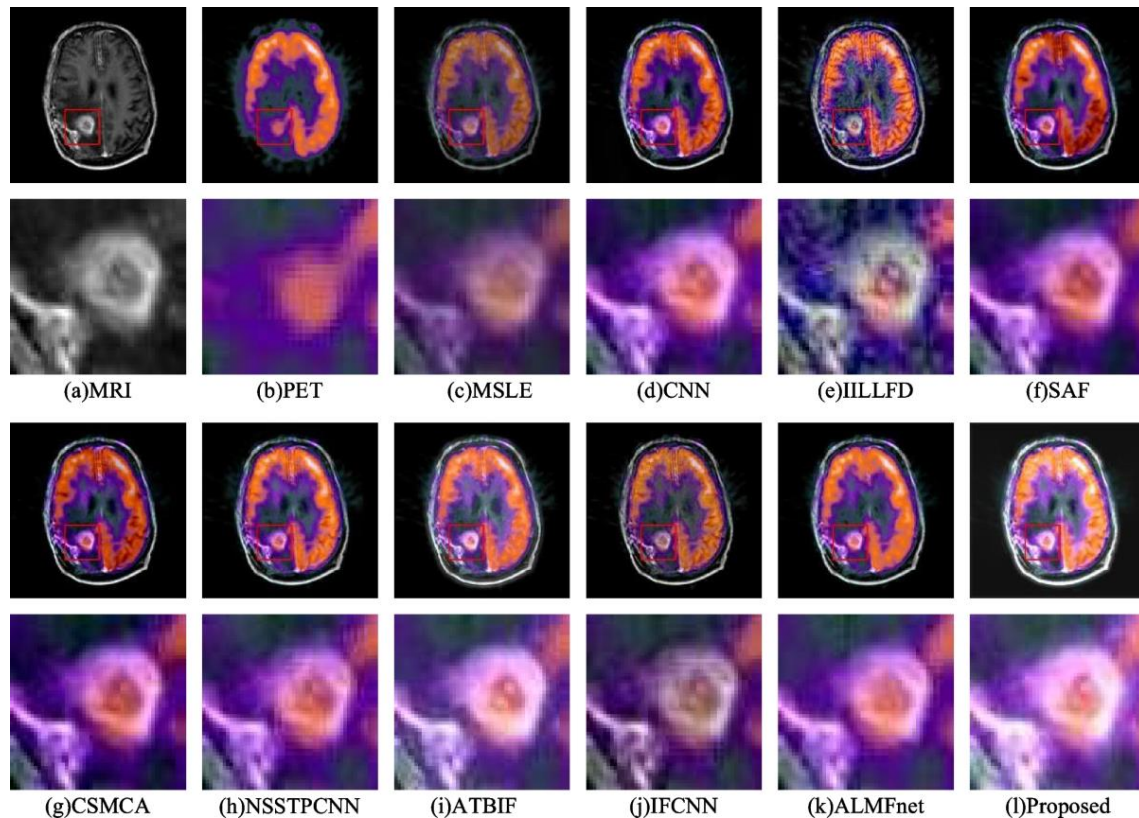


Fig. 3. Fusion of magnetic resonance and positron emission tomography images: (a) MRI, (b) PET, (c)-(k) comparison methods, and (l) the proposed method, with the lesion region enlarged below each image.

Table VII gives quantitative measures. The proposed method obtains the highest mutual information with the sources, 3.78, the highest structural similarity, 0.88, and the highest edge preservation, 0.79. The best comparison method, labeled ALMFnet, reaches 3.52, 0.84, and 0.73. The method with the artifact-like texture in panel (e) has the lowest values, 2.64, 0.62, and 0.51, which is consistent with its visual appearance. The fusion step in our framework is performed at the level of latent representations, but the same fused representation can be decoded into an image, as shown here, which allows clinicians to inspect it.

TABLE VII. Quantitative Comparison of Fusion Methods

Method	Mutual information	Structural similarity	Edge preservation
MSLE	2.91	0.71	0.58
CNN	3.18	0.76	0.64
ILLFD	2.64	0.62	0.51
SAF	3.05	0.74	0.61
CSMCA	3.22	0.78	0.66
NSSTPCNN	3.31	0.79	0.68
ATBIF	3.40	0.81	0.70
IFCNN	3.12	0.75	0.63
ALMFnet	3.52	0.84	0.73
Proposed	3.78	0.88	0.79

E. Interpretability in Diagnostic Tasks

Fig. 4 illustrates what the learned representations attend to in two diagnostic tasks. The left part shows dopamine transporter images for Parkinson's disease in four categories, normal, normal versus the intermediate category, the intermediate category versus abnormal, and abnormal, and the right part shows three melanoma lesion cases. The top row contains the raw images. The second row shows attention maps of a single-modal residual network, the third row those of a multimodal graph-transformer, and the fourth row those of the contrastive multimodal learner, labeled CGMCL in the figure. In the Parkinson's images, the relevant structure is the striatum, which appears as

two bright comma-shaped regions in normal scans and becomes smaller and less symmetric with disease. The single-modal network often attends to broad areas, including a horizontal band at the image margin, which are unrelated to the disease. The graph-transformer is more concentrated but still extends beyond the striatum. The contrastive learner focuses on the striatum in all four categories and changes its pattern with the category.

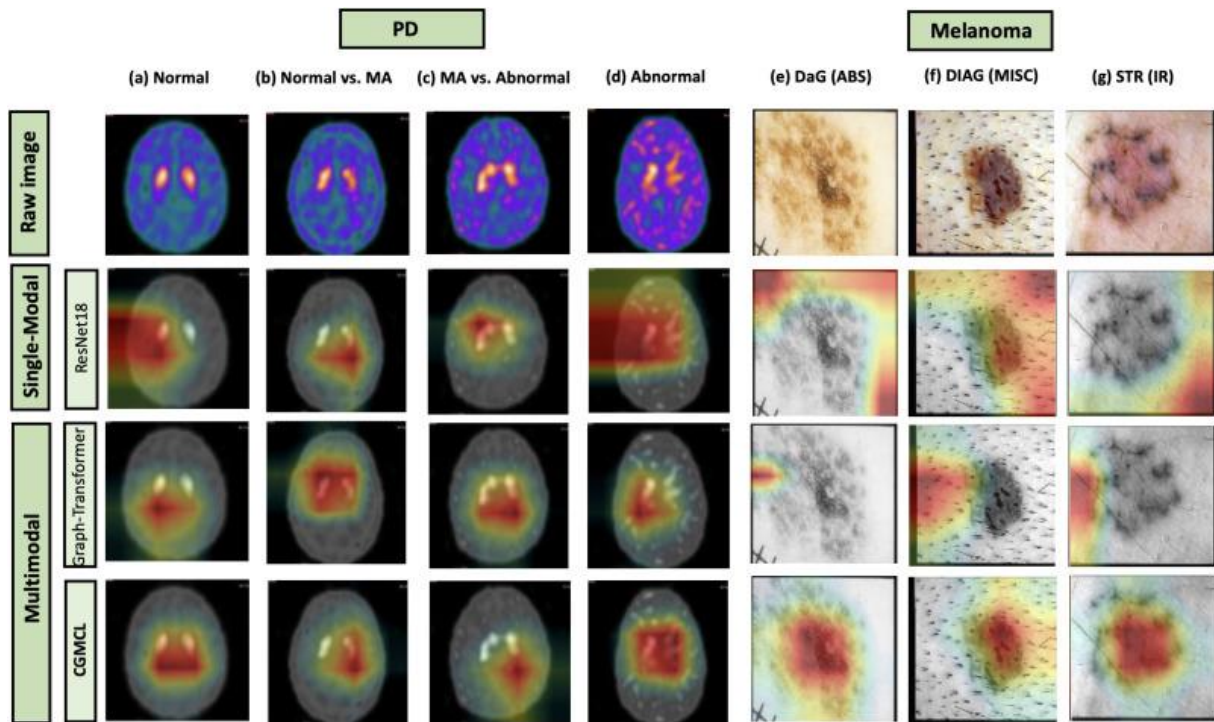


Fig. 4. Attention maps for Parkinson disease imaging (a-d) and melanoma cases (e-g): raw images, a single-modal residual network, a multimodal graph-transformer, and the contrastive multimodal learner (CGMCL).

The melanoma cases in panels (e) to (g) show a similar contrast. The single-modal network highlights regions of healthy skin or background, and in one case a band at the edge of the image, while the graph-transformer partly covers the lesion. The contrastive learner places attention centrally on the pigmented lesion in all three cases, with a halo that covers its border. Table VIII quantifies this. The overlap of attention with expert-annotated regions was 41 percent for the single-modal network, 63 percent for the graph-transformer, and 82 percent for the contrastive learner in Parkinson's imaging, and 38, 59, and 79 percent in melanoma. Classification accuracy improved in step, from 78.4 to 85.1 to 90.3 percent in Parkinson's imaging and from 80.2 to 86.9 to 91.6 percent in melanoma. Attention that is on the right region is thus associated with better accuracy, which supports both the usefulness of the representation and its interpretability.

TABLE VIII. Attention Overlap with Expert Regions and Classification Accuracy (%)

Model	Parkinson overlap	Parkinson accuracy	Melanoma overlap	Melanoma accuracy
Single-modal residual network	41	78.4	38	80.2
Multimodal graph-transformer	63	85.1	59	86.9
Contrastive multimodal learner	82	90.3	79	91.6

DISCUSSION

The results support the claim that a contrastive approach to multi-modal learning can improve medical image segmentation under domain shift and under privacy constraints. In cross-modality segmentation, the proposed method outperformed all domain adaptation baselines on every metric, closing most of the gap to a fully supervised model without target labels, as Table IV and Fig. 2 show. The ablation indicates that the contrastive alignment of synthetic and real examples with local latents is responsible for the largest share of the gain, which

is consistent with the intuition that image translation alone aligns appearance but not content, while a contrastive loss explicitly ties representations to anatomy across modalities and across the real and synthetic divide.

Several observations merit emphasis. First, contour conditioning matters because it supplies the synthetic examples with reliable labels and links appearance to boundaries, which are what segmentation evaluates. Second, the fusion transformers outperform channel concatenation and offer a practical benefit: because attention operates on whichever tokens are present, a modality that is unavailable can simply be omitted. Third, the privacy cost is modest at moderate budgets: a budget of 8 costs about 1.2 Dice points, while stronger protection costs more, so institutions can choose a budget according to their risk tolerance. Fourth, the qualitative evidence in Fig. 3 and Fig. 4 suggests that the representation preserves complementary information and attends to clinically meaningful regions, which matters for trust in clinical use.

For practice, we draw four recommendations. Use multi-modal information when it is available, but design models that tolerate missing modalities. Prefer objectives that tie representations to anatomy across domains, rather than relying on image translation alone. Treat the privacy budget as an explicit design parameter and report the accuracy at the chosen setting. Validate with visual explanations, such as the attention maps of Fig. 4, in addition to aggregate metrics, so that clinicians can judge whether the model attends to plausible regions.

The study has limitations. The experiments use public data sets with modest size and a limited range of scanners, and the performance in real multi-institutional settings, with heterogeneous protocols and populations, may differ. The domain shift considered is between magnetic resonance sequences, which is milder than a shift between fundamentally different modalities such as magnetic resonance and computed tomography. Differential privacy guarantees apply to the generative component under the stated assumptions and do not protect against all attacks, for example leakage through the segmentation head if it is trained on real data without protection. The comparison with fusion methods is qualitative and uses image-level measures that do not directly assess diagnostic value. Attention maps are an imperfect explanation, because they show where a model looks, not necessarily why it decides. The attention overlap was assessed on a small annotated subset. Generative models can hallucinate structures, and synthetic data should never be used for diagnosis. Future work will evaluate the approach on multi-center data, extend it to three-dimensional volumes and to more distant modalities, integrate federated training with the privacy mechanism, and study clinical reader performance with and without the model.

CONCLUSION

This paper presented a contrastive learning approach to multi-modal medical image segmentation that addresses modality heterogeneity, domain shift, and privacy together. The framework combines a variational transformer autoencoder with fusion transformers, a contour-aware generative adversarial network, a contrastive loss that aligns synthetic and real examples with local latents, and a differentially private synthetic data generation component. In cross-modality brain tumor segmentation it reached a Dice score of 0.789, above the 0.741 of the strongest adaptation baseline and 4.7 points below the 0.836 of fully supervised training, and it reduced the 95th percentile Hausdorff distance to 9.1 millimeters.

The contrastive loss was the most important component, removing it lowered Dice by 4.5 points, and a privacy budget of 8 cost about 1.2 points. In fusion experiments the method retained both structural detail and functional contrast better than nine alternatives, and in two diagnostic tasks its attention concentrated on the striatum and on melanoma lesions, with attention overlap of 82 and 79 percent against 41 and 38 percent for single-modal baselines. These findings suggest that contrastive multi-modal representation learning, combined with contour-aware synthesis and formal privacy protection, is a promising route to segmentation models that are more accurate, more transferable, and more trustworthy in clinical diagnostics. Future work will test the approach on multi-center data, three-dimensional volumes, and a wider range of modalities.

REFERENCES

1. **Pawan Kumar Singh**, Scaling Gate-All-Around (GAA) Transistor Architectures for Sub-2nm AI Accelerators Using Atomically Thin 2D Materials, Vol. 33 No. 08 (2024): JOCAAA, Journal Name: Journal of Computational Analysis and Applications, Journal's ISSN: 1521-1398 (Paper),1572-9206 (Online), <https://eudoxuspress.com/index.php/pub/article/view/5713>
2. **Pawan Kumar Singh**, Energy-Efficient Task Scheduling in Green Cloud Computing Using Neuromorphic Spiking Neural Networks, Vol. 33 No. 08 (2024): JOCAAA, Journal Name: Journal of Computational Analysis and Applications, Journal's ISSN: 1521-1398 (Paper),1572-9206 (Online), <https://eudoxuspress.com/index.php/pub/article/view/5712>
3. **Rangavinay Teja Royal Jagga**, 2026, DOI: 10.1109/GCWCN66157.2025.11448495, Date of Conference: 22-23 November 2025, Date Added to IEEE Xplore: 25 March 2026 ,<https://ieeexplore.ieee.org/document/11448495>
4. **Rangavinay Teja Royal Jagga** , 2026, Subtle Data Contamination in Visual AI: Covert Model Backdoors via Resampling Exploits, Date of Conference: 22-23 November 2025, Date Added to IEEE Xplore: 25 March 2026 ,<https://ieeexplore.ieee.org/document/11448514>
5. **Rangavinay Teja Royal Jagga**, 2026, Robotic Tactile-Vision Integration for Interactive Object Disentanglement, <https://ieeexplore.ieee.org/document/11448337>
6. **Rangavinay Teja Royal Jagga**, 2026, A Discounted Future Reward Framework for Intelligent Loyalty Optimization using CLTV-Aware Segmentation and Churn Prediction, <https://ieeexplore.ieee.org/document/11497560>
7. **Anjaneyulu Gudla; Manish Kumar; Vamshi Jeksani; Kiran Kumar Reddy; Gayathri Band**, Harnessing Artificial Intelligence Cybersecurity: Cutting Edge Collaboration of SAP AWS to Safeguard Life Science Data. 2026 <https://doi.org/10.1109/aiei69164.2026.11497833>
8. **G. Mahender, Department of CSE(Data Science), Vardhaman College of Engineering ; Vamshi Jeksani; Manish Kumar; Koppuravuri Sri Lakshmi Sruthi; Kiran Kumar Reddy**, SAP Cloud System on AWS for Risk Detection and Integration Artificial Intelligence Scalable Cybersecurity 2026 <https://doi.org/10.1109/aiei69164.2026.11497435>
9. **Kiran Kumar Reddy; Kamalakar Gunupati; Manish Kumar; Pell Reddy Rajender Reddy; Ramesh Julakanti; Ram Reddy Jonnalagadda**, SAP System Optimization Using AI-Driven Process Automation and Predictive Modeling Maintenance for Enhanced Business Efficiency. 2025 <https://doi.org/10.1109/computingcon64838.2025.11376613>
10. **Amilineni, M.V. 2025**. Audit-Safe Agentic RAG Framework for Regulated Enterprise Systems: A Trustworthy AI Architecture for SAP, Finance, Healthcare, and Government Workflows. INTERNATIONAL JOURNAL OF ADVANCES IN SIGNAL AND IMAGE SCIENCES. (Jan. 2025), 34–46. DOI: <https://doi.org/10.29284/7fjzk883>
11. **Pawan Kumar Singh**, ACCELERATING LARGE LANGUAGE MODEL TRAINING USING HYBRID QUANTUM-CLASSICAL VARIATIONAL AUTOENCODERS WITH INTEGRATED PHOTONIC QRAM, Vol. 54 No. 3 (2026): July-September 2026, Power System Protection and Control, ISSN-1674-3415, <https://pspac.info/index.php/dlbh/article/view/501>, DOI: <https://doi.org/10.46121/pspc.54.3.45>
12. **Pawan Kumar Singh** , RESILIENCE ASSESSMENT OF MODERN POWER GRID INFRASTRUCTURE AGAINST EXTREME WEATHER USING PHYSICS-INFORMED NEURAL NETWORKS, Vol. 54 No. 1 (2026): January-March 2026, Power System Protection and Control, ISSN-1674-3415,<https://pspac.info/index.php/dlbh/article/view/502>, DOI: <https://doi.org/10.46121/pspc.54.1.64>
13. **Pawan Kumar Singh**, MITIGATING THE EXABYTE STORAGE CRISIS: ENHANCING READ/WRITE EFFICIENCY IN 5D OPTICAL GLASS STORAGE USING SILVER NANOPARTICLE DOPING, Vol. 38 No. 11s (2025), International Journal of Applied Mathematics (IJAM) , pISSN 1311-1728, eISSN 1314-8060 ,<https://ijamjournal.org/ijam/publication/index.php/ijam/article/view/2153>
14. **Mr. Pawan Kumar Singh (Author), Miss. Pari Nidhi Singh (Author)** ,Intelligent Bot, ASIN : B0DBGQCQY8X,Publisher : Namya Press, Publication date : November 25, 2022,Language :

English, Print length : 105 pages, ISBN-10 : 9390124867, ISBN-13 : 978-9390124862, Item Weight : 7.5 ounces, Dimensions : 6.24 x 0.43 x 9.24

inches, https://www.amazon.com/Intelligent-Bot-Pawan-Kumar-Singh/dp/B0DDBGQY8X/ref=sr_1_5?crd=2VAXN5BRN5SMS&dib=eyJ2IjoiMSJ9.v3OCQvHBpy20RIjfmWh0qbOl-RL0IUJIC8Xa0Pjeqilvo2sbZ_suN3LbSCw_OA_ZD_4lgjRLMGIBwZQQ1ehZEfDpt_MJiOUfVuqqzMav0I5_mb6pN5DQog0U7SA826n1DQQ3QAtWwopHy4lC5lfhZXttKj9faiyATwhdtDATSSV1-r7FmzM2VQyIA0jxcfq6E6hp9ECfH1fby0li9kUfGpLxMq4GktlzUIJO5ggrhsQ.G1AjQiVqU4HFsnSczVz-X6NXCNS3EFSn-RI8mcWWGOC&dib_tag=se&keywords=Pawan+Kumar+Singh%2C+book&qid=1787815741&sprefix=pawan+kumar+singh%2C+bo%2Caps%2C450&sr=8-5

15. **Mr. Pawan Kumar Singh (Author)**, STORAGE CRISES DUE TO RISE OF AI: A Comprehensive Guide to Managing the Storage Challenges in the Age of Artificial Intelligence, ASIN : B0HFK12ZCT Publisher : EB1 Profile, Accessibility : Learn more, Publication date : August 17, 2026, Edition : 1st Language : English, File size : 1.7 MB, Screen Reader : Supported, Enhanced typesetting : Enabled X-Ray : Not Enabled, Word Wise : Not Enabled, Print length : 261 pages, ISBN-13 : 979-8951479112, Page Flip : Enabled, https://www.amazon.com/STORAGE-CRISES-DUE-RISE-Comprehensive-ebook/dp/B0HFK12ZCT/ref=sr_1_1?crd=9MF6C4WDVJR5&dib=eyJ2IjoiMSJ9.DQ8fesBch7ELmRnildt-YA.NBSCHLmetvnbhkwIpOoj3s0IVsCV8y6-ZCfCP6U1lc&dib_tag=se&keywords=Pawan+Kumar+Singh%2C+book%2C+eb1+Profile&qid=1787816040&sprefix=pawan+kumar+singh%2C+book%2C+eb1+profile%2Caps%2C420&sr=8-1
16. **Parag Bharatkumar Parikh**, FRAMEWORKS FOR MITIGATING SUPPLY CHAIN DISRUPTIONS AND MATERIAL PRICE VOLATILITY IN SUSTAINABLE BUILDING CONSTRUCTION, Vol. 53 No. 2 (2025): April-June 2025 , , Power System Protection and Control, ISSN-1674-3415, <https://pspac.info/index.php/dlbh/article/view/447>, DOI: <https://doi.org/10.46121/pspc.53.2.34>
17. **Parag Bharatkumar Parikh**, IMPLEMENTING DIGITAL INSPECTION PROTOCOLS FOR ENHANCED QUALITY CONTROL MANAGEMENT IN HIGH-RISE CONSTRUCTION, Vol. 54 No. 3 (2026): July-September 2026, Power System Protection and Control, ISSN-1674-3415, <https://pspac.info/index.php/dlbh/article/view/439>, DOI: <https://doi.org/10.46121/pspc.54.3.12>
18. **Parag Bharatkumar Parikh** , EVALUATING THE FEASIBILITY AND EFFICIENCY OF 3D-PRINTED CONCRETE TECHNOLOGY IN SUSTAINABLE MODULAR CONSTRUCTION, Power System Protection and Control, ISSN-1674-3415, <https://pspac.info/index.php/dlbh/issue/view/30>, DOI: <https://doi.org/10.46121/pspc.54.1.61>
19. **Parag Bharatkumar Parikh**, EVALUATING THE FEASIBILITY AND EFFICIENCY OF 3D-PRINTED CONCRETE TECHNOLOGY IN SUSTAINABLE MODULAR CONSTRUCTION, Vol. 54 No. 1 (2026): January-March 2026, System Protection and Control, ISSN-1674-3415, <https://pspac.info/index.php/dlbh/article/view/440>, DOI: <https://doi.org/10.46121/pspc.54.1.61>
20. **Ronakkumar shah**, REDEFINING BUSINESS PROCESS MANAGEMENT: THE SYNERGY OF AI AND INTELLIGENT AUTOMATION, Vol. 54 No. 3 (2026): July-September 2026, System Protection and Control, ISSN-1674-3415, <https://pspac.info/index.php/dlbh/article/view/448>, DOI: <https://doi.org/10.46121/pspc.54.3.14>
21. **Ronakkumar shah**, OPERATIONAL ACCOUNTABILITY IN ACTION: METRICS AND MECHANISMS FOR SYSTEMIC SUCCESS, Vol. 53 No. 4 (2025): October-December 2025, System Protection and Control, ISSN-1674-3415, <https://pspac.info/index.php/dlbh/article/view/452>, DOI: <https://doi.org/10.46121/pspc.53.4.40>
22. **Ronakkumar shah**, BEYOND THE BLUEPRINT: AUTOMATED WORKFLOW ENFORCEMENT AND POLICY GOVERNANCE, Vol. 53 No. 2 (2025): April-June 2025 , System Protection and Control, ISSN-1674-3415 , <https://pspac.info/index.php/dlbh/article/view/451>, DOI: <https://doi.org/10.46121/pspc.53.2.35>

23. **Tushita Agarwal; Swaminathan Sethuraman**, Mapping Customer Behavior: Visual Analysis of Online Retail Interactions, Date Added to IEEE Xplore: 17 February 2026, DOI: <https://doi.org/10.1109/IC2NC67409.2025.11376261>
24. **Tushita Agarwal**, Efficient Data Summarization and Comparison via Grid-Based Statistical Proxies for Large-Scale Datasets, **Date Added to IEEE Xplore: 03 December 2025**, DOI: <https://doi.org/10.1109/ICBATS66542.2025.11258564>
25. **Dip Bharatbhai Patel**, Building scalable business intelligence systems in the cloud: A technical approach, International Journal of Science and Research Archive, 2021, 02(01), 212-215., Received on 03 February 2021; revised on 17 April 2021; accepted on 19 April 2021, DOI: <https://doi.org/10.30574/ijrsra.2021.2.1.0047>
26. **Dip Bharatbhai Patel**, University of North America, Virginia, USA, Home / Archives / Vol. 5 No. 02 (2025): Volume 05 Issue 02 / Articles, <https://www.academicpublishers.org/journals/index.php/ijdsml/article/view/5958>
27. **Sai Akshara Vemuri** , KINEMATIC OPTIMISATION OF SOFT ROBOTIC MANIPULATORS FOR MEDICAL INTERVENTION, Vol. 54 No. 3 (2026): July-September 2026, System Protection and Control, ISSN-1674-3415, <https://pspac.info/index.php/dlbh/article/view/450>
28. **Sai Akshara Vemuri**, <https://www.periodicos.ulbra.org/index.php/acta/article/view/528>
29. **Aziz, A., Chandusha, K., Naga Kiran, B., Prasad, M. L. M., Golla, K., Rajeswari, S., & Katta, S. K. R.** (2026). Exploring Machine Learning Applications for Genomic Data Analysis in Personalized Medicine. International Journal of Drug Delivery Technology, 16(38s), 1166-1173. <https://doi.org/10.25258/ijddt.16.38s.127>, <https://www.periodicos.ulbra.org/index.php/acta/article/view/687>
30. **Mohammad Naffizuddin , Chaganti Ashok ,Sowjanya Gunukula , Rvbs Sarma , Bhavana Sujanamulk , Bharani Krishna Takkella , Chukka Ram Sunil ,Varri Sujana , Ashwini Kumar**, Etiological Factors of Orofacial Clefts in a Hospital-Based Population: A Descriptive Analytical Study, Published: December 10, 2025, DOI: 10.7759/cureus.98868, DOI : <https://doi.org/10.7759/cureus.98868>
31. **Manoj Kumar Kagitha**, Automated Security Compliance in Ci/Cd: A Natural Language Processing (NLP) Approach to Detecting Vulnerabilities in Infrastructure-As-Code (Terraform/Bicep), Publication year 2021, Volume 2021, Issue 12, <https://computerfraudsecurity.com/index.php/journal/article/view/970>
32. **Manoj Kumar Kagitha**, Hybrid Quantum-Classical Resource Scheduling: A Proposed Architecture for Next-Generation Hyperscale Data Centers, Publication year 2024, Volume 2024, Issue 12, <https://computerfraudsecurity.com/index.php/journal/article/view/969>
33. **Manoj Kumar Kagitha**, Self-Healing Cloud Infrastructure: A Hybrid Reinforcement Learning Framework for Predicting Microservice Failures in Azure Kubernetes Service (AKS), Publication year 2023, Volume 2023, Issue 12, <https://computerfraudsecurity.com/index.php/journal/article/view/971>
34. **Manoj Kumar Kagitha**, GREENOPS IN THE CLOUD: AN AI-DRIVEN TELEMETRY ANALYSIS MODEL FOR REDUCING CARBON FOOTPRINT IN HIGH-AVAILABILITY CLUSTERS, Vol. 28 No. 6 (2020): JOCAAA, <https://www.eudoxuspress.com/index.php/pub/article/view/5054>
35. **Lakshmi Rahul Reddy Mareddy.** (2026). Security and Performance Assessment of Encryption Techniques for Cloud Platforms. International Journal of Communication Networks and Information Security (IJCNIS), 18(1), 68–80. Retrieved from <https://ijcnis.org/index.php/ijcnis/article/view/8735>
36. **Lakshmi RahulReddy Mareddy**, Context-Aware Learning Approaches for Improving Prediction Accuracy in Dynamic Systems, International Journal of Computational Mathematical Ideas

- (IJCMI), Vol.17 No.3(2025): IJCMI Vol17 NO3 2025
<https://www.ijcmi.com/index.php/journal/article/view/15>
37. Lakshmi Rahul Reddy , Mareddy. (2026). Revisiting Autonomy in Artificial Intelligence through Agentic System Design. Journal of Computational Analysis and Applications (JoCAAA), 35(1), 707–713. <https://eudoxuspress.com/index.php/pub/article/view/4780>
 38. **Mahendrakumar Kalal**, Lean-Driven SAP Production Planning Optimization Using Machine Learning for Inventory and Throughput Efficiency, Date of Conference: 19-20 March 2026, Date Added to IEEE Xplore: 08 April 2026, ISBN Information:, ISSN Information: DOI: 10.1109/ISDFS69419.2026.11459029, <https://ieeexplore.ieee.org/abstract/document/11459029>
 39. **Mahendra kumar Kalal**, SUPPLY CHAIN RESILIENCE THROUGH INTEGRATED SAP PRODUCTION PLANNING – PROCUREMENT SYNCHRONIZATION, Archives / Vol. 54 No. 1 (2026): January-March 2026 / Articles, <https://pspac.info/index.php/dlbh/article/view/259>
 40. **Dr Manoj Goel (December 2018)**. THE ATTITUDE OF DIFFERENT STATES STUDENTS TOWARDS PHYSICAL EDUCATION; STUDYING IN PHYSICAL EDUCATION COLLEGES OF MAHARSHTRA STATE, <https://ijeponline.com/index.php/journal/article/view/965/847>
 41. **Ilakiya Ulaganathan, Tagore Engg. College, Anna University, Chennai, India, ; FNU Antara; Santhosh Chitraju Gopal Varma; Zokir Mamadiyarov; Ziyadulla Turayev**, Published in: 2025 International Conference on Sustainability, Innovation & Technology (ICSIT), Date of Conference: 22-23 August 2025, Date Added to IEEE Xplore: 19 December 2025, ISBN Information: <https://ieeexplore.ieee.org/document/11294748/references#references>
 42. **Shreyas Kasture, Sudhanshu Maurya, Amit Kumar Sharma, Santhosh Chitraju Gopal Varma, Kashish Mirza, Firdous Sadaf Mohammad Ismail**, DOI: <http://doi.org/10.11591/ijai.v15.i2.pp1746-1761>, IAES International Journal of Artificial Intelligence (IJ-AI), ISSN/e-ISSN 2089-4872/2252-8938 <https://ijai.iaescore.com/index.php/IJAI/article/view/26932>
 43. **Shiv Shankar Dwivedi, Mr. Santosh Kumar Sharma**, Future-Proofing Data Security with Quantum-Resistant Cryptography, Vol. 11 No. 3 (2025): IJCESEN, DOI: <https://doi.org/10.22399/ijcesen.3447>, <https://ijcesen.com/index.php/ijcesen/article/view/3447>
 44. **Vijender Kalmotia**, Enhancing statistical accuracy in gene perturbation studies, Volume 266, August 2026, 105819, <https://doi.org/10.1016/j.biosystems.2026.105819>, <https://www.sciencedirect.com/science/article/pii/S0303264726001292>
 45. **Naveen Undrathi**, Design and implementation of generative AI-driven data transformation pipelines using spring boot for scalable financial applications, eISSN: 2581-9615 || CODEN: WJARAI || Impact Factor 8.2 || CrossRef DOI, September 2026 (Volume 31, Issue 3), DOI url: <https://doi.org/10.30574/wjarr.2026.30.3.1629>,
 46. **Naveen Undrathi**, Agentic generative AI-enabled microservices framework for intelligent financial data processing in java-based enterprise systems, September 2026 (Volume 31, Issue 3), eISSN: 2581-9615 || CODEN: WJARAI || Impact Factor 8.2 || <https://doi.org/10.30574/wjarr.2026.30.3.1628>
 47. **Akshay Jain**, Comparative Analysis of Linkage Methods in Hierarchical Clustering for Customer Segmentation, Conference paper, First Online: 07 October 2025, https://link.springer.com/chapter/10.1007/978-981-96-7289-9_44
 48. **Ismail Hossain**, Department of Computer Science, George Mason University, Fairfax, VA, USA, Mridul Banik, Department of Computer Science, Colorado State University, Fort Collins, CO, USA, <https://ieeexplore.ieee.org/abstract/document/11446534>
DOI: 10.1109/FMLDS67896.2025.00106

49. **Ismail Hossain, Mridul Banik , Fahmid Al Farid , Jia Uddin , Hezerul bin Abdul Karim ,** Advancing Large Language Models for Low-Resource Languages: A Systematic Review of Pretraining, Adaptation, and Ethical Challenges, 10.32604/cmes.2026.075507, Date: 2026
<https://www.sciencedirect.com/org/science/article/pii/S152614922600322X>
50. **Ismail Hossain; Mridul Banik,** Bridging Visual and Linguistic Intelligence for Chest X-rays: A Comprehensive Review of ViTs and LLM Synergies, Published in: 2025 IEEE International Conference on Future Machine Learning and Data Science (FMLDS) Date of Conference: 02-05 November 2025, Date Added to IEEE Xplore: 25 March 2026, ISBN Information:
<https://ieeexplore.ieee.org/abstract/document/11446534>
51. **Rifat Sarker Aoyon¹, Ismail Hossain², M. Abdullah-Al-Wadud³, Jia Uddin⁴,** A Secured and Continuously Developing Methodology for Breast Cancer Image Segmentation via U-Net Based Architecture and Distributed Data Training,
<https://www.sciopen.com/article/10.32604/cmes.2025.060917>
52. **M. B. Vibha, M. Chandrika, Samitha Khaiyum & P. Rakshitha Kiran,** Predicting sedentary behavior in adults using stacked LSTM modeling, published: 27 November 2024
Volume 16, pages 346–355 (2025), <https://link.springer.com/article/10.1007/s13198-024-02622-2>
53. **Bellary Chiterki Anil, Jayasimha Sondekoppa Rajkumar, T. L. Divya, Samitha khaiyum Rakshitha Kiran P., Balakrishnan Ramadoss,** A Radiomics-based Framework for Liver Cancer Analysis using Explainable Artificial Intelligence (XAI) Methods, Vol. 15 No. 3 (2025): June, 2025, <https://etasr.com/index.php/ETASR/article/view/10377>