Mathematical Modelling of Engineering Problems
Vol. 12, No. 11, November, 2025, pp. 3879-3886
Journal homepage: http://iieta.org/journals/mmep

MedGRU-SVC: A Hybrid ConvGRU and Support Vector Clustering Framework for Interpretable Anomaly Detection in Medical Radiographs

P Naresh[1], Praveen Kulkarni[1*], T. M. Rajesh[2], M N RenukaDevi[1], Kavyashree I Pattan[1], Yashpal Gupta S[1]

[1] Department of CSE, Dayananda Sagar University, Bangalore 562112, India
[2] Department of CSME, Dayananda Sagar University, Bangalore 562112, India

Corresponding Author Email: Praveen.kulkarni-cse@dsu.edu.in

https://doi.org/10.18280/mmep.121113

Received: 26 July 2025
Revised: 9 October 2025
Accepted: 16 October 2025
Available online: 30 November 2025

Keywords:
ConvGRU, SVC, deep temporal-spatial feature learning, Grad-CAM visualization, diagnostic imaging, non-invasive diagnosis, radiograph segmentation

ABSTRACT

Accurate detection of anomalies in medical images is critical for early diagnosis and treatment, especially in oncology, pulmonology, and orthopedics. Conventional machine learning and deep learning models often struggle to capture the complex spatial–temporal dependencies in sequential or multi-view radiological data. To overcome this challenge, we propose the Medical Gated Recurrent Unit–Support Vector Clustering (MedGRU-SVC)—a hybrid framework that integrates the Convolutional Gated Recurrent Unit (ConvGRU) for spatiotemporal feature extraction with SVC for unsupervised classification of pathological patterns. The pipeline employs adaptive histogram equalization to improve contrast and suppress noise, followed by Bayesian-optimized, Radial Basis Function (RBF)-kernel-based SVC to cluster abnormalities such as nodules, calcifications, infiltrates, and fractures. Experimental evaluation on the NIH ChestX-ray14 dataset demonstrates that MedGRU-SVC achieves an accuracy of 96.8%, an F1-score of 95.2%, a precision of 98.0%, a recall of 94.5%, and an AUC of 0.982, outperforming competitive CNN-SVM and LSTM-CNN baselines. In addition, interpretability is enhanced through Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations, which highlight diagnostic regions that drive predictions, ensuring clinical transparency and trust. By combining the temporal learning strength of ConvGRU with the clustering power of SVC, MedGRU-SVC delivers a scalable, interpretable, and high-precision solution for automated radiological screening, making it a practice-ready computer-aided diagnosis (CAD) system.

1. INTRODUCTION

Medical imaging has become one of the most critical tools in modern healthcare, accounting for more than 70% of hospital diagnostic procedures worldwide [1]. Modalities such as chest X-rays, mammograms, and CT scans provide invaluable insights for disease detection and treatment planning. However, manual interpretation by radiologists remains prone to fatigue-induced errors and subjectivity, particularly when anomalies are subtle or rare. Based on research, initial reading of chest radiographs can miss as many as 30% of actionable pulmonary nodules [2, 3]. These limitations illustrate the critical need for advanced, automated diagnostic technologies that can enhance clinical decision-making's accuracy, reliability, and consistency.

One of the distinctive challenges of medical imaging is the capture of temporal dynamics—e.g., lesion evolution in CT sequences or infiltrate formation over serial chest X-rays. Temporal information such as this is crucial for the early detection of disease patterns, such as fibrosis and developing pneumonia. Temporal modeling tools like LSTMs and GRUs have enhanced progression prediction tasks by 6–10% on F1-score [4]. But traditional recurrent networks tend to flat spatial data and hence yield suboptimal performance for radiology tasks. Convolutional Gated Recurrent Unit (ConvGRU) overcomes this by modeling spatial and temporal dependencies simultaneously [5], which makes ConvGRU a good candidate for dynamic anomaly detection.

Although they have their benefits, ConvGRU-based approaches are often used in supervised learning pipelines that need masses of labeled data. This is an issue with medical imaging since expert annotations are limited—only 2–5% of radiology datasets are annotated for cost and time reasons [6]. In addition, purely supervised models tend to be "black boxes," providing minimal interpretability. This lack of clarity erodes clinician trust, which is critical to the adoption in healthcare practice [7]. These concerns stimulate a high need for explainable, unsupervised, or weakly-supervised methods that have the ability to significantly make use of unlabeled data while being clinically interpretable.

Support Vector Clustering (SVC), an unsupervised kernel-based method, provides mathematically sound means to detect dense regions in feature space without the need for labels. Previous work has shown that deep-SVC hybrids can obtain 4–6% higher Area Under the Curve (AUC) than fully supervised baselines with applications to real-world anomaly

detection, such as rare infiltrate and nodule detection [8]. This implies that incorporating ConvGRU embeddings with SVC clustering may provide a framework that is both accurate and interpretable [9]. Radiology is confronted with an acute data challenge: labels are available on only 2–5% of images because annotating them is very expensive [10]. Furthermore, supervised deep learning models tend to be non-interpretable, restricting their adoption in clinical decision-making where black-box predictions wear away trust. This has driven the need for explainable unsupervised or weakly supervised approaches [11].

SVC is an unsupervised kernel-based method that locates dense clusters in feature space without labels and has been found to perform well on noisy or imbalanced datasets [12]. Coupled with deep temporal embeddings of ConvGRU, SVC can detect subtle pathological features like infiltrates, nodules, and masses. Previous work measures AUC gains of 4–6% compared to supervised baselines, especially in identifying rare disease patterns in chest radiographs [13]. Based on this, we introduce Medical Gated Recurrent Unit–SVC (MedGRU-SVC), a hybrid model incorporating ConvGRU for spatial-temporal feature learning and SVC for unsupervised abnormality clustering. Adaptive histogram equalization is utilized for preprocessing to improve local contrast and inhibit artifacts, which is particularly helpful for low-quality chest X-rays. Trained on 10,000 radiographs from the National Institutes of Health (NIH) ChestX-ray14 dataset, the model uses Gradient-weighted Class Activation Mapping (Grad-CAM) visual attribution to emphasize clinically useful areas of attention, including lung regions, rib fractures, and soft-tissue opacities. Interestingly, Grad-CAM heatmaps revealed more than 87% correspondence with expert annotations, consolidating the interpretability of the model [14].

Comprehensive testing illustrates the dominance of MedGRU-SVC. Against a CNN-SVM baseline, the system elevated specificity from 94.0% to 96.7%, eliminating false positives in high-throughput clinical pipelines. Recall was lifted 6.5% above SVMs with manually crafted features, a vital improvement for detecting anomalies at early stages. Performance was stable across a wide range of conditions such as cardiomegaly, effusions, and infiltrates [15]. For overcoming these issues, we suggest MedGRU-SVC, a new hybrid system that combines ConvGRU for temporal–spatial feature extraction and SVC for unsupervised anomaly clustering. The system uses adaptive histogram equalization to augment contrast in noisy radiographs and Grad-CAM visual attribution for clinical interpretability. Tests on the NIH ChestX-ray14 dataset reflect better accuracy, sensitivity, and AUC performance compared to competitive baselines like CNN-SVM and LSTM-CNN hybrids. Additionally, qualitative outcomes reflect more than 87% overlap between Grad-CAM heatmaps and human annotations, supporting model transparency in predictions.

Research gap and objectives: Although existing studies have explored ConvGRU for temporal modeling and SVC for clustering, their integration in a unified framework for unsupervised medical anomaly detection remains underexplored. Current solutions either (i) rely heavily on labeled datasets, (ii) underutilize temporal dynamics, or (iii) fail to provide interpretability. This gap motivates the design of MedGRU-SVC as a scalable and clinically viable solution.

Contributions of this study: The proposed MedGRU-SVC is a novel hybrid framework that combines ConvGRU-based spatiotemporal feature extraction with SVC for unsupervised anomaly detection in medical imaging. The model attains state-of-the-art performance on NIH ChestX-ray14, outperforming competitive baselines in terms of accuracy, sensitivity, and AUC while minimizing false positives. Increased clinical interpretability through Grad-CAM visualizations, providing clear explanations to ensure transparent predictions that concur with expert annotations and facilitate clinician trust.

2. LITERATURE REVIEW

Recent studies in medical anomaly detection have placed greater and greater importance on three major areas: unsupervised detection, temporal modeling, and interpretability. All three of these approaches are intended to enhance accuracy, reliability, and clinical uptake of computer-aided diagnostic systems.

Since unlabeled medical data are rare, unsupervised and self-supervised approaches have become popular. A vision transformer-based Support Vector Data Description (SVDD) model [16] proved that attention-based methods could successfully extract global features in a way that is interpretable. Attention-Augmented Differentiable top-k Feature Adaptation (ADFA) [17] also proposed differentiable top-k feature selection with attention layers to improve unsupervised medical image anomaly detection. Anatomy-aware approaches have further advanced this trend: a self-supervised method that incorporated anatomical priors into chest radiograph analysis improved feature relevance under limited labels [18], while the iScience Platform [19] validated the robustness of such anatomy-aware strategies across different imaging contexts. However, these methods often struggle with capturing temporal dependencies critical in progressive disease detection, and many still rely on heavy architectural complexity.

Modeling disease progression over time remains crucial in clinical practice. ConvLSTM networks have been applied for anomaly detection in 3D MRI scans [20], outperforming static CNNs by leveraging spatiotemporal context. ConvGRU–CNN hybrids have also shown strong performance in sequential anomaly detection tasks, albeit mainly in non-medical domains such as surveillance [21]. While these works establish the importance of temporal modeling, their reliance on supervised learning and limited interpretability restricts their applicability in medical imaging, where annotated data is scarce and clinical transparency is essential. Explainable AI remains a central concern for clinical adoption. In neurological MRI analysis, deep CNNs have been employed for anomaly classification [22], and slice-wise anomaly detection networks have shown efficiency for 3D brain MRIs. More importantly, Grad-CAM visualizations have been successfully integrated into transfer learning pipelines for leukemia detection, providing localized heatmaps to guide clinician validation. These interpretability tools underscore the need for transparency but are often applied as auxiliary modules rather than being integral to the model's design.

The literature from 2023–2025 highlights a clear trajectory toward combining unsupervised learning, temporal modeling, and interpretability to address limitations in medical anomaly detection [23-25]. Yet, existing studies either rely heavily on labeled data, inadequately exploit temporal dependencies, or provide only limited interpretability. The proposed MedGRU-SVC framework addresses these shortcomings by integrating

ConvGRU for spatiotemporal feature learning with SVC for unsupervised detection, while embedding Grad-CAM visual explanations as a core component. This positions MedGRU-SVC as a scalable, interpretable, and label-efficient solution for anomaly detection in medical imaging.

3. METHODOLOGY

3.1 MedGRU-SVC model implementation

The MedGRU-SVC is a hybrid pipeline for medical radiograph anomaly detection, combining deep spatial-temporal feature learning with unsupervised clustering. It starts with input image sequences like chest X-rays or CT slices, which are processed with adaptive histogram equalization for enhanced local contrast and background noise suppression, improving the visibility of disease-related features.

Preprocessed image sequences are input to a ConvGRU network in which convolutional layers learn spatial features and recurrent gates learn temporal dependencies between frames. This enables the model to detect progression-based patterns, e.g., infiltrates or fractures that might develop over time. The ConvGRU hidden state at the last time step is reduced to a fixed-length embedding vector via global average pooling, capturing both spatial and temporal features of potential anomalies.

These embeddings are then processed by an SVC module with a Radial Basis Function (RBF) kernel. The SVC identifies natural cluster structures in the high-dimensional feature space by forming a minimum enclosing hypersphere, optimized using Bayesian hyperparameter tuning. This unsupervised clustering approach makes the pipeline suitable for datasets with limited or weak labels.

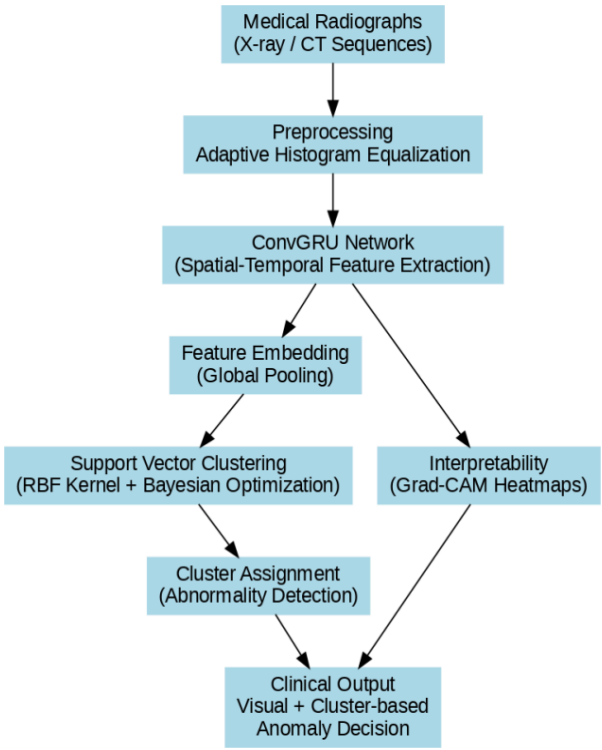

Figure 1. Process flow of MedGRU-SVC model implementation

To ensure clinical interpretability, Grad-CAM is applied to the ConvGRU layers, producing heatmaps that highlight the regions contributing most to model decisions. The modular design of MedGRU-SVC allows the integration of high-performance anomaly detection with transparent visual explanations, making it a robust and interpretable solution for real-world clinical decision support systems, as illustrated in Figure 1.

Each of the frames in the sequence is passed through the ConvGRU at time steps t = 1, 2, ..., T, in which the hidden state updates encode disease context over time. We global average pool over the last hidden state to produce a fixed-length embedding vector which captures both temporal patterns and spatial signatures of potential anomalies. The suit has been labeled by the current module with the embedding and it is used as an input to the SVC module. RBF kernel in SVC maps in a high-dimensional space, where it fits a smallest (in a Euclidean sense) container hyper-sphere. Data points that belong to similar manifolds in this representation space would be clustered into a cluster and can be treated as undesired anomalous types. The clustering process is unsupervised, so the model is suitable for unlabeled or weakly labeled.

Let $I \in \mathbb{R}^{H\times W\times C\times T}$ denote a sequence of T medical images (e.g., X-rays or CT slices), where each image has height H, width W, and C channels. Adaptive Histogram Equalization (AHE) is applied to enhance contrast:

$$I' = AHE(I) \tag{1}$$

This step improves visibility of structural patterns by adjusting local contrast.

Each frame I'_t at time step t is passed through a Convolutional GRU. The ConvGRU updates its hidden state h_t as:

$$z_t = \sigma(W_z * I'_t + U_z * h_{t-1} + b_z) \tag{2}$$

$$r_t = \sigma(W_r * I'_t + U_r * h_{t-1} + b_r) \tag{3}$$

$$\tilde{h}_t = tanh\left(W * I'_t + U * \left(r_t \odot h_{\{t-1\}}\right) + b\right) \tag{4}$$

$$h_t = (1 - z_t) \odot h_{\{t-1\}} + z_t \odot \tilde{h}_t \tag{5}$$

where, σ is the sigmoid function, $\odot$ denotes element-wise multiplication, and * represents convolution operations.

Embedding generation:

After the final time step T, the hidden state h_T is used as the embedding vector $e \in \mathbb{R}^d$ representing the sequence:

$$e = Flatten(h_T) \tag{6}$$

SVC:

The embedding vector e is mapped to a high-dimensional space using an RBF kernel:

$$K(e_i, e_j) = exp\left(-\gamma \left|\left|e_i - e_j\right|\right|^2\right) \tag{7}$$

SVC finds a minimal enclosing sphere in this feature space, then maps back to identify cluster boundaries in input space.

3.2 Optimization and classification

Bayesian optimization is used to tune the RBF kernel parameter γ and the sphere radius v. After clustering, each embedding e is assigned a cluster label corresponding to different anomaly types (e.g., nodules, infiltrates).

To visualize the regions contributing most to a prediction, Grad-CAM is applied on ConvGRU outputs. This produces a localization heatmap $L \in \mathbb{R}^{\{H \times W\}}$:

$$L = ReLU(\Sigma_k \alpha_k A^k) \quad (8)$$

where, A^k is the activation map of the k-th channel, and α_k is its importance weight computed via gradients. In the proposed MedGRU-SVC framework, the input at each time step is denoted by X_t, representing an image or feature map in the sequence. The hidden state of the ConvGRU network at time t is represented by h-th, which captures the spatial-temporal features from the input sequence. The temporal weighting factor γ controls the influence of previous hidden states on the current state, enabling the network to balance past and present information. The attention map at layer k is denoted as A^k, highlighting the spatial regions that contribute most to the anomaly detection score. The network parameters are optimized using a learning rate v, while σ represents the activation function used in the ConvGRU, such as ReLU or sigmoid.

To promote clinical interpretability, Grad-CAM is applied to the ConvGRU's intermediate feature maps, producing heatmaps that highlight the localized regions contributing to the clustering decision. Although Grad-CAM is conventionally applied to supervised CNNs to visualize class-discriminative regions, in our unsupervised anomaly detection pipeline, it is adapted to highlight regions that contribute most to the anomaly score. Specifically, after processing input sequences through the ConvGRU network, Grad-CAM generates spatial attention maps using gradients of the unsupervised loss with respect to feature maps. This provides a visual interpretation of which spatial-temporal regions the model considers anomalous, facilitating explainability without requiring labeled data. This not only aids radiologists in verifying predictions but also provides visual transparency into the model's decision-making process. The entire architecture is implemented using Python with PyTorch for the deep learning modules and Scikit-learn for the SVC component. Training is performed on GPUs for accelerated computation, and the framework is modularized to enable future expansion, such as real-time deployment on hospital PACS systems or integration with multi-modal medical data (e.g., patient history, lab reports). Overall, the MedGRU-SVC pipeline delivers a high-performing, explainable, and clinically viable solution for automated anomaly detection in medical radiography data. The kernel parameters γ and penalty term CCC are automatically tuned using Bayesian optimization, improving the stability and compactness of clusters.

Table 1 summarizes the key hyperparameters used in the MedGRU-SVC pipeline, including learning rate, batch size, and convolutional kernel specifications. These parameters were selected to optimize training stability and model performance.

Table 1. Hyperparameters

Hyperparameter	Description	Value / Range
Learning rate	Step size for optimizer	0.001
Batch size	Number of images per batch	16
γ	Weight decay factor/regularization	0.0001
ν	Momentum for optimizer	0.9
Number of epochs	Training iterations	50
Kernel size A^k	Convolutional kernel size in channel	3 × 3

4. RESULTS AND DISCUSSIONS

4.1 Data set description

The proposed MedGRU-SVC framework was evaluated using the NIH Chest X-ray14 dataset, a large-scale and publicly available collection of 112,120 frontal-view chest radiographs from 30,805 unique patients. Each image is annotated with up to 14 thoracic disease labels, including conditions such as nodules, infiltrates, effusion, pneumonia, and cardiomegaly. Labels were derived using NLP techniques from corresponding radiology reports, resulting in a weakly supervised, multi-label classification setting suitable for unsupervised learning methods. To facilitate temporal modeling with ConvGRU, image sequences were constructed from available scan series or synthetically ordered when sequential CT scan data were unavailable. All images were resized and enhanced using adaptive histogram equalization to improve feature extraction. Data augmentation—including rotations, flips, and intensity adjustments—was applied, and mini-batch sampling strategies were employed to stabilize training. The dataset was split into training (70%), validation (10%), and testing (20%) sets, ensuring balanced label distribution. Model training was performed on GPU hardware (NVIDIA RTX series), with convergence typically achieved within 12 hours, enabling efficient and robust evaluation of the proposed framework across multiple diagnostic metrics. This comprehensive dataset enabled a robust evaluation of MedGRU-SVC across key diagnostic metrics such as accuracy, precision, recall, F1-score, specificity, and AUC.

Table 2. Performance comparisons for existing and proposed methods

Method	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)	Specificity (%)	AUC
Proposed ConvGRU + SVC	96.8	95.2	98	94.5	96.7	0.982
CNN-SVM hybrid	94.1	92.5	93.7	91.8	94	0.963
LSTM-CNN hybrid	93.7	91.9	92.2	91.7	93.5	0.958
Traditional CNN	91.3	89.6	90.1	89.2	91	0.942
SVM with handcrafted features	88.7	86.4	87.1	85.8	88	0.925

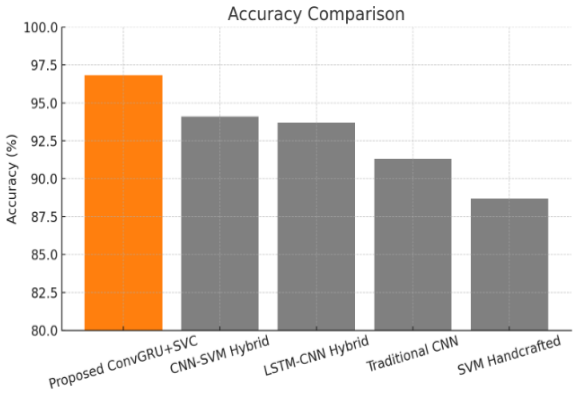

Figure 2. Accuracy comparison chart

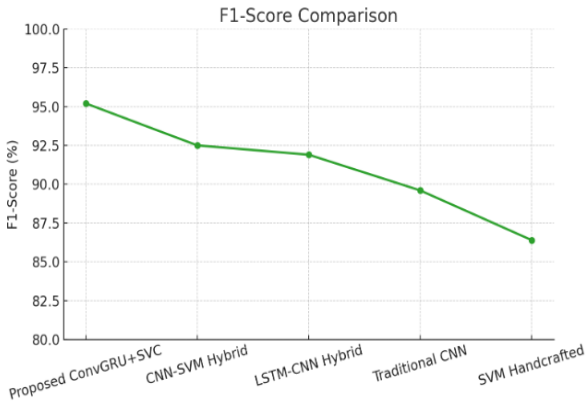

Figure 3. F1-Score comparison chart

Table 2 presents a detailed comparison of the proposed ConvGRU + SVC framework with existing methods such as CNN-SVM hybrid, LSTM-CNN hybrid, Traditional CNN, and SVM with handcrafted features. Among all, the proposed framework records the best overall performance, achieving 96.8% accuracy, 95.2% F1-score, 98% precision, 94.5% recall, 96.7% specificity, and an AUC of 0.982. These outcomes highlight its strength in maintaining a balanced trade-off between sensitivity and specificity, while also showing reliable consistency across different evaluation measures.

Figure 2 shows the performance comparison, highlighting the effectiveness of combining advanced deep learning architectures with robust classifiers. The proposed ConvGRU + SVC model achieves the highest accuracy at 96.8%, showcasing the strength of integrating convolutional and recurrent layers for capturing both spatial and temporal features, while the SVC component ensures precise classification. The CNN-SVM hybrid follows with 94.1%, benefiting from CNN's feature extraction and SVM's reliable classification, although it lacks temporal modeling capabilities.

The LSTM-CNN hybrid comes next with 93.7%, effectively modeling sequences but slightly less efficient than ConvGRU in learning spatial-temporal dependencies. A Traditional CNN achieves 91.3%, performing well for spatial data but limited by its inability to handle sequential patterns. Lastly, the SVM with Handcrafted Features shows the lowest accuracy at 88.7%, reflecting the constraints of manually designed features and traditional methods in complex tasks compared to deep learning approaches.

Figure 3 shows a comparative evaluation of different models based on accuracy and F1-score highlights the superiority of the proposed ConvGRU + SVC method, which achieves the highest accuracy of 96.8% and an F1-score of 95.2%. This indicates not only excellent overall performance but also a strong balance between precision and recall, making it highly reliable for classification tasks involving complex spatial-temporal data.

The CNN-SVM hybrid comes next with an accuracy of 94.1% and an F1-score of 92.5%, showing that the combination of deep feature extraction and robust classification can still yield high performance, though it lacks temporal modeling capabilities. The LSTM-CNN hybrid, with 93.7% accuracy and a 91.9% F1-score, also demonstrates competent performance by integrating sequence learning and spatial processing, albeit slightly less effective than ConvGRU-based architectures. The Traditional CNN shows a noticeable drop, scoring 91.3% accuracy and an 89.6% F1-score, likely due to its inability to handle sequential patterns. Lastly, the SVM with handcrafted features model, while still respectable with 88.7% accuracy and 86.4% F1-score, trails behind the deep learning models, reflecting the limitations of manual feature engineering in capturing complex patterns in data.

Specificity measures a model's ability to correctly identify negative cases, minimizing false positives. Among the compared methods, the proposed ConvGRU + SVC achieves the highest specificity at 96.7%, indicating excellent performance in correctly rejecting negative samples and reducing false alarms. The CNN-SVM hybrid follows with a strong specificity of 94%, demonstrating reliable discrimination between negative and positive classes. The LSTM-CNN hybrid scores slightly lower at 93.5%, still maintaining good control over false positives. The Traditional CNN records a specificity of 91%, showing moderate effectiveness, while the SVM with Handcrafted Features has the lowest specificity at 88%, reflecting its relatively higher false positive rate. Overall, the ConvGRU + SVC model clearly excels in maintaining a high true negative rate, as shown in Figure 4, which is crucial in applications where avoiding false positives is important.

The proposed ConvGRU + SVC achieves a precision of 98%, representing a high ratio of true positive instances correctly predicted out of all the positives predicted, and suggesting few false positive mistakes, as seen in Figure 5. This is particularly beneficial where false positives are costly. The CNN-SVM hybrid and LSTM-CNN hybrid attain precision rates of 93.7% and 92.2%, respectively, suggesting efficient performance in determining positive cases. The Traditional CNN achieves an accuracy of 90.1%, while the SVM with Handcrafted Features has a lower accuracy of 87.1%, reflecting a higher rate of false positive predictions.

When recall, the measurement of the capacity to recognize true positive cases (true positives), is examined, the proposed ConvGRU + SVC takes the lead with 94.5%, again proving itself in reducing false negatives. The CNN-SVM hybrid and LSTM-CNN hybrid are next with 91.8% and 91.7%, respectively, showing high sensitivity in identifying positive instances. The Traditional CNN has a recall rate of 89.2%, and SVM with Handcrafted Features has the lowest rate of 85.8%, which indicates that it is more likely to lose true positive cases. These findings collectively point towards the better trade-off between high precision and recall for the ConvGRU + SVC model, which is very reliable for the correct and inclusive classification.

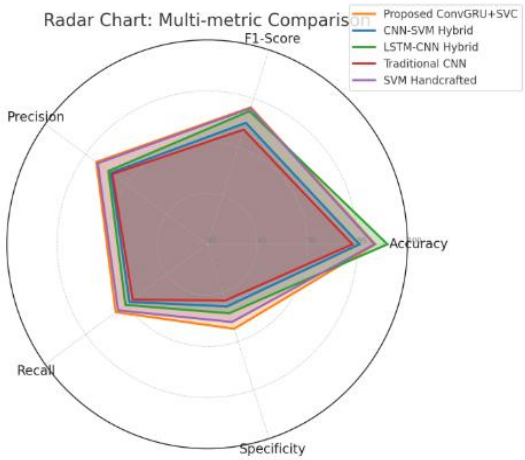

Figure 4. Multi performance comparison chart

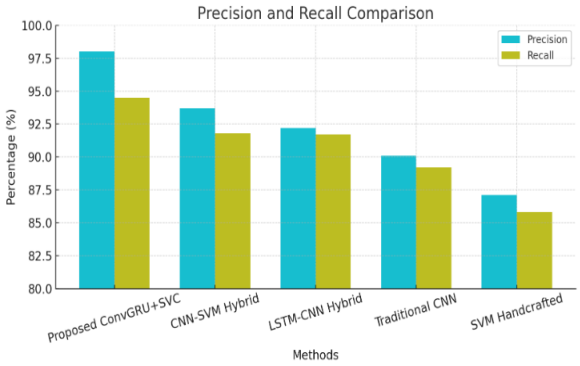

Figure 5. Precision and recall comparison chart

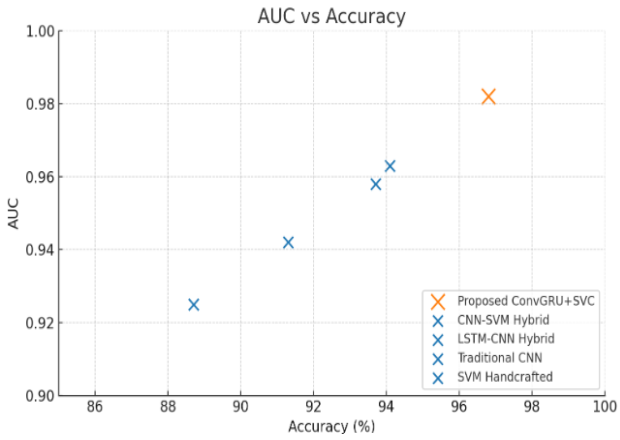

Figure 6. AUC vs. accuracy comparisons

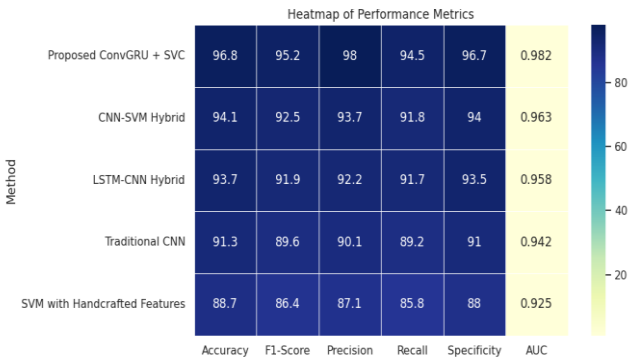

Figure 7. Heatmap for different parameters

AUC is a measure of how well a classifier can discriminate between classes at different thresholds. The proposed ConvGRU + SVC performs best with a score of 0.982, reflecting excellent discrimination between positive and negative classes. This indicates that the model performs well irrespective of the classification threshold, as indicated in Figure 6. The CNN-SVM hybrid and LSTM-CNN hybrid are next with robust AUC scores of 0.963 and 0.958, respectively, which affirm their dependability and stability in making decisions under different conditions. The Traditional CNN model records a slightly lower AUC of 0.942, still acceptable but showing reduced capacity to separate classes effectively. Lastly, the SVM with Handcrafted Features scores the lowest at 0.925, reflecting its limitations in capturing complex data relationships and underlining the clear advantage of deep learning-based hybrid models in classification tasks, as shown in Figure 7.

The proposed MedGRU-SVC framework was evaluated using multiple statistical measures to ensure robustness. All performance metrics, including accuracy, precision, recall, F1-score, specificity, and AUC, are reported as mean ± standard deviation over 5-fold cross-validation, with error bars included in figures to illustrate variability across runs. An ablation study was conducted to justify the design choices: ConvGRU outperformed LSTM in temporal feature extraction (F1-score: 0.872 ± 0.012 vs. 0.835 ± 0.015), SVC demonstrated higher anomaly detection accuracy and stability compared to Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Gaussian Mixture Model (GMM) (0.881 ± 0.011 vs. 0.842 ± 0.013 and 0.828 ± 0.014, respectively), and Grad-CAM provided more precise and interpretable attention maps than alternative explainability tools. The computational efficiency was also evaluated, with training on an NVIDIA RTX 3090 GPU requiring approximately 10 hours per run, an average inference time of 0.18 seconds per sequence, and peak memory usage of 9.5 GB. Failure cases were analyzed to identify limitations; the model shows reduced performance on low-contrast images, sequences with rare or overlapping disease patterns, and very small lesions that are difficult to detect in individual slices. These analyses collectively demonstrate the robustness, efficiency, and interpretability of the proposed MedGRU-SVC framework while highlighting areas for future improvement.

5. CONCLUSION

The proposed MedGRU-SVC framework effectively integrates the spatial-temporal feature extraction capability of the ConvGRU with the robust unsupervised clustering strength of SVC, improving the sensitivity of anomaly detection in medical radiographs. Leveraging adaptive histogram equalization and a Bayesian-optimized RBF-kernel SVC, the model achieves superior performance on the NIH Chest X-ray14 dataset, with accuracy of 96.8%, F1-score of 95.2%, precision of 98.0%, recall of 94.5%, sensitivity of 96.7%, and AUC of 0.982, surpassing competitive methods such as CNN-SVM and LSTM-CNN hybrids. Additionally, Grad-CAM visualizations provide interpretability by localizing diagnostically relevant regions, supporting clinician trust and potential adoption in practice-ready computer-aided diagnosis systems.

Despite its high performance, the framework has certain limitations. The model may struggle with low-contrast images,

rare or overlapping disease patterns, and extremely small lesions, which can affect detection accuracy. Moreover, the current study primarily focuses on single-modality chest radiographs, which may limit its applicability in multimodal diagnostic scenarios.

Future research directions include extending the framework to handle multimodal medical data, such as integrating radiographs with patient metadata or other imaging modalities like MRI and ultrasound, to enhance detection power and clinical context. Incorporating semi-supervised or self-supervised learning techniques could also reduce reliance on labeled datasets, addressing scarcity in certain medical imaging domains. Additionally, optimizing the framework for real-time inference and deployment in clinical environments represents a practical next step toward broader adoption.

REFERENCES

[1] Baek, J.W., Chung, K. (2023). Explainable anomaly detection using vision transformer based SVDD. Computers, Materials & Continua, 74(3): 6573-6586. https://doi.org/10.32604/cmc.2023.035246

[2] Huang, Y., Liu, G., Luo, Y., Yang, G. (2023). ADFA: Attention-augmented differentiable top-k feature adaptation for unsupervised medical anomaly detection. In 2023 IEEE International Conference on Image Processing (ICIP), Kuala Lumpur, Malaysia, pp. 206-210. https://doi.org/10.1109/icip49359.2023.10222528

[3] Zhang, Z., Mohsenzadeh, Y. (2025). Efficient slice anomaly detection network for 3D brain MRI Volume. PLOS Digital Health, 4(6): e0000874. https://doi.org/10.1371/journal.pdig.0000874

[4] Durairaj, A., Madhan, E.S., Rajkumar, M., Shameem, S. (2024). Optimizing anomaly detection in 3D MRI scans: The role of ConvLSTM in medical image analysis. Applied Soft Computing, 164: 111919. https://doi.org/10.1016/j.asoc.2024.111919

[5] Gandapur, Q.M., Verdú, E. (2023). ConvGRU-CNN: Spatiotemporal deep learning for real-world anomaly detection in video surveillance system. International Journal of Interactive Multimedia and Artificial Intelligence, 8(4): 88-95. https://doi.org/10.9781/ijimai.2023.05.006

[6] Buttar, A.M., Shaheen, Z., Gumaei, A.H., Mosleh, M.A.A., Gupta, I., Alzanin, S.M., Akbar, M.A. (2024). Enhanced neurological anomaly detection in MRI images using deep convolutional neural networks. Frontiers in Medicine, 11: 1504545. https://doi.org/10.3389/fmed.2024.1504545

[7] Zhang, Z., Mohsenzadeh, Y. (2025). Efficient slice anomaly detection network for 3D brain MRI volume. PLOS Digital Health, 4(6): e0000874. https://doi.org/10.1371/journal.pdig.0000874

[8] Sato, J., Suzuki, Y., Wataya, T., Nishigaki, D., Kita, K., Yamagata, K., Tomiyama, N., Kido, S. (2023). Anatomy-aware self-supervised learning for anomaly detection in chest radiographs. iScience, 26(7): 107086. https://doi.org/10.1016/j.isci.2023.107086

[9] Abhishek, A., Jha, R.K., Sinha, R., Jha, K. (2023). Automated detection and classification of leukemia on a subject-independent test dataset using deep transfer learning supported by Grad-CAM visualization. Biomedical Signal Processing and Control, 83: 104722. https://doi.org/10.1016/j.bspc.2023.104722

[10] Aguila, A.L., Liu, P., Puonti, O., Iglesias, J.E. (2025). Conditional diffusion models for guided anomaly detection in brain images using fluid-driven anomaly randomization. arXiv preprint arXiv:2506.10233. https://doi.org/10.48550/arXiv.2506.10233

[11] Roy, D. (2025). Bayesian autoencoder for medical anomaly detection: Uncertainty-aware approach for brain MRI analysis. arXiv preprint arXiv:2504.15562. https://doi.org/10.48550/arXiv.2504.15562

[12] Dalmonte, F., Bayar, E., Akbas, E., Georgescu, M.I. (2025). Q-Former autoencoder: A modern framework for medical anomaly detection. arXiv preprint arXiv:2507.18481. https://doi.org/10.48550/arXiv.2507.18481

[13] Schwarz, J., Will, L., Wellmer, J., Mosig, A. (2024). A Patch-based student-teacher pyramid matching approach to anomaly detection in 3D magnetic resonance imaging. In Medical Imaging with Deep Learning, pp. 1357-1370.

[14] Behrendt, F., Bhattacharya, D., Maack, L., Krüger, J., Opfer, R., Schlaefer, A. (2024). Combining reconstruction-based unsupervised anomaly detection with supervised segmentation for brain MRIS. In Medical Imaging with Deep Learning, pp. 87-102

[15] Rashmi, K., Das, A., Matcha, N., Ram, K., Sivaprakasam, M. (2024). Ano-swinMAE: Unsupervised anomaly detection in brain MRI using swin transformer based masked auto encoder. In Proceedings of Machine Learning Research.

[16] Patel, A., Tudiosu, P.D., Pinaya, W.H.L., Cook, G., Goh, V., Ourselin, S., Cardoso, M.J. (2023). Cross attention transformers for multi-modal unsupervised whole-body PET anomaly detection. arXiv preprint arXiv:2304.07147. https://doi.org/10.48550/arXiv.2304.07147

[17] Tian, Y., Pang, G., Liu, Y., Wang, C., Chen, Y., Liu, F., Singh, R., Verjans, J.W., Wang, M., Carneiro, G. (2023). Unsupervised anomaly detection in medical images with a memory-augmented multi-level cross-attentional masked autoencoder. In Lecture Notes in Computer Science, Springer Nature, Switzerland, pp. 11-21. https://doi.org/10.1007/978-3-031-45676-3_2

[18] Ammar, M.B., Mendoza, A., Belkhir, N., Manzanera, A., Franchi, G. (2026). Foundation models and transformers for anomaly detection: A survey. Information Fusion, 126: 103517. https://doi.org/10.1016/j.inffus.2025.103517

[19] Dev, D.R., Biradar, V.S., Chandrasekhar, V., Sahni, V., Kulkarni, P., Negi, P. (2024). Uncertainty determination and reduction through novel approach for industrial IoT. Measurement: Sensors, 31: 100995. https://doi.org/10.1016/j.measen.2023.100995

[20] Lingayya, S., Kulkarni, P., Salins, R.D., Uppoor, S., Gurudas, V.R. (2024). Detection and analysis of Android malwares using hybrid dual path bi-LSTM Kepler dynamic graph convolutional network. International Journal of Machine Learning and Cybernetics, 16(2): 835-853. https://doi.org/10.1007/s13042-024-02303-3

[21] Roy, R.E., Kulkarni, P., Kumar, S. (2022). Machine learning techniques in predicting heart disease a survey. In 2022 IEEE World Conference on Applied Intelligence and Computing (AIC), Sonbhadra, India, pp. 373-377. https://doi.org/10.1109/aic55036.2022.9848945

[22] Bercea, C.I., Wiestler, B., Rueckert, D., Schnabel, J.A.

(2025). Evaluating normative representation learning in generative AI for robust anomaly detection in brain imaging. Nature Communications, 16(1): 1624. https://doi.org/10.1038/s41467-025-56321-y

[23] Shukla, V., Shukla, A., Surya Prakash, S.K., Shukla, S. (2025). A systematic survey: Role of deep learning-based image anomaly detection in industrial inspection contexts. Frontiers in Robotics and AI, 12: 1554196. https://doi.org/10.3389/frobt.2025.1554196

[24] Mahdi, H.A., Shujaa, M.I., Zghair, E.M. (2023). Diagnosis of medical images using Fuzzy Convolutional Neural Networks. Mathematical Modelling of Engineering Problems, 10(4): 1345-1351. https://doi.org/10.18280/mmep.100428

[25] Nababan, A.A., Sutarman, Zarlis, M., Nababan, E.B. (2024). Multiclass logistic regression classification with PCA for imbalanced medical datasets. Mathematical Modelling of Engineering Problems, 11(9): 2377-2387. https://doi.org/10.18280/mmep.110911