Toward Neuro-Symbolic and Reservoir-Inspired Medical Imaging: A BD-CeNN Autoencoder with ASP Rule Mining for Robust and Explainable Interpretation of Grayscale and Color Images

MUGISHA MUHIRE JEAN-PIERRE[1], KASRA MORTAZAVI[2],
WITESYAVWIRWA VIANNEY KAMBALE[3], MAHMOUD HAMED[2],
SAMUEL MATIA KANGONI[1], KYANDOGHERE KYAMAKYA[1,2]
[1]Faculté Polytechnique, Université de Kinshasa (UNIKIN),
Kinshasa,
DEMOCRATIC REPUBLIC OF CONGO

[2]Institute of Smart Systems Technologies, Alpen-Adria-Universität Klagenfurt,
Universitätsstraße 65-67, 9020 Klagenfurt,
AUSTRIA

[3]Faculty of Information and Communication Technology,
Tshwane University of Technology,
Pretoria,
SOUTH AFRICA

Abstract: - We present a neuro-symbolic framework for medical image analysis that integrates a Binary Discrete Cellular Neural Network (BD-CeNN) autoencoder, a reservoir-computing–inspired BD-CeNN refinement stack, and automatically mined Answer Set Programming (ASP) rules. The autoencoder converts grayscale and color inputs (CT, MRI, histopathology, dermatology) into discrete, symbolic latent codes, which are iteratively refined to improve robustness and diagnostic discrimination. From annotated cases, ASP rules capture human-readable relations and constraints, enabling transparent, auditable reasoning over the learned symbols while maintaining predictive performance. The hybrid design targets resource-constrained clinical environments where trust, explainability, and adaptability are essential. This paper details the conceptual architecture, motivation, and deployment feasibility; extensive benchmarking is left for future work, laying the groundwork for accessible, interpretable AI in medical imaging.

Key-Words: Neuro-symbolic artificial intelligence, explainable medical imaging, Binary Discrete Cellular Neural Networks (BD-CeNN), reservoir computing, Answer Set Programming (ASP), multimodal symbolic encoding, symbolic feature extraction, temporal symbolic reasoning, interpretable clinical decision support.

Received: May 19, 2025. Revised: August 21, 2025. Accepted: September 20, 2025. Published: April 22, 2026.

1 Introduction

The rapid evolution of artificial intelligence (AI) has significantly transformed medical imaging by enhancing diagnostic accuracy and automation. In particular, deep learning techniques most notably convolutional neural networks (CNNs) have achieved human-level or even superior performance in various tasks such as lesion detection, disease classification, and image segmentation across a wide range of modalities, including radiography, magnetic resonance imaging (MRI), computed tomography (CT), ultrasound, and digital pathology, [1], [2], [3]. However, despite these advancements, the inherent opacity of such models continues to hinder their integration into clinical workflows, where interpretability, reliability, and clinician trust are essential for decision support, [2], [3].

1.1 Limitations of Existing Methods

Despite their success, black-box deep learning models often lack the explanatory clarity necessary for informed clinical decision-making. Although post-hoc interpretability techniques, such as Grad-CAM, SHAP, and LIME, offer visual explanations, these methods often fall short in terms of robustness, clinical reliability, and alignment with human reasoning processes, [2], [4], [5]. Contemporary surveys emphasize that genuinely trustworthy AI systems must be intrinsically interpretable, clinically grounded, and designed with end-user needs in mind, [2], [6]. Moreover, conventional static feedforward architectures are ill-equipped to replicate human-like cognitive processes. Clinicians often revisit regions of interest, integrate domain knowledge, and iteratively refine their interpretation capabilities, which are largely absent in standard deep learning systems, [7], [8]. Another primary concern involves the high computational demands of state-of-the-art models. Many top-performing AI systems rely on vast datasets, powerful GPUs, and energy-intensive infrastructure, rendering them unsuitable for

deployment in resource-constrained settings, such as rural hospitals and mobile diagnostic units, [9], [10].

1.2 Toward Neuro-Symbolic and Temporal Dynamics

To bridge these limitations, we introduce a neuro-symbolic framework that integrates:

1. A BD-CeNN-based autoencoder to extract symbolic representations from grayscale and color medical images.
2. A reservoir-style BD-CeNN stack to capture temporal and contextual dynamics.
3. An ASP (Answer Set Programming) engine to enable interpretable, rule-based diagnostic reasoning.

BD-CeNN Autoencoding

Binary Discrete Cellular Neural Networks (BD-CeNNs), which generalize the classical Cellular Neural Network paradigm, support local binary computation conducive to symbolic abstraction and efficient hardware implementation, [11], [12]. These networks serve as interpretable autoencoders capable of aligning extracted features with clinically meaningful concepts while maintaining high computational efficiency. Their binary nature enhances compatibility with edge-computing devices and supports real-time deployment scenarios, [13].

Reservoir-Style BD-CeNN

Inspired by the reservoir computing paradigm previously applied in temporal domains such as EEG classification and radiographic sequence analysis, we adapt BD-CeNN layers to function as symbolic reservoirs. These stacked layers simulate cognitive reasoning cycles by embedding temporal dependencies and integrating iterative patterns, [14], [15]. This approach addresses a critical shortcoming in traditional deep architectures: the inability to emulate dynamic inference pathways essential for nuanced clinical interpretation.

ASP Rule-Based Reasoning

Answer Set Programming (ASP) is a form of declarative logic programming that has gained traction for its formal rigor and human-auditable output, [16]. Our ASP module provides an interpretable reasoning layer on top of neural feature extraction by learning symbolic rules from labeled datasets. This logic-based engine enables transparent diagnostic pathways and supports clinical audibility, essential features often missing in deep learning systems. While ASP has roots in expert systems, its integration with neuro-symbolic models for visual diagnostic tasks remains largely unexplored, [17], [18].

1.3 Innovation and Scope

This triadic architecture, comprising discrete symbolic autoencoding, temporal reasoning via BD-CeNN reservoirs, and rule-based logic, systematically addresses four major limitations in existing medical AI systems:

1. Inherent interpretability: The entire pipeline is designed for transparency, addressing critical demands for explainable AI in clinical contexts, [2], [4], [19].
2. Multimodal flexibility: It supports both grayscale and color modalities, enabling applications across radiology, dermatology, pathology, and beyond, [1], [3], [11].
3. Temporal reasoning capability: Through the symbolic reservoir stack, our system mimics iterative human reasoning for dynamic diagnostic contexts, [14], [15].
4. Edge deployability: The lightweight nature of BD-CeNNs, combined with the logic-driven ASP layer, enables AI diagnostics in low-resource environments without sacrificing explainability or accuracy, [12], [13], [20].

Our previous works, [21], [22], demonstrate a long-standing and diverse expertise in cellular neural networks (CeNNs): the authors have published a series of peer-reviewed studies spanning 2011 to 2019 showing CeNNs as ultrafast, flexible solvers for stiff ordinary and partial differential equations, real-time computational engineering, local traffic-signal control, time-series modelling and forecasting in transportation, and even raindrop detection for advanced driver-assistance systems. This experience builds on established CeNN theory and early architectures, [23], [24], [25], and collectively establishes a solid, multidisciplinary foundation for designing CeNN-based algorithms that operate in real-time and address practical engineering problems.

A comparative summary of the main limitations in current medical imaging AI and how our approach addresses them is provided in Table 1 (Appendix). This overview also motivates the research gaps discussed in Section 2.

2 A Critical Review of the Related State-of-the-Art

This section presents a comprehensive review of existing research, emphasizing the scientific gaps this

work seeks to address and highlighting the resulting innovation potential.

2.1 Explainable AI in Medical Imaging

The demand for explainable AI (XAI) in medical diagnostics is steadily increasing due to its implications for clinical safety, regulatory compliance, and trust, [2], [4], [6]. Despite progress, current XAI strategies typically fall into two main categories:

- Post-hoc methods such as Grad-CAM, SHAP, and LIME, while broadly applicable, often lack fidelity and stability in complex medical settings, [2], [5], [19].
- Self-explainable (S-XAI) or intrinsically interpretable models are more desirable but remain scarce in real-world clinical workflows due to their architectural limitations and trade-offs in performance, [7], [19].

Leading reviews emphasize that for medical adoption, XAI must be truthful, clinically relevant, and resource-efficient, [6], [26]; yet, few existing methods meet these criteria under operational constraints.

2.2 Hybrid and Neuro-Symbolic AI

Neuro-symbolic systems offer a promising fusion of sub-symbolic learning and symbolic reasoning, aiming to combine neural perception with logical inference, [17], [27]. For example, frameworks like NeuroSymAD utilize CNNs in conjunction with ASP-based logic rules for Alzheimer's diagnosis, [26], offering interpretable justifications that align with clinical reasoning. However, most of these systems are limited to grayscale neuroimaging, rely on floating-point CNNs, and lack support for binary symbolic processing or color image integration, which restricts their generalizability across imaging domains.

2.3 Reservoir Computing in Medical Imaging

Reservoir computing (RC), including Echo State Networks (ESNs) and Liquid State Machines (LSMs), has proven effective in temporal signal processing, particularly in applications such as EEG-based emotion recognition and seizure detection, [14], [15]. Despite its potential to model temporal dependencies and iterative reasoning, RC has not yet been systematically applied to static medical images, such as CT scans or histopathology images. This presents an opportunity to adapt RC principles for dynamic feature refinement in image analysis.

2.4 Discrete Neural Networks and BD-CeNNs

Binary Discrete Cellular Neural Networks (BD-CeNNs) extend traditional CNNs by enabling localized binary computations, facilitating high interpretability, low complexity, and edge deployability, [11], [12], [13]. These networks have been utilized for image segmentation and denoising; however, their integration into reservoir architectures or neuro-symbolic frameworks for medical imaging remains unexplored. Moreover, no current system fully leverages discretized symbolic encoding across grayscale and color modalities in a unified pipeline.

2.5 ASP Rule Mining in Diagnostics

Answer Set Programming (ASP) supports logic-based reasoning with auditable decision paths, critical for trustworthy medical AI, [16], [17], [18]. Prior use cases in diagnostic systems often required manually crafted rules, limiting scalability. While some recent works automatically extract rules from CNN activations, [26], they often lack symbolic abstraction and are rarely based on binary or discrete representations. No known framework currently auto-mines ASP rules from BD-CeNN symbolic encodings nor applies them across diverse imaging types.

2.6 Scientific Gaps & Innovation Potential

To position our contribution within the broader landscape of AI in medical imaging, Table 1 provides a comparative summary of prevailing approaches, unresolved challenges, and the innovations introduced by our proposed framework. While significant progress has been made in deep learning–based diagnostics, the field remains hindered by substantial shortcomings that impact explainability, generalizability, and clinical integration. Our approach addresses six identified limitations directly, combining symbolic reasoning, discrete computation, dynamic inference, and a lightweight architecture in a unified neuro-symbolic system.

2.6.1 Interpretability

One of the most persistent gaps in medical AI is the lack of inherently interpretable models. Current systems rely heavily on post-hoc explanation methods, such as Grad-CAM, SHAP, and LIME, which, despite their visual appeal, frequently yield explanations that are neither clinically validated nor robust across samples, [2], [5]. These models operate as black boxes, offering no guarantees that the explanations correspond to genuine causal

factors. Furthermore, their outputs are difficult to audit or defend in medico-legal contexts. Our solution addresses this by embedding interpretability at the architectural level, utilizing a BD-CeNN autoencoder to generate symbolic features, which are then interpreted through ASP-based logic reasoning. This combination ensures that explanations are not reverse-engineered but produced as part of the inference process, offering traceable, verifiable, and human-auditable justifications, [2], [6], [19].

2.6.2 Symbolic Integration

Symbolic reasoning enables clinicians and systems to follow rules, validate logical consistency, and generalize across domains. Yet, most deep learning pipelines lack symbolic integration. Even neuro-symbolic approaches often isolate the symbolic component to a post-processing step, detached from the learned representations. This disconnect limits their effectiveness and interpretability. By contrast, our model natively encodes symbolic structure through BD-CeNNs and maintains symbolic continuity by coupling this with ASP-based rule-based inference. This tight integration bridges neural perception with logic-driven decision-making, bringing AI workflows closer to the structured, rule-informed reasoning used by clinicians, [17], [27].

2.6.3 Temporal Inference

Diagnostic reasoning is often iterative and temporal in nature. Clinicians revisit regions of interest, adjust hypotheses, and incorporate contextual cues over time. However, most current AI systems process images in static, one-pass feedforward models, ignoring temporal or sequential aspects of interpretation. Inspired by reservoir computing (RC), our system introduces BD-CeNN reservoir layers that preserve prior activations and evolve symbolic features across multiple inference cycles. This mimics the temporal dynamics and recursive attention mechanisms found in expert-level clinical practice, [14], [15], marking a novel use of RC principles in symbolic medical image analysis.

2.6.4 Automated Rule Induction

While ASP has long been used for expert systems, traditional rule engines require manually defined rule sets, which are difficult to scale, maintain, and validate, [16], [17], [18]. Some recent approaches attempt to extract rules from CNN activations, but they typically lack symbolic fidelity or operate on fuzzy, uninterpretable features. Our approach automates rule mining directly from the discretized symbolic features produced by the BD-CeNN. This facilitates the creation of meaningful, structured ASP rules with strong semantic alignment to image content and clinical categories, [18], [26]. The resulting system produces both interpretable and data-driven explanations, eliminating the bottleneck of manual rule definition.

2.6.5 Multimodal Support

Medical imaging encompasses a diverse range of modalities, including grayscale images such as X-rays and MRIs, as well as color-rich images like dermatological scans or histopathological slides. Many XAI and neuro-symbolic systems are restricted to grayscale modalities, especially in the neurological imaging domain, [26]. Our framework generalizes across grayscale and color inputs, supporting a wider range of clinical applications. This cross-modality compatibility enhances usability in mixed-imaging workflows (e.g., oncology, dermatology, pathology), expanding the potential for real-world deployment across departments and specialties.

2.6.6 Resource-Efficiency and Edge Deployability

High-performing CNNs require significant computational resources, including GPUs, cloud access, and energy consumption. These constraints limit their adoption in rural clinics, point-of-care devices, and mobile diagnostic units, where computing capacity and internet access may be limited, [9], [10], [20]. In contrast, our use of BD-CeNNs, which operate with binary, localized computations, drastically reduces hardware requirements and power consumption, [11], [13]. When coupled with symbolic reasoning via ASP, which is lightweight and hardware-efficient, the result is a model inherently suited for edge deployment in resource-constrained environments. The proposed framework addresses six significant gaps in the current literature and technology stack for medical AI systems, as shown in Table 1 (Appendix).

2.7 Why This Matters

By embedding interpretability into the core architecture, this framework overcomes the brittleness of post-hoc XAI. As an S-XAI model, it aligns with clinical demands for auditability, transparency, and efficiency, [4], [6], [19]. Furthermore, the introduction of BD-CeNN reservoirs offers a novel approach to iterative reasoning, simulating multi-pass diagnostic reasoning used by clinicians, [8], [27]. Crucially, this system generalizes to both grayscale and color imaging, contrasting with the narrow scope of most existing XAI systems and leverages binary symbolic encoding to optimize for speed, power consumption, and interpretability, [11], [13], [20]. To our knowledge, this is the first framework to integrate:

1. Symbolic BD-CeNN autoencoding.
2. Temporal refinement through a reservoir-like structure.
3. ASP-based logic reasoning for diagnosis.

This unified design lays a foundation for the next generation of neuro-symbolic AI tailored for transparent, adaptive, and deployable medical diagnostics.

3 Our System Architecture

The proposed pipeline blends data-driven and logic-driven mechanisms across four stages. We begin with the symbolic latent encoding module (Figure 1 (Appendix)), which provides the cognitive interface between machine perception and human-level reasoning. The complete four-stage workflow is summarized in Figure 2 (Appendix) and detailed in the following subsections.

3.1 Stage 1: BD-CeNN Autoencoder – Symbolic Latent Encoding

This first stage transforms raw medical images, whether grayscale (e.g., CT, MRI) or color (e.g., histology, dermatology), into compact, symbolic, and interpretable representations (Figure 1 (Appendix)). These representations serve as the cognitive interface between machine perception and human-level reasoning. This is achieved via a Binary Discrete Cellular Neural Network (BD-CeNN)-based autoencoder, a biologically and symbolically inspired system designed to extract clinically relevant features as logical, structured, and composable symbolic tokens. This stage forms the semantic backbone of the neuro-symbolic pipeline, defining how visual data enters a logic-driven, interpretable AI ecosystem.

3.1.1 From Visual Data to Symbolic Knowledge

Medical images are traditionally analyzed by clinicians in conceptual terms:

- ”The lesion has spiculated margins and central necrosis.”
- ”This pattern resembles a honeycombing seen in pulmonary fibrosis.”
- ”The nucleus is hyperchromatic with irregular contours.”

AI systems based on convolutional neural networks (CNNs) often fail to replicate this abstraction ability, producing uninterpretable numerical embeddings that are even unintelligible to experts. The BD-CeNN autoencoder addresses this by:

- Translating visual data directly into symbolic assertions.
- Capturing what is present in an image, where it is located, and how it behaves.
- Organizing this information into discrete logic-compatible structures suitable for reasoning, feedback, and human-machine collaboration.

3.1.2 Preprocessing and Harmonization Across Modalities

To handle diverse imaging modalities, a harmonization process ensures a modality-agnostic symbolic encoding:

- Grayscale images (e.g., CT scans): focus on density gradients, edge structure, and anatomical location.
- Color images (e.g., histology, skin lesions): encode hue, saturation, and color contrast into symbolic units (e.g., ”reddish center,” ”blue cytoplasm”).

Rather than treat each modality separately, BD-CeNN layers normalize and encode clinical visual cues into a unified symbolic vocabulary, enabling cross-modal logic and diagnostics.

3.1.3 BD-CeNN Architecture: Structure and Function

A BD-CeNN consists of a 2D cellular grid, where each cell is a binary unit (0 or 1), and its next state depends on:

- The current state of itself and its neighbors.
- A discrete update rule (expressed as a Boolean or symbolic function).
- Local topology (typically Moore or von Neumann neighborhoods).

This architecture is inspired by biological neural fields and cellular automata, providing:

- Spatial pattern emergence without convolutions.
- Discrete semantics from local rules.
- Composability and modularity are ideal for symbolic AI.

Each BD-CeNN layer acts as a symbolic transformation stage, gradually abstracting away pixel-level data into structured patterns.

3.1.4 Mathematical Formulation of the BD-CeNN Autoencoder

Let the input image be $I \in [0,1]^{H\times W\times C}$, where $C \in \{1,3\}$ denotes grayscale or color channels. After preprocessing and binarization, we obtain an input field $u \in \{0,1\}^{H\times W\times F_0}$ (e.g., thresholded intensity/edge/color cues), where F_0 is the number of binary input feature maps.

Cell grid and neighborhood topology: A BD-CeNN layer is defined over a 2D grid of cells $\mathcal{V} = \{1,\dots,H\}\times\{1,\dots,W\}$. Each cell $(i,j)\in\mathcal{V}$ has a binary state $x_{i,j}^{(\ell,t)} \in \{0,1\}$ at discrete time t in encoder layer ℓ. We formally define common neighborhood systems with radius r in Eqs. (1)–(2):

$$\mathcal{N}_r^{\text{Moore}}(i,j) = \{(p,q) : \max(|p-i|,|q-j|) \le r\}, \tag{1}$$

$$\mathcal{N}_r^{\text{vN}}(i,j) = \{(p,q) : |p-i| + |q-j| \le r\}. \tag{2}$$

In our implementation, $\mathcal{N}(i,j)$ is chosen as either $\mathcal{N}_r^{\text{Moore}}$ or $\mathcal{N}_r^{\text{vN}}$ and kept fixed for all updates.

Explicit BD-CeNN state-update equation: Given local templates $A^{(\ell)} \in \mathbb{R}^{(2r+1)\times(2r+1)}$ and $B^{(\ell)} \in \mathbb{R}^{(2r+1)\times(2r+1)}$, bias $\beta^{(\ell)} \in \mathbb{R}$ and threshold $\theta^{(\ell)} \in \mathbb{R}$, the synchronous binary update is given in Eq. (3):

$$x_{i,j}^{(\ell,t+1)} = \mathbb{H}\Bigg(\sum_{(p,q)\in\mathcal{N}(i,j)} A_{p-i,q-j}^{(\ell)}\, x_{p,q}^{(\ell,t)} + \sum_{(p,q)\in\mathcal{N}(i,j)} B_{p-i,q-j}^{(\ell)}\, u_{p,q}^{(\ell)} + \beta^{(\ell)} - \theta^{(\ell)} \Bigg), \tag{3}$$

where $\mathbb{H}(\cdot)$ is the Heaviside step function ($\mathbb{H}(s)=1$ if $s \ge 0$, else 0) and $u^{(\ell)}$ is the binary input to layer ℓ (for $\ell = 1$, $u^{(1)} = u$; for deeper layers $u^{(\ell)}$ can be set to the previous layer output). To encourage edge deployability and interpretability, the templates may be restricted to low-precision values (e.g., $A_{a,b}^{(\ell)}, B_{a,b}^{(\ell)} \in \{-1,0,1\}$), which makes each update a transparent local vote/logic aggregation.

Encoder and decoder as explicit functions: Let $\Phi^{(\ell)}$ denote T_ℓ repeated applications of Eq. (3) in layer ℓ. The encoder is the composition in Eq. (4):

$$E_\phi(I) = z = g\Big(\Phi^{(L)} \circ \cdots \circ \Phi^{(1)}\big(u(I)\big)\Big) \in \{0,1\}^M, \tag{4}$$

where $u(I)$ is the preprocessing/binarization operator, L is the number of BD-CeNN encoder layers, and $g(\cdot)$ maps the final binary cell field to a symbolic latent vector z. A simple, reproducible g is regional pooling with thresholds, defined in Eq. (5): for concepts $m = 1,\dots,M$ associated with regions $\mathcal{R}_m \subseteq \mathcal{V}$ and thresholds κ_m,

$$z_m = \mathbb{I}\left[\sum_{(i,j)\in\mathcal{R}_m} x_{i,j}^{(L,T_L)} \ge \kappa_m \right]. \tag{5}$$

The decoder reconstructs $\hat{I}$ from z via prototypes (or learned sparse bases) P_m as in Eq. (6):

$$\hat{I} = D_\psi(z) = \sigma\left(\sum_{m=1}^{M} z_m\, P_m \right), \tag{6}$$

where $\sigma(\cdot)$ is a squashing function (e.g., sigmoid) applied elementwise.

Symbolic latent space and semantics: We define a concept dictionary $\mathcal{D} = \{c_1,\dots,c_M\}$ (e.g., `border_irregular`, `hyperdense_core`, `radial_streaks`). The symbolic latent space for image I is the set in Eq. (7):

$$\mathcal{S}(I) = \{\, c_m \;:\; z_m = 1 \,\}, \tag{7}$$

which is directly mapped to ASP facts (Stage 3) using predicates such as `active(img_id, `c_m`)`. Sparsity and interpretability can be quantified by the fraction of active symbols in Eq. (8):

$$\text{spar}(z) = \|z\|_0 / M. \tag{8}$$

Training objective and reproducibility: With a dataset $\{(I^{(n)}, y^{(n)})\}_{n=1}^N$, the autoencoder parameters ϕ,ψ can be learned by minimizing Eq. (9):

$$\min_{\phi,\psi} \sum_{n=1}^{N} \Big(\|I^{(n)} - D_\psi(E_\phi(I^{(n)}))\|_1 + \lambda_s \|E_\phi(I^{(n)})\|_0 \Big), \tag{9}$$

where the first term enforces reconstruction fidelity and the second enforces a sparse, symbolic code. Since the system is discrete, optimization can be carried out using standard discrete relaxations (e.g., straight-through estimators for $\mathbb{H}$) or template selection over a finite template library.

Dynamics, convergence notion, and computational complexity: Because the state space is finite (2^{HW} per feature map), the synchronous update induces a finite-state dynamical system; thus, every trajectory is eventually periodic. In this work we define *convergence* as reaching a

fixed point within T_ℓ updates, i.e., the Hamming stability condition in Eq. (10) holds for some $t \leq T_\ell$:

$$\|X^{(\ell,t+1)} - X^{(\ell,t)}\|_0 = 0. \tag{10}$$

The per-layer computational cost is given in Eq. (11):

$$\mathcal{O}(T_\ell \, H \, W \, |\mathcal{N}|), \tag{11}$$

where $|\mathcal{N}|$ is the neighborhood size; this yields a predictable, edge-friendly runtime bound.
The BD-CeNN autoencoder mechanism is fully specified as derived from Eqs. (1)–(11).

3.1.5 Encoding Pipeline: Layers of Symbolic Abstraction

Let's break down the encoder's symbolic transformation process:

Layer 1: Low-Level Perceptual Abstraction

- Detects contrasts, blobs, gradients, and boundary primitives.
- Outputs binary maps such as "edge present," "contrast peak," and "chromatic variance high."

Layer 2–3: Mid-Level Structure Encoding

- Aggregates patterns across spatial and chromatic neighborhoods.
- Encodes regional features like "ring shape," "central void," and "bilateral symmetry."

Top Layer: Symbolic Mapping

- Maps latent spatial patterns to symbolic units using a pre-learned or manually designed dictionary.
- Outputs symbolic labels:

```
lesion_border = irregular,
intensity_core = high,
color_pattern = reticular.
```

3.1.6 Symbolic Latent Space: Format and Semantics

The symbolic latent space is the output of the encoder and the input to all downstream logic. It is:

- Sparse: Only meaningful concepts are activated.
- Discrete: No ambiguous floating-point values, just true/false or categorical symbols.
- Semantic: Each symbol is meaningful, editable, and explainable.
- Relational: Each symbol can participate in logic rules (e.g., "IF lesion_size=large AND core_density=low THEN diagnosis=suspect").

This symbolic output can be directly parsed by logic engines, temporal propagators, or knowledge graphs, making it a powerful foundation for hybrid AI.

3.1.7 Optional Decoder: Auditing and Interactive Refinement

The decoder is used during training, development, or clinician-in-the-loop workflows to:

- Validate that the symbolic code retains diagnostic visual essence.
- Provide symbolic saliency maps, showing which tokens affect which image regions.
- Allow interactive editing, where a physician can turn symbolic indicators on and off and observe the reconstructed implications.

In effect, the decoder acts as a symbolic visualizer and auditor. This is essential for:

- Explaining predictions to clinicians or regulators.
- Training users or students on symbolic features.
- Debugging symbolic bottlenecks.

3.1.8 Clinical Example: CT Chest Scan

```
1. BD-CeNN layer 1: detects
 ring-shaped high-intensity regions
 (potential lesions).
2. Layer 2-3: characterizes the
 boundary as "spiculated" and
 location as "peripheral."
3. Final layer:
 outputs symbolic vector:
    (1) lesion_shape = irregular
    (2) density_core = hyperdense
    (3) boundary_type = spiculated
    (4) position = peripheral
4. The decoder reconstructs a
 coarse symbolic overlay.
5. ASP in later
 stages interprets:
    (1) IF lesion_shape = irregular
    AND density_core = hyperdense
    AND boundary_type = spiculated
    THEN risk = high_malignancy.
```

This shows the full semantic interpretability path from pixels to logic to diagnosis.

3.1.9 Scientific and Practical Benefits

The proposed framework introduces several distinctive features supporting transparency, adaptability, and clinical integration. First, interpretability is ensured by the symbolic nature of the model: each token in the output space corresponds to a clearly defined concept, allowing direct semantic mapping. Second, the architecture supports composability, meaning that new medical concepts or rules can be introduced incrementally without requiring full retraining of the model, an essential trait for evolving clinical environments. Third, the system emphasizes explainability by constructing human-traceable inference chains, making each diagnostic decision auditable and reproducible. Additionally, edge deployment is enabled through the lightweight computational design of BD-CeNN logic, which allows the model to function efficiently even on resource-constrained devices. The framework is also designed with ontology compatibility in mind; its symbolic outputs can be seamlessly mapped to standardized clinical coding systems, such as SNOMED or ICD, thereby enhancing interoperability with electronic health records. Lastly, the architecture supports interactive feedback loops, allowing clinicians to directly modify symbolic codes to correct or refine model outputs, thereby fostering a collaborative human–AI diagnostic workflow.

3.1.10 Theoretical Foundations and Inspirations

This approach draws inspiration from multiple foundational paradigms. From symbolic artificial intelligence, it inherits the capacity to represent and manipulate knowledge using logic-based systems. Cognitive psychology contributes to the notion that human perception functions through abstraction and feature selection, a process mirrored in the system's encoding mechanisms. The architectural design also reflects principles from cellular automata, where complex global behavior emerges from local interactions, echoing the BD-CeNN's spatially distributed processing. Furthermore, the framework aligns with the goals of neuromorphic computing, prioritizing energy-efficient, logic-driven architectures ideal for deployment in edge AI scenarios. At the core of this methodology lies the BD-CeNN autoencoder, which synthesizes these interdisciplinary influences into a unified, medically grounded application. This component does more than reduce dimensionality; it transforms raw visual input into a symbolic representation that serves as the foundation for logic-based reasoning. It captures diagnostically salient features from both grayscale and color medical images, encoding them in a form that is:

- Compact, interpretable, and inherently transparent.
- Optimized for symbolic manipulation and logical inference.
- Resilient to variation while remaining clinically meaningful.

This symbolic abstraction stage is not merely a precursor to reasoning; it is the semantic engine of the entire neuro-symbolic framework. By bridging perceptual and cognitive layers, the BD-CeNN autoencoder enables a shift from opaque prediction to interpretable and traceable diagnostic intelligence, marking a decisive advancement in designing explainable, deployable AI systems in healthcare.

3.2 Stage 2: BD-CeNN Reservoir – Temporal Symbolic Refinement

After symbolic features have been extracted via the BD-CeNN autoencoder, the next critical processing step is their refinement and dynamic enhancement through a specialized BD-CeNN reservoir configured in accordance with the principles of Reservoir Computing (RC). This architectural layer is designed to retain key spatial and symbolic cues and emulate the temporal, iterative analytical behavior characteristic of human diagnostic reasoning. While traditional image-processing models employ single-pass feedforward logic, our reservoir architecture enables symbolic information to evolve through multiple discrete time steps, with each layer serving as a symbolic reasoning iteration. This gives rise to a system that thinks not just deeply but repeatedly and contextually.

Architecture and Function of the BD-CeNN Reservoir.

The BD-CeNN reservoir comprises a stack of discretized binary processing layers, each modeled as a Cellular Neural Network (CeNN) with local, cell-based interactions. Unlike typical CNNs, these networks operate on binary states with discrete update rules, resulting in symbolic transformations rather than floating-point tensor convolutions.

Each layer in the stack performs the following:

- Applies fixed local update rules to symbolic feature maps (e.g., cell flips based on logical templates),
- Transmits refined symbolic states to the next layer (analogous to virtual time steps),
- Enables non-linear transformations over symbolic codes via binary cellular interactions.

The result is a computationally lightweight, energy-efficient mechanism for temporal-like symbolic evolution.

3.2.1 Reservoir-Style BD-CeNN as a Discrete Dynamical System

Let $Z \in \{0,1\}^{H\times W\times F}$ denote the symbolic feature maps produced by Stage 1 (before ASP), and let $R^{(t)} \in \{0,1\}^{H\times W\times F}$ denote the reservoir state at virtual time t (or equivalently, at depth t in the BD-CeNN reservoir stack). We initialize the reservoir as in Eq. (12):

$$R^{(0)} = Z. \tag{12}$$

For $t = 0,\dots,T_R-1$, the reservoir evolves via a fixed (non-trained) BD-CeNN transition given in Eq. (13):

$$R^{(t+1)}_{i,j,f} = \mathbb{H}\Bigg(\sum_{(p,q)\in\mathcal{N}(i,j)} \sum_{f'=1}^{F} \widetilde{A}_{f,f'}(p-i,q-j)\, R^{(t)}_{p,q,f'} + \sum_{(p,q)\in\mathcal{N}(i,j)} \sum_{f'=1}^{F} \widetilde{B}_{f,f'}(p-i,q-j)\, Z_{p,q,f'} + \widetilde{\beta}_f - \widetilde{\theta}_f \Bigg). \tag{13}$$

where $\widetilde{A}, \widetilde{B}$ are sparse local connectivity templates (fixed after initialization) and $\mathcal{N}$ is the same formal neighborhood family as in Eq. (3). Equation (13) explicitly formalizes the paper's statements that the reservoir "refines" symbols by local propagation and suppression.

Connectivity structure and initialization: To instantiate a reproducible reservoir, we set each template coefficient to a small discrete value with prescribed sparsity ρ as derived from Eqs. (13)–(14):

$$\widetilde{A}_{f,f'}(\cdot), \widetilde{B}_{f,f'}(\cdot) \sim \mathrm{Unif}\{-1,0,1\}, \qquad \Pr(\neq 0) = \rho. \tag{14}$$

To promote stable "echo-state-like" behavior in the discrete setting, we enforce the bounded local influence condition in Eq. (15):

$$\max_{i,j,f} \sum_{(p,q)\in\mathcal{N}(i,j)} \sum_{f'=1}^{F} \left|\widetilde{A}_{f,f'}(p-i,q-j)\right| \le \Gamma, \tag{15}$$

with a small Γ (e.g., $\Gamma \le 1$), which limits how strongly a single update can amplify disagreements. In analogy to the echo-state property in classical reservoir computing (often enforced via spectral radius control), Eq. (15) plays the role of a discrete contractivity condition by limiting local amplification under a Hamming-like metric.

Formal compatibility and noise suppression: We quantify symbolic *compatibility* between neighboring cells by a disagreement energy on the binary field defined in Eq. (16):

$$\mathcal{E}(R) = \sum_{(v,v')\in\mathcal{E}_{\mathcal{N}}} w_{v,v'}\, \|R_v - R_{v'}\|_1 \; - \; \sum_{v} h_v^{\top} R_v, \tag{16}$$

where $\mathcal{E}_{\mathcal{N}}$ is the set of neighborhood edges induced by $\mathcal{N}$, $R_v \in \{0,1\}^F$ is the feature vector at cell $v=(i,j)$, $w_{v,v'} \ge 0$ penalizes incompatible neighbor symbols, and h_v encodes unary preferences (e.g., keep strong activations). Operationally, "suppressing weak/erroneous activations" corresponds to updates that reduce $\mathcal{E}$ by removing isolated activations that increase neighborhood disagreement, consistent with Eqs. (13) and (16).

Convergence criterion: We define reservoir convergence at time t by the Hamming stability condition in Eq. (17):

$$\Delta(t) = \|R^{(t+1)} - R^{(t)}\|_0 = 0, \tag{17}$$

or by an energy plateau $\mathcal{E}(R^{(t+1)}) = \mathcal{E}(R^{(t)})$ as derived from Eq. (16). Because $R^{(t)}$ is finite-state, trajectories are eventually periodic; in deployment we cap T_R and optionally stop early when Eq. (17) holds.

Computational complexity: The reservoir update cost is given in Eq. (18) for the general multi-feature coupling case, or in Eq. (19) if features are updated independently:

$$\mathcal{O}(T_R\, H\, W\, |\mathcal{N}|\, F^2), \tag{18}$$

$$\mathcal{O}(T_R\, H\, W\, |\mathcal{N}|\, F). \tag{19}$$

Virtual Time Steps in Static Imaging

Even though medical images are static snapshots (e.g., CT slices and dermatoscopic scans), the BD-CeNN reservoir treats them as temporally unfolding symbolic events. Each layer performs an incremental symbolic operation akin to a diagnostic iteration. In this sense, the architecture simulates symbolic cognitive dynamics without needing recurrent loops or continuous-time processing. Example analogies:

- A radiologist may first scan an X-ray for gross abnormalities, then focus on finer details like vessel branching or asymmetries. The reservoir mimics this layered focus and refinement.
- In histopathology, coarse pattern recognition is often followed by texture-specific inspections and edge confirmations. The reservoir reflects such attention-guided symbolic sharpening.

Thus, although the image is static, its symbolic meaning evolves through a structured transformation.

Key Functional Benefits:

1. Contextual Awareness: Later layers integrate broader symbolic patterns and relationships. For example, adjacent activations for ”mass-like density” and ”pleural contact” may combine into a more semantically rich code such as ”tumor extension.”
2. Symbolic Consistency Enforcement: The reservoir gradually enforces compatibility between symbolic neighbors. For instance, if a symbolic pattern suggests a ”lobular shape” in one part but a ”diffuse spread” in another, multiple BD-CeNN layers help resolve this contradiction via symbolic propagation dynamics.
3. Noise Suppression and Feature Reinforcement: Erroneous or weak symbolic activations (e.g., due to visual artifacts or scanner noise) are suppressed as consistent high-confidence patterns stabilize across layers.
4. Diagnostic Focus Emergence: As clinicians zoom in on diagnostically suspicious regions, the reservoir helps the model converge on critical symbolic features, enabling stronger downstream inferences.

Memory-Like Behavior without Backpropagation:

One of the hallmarks of Reservoir Computing is that the core network, the ”reservoir,” is not trained in the traditional sense. Similarly, the BD-CeNN reservoir uses fixed transition rules, with learning occurring only at later stages (e.g., symbolic reasoning or feedback refinement). This design allows:

- Fast deployment and reusability across tasks,
- Better explainability (since rules are known and interpretable),
- Low computational cost (no need for GPU-intensive optimization),
- Compatibility with neuromorphic or logic-in-memory hardware.

Despite this, the system exhibits emergent memory: earlier symbolic decisions influence later ones, supporting temporal integration and symbolic accumulation.

Multi-Pass Symbolic Understanding:

The reservoir performs a multi-pass symbolic understanding through its layered design; each BD-CeNN layer acts as a re-evaluation mechanism. This reflects clinical reality, where decisions are rarely made from a single glance. Instead, doctors iterate through differential diagnoses, refine mental models, and cross-check visual evidence.

The BD-CeNN reservoir is similarly:

- Revisits ambiguous symbolic regions (e.g., low-confidence edges),
- Confirms symbolic predictions via iterative consensus,
- Aggregates micro-patterns into macro-symbolic structures.

In effect, it creates symbolic momentum features that co-exist and evolve toward diagnostic convergence.

Feeding into Higher-Level Reasoning:

By the end of the BD-CeNN reservoir sequence, the symbolic representation is:

- Sharper and more stable,
- Semantically richer and less noisy,
- Structured in a way that logic-based systems can act upon.

This refined symbolic state serves as the input to the ASP rule engine in Stage 3, ensuring that logical rule matching occurs over data that is both perceptually compressed and cognitively enhanced.

Theoretical Significance:

The BD-CeNN reservoir represents a shift from pure neural architectures toward symbolic dynamical systems, combining:

- Discrete logic-based cell behavior,
- Spatial-temporal propagation of symbolic states,
- Non-learned, yet adaptive computation.

It occupies a novel space between:

- Neural fields (continuous, learned, opaque),
- Rule-based expert systems (static, inflexible),
- And interpretable dynamical symbolic systems (adaptive, structured, explainable).

This contributes to the interpretability discourse and opens pathways for AI models that reason, refine, and revise like human experts.

Stage 2 is the heart of the pipeline’s symbolic cognition, introducing structured, iterative, memory-like transformation to symbolic inputs. It prepares the symbolic field for high-confidence reasoning and supports future integration with clinician-defined rules, continuous learning, and symbolic feedback.

3.3 Stage 3: ASP Rule Mining – Symbolic Knowledge Extraction

In this critical stage of the neuro-symbolic architecture, the system transitions from sub-symbolic processing to explicit symbolic reasoning. The symbolic feature maps produced by the BD-CeNN autoencoder and reservoir layers serve as the foundation for constructing logical rules using Answer Set Programming (ASP), a powerful declarative programming paradigm tailored for knowledge representation and non-monotonic reasoning. This stage equips the system to reason like a domain expert while remaining transparent, auditable, and evolvable.

Symbolic Feature Space as a Basis for Logic. After passing through the earlier stages, each medical image is reduced to a structured symbolic representation of labeled, discrete features extracted from the input. These features may include:

- Morphological descriptors: e.g., size, compactness, border sharpness, irregularity.
- Textural properties: e.g., granularity, contrast, smoothness.
- Spatial relationships: e.g., central vs peripheral, multi-region vs focal.
- Contextual metadata: e.g., patient age group, anatomical zone.

These symbolic descriptors abstract away low-level pixel information, forming a semantically meaningful basis for the reasoning engine. Importantly, these are not learned end-to-end solely through backpropagation; they are explicitly interpretable and can be annotated, verified, or supplemented by human experts.

Automated Rule Mining from Annotated Data. The system utilizes supervised training data to induce diagnostic rules automatically. The symbolic features are paired with corresponding diagnosis labels for each annotated image in the training set. A rule mining algorithm built on ASP solvers searches for consistent, minimal, and discriminative rules that connect symbolic patterns to diagnoses.

3.3.1 Formal ASP Rule Mining Specification

We make Stage 3 fully explicit by defining:

(i) The mapping from symbolic features to predicates,

(ii) The rule hypothesis space (search space),

(iii) the objective function for rule selection,

(iv) conflict handling.

Symbol-to-predicate mapping: For each training image n with symbolic code $z^{(n)} \in \{0,1\}^M$ (Eq. (4)), we create ASP facts as in Eqs. (20)–(22):

$$\texttt{img}(n). \tag{20}$$

$$\texttt{active}(n, c_m) \text{ for all } m \text{ with } z_m^{(n)} = 1, \tag{21}$$

$$\texttt{label}(n, y^{(n)}). \tag{22}$$

Thus, every symbolic feature $c_m \in \mathcal{D}$ becomes a logical predicate instance `active(n,` c_m`)` with a precise, reproducible construction (Eqs. (20)–(21)).

Rule language and hypothesis space: We restrict to interpretable, diagnosis-predictive rules of the form in . (23):

$$\texttt{pred}(n, c) \leftarrow a_1(n), \ldots, a_k(n), \texttt{not } b_1(n), \ldots, \texttt{not } b_\ell(n), \tag{23}$$

where each $a_i(n)$ and $b_j(n)$ is an `active(n,` c`)` literal and $c \in \mathcal{Y}$ is a diagnosis label. The *search space* is all rules with body length at most K (i.e., $k + \ell \leq K$), with a user-defined limit $R_{\max}$ on the number of selected rules, providing a concrete, finite hypothesis space for rule mining as derived from Eq. (23).

Consistency, minimality, and discriminativeness: Given a candidate rule set $\mathcal{R}$, we define:

- *Consistency:* $\mathcal{R}$ is consistent on the training set if no image is assigned two different diagnoses, i.e., it is *not* the case that both `pred(n,`c_1`)` and `pred(n,`c_2`)` hold for $c_1 \neq c_2$ in any stable model induced by the facts `active/2` (Eqs. (21)–(22)) and rules $\mathcal{R}$ (Eq. (23)).
- *Minimality (rule-level):* a rule r is body-minimal if removing any literal from its body strictly decreases its score (Eq. (26)) or introduces inconsistency (as defined above from Eq. (23)).
- *Discriminativeness:* a rule should separate its target class from others; we quantify this via precision (confidence) and coverage (Eq. (24)).

Objective function, confidence, and a bound on rule quality: For a rule r predicting class c, let TP_r and FP_r be the number of training images whose bodies are satisfied and whose labels are c (true positives) or not c (false positives). We define *confidence* and *coverage* in Eq. (24):

$$\text{conf}(r) = \frac{\text{TP}_r + \varepsilon}{\text{TP}_r + \text{FP}_r + 2\varepsilon}, \qquad \text{cov}(r) = \frac{\text{TP}_r}{N}, \tag{24}$$

with a small $\varepsilon > 0$ for smoothing. If a rule has support $S_r = \mathrm{TP}_r + \mathrm{FP}_r$, then Hoeffding's inequality yields a simple generalization bound as in Eq. (25):

$$\Pr\Big(\big|\mathrm{conf}_{\mathrm{true}}(r) - \mathrm{conf}(r)\big| > \delta\Big) \;\le\; 2\exp(-2S_r\delta^2), \tag{25}$$

making the confidence notion formally analyzable and explicitly support-dependent via Eqs. (24) and (25). We score each rule by Eq. (26):

$$\mathrm{score}(r) = \alpha\,\mathrm{conf}(r) + \beta\,\mathrm{cov}(r) - \gamma\,\mathrm{len}(r), \tag{26}$$

where $\mathrm{len}(r) = k + \ell$ is body length and $\alpha, \beta, \gamma \ge 0$ control the precision–coverage–simplicity trade-off (Eqs. (24)–(26)). The ASP mining task is posed as a global optimization: select a rule set $\mathcal{R}$ maximizing $\sum_{r\in\mathcal{R}} \mathrm{score}(r)$ subject to consistency, body-length constraints (K), and a rule budget ($R_{\max}$), which is directly grounded in Eqs. (23) and (26). This objective can be encoded in ASP via weak constraints that:

(i) minimize training misclassification,

(ii) minimize total rule length/number of rules,

(iii) maximize summed scores (Eq. (26)).

Learning curve definition (Figure 1): The plotted "average confidence" at iteration t is defined in Eq. (27):

$$\overline{\mathrm{conf}}(t) = \frac{1}{|\mathcal{R}_t|} \sum_{r\in\mathcal{R}_t} \mathrm{conf}(r), \tag{27}$$

where $\mathcal{R}_t$ denotes the rule set mined at iteration t (e.g., after processing an additional batch of labeled examples or after a re-mining pass), and $\mathrm{conf}(r)$ is as in Eq. (24).

Conflict resolution: When multiple rules fire for the same image, we resolve conflicts by:

(i) preferring higher-confidence rules (Eq. (24)),

(ii) if ties remain, preferring shorter rules (Eq. (26)).

This policy is explicit and auditable because every `pred/2` fact can be traced to the winning rule and its confidence, as derived from Eqs. (23)–(26).

Algorithmic outline: A reproducible implementation follows:

```
Algorithm 1 (ASP rule induction, summary):
Input: {(z^(n), y^(n))}_{n=1}^N,
max body length K, max rules Rmax.
1) Emit ASP facts:
     img(n), active(n,c_m), label(n,y^(n)).
```

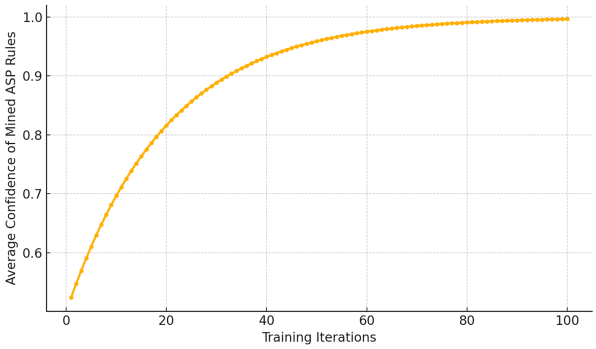


Fig.. 1: ASP Rule Learning Curve – Confidence vs Training Iterations. This figure illustrates how the average confidence of mined ASP (Answer Set Programming) rules improves progressively over 100 training iterations. The curve models an exponential increase, reflecting the system's growing accuracy in symbolic reasoning as it encounters more labeled examples and iteratively refines its symbolic inference engine. For this experiment, a sample of artificial data was generated to serve as a proof of concept. Confidence is defined in Eq. (24); the learning curve reports $\overline{\mathrm{conf}}(t)$ over mined rule sets $\mathcal{R}_t$(see Stage 3). *Source: created by the authors.*

```
2) Generate candidate bodies up to length K
   from dictionary D.
3) Use ASP optimization to select
    a consistent rule set R:
   - minimize training misclassification
   - minimize |R| and total body length
   - maximize sum score(r) (Eq. rule_score).
4) Output:
    mined rules +
    conf(r) (Eq. rule_conf) for auditing.
```

Because this stage is encoded as a solver-level optimization, the selected rules are globally optimal with respect to the stated objective and constraints (subject to solver completeness), enabling theoretical analysis of rule quality under the defined score.
Example rules may include:

```
IF border_irregularity = true
AND density = hyperdense
AND region = upper_lobe
THEN diagnosis = suspicious_nodule
```

Or:

```
IF asymmetry = high
AND lesion_color = dark_brown
AND pattern = radial_streaks
THEN diagnosis = likely_melanoma
```

These rules encapsulate domain-relevant diagnostic logic, grounded in real data and structured

to mirror how clinicians think and make decisions. The resulting rule base is:

- Explicitly encoded (no hidden weights or uninterpretable embeddings),
- Modular and editable, and
- Executable via ASP inference engines such as Clingo.

ASP Rule Execution for Diagnosis.
During inference, the system processes new symbolic feature maps through the ASP engine, which matches known rules against the current symbolic configuration. If one or more rules apply, the engine generates a diagnosis and a traceable justification chain, a sequence of rule firings explaining how and why the diagnosis was reached. Unlike typical neural networks that provide probabilistic outputs, this ASP-based symbolic layer supports:

- Logical entailment (e.g., deduce all consistent outcomes),
- Conflict resolution (e.g., rule prioritization or preferences),
- Explanation generation (e.g., which features triggered which rules), and
- Diagnostic traceability (vital for clinical validation and accountability).

Key Advantages of the Symbolic Layer:

1. Transparency: Each prediction is backed by a clear, human-readable reasoning chain.
2. Modifiability: Domain experts can edit or add new rules without retraining the entire model.
3. Interdisciplinary adaptability: Rules can incorporate non-image data (e.g., lab results, patient history) for richer reasoning.
4. Auditability and compliance: The logic can be audited and verified for safety, trustworthiness, and compliance with ethical AI regulations (e.g., EU AI Act, FDA GMLP).
5. Medical pedagogy: Rules and reasoning traces can serve as educational material, helping students and residents understand expert decision logic.
6. Patient personalization: Rules can be dynamically adjusted to consider individual characteristics (e.g., age, risk group, genetic markers).

Incremental Rule Evolution and Lifelong Learning.
The ASP module supports lifelong learning by allowing the rule base to grow and evolve. New rules can be discovered from updated training data, physician feedback, or post-deployment corrections. Furthermore, incorrect or outdated rules can be pruned or deprioritized without disrupting the system's performance. This positions the symbolic layer as a static classifier and an adaptive reasoning agent capable of integrating human supervision and evolving domain knowledge into its logic framework.
Integration in the Hybrid Pipeline.
As part of the larger neuro-symbolic architecture, the ASP reasoning layer operates on top of symbolic encodings generated by the BD-CeNN reservoir. This modular structure ensures a clear separation between:

- Perception (feature extraction and compression),
- Abstraction (symbolic transformation), and
- Cognition (rule-based reasoning and inference).

This layered approach mirrors human diagnostic processes and supports robust, interpretable, collaborative AI.

In summary, Stage 3 establishes the core of symbolic intelligence within the system. Rule mining, knowledge representation, and auditable inference enable a data-driven and clinically grounded decision-making mode. This represents a decisive shift from conventional deep learning toward hybrid reasoning architectures that can think, explain, and learn like real-world experts.

3.4 Stage 4: Explainable Output + Symbolic Feedback

The culmination of the proposed neuro-symbolic pipeline is the generation of explainable, symbolic outputs that are diagnostically meaningful and fully auditable. This stage translates the compressed symbolic representations refined via BD-CeNN and reservoir layers into actionable conclusions through a chain of logic-based reasoning. At this point, the system becomes a diagnostic tool and a transparent cognitive partner to human clinicians.
Structured Diagnostic Output.
At the core of this stage is the diagnostic prediction, which reflects the system's final clinical assessment of the input image. This may include outcomes such as:

- "Possible early-stage tumor detected in the upper right quadrant,"
- "Normal anatomical pattern with no detected irregularities," or

- ”Presence of hyperpigmented region consistent with dermal inflammation.”

Each diagnosis is derived from the symbolic features encoded in the latent space and decoded via logical inferences using the Answer Set Programming (ASP) engine. Importantly, these predictions are not made in isolation; they are grounded in structured reasoning pathways, rather than the product of opaque neural activation patterns, as in traditional deep learning systems.

Symbolic Reasoning Trace.

Alongside the diagnostic prediction, the system provides a reasoning trace and a transparent, step-by-step account of the rule chain that led to the final decision. For example, the system may display a chain such as:

1. IF lesion_size = large AND border_irregularity = true THEN suspicion_level = high;
2. IF suspicion_level = high AND patient_age $>$ 50 THEN diagnosis = “possible malignant tumor.”

This level of transparency enables clinicians to audit the system’s logic, question it where appropriate, and gain insights into the machine’s interpretation process. The symbolic reasoning trace effectively transforms the system from a black box into a white-box diagnostic collaborator, promoting confidence and usability, especially in sensitive or high-stakes medical scenarios.

Bi-directional Symbolic Feedback Loop.

Crucially, this stage includes an optional symbolic feedback mechanism that allows the output of the ASP-based reasoning engine to influence upstream processing within the BD-CeNN reservoir. This feedback loop enables:

- State reconfiguration of symbolic activations within the reservoir based on rule-derived insights,
- Iterative reasoning, where the symbolic output prompts a re-analysis of ambiguous features, and
- Adaptive refinement, allowing the network to adjust to new symbolic constraints or priorities in real-time.

Such a loop bridges the classical division between perception and cognition. Instead of unidirectional inference, the model operates as a closed-loop system, capable of dynamic reinterpretation and knowledge-guided evolution. This aligns with recent advances in active inference, symbolic constraint satisfaction, and self-regulating neural-symbolic systems.

Clinical and Educational Relevance:

This stage of the neuro-symbolic framework is particularly impactful for Clinical Decision Support Systems (CDSS), where both stringent regulatory standards and foundational principles of medical ethics require that AI tools provide accurate outputs and transparent and justifiable reasoning. Regulations such as the European Union’s AI Act and the U.S. FDA’s Good Machine Learning Practice (GMLP) guidelines increasingly mandate that any clinical AI system must be explainable, auditable, and fair in its diagnostic conclusions. The BD-CeNN framework addresses these demands by embedding interpretability directly into its architecture through symbolic processing. Rather than relying on opaque probability scores or uninterpretable feature maps, the system outputs explicit symbolic representations and logic-based reasoning paths that physicians can scrutinize. This allows clinicians to view the diagnostic outcome and examine the reasoning behind it, verify its clinical appropriateness, and, where necessary, intervene or override it based on contextual expertise. Such capacity for human-in-the-loop oversight is not just a technical feature; it is a prerequisite for building trust, accountability, and real-world clinical utility in AI-assisted diagnostics. From an educational standpoint, the system’s symbolic trace is a powerful pedagogical tool. In contrast to black-box neural networks that provide no insight into their internal operations, the BD-CeNN autoencoder and its symbolic outputs make the AI’s ”thought process” fully visible and traceable. Medical students and junior clinicians can study how specific visual features such as lesion shape, color variation, or edge irregularity map to discrete symbolic indicators and how these, in turn, trigger diagnostic logic rules. This structured transparency helps learners internalize diagnostic reasoning patterns that mirror expert-level thinking. Moreover, it fosters a deeper conceptual understanding of medical image interpretation by explicitly linking visual cues to clinical meaning. The framework’s explainable nature thus supports machine accountability and human learning, making it ideally suited for integration into medical curricula and training programs. The same interpretability that enables clinical validation also transforms the system into an educational scaffold bridging the gap between AI-driven automation and human medical reasoning.

Customizability and Human-in-the-Loop Extension.

One of the defining strengths of this stage lies in its inherent support for dynamic customization and interactive refinement, which are essential for real-world clinical deployment. Unlike conventional deep learning models that require

full retraining to incorporate new knowledge or correct misclassifications, the symbolic architecture allows domain experts to intervene directly in the reasoning layer. Clinicians and medical informaticians can modify existing rules, introduce new ones, or deactivate outdated logic patterns within the ASP engine, tailoring the diagnostic process to reflect evolving medical knowledge, patient-specific nuances, or institutional standards of care. This flexibility is especially valuable when medical guidelines are regularly updated, rare or emerging conditions must be quickly integrated into diagnostic protocols, or when local practice patterns differ from those embedded in generic datasets. It also enables the system to accommodate individualized constraints, such as comorbidities, demographics, or family history, which often influence diagnostic interpretation but are difficult to encode in purely data-driven models. By opening the symbolic layer to expert interaction, the framework invites collaborative intelligence, allowing AI to adapt responsively rather than remain rigid and static. Beyond its technical flexibility, this human-in-the-loop capability establishes a philosophical alignment with clinician-centered design. It ensures that the AI remains a tool that complements medical reasoning rather than replaces it, supporting accountability, transparency, and continuous learning. Notably, the system's architecture also enables bi-directional adaptation, allowing feedback from symbolic reasoning to influence upstream symbolic states in the reservoir, thereby fostering a feedback loop between inference and representation. This interactive feedback mechanism enables the real-time re-evaluation of ambiguous cases, allowing the AI system to learn from both data and human corrections and preferences. In essence, Stage 4 serves as the cognitive core of the entire pipeline, where latent symbolic patterns extracted and refined through earlier BD-CeNN layers are transformed into structured, semantically meaningful diagnostic outputs. By enabling explainable outcomes, traceable logic paths, and adaptive integration of human insight, this stage exemplifies the foundational principles of neuro-symbolic AI: reasoning and interpretability, contextual sensitivity, and sustained human alignment across the life cycle of medical decision-making.

3.5 Reproducibility and Theoretical Remarks

The pipeline is fully determined by:

(i) neighborhood family and radius $(\mathcal{N}, r)$;

(ii) encoder depth L and update steps $\{T_\ell\}$;

(iii) encoder templates and thresholds $\{A^{(\ell)}, B^{(\ell)}, \beta^{(\ell)}, \theta^{(\ell)}\}$ (Eq. (3));

(iv) symbolic dictionary $\mathcal{D}$ and pooling map g (Eq. (5));

(v) reservoir depth T_R and initialization sparsity ρ with bounded influence Γ (Eq. (15));

(vi) ASP mining constraints $(K, R_{\max})$ and objective weights (α, β, γ) (Eq. (26)).

Reporting this tuple makes the method reproducible end-to-end.

Convergence and stability: Both the BD-CeNN encoder (Eq. (3)) and the reservoir (Eq. (13)) are finite-state dynamical systems; therefore, trajectories are eventually periodic. We define a practical notion of convergence as fixed-point reachability within a budgeted number of updates (Hamming stability), which is testable exactly. The bounded-influence condition (Eq. (15)) provides an explicit stability knob limiting local amplification of disagreements and promoting short transient dynamics in deployment.

Computational complexity (stage-wise bounds): Stage 1 (encoder) has cost $\mathcal{O}(\sum_{\ell=1}^{L} T_\ell \, H \, W \, |\mathcal{N}|)$. Stage 2 (reservoir) has cost $\mathcal{O}(T_R \, H \, W \, |\mathcal{N}| \, F^2)$ in the general multi-feature coupling case (or $\mathcal{O}(T_R \, H \, W \, |\mathcal{N}| \, F)$ for per-feature updates). Stage 3 (ASP mining) is NP-hard in general due to combinatorial rule selection, but it is *exactly* defined as an optimization over the finite rule hypothesis space (Eq. (23)); complexity is controlled explicitly by $(K, R_{\max})$ and by the solver's pruning/optimization strategies.

Limitations and failure modes: The discrete dynamics can enter short cycles if templates are poorly chosen; we therefore cap T_ℓ and T_R and optionally stop early when Hamming stability is reached. The symbolic dictionary may be incomplete (missing clinically relevant concepts), limiting rule coverage; this motivates domain-guided dictionary expansion and ablations over $|\mathcal{D}|$. Finally, ASP rule mining can overfit when the dataset is small or when K is too large; Eq. (26) explicitly penalizes rule length, and confidence can be paired with its support-dependent bound (Stage 3) to filter low-support rules.

4 Key Contributions

This work introduces a novel, unified neuro-symbolic framework combining discrete neural computation,

temporal dynamics, and symbolic logic for medical image interpretation. It offers several key innovations that distinguish it from existing AI-based diagnostic systems, particularly in explainability, adaptability, and deployment feasibility. First, we propose a novel BD-CeNN-based autoencoder designed explicitly for the symbolic encoding of medical images, capable of processing both grayscale (e.g., CT, X-ray, MRI) and color domains (e.g., histopathology, dermatology). Unlike conventional autoencoders that rely on continuous-valued latent representations, the BD-CeNN model encodes discrete symbolic states that are directly compatible with logic-based reasoning. This symbolic abstraction facilitates interpretable downstream processing while reducing computational complexity. Second, the framework introduces a reservoir computing–inspired stack of BD-CeNN layers that enables dynamic symbolic reasoning over image content. These layers operate sequentially in a pseudo-temporal fashion, refining and evolving symbolic feature maps as they propagate through the architecture. This design simulates a form of iterative, memory-based processing akin to the diagnostic reasoning used by clinicians, thereby bridging the gap between static neural inference and temporally aware, context-sensitive analysis. Third, the architecture seamlessly integrates perception and cognition, where the BD-CeNN-based feature extraction layers (perception) are tightly coupled with an ASP-based rule engine (cognition). This coupling enables symbolic features to be directly mapped onto formalized diagnostic logic, supporting transparent, auditable, and expert-verifiable decision-making. The process remains traceable from raw input to symbolic output, a rare characteristic in modern AI pipelines. Finally, the framework is optimized for real-world applicability in constrained environments. Its discrete, low-resource BD-CeNN components and symbolic inference engine are well-suited for edge deployment in settings where computational resources, stable connectivity, and technical support may be limited. This includes rural health centers, mobile diagnostic platforms, emergency care units, and regulatory-sensitive environments. The model's interpretable nature aligns it with global trends toward ethical, human-centric healthcare AI, supporting accountability and clinical adoption. Together, these contributions establish a strong foundation for advancing the field of explainable and accessible medical AI, particularly at the intersection of neuro-symbolic reasoning and practical healthcare deployment.

5 Real-World Impact

The proposed system holds substantial promise for real-world implementation across various healthcare domains, ranging from direct clinical diagnostics to medical education and public health support. Its hybrid neuro-symbolic design ensures high diagnostic accuracy and delivers explainable, traceable, and efficient decision-making, making it particularly well-suited to modern demands for transparent AI in medicine. In radiology, the system supports the analysis of grayscale imaging modalities, including X-rays, CT scans, and MRI, by extracting and overlaying symbolic features such as geometric boundaries, intensity gradients, and structural anomalies. These features are mapped to logic-based rules that generate interpretable diagnostic suggestions, which clinicians can validate and trace back to symbolic reasoning paths. This augmented explainability builds trust in automated tools, reduces diagnostic ambiguity, and enhances acceptance among radiologists, especially in complex cases involving subtle abnormalities or multi-region pathologies. In fields such as dermatology and pathology, where diagnostic accuracy depends on color variation, texture, and morphological patterns, the system excels by providing a symbolic representation of color-specific and shape-sensitive features. It enables robust classification of skin lesions (e.g., melanoma detection) or identification of cell types in histopathological slides. This makes it especially useful for early detection of cancers, infections, and autoimmune diseases, where minor visual deviations carry significant diagnostic implications. Additionally, the logic-based interpretability allows pathologists to audit decisions, a feature often lacking in traditional CNN-based systems. The framework is also ideally suited for telemedicine, a growing field that demands reliable, low-latency diagnostic tools. The system can transmit symbolic encodings of medical images rather than full-resolution data in remote or bandwidth-constrained environments such as rural clinics, field hospitals, or mobile diagnostic units. These compressed representations preserve critical diagnostic cues while dramatically reducing data size, making real-time AI-assisted diagnostics feasible even on mobile or embedded devices. The system is an interactive training tool for students and healthcare professionals in medical education. The system cultivates a deeper understanding of medical logic by exposing the symbolic structures underlying each diagnostic output and the rule-based reasoning chain that leads to a conclusion. Trainees see what the AI concludes and why it arrives at that conclusion. This promotes conceptual clarity, supports diagnostic skill-building, and fosters a culture of accountable AI-assisted practice. It also opens possibilities for curriculum integration, particularly in radiology, digital pathology, and

clinical decision-making courses. Moreover, the framework's symbolic foundation supports clinical auditability and compliance with regulatory requirements, such as those proposed in the EU AI Act and the FDA's Good Machine Learning Practice (GMLP) guidelines. The system aligns with emerging standards for human-centered, trustworthy medical AI in this context. In summary, the proposed architecture offers a flexible and future-ready solution to challenges in diagnostic automation, clinical transparency, and medical training. Its real-world utility spans primary care, specialty medicine, education, and telehealth, providing a scalable model for integrating explainable AI into the daily fabric of healthcare delivery.

6 Conclusion and Future Work

We have presented a reservoir-enhanced neuro-symbolic AI architecture that fuses discrete symbolic encoding, dynamic feature propagation, and logic-based reasoning to optimize diagnostic performance and model interpretability jointly. At the heart of the system lies a Binary Discrete Cellular Neural Network (BD-CeNN), which encodes grayscale and color medical images into evolving, compact symbolic representations. These symbolic embeddings are refined through reservoir computing dynamics and interpreted via Answer Set Programming (ASP) rules automatically derived from expert-labeled datasets, thereby enabling logic-grounded, auditable decision-making, [11], [14], [16]. This architecture directly tackles several long-standing limitations in medical AI systems. First, it delivers inherent interpretability by embedding explainability within the model structure, moving beyond the limitations of post-hoc techniques such as Grad-CAM or SHAP, which often lack reliability and clinical alignment, [2], [4], [6]. Second, reservoir-inspired symbolic refinement enables context-aware, temporally structured feature evolution, a feature missing from conventional static CNN-based models, [3], [14], [15]. Third, due to its binary and symbolic design, the BD-CeNN enables energy-efficient inference and is well-suited for deployment on edge devices in resource-constrained clinical environments, [9], [10], [13], [20]. Lastly, the architecture supports a broad spectrum of imaging modalities, including both grayscale modalities such as MRI and CT, and color-rich domains such as histopathology and dermatology, thus overcoming the narrow specialization seen in many current models, [1], [3], [8].

Future Directions

1. Symbolic Feedback into the Reservoir: A promising direction for future research is the development of feedback mechanisms that dynamically influence the state evolution of the BD-CeNN reservoir through ASP-derived symbolic rules. This bi-directional coupling would strengthen semantic alignment between logical inference and symbolic feature dynamics, enabling adaptive and context-sensitive reasoning, [14], [15].
2. Multimodal Data Integration: Another avenue involves extending the framework to incorporate non-visual clinical data, such as structured EHR entries, sensor data, or physician annotations, alongside image inputs. Multimodal fusion has been shown to significantly enhance the contextual richness and diagnostic accuracy of AI systems in healthcare, [3], [7], [28].
3. Clinical Validation and Usability Studies: To ensure real-world utility, we plan to engage in clinician-in-the-loop evaluations, usability testing, and pilot deployments. These efforts align with the growing emphasis on human-centered and explainable AI design in medicine, ensuring systems are interpretable, trustworthy, and practically deployable, [2], [6], [28].

In summary, this work contributes to the development of a new generation of transparent, resource-efficient, and semantically grounded AI systems for medical imaging. By uniting BD-CeNN-based symbolic encoding, reservoir computing, and ASP-based logical inference, it advances the theoretical foundations and practical applicability of explainable neuro-symbolic AI in clinical environments.

References

[1] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017. DOI: 10.1016/j.media.2017.07.005.

[2] Erico Tjoa and Cuntai Guan, "A survey on explainable artificial intelligence (xai): Toward medical xai," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 11, pp. 4793–4813, 2020. DOI: 10.1109/TNNLS.2020.3027314.

[3] Alexander Selvikvåg Lundervold and Arvid Lundervold, "An overview of deep learning in medical imaging focusing on mri," *Zeitschrift fuer medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019. DOI: 10.1016/j.zemedi.2018.11.002.

[4] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang, "Xai—explainable artificial intelligence," *Science robotics*, vol. 4, no. 37, eaay7120, 2019. DOI: 10.1126/scirobotics.aay7120.

[5] Scott M Lundberg and Su-In Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, 2017. DOI: 10.48550/arXiv.1705.07874.

[6] Didier Dubois, Henri Prade, and Steven Schockaert, "Generalized possibilistic logic: Foundations and applications to qualitative reasoning about uncertainty," *Artificial Intelligence*, vol. 252, pp. 139–174, 2017. DOI: 10.1016/j.artint.2017.08.001.

[7] Andreas Holzinger, André Carrington, and Heimo Müller, "Measuring the quality of explanations: The system causability scale (scs) comparing human and machine explanations," *KI-Künstliche Intelligenz*, vol. 34, pp. 193–198, 2020. DOI: 10.1007/s13218-020-00636-z.

[8] Eric J Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature medicine*, vol. 25, pp. 44–56, 2019. DOI: 10.1038/s41591-018-0300-7.

[9] Xubin Wang, Zhiqing Tang, Jianxiong Guo, Tianhui Meng, Chenhao Wang, Tian Wang, and Weijia Jia, "Empowering edge intelligence: A comprehensive survey on on-device ai models," *ACM Computing Surveys*, vol. 57, no. 9, pp. 1–39, 2025, Art. no. 228. DOI: 10.1145/3724420.

[10] Song Han, Huizi Mao, and William J Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding," *arXiv preprint arXiv:1510.00149*, 2015. DOI: 10.48550/arXiv.1510.00149.

[11] Gabriele Manganaro, Paolo Arena, and Luigi Fortuna, *Cellular neural networks: chaos, complexity and VLSI processing*. Springer Science & Business Media, 2012, vol. 1, ISBN: 978-3-642-60044-9. DOI: 10.1007/978-3-642-60044-9.

[12] Mariofanna Milanova and Ulrich Büker, "Object recognition in image sequences with cellular neural networks," *Neurocomputing*, vol. 31, no. 1-4, pp. 125–141, 2000. DOI: 10.1016/S0925-2312(99)00177-0.

[13] Huaqing Li, Xiaofeng Liao, Chuandong Li, Hongyu Huang, and Chaojie Li, "Edge detection of noisy images based on cellular neural networks," *Communications in Nonlinear Science and Numerical Simulation*, vol. 16, no. 9, pp. 3746–3759, 2011. DOI: 10.1016/j.cnsns.2010.12.017.

[14] Herbert Jaeger and Harald Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *science*, vol. 304, no. 5667, pp. 78–80, 2004. DOI: 10.1126/science.1091277.

[15] Mantas Lukoševičius and Herbert Jaeger, "Reservoir computing approaches to recurrent neural network training," *Computer science review*, vol. 3, no. 3, pp. 127–149, 2009. DOI: 10.1016/j.cosrev.2009.03.005.

[16] Michael Gelfond and Vladimir Lifschitz, "The stable semantics for logic programs," in *Proceedings of the 5th International Conference on Logic Programming*, Access Date: 23-06-2025, 1988, pp. 1070–1080. [Online]. Available: https://dblp.org/rec/conf/iclp/GelfondL88.

[17] Gerhard Brewka, Thomas Eiter, and Mirosław Truszczyński, "Answer set programming at a glance," *Communications of the ACM*, vol. 54, no. 12, pp. 92–103, 2011. DOI: 10.1145/2043174.2043195.

[18] Roland Kaminski and Torsten Schaub, "On the foundations of grounding in answer set programming," *Theory and Practice of Logic Programming*, vol. 23, no. 6, pp. 1138–1197, 2023. DOI: 10.1017/S1471068422000308.

[19] Andreas Holzinger, Georg Langs, Helmut Denk, Kurt Zatloukal, and Heimo Müller, "Causability and explainability of artificial intelligence in medicine," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 9, e1312, 2019. DOI: 10.1002/widm.1312.

[20] Vahideh Hayyolalam, Moayad Aloqaily, Öznur Özkasap, and Mohsen Guizani, "Edge intelligence for empowering iot-based healthcare systems," *IEEE Wireless*

Communications, vol. 28, no. 3, pp. 6–14, 2021. DOI: 10.1109/MWC.001.2000345.

[21] Jean Chamberlain Chedjou and Kyandoghere Kyamakya, "A universal concept based on cellular neural networks for ultrafast and flexible solving of differential equations," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 4, pp. 749–762, 2014. DOI: 10.1109/TNNLS.2014.2323218.

[22] Jean Chamberlain Chedjou and K Kyamakya, "Cellular neural networks based local traffic signals control at a junction/intersection," *IFAC Proceedings Volumes*, vol. 45, no. 4, pp. 80–85, 2012. DOI: 10.3182/20120403-3-DE-3010.00059.

[23] Leon O Chua and Lin Yang, "Cellular neural networks: Theory," *IEEE Transactions on circuits and systems*, vol. 35, no. 10, pp. 1257–1272, Oct. 1988. DOI: 10.1109/31.7600.

[24] Leon O Chua and Lin Yang, "Cellular neural networks: Applications," *IEEE Transactions on circuits and systems*, vol. 35, no. 10, pp. 1273–1290, 1988. DOI: 10.1109/31.7601.

[25] Tamas Roska and Leon O Chua, "The cnn universal machine: An analogic array computer," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 40, no. 3, pp. 163–173, 1993. DOI: 10.1109/82.222815.

[26] Yexiao He, Ziyao Wang, Yuning Zhang, Tingting Dan, Tianlong Chen, Guorong Wu, and Ang Li, *Neurosymad: A neuro-symbolic framework for interpretable alzheimer's disease diagnosis*, 2025. arXiv: 2503.00510 [eess.IV].

[27] Tarek R Besold, Sebastian Bader, Howard Bowman, Pedro Domingos, Pascal Hitzler, Kai-Uwe Kühnberger, Luis C Lamb, Priscila Machado Vieira Lima, Leo de Penning, Gadi Pinkas, et al., "Neural-symbolic learning and reasoning: A survey and interpretation 1," in *Neuro-symbolic artificial intelligence: The state of the art*, IOS press, 2021, pp. 1–51, ISBN: 978-1-64368-244-0. DOI: 10.3233/FAIA210348.

[28] Ibomoiye Domor Mienye, George Obaido, Nobert Jere, Ebikella Mienye, Kehinde Aruleba, Ikiomoye Douglas Emmanuel, and Blessing Ogbuokiri, "A survey of explainable artificial intelligence in healthcare: Concepts, applications, and challenges," *Informatics in Medicine Unlocked*, vol. 51, p. 101 587, 2024. DOI: 10.1016/j.imu.2024.101587.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

APPENDIX

Table 1: Comparison of Challenges, Existing Approaches, Unmet Needs, and Our Contributions *Source: created by the authors*.

Challenge	Existing Approaches	Unmet Need	Our Contribution
Interpretability	Post-hoc XAI, self-explainable models, [2], [6], [19]	Lack of intrinsic, auditable interpretability	BD-CeNN + ASP pipeline with built-in symbolic explanations
Symbolic Integration	Detached symbolic layers or none, [17], [27]	Disconnection from learned features	Symbolic encoding and reasoning integrated throughout the pipeline
Temporal Inference	Static feedforward models; RC only in EEG/time series, [14], [15]	No iterative reasoning for static images	BD-CeNN reservoir mimicking temporal diagnostic reasoning
Rule Induction	Manual rule design or fuzzy CNN-based rules, [16], [18], [26]	No scalable, discrete rule mining from interpretable features	Automatic ASP rule generation from symbolic BD-CeNN features
Multimodal Support	Limited to grayscale or narrow domains, [26]	Lack of generalization across imaging types	Full support for grayscale and color imaging
Edge Deployment	Deep CNNs: computationally intensive, [9], [10], [20]	Inaccessible to low-resource or mobile clinical settings	Efficient, interpretable, and low-power model suitable for on-device deployment

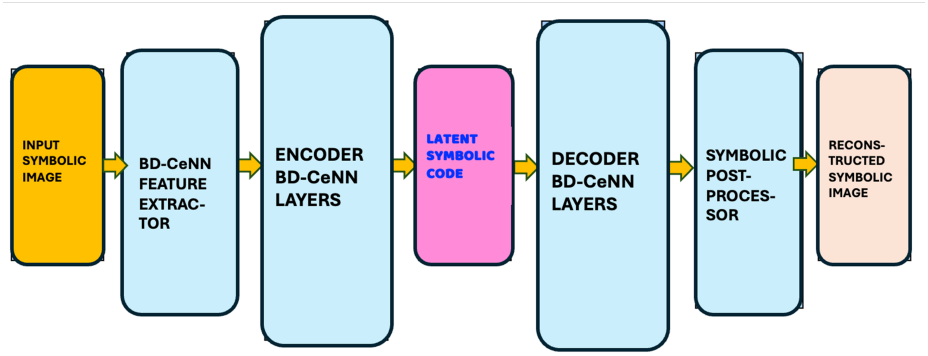
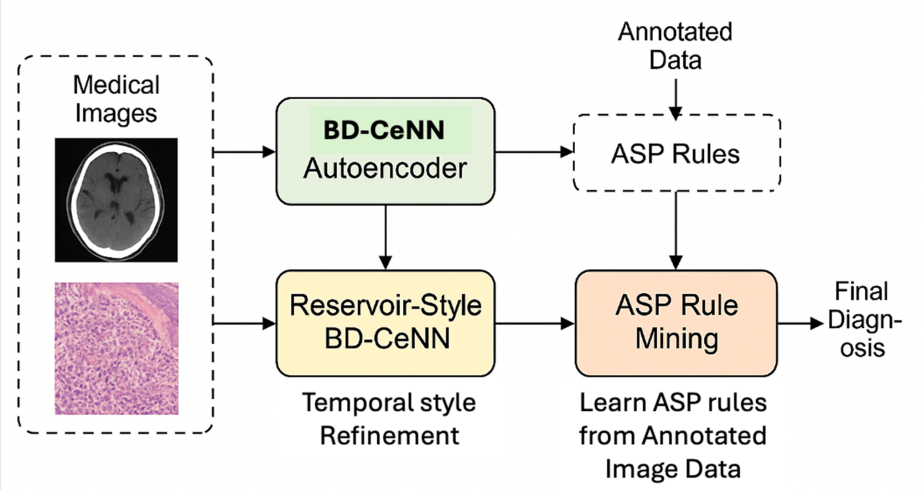
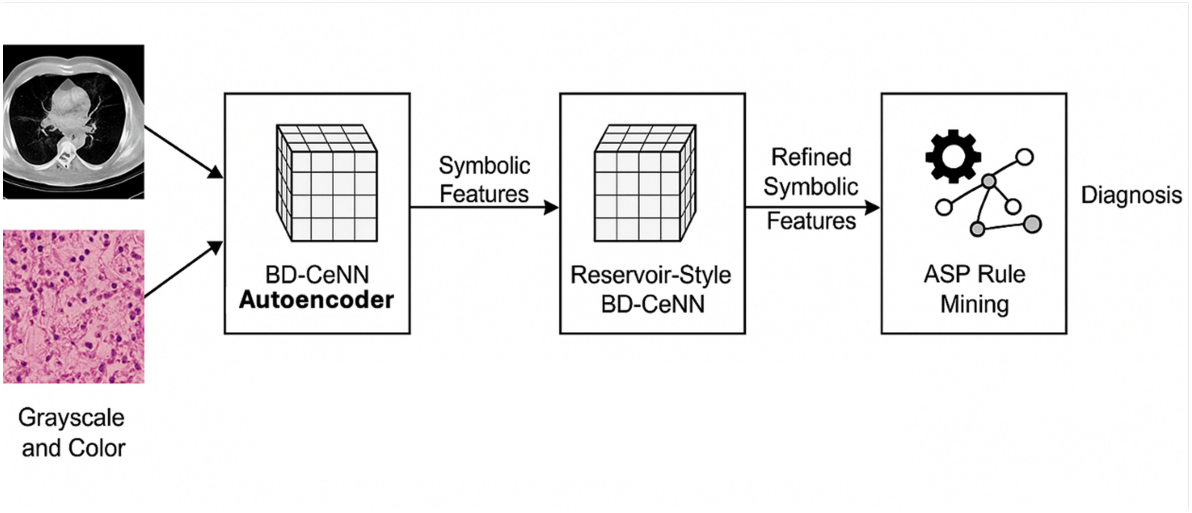


Fig. 1: BD-CeNN Autoencoder – visual representation of encoding/decoding symbolic features. *Source: created by the authors.*



(a) Training Phase: BD-CeNN Autoencoder with ASP Rule Learning



(b) Inference Phase: Full Neuro-Symbolic Pipeline

Fig.: Full Neuro-Symbolic Pipeline – from input image to explainable output. *Source: created by the authors.*