

NONPARAMETRIC UNSUPERVISED DATA CONDENSATION FOR GIGAPIXEL HISTOLOGICAL IMAGES

005 **Anonymous authors**

006 Paper under double-blind review

ABSTRACT

011 Histological whole-slide images (WSIs) are central to computational pathology
 012 but are extremely large, often several gigabytes, making them infeasible for di-
 013 rect use in standard vision pipelines. Prior approaches reduce training cost by
 014 condensing WSIs into a fixed number of representative features (prototypes), but
 015 this approach overlooks the varying complexity and diversity of WSIs, leading
 016 to loss of critical information. To this end, we propose **NICER**, a probabilistic
 017 data condensation framework that decomposes each WSI into feature patterns to
 018 capture heterogeneity and concept prototypes to ensure compactness. By reformu-
 019 lating prototype construction as a nonparametric condensation problem, NICER
 020 adapts the number of prototypes to slide complexity while preserving relevant
 021 information. Experiments on four histological datasets show that NICER outper-
 022 forms prior methods, yielding up to 90% performance gains and superior effi-
 023 ciency trade-offs, setting a new paradigm for histological representation learning.

1 INTRODUCTION

027 Histological whole-slide images (WSIs) are high-resolution digital scans of tissue slides and have
 028 become central to Computational Pathology (CPath) (Song et al., 2024; 2023), enabling tasks such
 029 as classification (Xiang & Zhang, 2023; Shao et al., 2021), segmentation (Graham et al., 2023;
 030 Guo et al., 2023), and survival prediction (Fan et al., 2023). However, their enormous resolution,
 031 often exceeding $100,000 \times 100,000$ pixels (hundreds of gigapixels and several gigabytes per slide),
 032 makes full-slide processing infeasible. For example, even a single WSI cannot fit into a multi-head
 033 self-attention (MHSA) unit due to its quadratic memory complexity.

034 **Challenge.** Multiple Instance Learning (MIL) addresses WSI scale by partitioning each slide into
 035 thousands of patches (e.g., $> 10,000$), embedding them with a pre-trained encoder, and aggregat-
 036 ing them into a slide-level representation (Tang et al., 2023; Nguyen et al., 2025b; Xiang & Zhang,
 037 2023). Since the full processing and storing of patches is costly (Jin et al., 2025; Sacco et al., 2020),
 038 recent studies reveal the strong morphological redundancy of WSIs (Song et al., 2024; Vu et al.,
 039 2023), summarize them into compact, representative prototype sets that are transferable for down-
 040 stream tasks (Song et al., 2024; Jin et al., 2025). However, the high variability of WSIs undermines
 041 the representativeness of the prototypes. For instance, some slides contain large homogeneous re-
 042 gions, while others show highly heterogeneous tumor areas requiring denser sampling (see Fig. 1).

043 **Limitation of Prior Work.** Existing methods (Vu et al., 2023; Claudio Quiros et al., 2024; Song
 044 et al., 2024) operate under the restrictive assumption that a fixed set of prototypes can adequately
 045 represent all slides, regardless of their complexity. While conceptually simple, this assumption
 046 overlooks the wide variability in morphological redundancy and structural complexity across slides
 047 and institutions, leading to either redundant prototypes or information loss (see Fig. 1). Failing to
 048 adapt to this variability forces a trade-off between accuracy and efficiency, with most approaches
 049 sacrificing the former for the latter, as shown in Fig. 2.

050 **Fundamental Gap.** In hindsight, what is missing from existing approaches is a mechanism to
 051 balance aggressive feature reduction with information preservation in an unsupervised manner. Be-
 052 cause feature distributions and complexity levels vary widely across slides, using a prototype set
 053 with fixed capacity may achieve efficiency but risks losing critical information or introducing re-
 054 dundancy. Increasing the prototype set size can mitigate information loss but at the cost of reduced
 055 efficiency for the entire WSI pipeline. This raises a fundamental question: *How can we identify and*

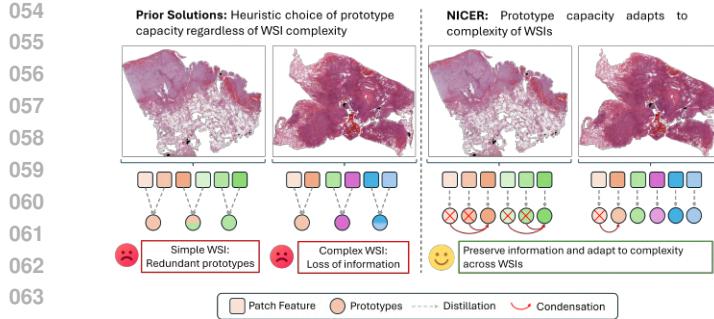


Figure 1: Conceptual illustration highlighting the key distinction between our NICER and prior work. When WSI’s complexity exceeds the prototypes’ capacity, different features (different colors) might collapse into less information prototypes, and simple WSIs might suffer from redundancy.

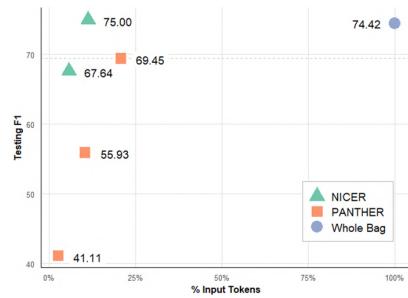


Figure 2: On the same task and model, NICER achieves higher compression rate and better downstream performance (F1) than other baselines. (see Sec. 3).

model the varying complexity of WSIs during condensation? Tackling this challenge hence calls for a new approach that adaptively balances information preservation and efficiency on a per-slide basis, enabling flexible and optimal slide-level representations.

Solution Vision. To address the above question, our solution insight is: effective WSI condensation should not begin with aggressive reduction, but rather with intentional redundancy to preserve rare and heterogeneous signals which vary significantly across WSIs. Redundancy is then adaptively removed to restore efficiency, allowing each slide to determine its own capacity based on complexity. This design avoids early information loss, achieves a principled balance between preservation and efficiency, and provides a flexible foundation for robust, slide-adaptive learning.

Technical Contribution. To realize this vision, we introduce **NICER**, a novel *NonparametricIC unsupervised data condensation framework* that reformulates prototype construction as an unsupervised condensation problem. NICER first learns a high-capacity set of *feature patterns* to preserve diverse and heterogeneous information from each slide redundantly, and then condense them into a compact set of *feature concepts*. Redundant concepts are pruned, and the number of retained concepts adapts automatically to slide complexity, making the process nonparametric and slide-adaptive. The entire procedure is formalized through a generative formulation governed by learnable parameters. Our main contributions are as follows:

1. We cast prototype construction as an unsupervised data condensation task, formulated as a hierarchical optimization problem. Prototypical information is first distilled from the WSI feature bag into a set of patterns, which are then condensed into a compact set of concepts. This design adapts the concept set capacity to the complexity of each WSI, achieving a balance between information preservation and efficiency (see Section 2.3).
2. We develop an algorithm that identifies the most probable associations between patterns and condensed concepts. Framed as a latent variable in our generative model, this association is efficiently inferred in a probabilistic view, enabling the proposed approach to be practical and applicable across diverse real-world medical settings (see Section 2.4).
3. We evaluate the performance of NICER against existing baselines through extensive experiments on cancer subtyping and survival prediction tasks, spanning four benchmark datasets. The results demonstrate that NICER consistently surpasses competing methods across diverse settings, establishing new state-of-the-art performance in unsupervised prototype construction (see Section 3).

2 METHODOLOGY

2.1 PROBLEM FORMULATION AND METHOD OVERVIEW

Unlike prior work (Vu et al., 2023; Song et al., 2024), which prioritizes efficiency over information preservation by fixing prototype capacity across all WSIs, our goal is to balance the two in an un-

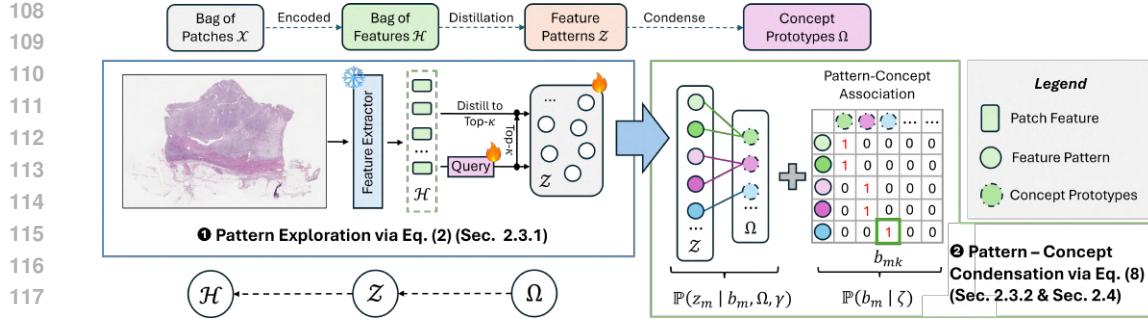


Figure 3: **Overview of NICER.** NICER decouples preservation and efficiency into patterns \mathcal{Z} and concepts Ω , learned via two stages: *pattern exploration* to capture diverse information, and *condensation* to merge redundancies into compact concepts.

supervised and slide-adaptive manner. To this end, we introduce a hierarchical formulation wherein each WSI is represented by a pattern set preserving the diversity of the original feature bag, which is subsequently summarized into more compact concepts as follows:

Problem Formulation. Formally, a WSI is tessellated into non-overlapping patches $\mathcal{X} = \{x_1, \dots, x_N\}$, where each patch $x_i \in \mathbb{R}^{H \times W \times 3}$. A pretrained encoder $f_{\text{enc}}(\cdot)$ maps each patch x_i to a latent embedding $h_i \in \mathbb{R}^d$, yielding a feature bag $\mathcal{H} = \{h_1, \dots, h_N\}$. The objective is to condense \mathcal{H} into a smaller concept (prototype) set $\Omega = \{\omega_k\}_{k=1}^K$ with $K \ll N$, i.e., maximizing $\mathbb{P}(\mathcal{H} \mid \Omega)$ under Bayesian lens. We rewrite this problem under the hierarchical abstraction $\mathcal{H} \leftarrow \mathcal{Z} \leftarrow \Omega$, where patterns \mathcal{Z} approximate features \mathcal{H} and concepts Ω are their underlying generators, leading to maximization of $\mathbb{P}(\mathcal{H}, \mathcal{Z} \mid \Omega)$.

This two-level abstraction provides a principled framework for efficient WSI condensation and explicitly balances accuracy and efficiency, addressing a key limitation in previous researches.

Method Overview. Fig. 3 illustrates the NICER framework, which operates in two iterative stages. In the *pattern exploration*, patterns \mathcal{Z} are learned to capture the diversity of \mathcal{H} through selective interactions between patches and patches (Sec. 2.3). In the *condensation stage*, concepts Ω are introduced to enforce compactness by modeling patterns as samples generated from a smaller concept set (Sec. 2.3.2). During condensation, concepts that do not contribute to generating observed patterns are pruned, imparting NICER with a nonparametric nature and scalability to the complexity of patterns. In summary, patterns ensure information preservation, while concepts enforce efficiency.

2.2 PROBABILISTIC NONPARAMETRIC DATA CONDENSATION

Intuitively, using proposed hierarchical formulation, NICER enriches the standard WSI condensation framework by inserting an intermediate random variable that helps preserve the diversity of features in the original WSI. Instead of collapsing the whole slide image into a single compressed set directly, we first allow each WSI to generate a pattern set of size M . These patterns (\mathcal{Z}) retain the originality of the data (\mathcal{H}) while naturally containing some redundancy. The condensation phase then acts like a sculptor, carefully shaving away overlaps and compressing the patterns into a compact, information-rich concept set Ω .

Data Condensation Model. Given a hierarchical probabilistic model $\mathcal{H} \leftarrow \mathcal{Z} \leftarrow \Omega$, the data condensation model is factorized by:

$$\log \mathbb{P}(\mathcal{H}, \mathcal{Z} \mid \Omega) = \log \mathbb{P}(\mathcal{H} \mid \mathcal{Z}, \Omega) + \log \mathbb{P}(\mathcal{Z} \mid \Omega). \quad (1)$$

Data condensation is then achieved via fitting the parameterization of this probabilistic model to the pre-trained feature observation \mathcal{H} , followed by the most probable concept set $\Omega = \arg \max_{\Omega'} \max_{\mathcal{Z}} \mathbb{P}(\mathcal{H}, \mathcal{Z} \mid \Omega')$ via principled probabilistic inference.

162 2.3 PARAMETERIZATION
163

164 We parameterize the conditional terms in the condensation model (Eq. 1), interpreting $\mathbb{P}(\mathcal{H} \mid \mathcal{Z}, \Omega)$
165 as *pattern exploration*, where diverse patterns (\mathcal{Z} , of size M) are extracted, and $\mathbb{P}(\mathcal{Z} \mid \Omega)$ as *conden-*
166 *sation*, modeling patterns with an unknown concept set Ω . This formulation enables nonparametric
167 condensation and slide-wise adaptability.

168 2.3.1 PATTERN EXPLORATION
169

170 In the pattern exploration stage, we use $\mathbb{N}(h_i \mid \mu_i, \mathbf{I})$ to denote the density of h_i , which is modeled as
171 a Gaussian variable $h_i \sim \mathbb{N}(\mu_i, \mathbf{I})$, with the mean $\mu_i \triangleq \mu_i(\mathcal{Z}; \theta)$ maps the shared pattern set \mathcal{Z} into
172 a representation best aligned with h_i . Conceptually, this is akin to a retrieval process: each feature
173 selects its best-matching pattern, $\mu_i(\mathcal{Z}; \theta) = z_{(i)}^* = \mathcal{Z}[m_i^*]$ where $\mathcal{Z}[m_i^*]$ denotes the selected
174 pattern for h_i from \mathcal{Z} . When h_i and z_m are ℓ_2 -normalized, the log-likelihood reduces to

$$175 \log \mathbb{P}(\mathcal{H} \mid \mathcal{Z}, \theta) = \sum_{i=1}^N \log \mathbb{N}(h_i \mid z_{(i)}^*, \sigma^2 \mathbf{I}) \approx -\frac{1}{2\sigma^2} \sum_{i=1}^N (2 - 2\langle h_i, z_{(i)}^* \rangle). \quad (2)$$

179 **Remark.** Eq. 2 reveals an intuitive principle: maximizing the likelihood of the feature bag \mathcal{H}
180 reduces to aligning features with their closest patterns. Thus, pattern learning naturally emerges as
181 a retrieval-style process, but one firmly anchored in a probabilistic framework.

182 **Design and Learning.** Based on the above probabilistic analysis, we distill prototypical information
183 from the feature bag \mathcal{H} into the pattern set \mathcal{Z} by framing distillation as a retrieval-based selection
184 process. Each feature $h_i \in \mathcal{H}$ associates with pattern $z_m \in \mathcal{Z}$ via cosine similarity,

$$185 \Gamma(h_i, z_m) \triangleq \langle h_i, z_m \rangle. \quad (3)$$

186 Extending the baseline formulation in Eq. 2, we allow each feature to distribute its information
187 across its top- κ most relevant patterns ($\kappa \ll M$), which improves expressiveness while preventing
188 over-dispersion that weakens discriminative power. Maximizing $\Gamma(\cdot, \cdot)$ then drives an adaptive se-
189 lection process, yielding a pattern set \mathcal{Z} that is both compressed and information-preserving, diverse
190 enough to capture the variability of \mathcal{H} , and efficient for the subsequent condensation stage.

191 2.3.2 PATTERN-CONCEPT CONDENSATION

193 In this section, we derive the condensation term in Eq. 1 and present the construction of the concept
194 set Ω . While the pattern set \mathcal{Z} provides a comprehensive view of \mathcal{H} , it remains tied to local feature
195 variations and often carries redundancy. To move beyond this, we introduce Ω as a higher-level
196 *semantic abstraction*, a compact set of concept prototypes that captures only the essential struc-
197 tures of the WSI, discarding spurious noise and redundant information. This step is formalized as
198 a generative process, where each pattern in \mathcal{Z} is modeled as a probabilistic sample from a con-
199 cept prototype in Ω , and the pattern-concept associations are inferred by maximizing the posterior
200 distribution (Sec. 2.4 for details).

201 **Pattern Prior.** In particular, we model this condensation process as a generative model based
202 on a nonparametric point process. In this view, the point process distributed a prior set of latent
203 concepts, and each WSI pattern is considered as a sample from a pattern-generation distribution
204 parameterized with a particular concept. We hence enforce that a pattern $z_m \in \mathcal{Z}$ must be generated
205 by a distribution governed by exactly one concept prototype $\omega_k \in \Omega$ with ω_k indicates the k -th
206 concept. Concretely, each $z_m \in \mathcal{Z}$ is treated as an independent sample from a Gaussian distribution:

$$208 z_m \sim \mathbb{N}(\psi_m(\Omega), \Sigma_m(\Omega, \gamma)) \quad (4)$$

209 where ψ_m is an assignment neural function that determines if the concept prototype ω_k responsible
210 for generating z_m and $\Sigma_m(\cdot, \gamma)$ is neural function estimating corresponding covariance matrix with
211 parameters γ . Since this generative model aims to perform condensation, each pattern z_m must be
212 assigned to exactly one concept prototype $\omega_k \in \Omega$. This prior is implemented explicitly on the as-
213 signment function ψ_m via an introduction of novel assignment variables $b_m \triangleq (b_{m1}, b_{m2}, \dots, b_{mK})$
214 such as $b_{mk} \in \{0, 1\}$, $\forall m = 1 \dots M, k \dots K$ and $\sum_k b_{mk} = 1$. Further diagonalizing the parame-
215 terized covariance matrix, we can rewrite Eq. 4 as follows:

$$215 z_m \mid b_m \sim \mathbb{N}(\psi_m, \text{diag}(\delta(\psi_m; \gamma))), \text{ where: } \psi_m \triangleq b_{m1} \cdot \omega_1 + b_{m2} \cdot \omega_2 + \dots + b_{mK} \cdot \omega_K. \quad (5)$$

216 where $\Sigma_m(\Omega, \gamma)$ is rewritten to $\text{diag}(\delta(\Omega, \gamma))$ with $\delta(\Omega; \gamma)$ as a neural function with parameter γ .
 217 Herel, b_{mk} indicates whether m -th pattern is generated by k -th concept. This reformulation shows
 218 that each pattern associate with one concept in the learnable set Ω and their assignment is governed
 219 by binary variables b_{mk} , $\forall m \in \{1, \dots, M\}$ and $\forall k \in \{1, \dots, K\}$.
 220

221 **Assignment Prior.** Using Eq. 4, Eq. 5 and the definition of our new assignment variable $b \triangleq$
 222 $\{b_m\}_{m=1}^M$, we can now solve Eq. 1 with respect to $\mathbb{P}(\mathcal{Z}, b \mid \Omega)$ instead of $\mathbb{P}(\mathcal{Z} \mid \Omega)$ as follows,
 223

$$224 \log \mathbb{P}(\mathcal{Z}, b \mid \Omega) = \log \left\{ \prod_{m=1}^M \mathbb{P}(z_m, b_m \mid \Omega, \gamma, \zeta) \right\} = \sum_{m=1}^M \log \mathbb{P}(z_m \mid \Omega, \gamma) + \log \mathbb{P}(b_m \mid \zeta), \quad (6)$$

226 where $\mathbb{P}(b_m \mid \zeta)$ imposes an assignment prior governed by parameters ζ . To ensure that every
 227 pattern $z_m \in \mathcal{Z}$ is consistently tied with exactly one concept $\omega_k \in \Omega$, we enforce a categorical
 228 distribution over the assignment variables b_m as,
 229

$$230 \mathbb{P}(b_m \mid \zeta) \triangleq \prod_{k=1}^K \pi_k^{b_{mk}}, \text{ where: } \pi_k \triangleq \frac{\exp(\alpha(\omega_k; \zeta))}{\sum_k \exp(\alpha(\omega_k; \zeta))} \quad (7)$$

231 where $\alpha(\cdot)$ is a deep network parameterized by ζ . This prior serves two complementary purposes.
 232 First, the categorical form enforces that each pattern $z_m \in \mathcal{Z}$ is associated with exactly one concept
 233 prototype, thereby encouraging compactness in the condensation process. Second, by parameterizing
 234 the assignment probabilities through learnable logits, NICER adapts the allocation of patterns to
 235 concepts dynamically, ensuring that assignments reflect the diverse information captured in \mathcal{Z} .
 236

237 2.4 CONDENSATION LEARNING

239 Given the nonparametric pattern-generating story above, our original objective which maximizes the
 240 joint likelihood of \mathcal{H}, \mathcal{Z} given Ω (see Eq. 1) now reduces to the pattern condensation problem as,
 241

$$242 \max_{\Omega, \gamma, \zeta, b} \left\{ \sum_{m=1}^M \log \mathbb{P}(z_m, b_m \mid \gamma, \zeta, \Omega) \right\} = \max_{\Omega, \gamma, \zeta, b} \sum_{m=1}^M \left\{ \log \mathbb{P}(z_m \mid b_m, \gamma, \Omega) + \log \mathbb{P}(b_m \mid \zeta) \right\} \quad (8)$$

244 which is directly computable when using Eq. 5 and Eq. 7 (see Appendix E for more details). Solving
 245 Eq. 8 is however not trivial due to its mixed set of discrete/continuous variables. To sidestep this
 246 intractability, we instead solve Eq. 8 via alternating between (1) optimizing (γ, ζ, Ω) while fixing
 247 b ; and (2) optimizing b given (γ, ζ, Ω) . The first optimization sub-problem is straightforward as it
 248 reduces to derivations from Eq. 5, while the latter is less trivial due to the discrete nature of the
 249 optimizing variables b . Fortunately, we must recall that b_m is exactly one-hot vector. This constraint
 250 is important in the condensation settings because it allows us to recast the non-linear log probability
 251 function to a linear form that can be solved effectively, as shown in Lemma E.1.
 252

253 Given (γ, ζ, Ω) , this observation allows us to derive the linear form of Eq. 5 and Eq. 7 as direct
 254 consequences (see Appendix E for details), which reformulates Eq. 8 as follows:
 255

$$256 b^* = \arg \max_b \left\{ \sum_{m=1}^M \log \mathbb{P}(z_m, b_m \mid \gamma, \zeta, \Omega) \right\} = \arg \max_b \left\{ R_1(b) + R_2(b) \right\}$$

$$257 \text{where: } R_1(b) = \sum_{m=1}^M \sum_{k=1}^K b_{mk} \cdot \log \mathbb{N}(z_m \mid \omega_k; \text{diag}(\delta(\omega_k; \gamma))) \quad (9)$$

$$258 R_2(b) = \sum_{m=1}^M \sum_{k=1}^K b_{mk} \cdot \log \left(\frac{\exp(\alpha(\omega_k; \zeta))}{\sum_k \exp(\alpha(\omega_k; \zeta))} \right)$$

262 which is a weighted linear optimization task. Here, $R_1(b)$ is derived from log likelihood function
 263 of our pattern prior (see Eq. 5) while $R_2(b)$ originates from the log likelihood of our categorical
 264 assignment prior (see Eq. 7) using results of Lemma E.1. Here, we emphasize the many-to-one
 265 nature of the condensation problem, i.e., many patterns can be assigned to a single concept, by
 266 iteratively optimizing each b_m while holding the remaining assignments b_{-m} fixed. This reduces
 267 Eq. 9 to a maxima search problem that can be solved with linear complexity $\mathcal{O}(M)$. While solving
 268 Eq. 9, concepts that do not contribute to generating any observed patterns are treated as redundant
 269 and removed. This pruning mechanism gives NICER its nonparametric nature, enabling it to adapt
 to the varying complexity levels of WSIs. Pseudocode for NICER can be found in Appendix B

270 Table 1: Performance of baselines on Condensation Ability tasks. The best and second-best results
 271 are highlighted in **bold red**, and **blue**, respectively.

273 Method	274 Decoder	275 Cancer Subtyping						276 Survival Prediction		
		277 PANDA			278 NSCLC			279 LUAD	280 BRCA	281
281	282	Kappa	283 Accuracy	284 F1	Kappa	285 Bal. Acc.	286 F1	287 C-Index	288 C-Index	289
Whole Bag	ABMIL	91.93 \pm 0.48	76.21 \pm 1.53	76.37 \pm 1.38	90.31 \pm 1.65	94.52 \pm 1.19	95.19 \pm 0.78	62.12 \pm 1.27	78.52 \pm 3.82	290
DeepSets		57.26 \pm 38.13	51.60 \pm 17.98	46.42 \pm 23.91	79.51 \pm 1.80	89.82 \pm 0.88	89.73 \pm 0.91	59.89 \pm 5.34	49.23 \pm 3.59	291
ProtoCount		0.83 \pm 9.55	24.24 \pm 1.21	11.77 \pm 1.28	10.69 \pm 3.54	55.34 \pm 1.78	53.86 \pm 3.15	51.91 \pm 5.75	56.47 \pm 12.03	292
H2T		75.03 \pm 1.08	53.91 \pm 1.10	50.66 \pm 1.01	79.45 \pm 1.80	89.67 \pm 0.88	89.72 \pm 0.90	51.83 \pm 2.59	45.86 \pm 4.29	293
OT		41.92 \pm 11.16	35.50 \pm 3.22	29.98 \pm 4.49	80.79 \pm 4.69	90.39 \pm 2.36	90.37 \pm 2.37	54.07 \pm 3.71	66.61 \pm 6.30	294
InfiniteGPFA		0.00 \pm 0.00	26.85 \pm 0.00	11.36 \pm 0.00	6.61 \pm 2.49	53.29 \pm 1.25	50.61 \pm 3.93	50.48 \pm 4.35	46.65 \pm 10.52	295
PANTHER		65.52 \pm 12.62	43.77 \pm 0.76	42.52 \pm 1.32	83.98 \pm 2.39	92.02 \pm 1.18	92.01 \pm 1.21	58.92 \pm 2.21	72.07 \pm 6.17	296
NICER (Ours)		90.98 \pm 0.97	73.52 \pm 1.27	73.22 \pm 1.55	89.73 \pm 1.81	94.84 \pm 0.92	94.87 \pm 0.91	64.67 \pm 2.11	75.69 \pm 2.02	297
Whole Bag	DSMIL	90.93 \pm 0.43	74.05 \pm 1.38	74.11 \pm 1.31	84.60 \pm 4.16	92.27 \pm 2.08	92.30 \pm 2.08	63.32 \pm 4.02	67.96 \pm 7.12	298
DeepSets		85.53 \pm 3.36	62.26 \pm 1.77	62.88 \pm 1.73	25.30 \pm 36.76	62.68 \pm 18.46	59.10 \pm 20.77	56.86 \pm 6.41	48.47 \pm 15.91	299
ProtoCount		14.78 \pm 8.52	22.52 \pm 2.11	14.47 \pm 0.53	-0.90 \pm 13.16	49.52 \pm 6.63	44.24 \pm 10.37	52.71 \pm 4.47	41.13 \pm 8.69	300
H2T		60.56 \pm 1.57	41.83 \pm 0.84	34.73 \pm 5.23	44.90 \pm 3.95	72.61 \pm 2.03	70.10 \pm 2.64	50.00 \pm 0.00	49.82 \pm 0.26	301
OT		55.06 \pm 7.77	37.66 \pm 3.74	36.60 \pm 3.94	48.26 \pm 8.11	74.20 \pm 4.00	73.52 \pm 4.49	58.17 \pm 4.11	70.84 \pm 0.23	302
InfiniteGPFA		17.72 \pm 9.52	24.76 \pm 2.19	20.56 \pm 2.03	-2.70 \pm 12.64	48.62 \pm 6.31	46.52 \pm 7.29	46.25 \pm 9.89	46.16 \pm 6.69	303
PANTHER		71.74 \pm 2.46	46.09 \pm 3.02	46.32 \pm 3.28	75.05 \pm 8.12	87.60 \pm 4.04	87.46 \pm 4.13	54.15 \pm 4.88	69.96 \pm 2.49	304
NICER (Ours)		87.81 \pm 0.69	66.07 \pm 1.73	65.19 \pm 1.65	84.61 \pm 1.56	92.30 \pm 0.74	91.34 \pm 0.79	62.12 \pm 1.41	74.53 \pm 1.73	305
Whole Bag	ILRA	93.46 \pm 0.24	78.90 \pm 1.76	79.26 \pm 2.05	87.17 \pm 0.90	93.57 \pm 0.41	93.58 \pm 0.45	67.22 \pm 2.12	81.46 \pm 1.46	306
DeepSets		75.49 \pm 2.64	58.31 \pm 1.12	54.52 \pm 2.01	28.20 \pm 39.89	64.10 \pm 19.94	53.03 \pm 27.79	56.72 \pm 5.95	48.47 \pm 15.91	307
ProtoCount		8.65 \pm 2.82	25.88 \pm 0.83	14.98 \pm 2.48	13.26 \pm 8.39	56.63 \pm 4.19	51.57 \pm 10.17	55.02 \pm 12.35	58.83 \pm 6.48	308
H2T		35.20 \pm 17.79	35.05 \pm 5.32	30.69 \pm 6.22	77.50 \pm 3.31	88.65 \pm 1.69	88.69 \pm 1.71	51.28 \pm 1.81	49.82 \pm 0.26	309
OT		39.79 \pm 7.28	37.73 \pm 3.27	31.92 \pm 1.93	83.98 \pm 0.89	92.02 \pm 0.41	91.98 \pm 0.46	63.16 \pm 3.32	62.86 \pm 7.46	310
InfiniteGPFA		0.62 \pm 2.38	27.00 \pm 0.56	13.28 \pm 2.61	4.50 \pm 5.49	52.26 \pm 2.77	45.92 \pm 8.34	51.67 \pm 5.47	57.03 \pm 8.21	311
PANTHER		69.08 \pm 3.63	48.62 \pm 5.28	47.91 \pm 5.45	81.41 \pm 2.38	90.72 \pm 1.14	90.69 \pm 1.21	63.00 \pm 9.98	69.18 \pm 1.35	312
NICER (Ours)		88.92 \pm 2.37	69.73 \pm 3.02	69.56 \pm 3.59	88.47 \pm 1.56	94.28 \pm 0.76	94.23 \pm 0.79	68.06 \pm 4.69	76.99 \pm 3.95	313

3 EXPERIMENTS

This section empirically evaluates NICER on four datasets across two key pathology tasks: cancer subtyping and survival prediction. Dataset details and baselines are provided in Sec. 3.1, while experimental results are reported in Sec. 3.2.

3.1 EXPERIMENT SETTINGS

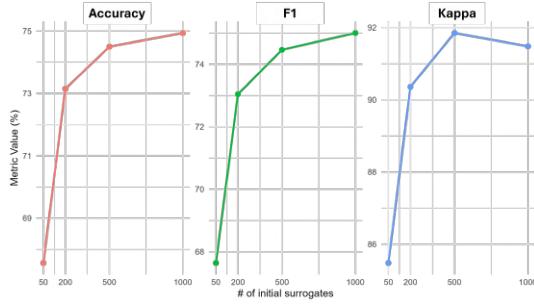
Datasets and Evaluation Metrics. For cancer subtyping, we evaluate NICER on two different tasks: NSCLC subtyping on TCGA (2 classes), and ISUP grading based on PANDA challenge (6 classes) (Bulten et al., 2022; 2020). In survival prediction, we evaluate NICER on TCGA across two cancer repositories: BRCA and LUAD. Following prior work (Song et al., 2024), we evaluate the cancer subtyping tasks using Cohen’s Kappa (Vieira et al., 2010), accuracy, and weighted F1, replacing accuracy with balanced accuracy for NSCLC due to class imbalance. For survival tasks, we report the concordance index (C-Index) (Alabdallah et al., 2024).

Evaluation Settings. We view unsupervised prototype construction as both a condensation framework and a representation learning approach, and evaluate along two axes: (i) condensation ability, by applying NICER and baselines on the training set and testing against the original WSI feature bag; and (ii) morphological prototyping, by applying methods to both training and test sets.

Baselines. We consider two baseline categories: 1) *unsupervised prototyping* methods, which learn unsupervised representations followed by a task-specific neural predictor, and 2) *MIL-based predictors*, which construct supervised slide-level prototypical representations. For *unsupervised prototyping*, we compare NICER with **DeepSets** (Zaheer et al., 2017), **ProtoCounts** (Claudio Quiros et al., 2024), **H2T** (Yu et al., 2023), **InfiniteGPFA** (Yu et al., 2025), **Optimal Transport (OT)** (Mialon et al., 2021), and **PANTHER** (Song et al., 2024). Specifically, DeepSets, ProtoCounts, and H2T build prototypes from histological information and distance-based clustering (e.g., K-Means); InfiniteGPFA is adapted to perform latent factor analysis on each WSI; and OT and PANTHER adopt Gaussian Mixture Models for soft prototypical assignment. For *MIL-based predictors*, we adopt three supervised baselines: attention-based MIL (ABMIL) (Ilse et al., 2018), dual-stream MIL (DSMIL) (Li et al., 2021a), and low-rank MIL (ILRA) (Xiang & Zhang, 2023). These models are trained on unsupervised prototypes and evaluated on original feature bags to measure condensation ability, or used directly to assess NICER’s effectiveness in producing slide-level unsupervised representations. Further implementation details are provided in Appendix C.

324
 325
 326
 327
 328
 Table 2: Performance of baselines on Morphological
 Prototyping tasks. ”TrL”/”TrM” is a linear/nonlinear
 transformer. The best and second-best results are high-
 lighted in **bold red**, and **blue**, respectively.

329	Method	Cancer Subtyping			Survival Prediction	
		PANDA		NSCLC		LUAD
		Accur.	F1	Bal. Acc.	F1	BRCA
330	MIL	ABMIL	74.05	74.42	94.19	94.23
		DSMIL	72.48	72.52	95.17	95.19
		ILRA	76.96	77.11	93.21	93.26
	331 Unsupervised	DeepSets	61.52	60.51	86.50	86.53
		ProtoCount	27.52	20.84	55.55	47.03
		H2T	55.93	53.81	77.75	77.75
		OT	73.15	72.87	88.42	88.45
		InfiniteGPFA	14.09	3.48	50.00	34.41
		PANTHER	70.47	69.98	82.69	82.68
		PANTHER _{TrM}	70.02	70.06	88.53	89.45
		NICER _{TrL}	72.48	72.65	94.27	94.23
		NICER _{TrM}	76.96	77.12	95.17	95.19
						70.57
						76.98
						81.40



340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 Figure 4: Impact of initial patterns number on
 350
 NICER’s overall performance.
 351

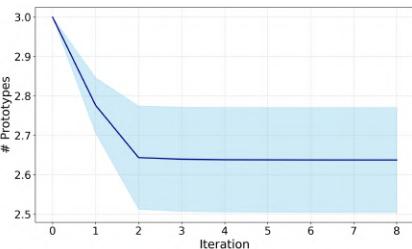
3.2 PERFORMANCE ON CONDENSATION ABILITY

354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
Results on cancer subtyping. Tab. 1 highlights that NICER consistently has superior performance across different predictors for cancer subtyping. On the PANDA dataset, when paired with complex predictors such as DSMIL and ILRA, NICER achieves improvements of 2.28–15.04% across evaluation metrics, with a particularly notable gain of 15.04% in F1 score over DeepSets, when using DSMIL. This advantage becomes even more pronounced with simpler architectures such as ABMIL, where the performance margin widens to 33.72% relative to the next-best baseline. Importantly, NICER sustains strong performance with only a minor drop from Whole Bag ($\sim 2.87\%$ in Kappa), offering a favorable trade-off between accuracy and efficiency. Consistent patterns are observed on the NSCLC dataset, where NICER surpasses competing methods by up to 5.27% in balanced accuracy. This is due to NICER’s nonparametric design, which condenses WSIs based on their complexity, yielding robust representations across architectures.

365
 366
 367
 368
 369
 370
 371
 372
 373
Results on survival prediction. In survival prediction tasks, NICER consistently outperforms other baselines across diverse predictor architectures, achieving performance gains of up to 79.73%. On BRCA, NICER improves the C-Index by as much as 33.40% over baselines, with an average margin of 4.98% compared to the closest competitors. On LUAD, similar improvements are observed, with margins of 2.82% and 9.56% when combined with ABMIL and DSMIL predictors, respectively. Remarkably, when paired with the more complex ILRA predictor, NICER, along with methods such as PANTHER and ProtoCount, can surpass the Whole Bag upper bound, indicating that unsupervised feature construction can denoise raw feature bags and enhance downstream predictive performance. These results highlight both the robust effectiveness of NICER and its model-agnostic generality, underscoring its potential as a broadly applicable framework for histological analysis.

3.3 PERFORMANCE ON MORPHOLOGICAL PROTOTYPING

374
 375
 376
 377
 As shown in Tab. 2, NICER combined with a TrM consistently outperforms other unsupervised methods across all scenarios, demonstrating its ability to preserve semantic information from the



378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533
 534
 535
 536
 537
 538
 539
 540
 541
 542
 543
 544
 545
 546
 547
 548
 549
 550
 551
 552
 553
 554
 555
 556
 557
 558
 559
 560
 561
 562
 563
 564
 565
 566
 567
 568
 569
 570
 571
 572
 573
 574
 575
 576
 577
 578
 579
 580
 581
 582
 583
 584
 585
 586
 587
 588
 589
 590
 591
 592
 593
 594
 595
 596
 597
 598
 599
 600
 601
 602
 603
 604
 605
 606
 607
 608
 609
 610
 611
 612
 613
 614
 615
 616
 617
 618
 619
 620
 621
 622
 623
 624
 625
 626
 627
 628
 629
 630
 631
 632
 633
 634
 635
 636
 637
 638
 639
 640
 641
 642
 643
 644
 645
 646
 647
 648
 649
 650
 651
 652
 653
 654
 655
 656
 657
 658
 659
 660
 661
 662
 663
 664
 665
 666
 667
 668
 669
 670
 671
 672
 673
 674
 675
 676
 677
 678
 679
 680
 681
 682
 683
 684
 685
 686
 687
 688
 689
 690
 691
 692
 693
 694
 695
 696
 697
 698
 699
 700
 701
 702
 703
 704
 705
 706
 707
 708
 709
 710
 711
 712
 713
 714
 715
 716
 717
 718
 719
 720
 721
 722
 723
 724
 725
 726
 727
 728
 729
 730
 731
 732
 733
 734
 735
 736
 737
 738
 739
 740
 741
 742
 743
 744
 745
 746
 747
 748
 749
 750
 751
 752
 753
 754
 755
 756
 757
 758
 759
 760
 761
 762
 763
 764
 765
 766
 767
 768
 769
 770
 771
 772
 773
 774
 775
 776
 777
 778
 779
 780
 781
 782
 783
 784
 785
 786
 787
 788
 789
 790
 791
 792
 793
 794
 795
 796
 797
 798
 799
 800
 801
 802
 803
 804
 805
 806
 807
 808
 809
 810
 811
 812
 813
 814
 815
 816
 817
 818
 819
 820
 821
 822
 823
 824
 825
 826
 827
 828
 829
 830
 831
 832
 833
 834
 835
 836
 837
 838
 839
 840
 841
 842
 843
 844
 845
 846
 847
 848
 849
 850
 851
 852
 853
 854
 855
 856
 857
 858
 859
 860
 861
 862
 863
 864
 865
 866
 867
 868
 869
 870
 871
 872
 873
 874
 875
 876
 877
 878
 879
 880
 881
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899
 900
 901
 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917
 918
 919
 920
 921
 922
 923
 924
 925
 926
 927
 928
 929
 930
 931
 932
 933
 934
 935
 936
 937
 938
 939
 940
 941
 942
 943
 944
 945
 946
 947
 948
 949
 950
 951
 952
 953
 954
 955
 956
 957
 958
 959
 960
 961
 962
 963
 964
 965
 966
 967
 968
 969
 970
 971
 972
 973
 974
 975
 976
 977
 978
 979
 980
 981
 982
 983
 984
 985
 986
 987
 988
 989
 990
 991
 992
 993
 994
 995
 996
 997
 998
 999
 1000
 1001
 1002
 1003
 1004
 1005
 1006
 1007
 1008
 1009
 1010
 1011
 1012
 1013
 1014
 1015
 1016
 1017
 1018
 1019
 1020
 1021
 1022
 1023
 1024
 1025
 1026
 1027
 1028
 1029
 1030
 1031
 1032
 1033
 1034
 1035
 1036
 1037
 1038
 1039
 1040
 1041
 1042
 1043
 1044
 1045
 1046
 1047
 1048
 1049
 1050
 1051
 1052
 1053
 1054
 1055
 1056
 1057
 1058
 1059
 1060
 1061
 1062
 1063
 1064
 1065
 1066
 1067
 1068
 1069
 1070
 1071
 1072
 1073
 1074
 1075
 1076
 1077
 1078
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088
 1089
 1090
 1091
 1092
 1093
 1094
 1095
 1096
 1097
 1098
 1099
 1100
 1101
 1102
 1103
 1104
 1105
 1106
 1107
 1108
 1109
 1110
 1111
 1112
 1113
 1114
 1115
 1116
 1117
 1118
 1119
 1120
 1121
 1122
 1123
 1124
 1125
 1126
 1127
 1128
 1129
 1130
 1131
 1132
 1133
 1134
 1135
 1136
 1137
 1138
 1139
 1140
 1141
 1142
 1143
 1144
 1145
 1146
 1147
 1148
 1149
 1150
 1151
 1152
 1153
 1154
 1155
 1156
 1157
 1158
 1159
 1160
 1161
 1162
 1163
 1164
 1165
 1166
 1167
 1168
 1169
 1170
 1171
 1172
 1173
 1174
 1175
 1176
 1177
 1178
 1179
 1180
 1181
 1182
 1183
 1184
 1185
 1186
 1187
 1188
 1189
 1190
 1191
 1192
 1193
 1194
 1195
 1196
 1197
 1198
 1199
 1200
 1201
 1202
 1203
 1204
 1205
 1206
 1207
 1208
 1209
 1210
 1211
 1212
 1213
 1214
 1215
 1216
 1217
 1218
 1219
 1220
 1221
 1222
 1223
 1224
 1225
 1226
 1227
 1228
 1229
 1230
 1231
 1232
 1233
 1234
 1235
 1236
 1237
 1238
 1239
 1240
 1241
 1242
 1243
 1244
 1245
 1246
 1247
 1248
 1249
 1250
 1251
 1252
 1253
 1254
 1255
 1256
 1257
 1258
 1259
 1260
 1261
 1262
 1263
 1264
 1265
 1266
 1267
 1268
 1269
 1270
 1271
 1272
 1273
 1274
 1275
 1276
 1277
 1278
 1279
 1280
 1281
 1282
 1283
 1284
 1285
 1286
 1287
 1288
 1289
 1290
 1291
 1292
 1293
 1294
 1295
 1296
 1297
 1298
 1299
 1300
 1301
 1302
 1303
 1304
 1305
 1306
 1307
 1308
 1309
 1310
 1311
 1312
 1313
 1314
 1315
 1316
 1317
 1318
 1319
 1320
 1321
 1322
 1323
 1324
 1325
 1326
 1327
 1328
 1329
 1330
 1331
 1332
 1333
 1334
 1335
 1336
 1337
 1338
 1339
 1340
 1341
 1342
 1343
 1344
 1345
 1346
 1347
 1348
 1349
 1350
 1351
 1352
 1353
 1354
 1355
 1356
 1357
 1358
 1359
 1360
 1361
 1362
 1363
 1364
 1365
 1366
 1367
 1368
 1369
 1370
 1371
 1372
 1373
 1374
 1375
 1376
 1377
 1378
 1379
 1380
 1381
 1382
 1383
 1384
 1385
 1386
 1387
 1388
 1389
 1390
 1391
 1392
 1393
 1394
 1395
 1396
 1397
 1398
 1399
 1400
 1401
 1402
 1403
 1404
 1405
 1406
 1407
 1408
 1409
 1410
 1411
 1412
 1413
 1414
 1415
 1416
 1417
 1418
 1419
 1420
 1421
 1422
 1423
 1424
 1425
 1426
 1427
 1428
 1429
 1430
 1431
 1432
 1433
 1434
 1435
 1436
 1437
 1438
 1439
 1440
 1441
 14

378
 379 Table 3: Performance-Efficiency trade-
 380 offs comparison of NICER and the SOTA
 381 method, on PANDA using ABMIL.

Method	Config	% Input	F1
Whole Bag (Upperbound)	-	100%	74.42
PANTHER	K=16	2.58%	41.11
	K=64	10.39%	55.93
	K=128	20.78%	69.45
NICER (Ours)	M=50	5.79%	67.64
	M=200	11.32%	74.46
	M=500	11.56%	74.50
	M=1000	11.31%	75.00

Table 4: Impact of top- κ in pattern distillation on NICER’s performance. Conducted on PANDA with ABMIL predictor

Task	κ	Kappa	Acc.	F1
Condensation	1	90.18	70.02	69.87
	3	91.48	74.94	75.00
	5	90.47	72.93	72.72
	10	89.52	70.47	70.49
Prototyping	1	92.89	75.17	75.47
	3	93.07	77.85	77.96
	5	92.77	75.62	75.80
	10	92.89	73.83	73.82

393 original noisy feature bag. Due to its nonparametric learning process and decomposition of di-
 394 versity and compactness, NICER constructs representations that achieve competitive performance,
 395 exhibit only minimal drops across diverse tasks and datasets, underscoring its model-agnostic na-
 396 ture. In contrast, the performance of conventional MIL methods is strongly dependent on the chosen
 397 architecture.

3.4 ABLATION STUDIES

400 **Impact of Initial Number of Patterns.** Fig. 4 illustrates the effect of the number of patterns ini-
 401 tialized at the start of the NICER algorithm on feature construction performance. Experiments are
 402 conducted on the PANDA dataset and evaluated on the Condensation Ability task using ABMIL as
 403 the predictor. As shown, NICER achieves substantial performance gains as the number of patterns
 404 increases, but the improvement begins to saturate beyond a certain point (e.g., ~ 200). This trend
 405 arises because a larger pattern set provides greater distillation capacity, while the plateau reflects the
 406 condensation process, which removes redundancy and converges toward stable representations.

407 **Performance-Efficiency Trade-offs Comparison.** To assess the effectiveness of NICER in the
 408 condensation problem, we compare its performance-efficiency trade-off against PANTHER on the
 409 cancer subtyping task using the PANDA dataset across varying condensation levels. The condensa-
 410 tion level is controlled by the number of prototypes K in PANTHER and the number of initial
 411 patterns M in NICER, where M serves as an upper bound on the number of final concepts. As
 412 reported in Tab. 3, both methods exhibit an upward trend in F1 as the % Input increases. Notably,
 413 at comparable condensation levels (e.g., PANTHER with 10.39% Input vs. NICER with 11.32% In-
 414 put), NICER surpasses PANTHER by a substantial margin of nearly 19%, underscoring its superior
 415 ability to preserve relevant information during condensation.

416 **Effectiveness of Condensation Stage.** Tab. 3 shows how M influences the capacity of the con-
 417 densed prototype set. With small M (e.g., 50), the final set is limited by the initial pattern pool. As
 418 M grows (e.g., beyond 200), the number of prototypes increases but quickly stabilizes, as seen in the
 419 (% Input) column. Downstream performance follows the same trend, with F1 scores plateauing once
 420 stabilization occurs. This confirms that NICER’s condensation stage effectively merges redundant
 421 patterns into a compact yet informative concept set.

422 **Impact of Top- κ in Pattern Learning.** We perform a sensitivity analysis on the number of patterns
 423 selected per patch feature (top- κ) across both Condensation Ability and Morphological Prototyp-
 424 ing tasks. As shown in Tab. 4, performance steadily improves across all quantitative metrics as κ
 425 increases, but begins to plateau or even slightly decline beyond a certain point (e.g., $\kappa = 3$). This oc-
 426 curs because excessively large κ values cause each patch’s information to be distributed too broadly
 427 across patterns, thereby reducing the effectiveness of the distillation process. These findings are
 428 consistent with our earlier observation in Fig. 4 and align with the design discussed in Sec. 2.3.

429 **Concept Prototypes Diversity.** To analyze the behavior of our condensation process, which maps
 430 patterns into concepts, we visualize 2D t-SNE embeddings of the concept set on the NSCLC dataset
 431 over the first five iterations. As shown in Fig. 6, the learned concepts become increasingly dispersed
 (brighter points) as training progresses, reflecting convergence toward diverse and specific infor-

432 mation. This demonstrates the effectiveness of the condensation process in reducing overlap and
 433 redundancy while preserving diversity, a key strength of NICER. More experimental results can be
 434 found in Appendix F

436 437 4 RELATED WORK

438
 439 **Multiple Instance Learning.** While initial histology-based outcome prediction was centered on
 440 pathologist-annotated region-of-interests (Bychkov et al., 2018; Kather et al., 2019; Mobadersany
 441 et al., 2018), later works have utilized WSIs for clinical prediction tasks with MIL (Campanella
 442 et al., 2019; Chen et al., 2022; Tang et al., 2023; Nguyen et al., 2025b;a). There is a sustained effort
 443 for new MIL schemes, with a focus on developing new patch aggregation strategies to learn more
 444 representative and task-specific embedding, towards better predictive accuracy (Li et al., 2021b; Lu
 445 et al., 2023; Shao et al., 2021; Tang et al., 2023; Xiang & Zhang, 2023) or interpretability (Javed
 446 et al., 2022; Thandiackal et al., 2022). Recent MIL proposals further enhance efficiency during
 447 training and inference by adopting low-rank property of histological images (Xiang & Zhang, 2023)
 448 or using sparse coding model as a regularization in an attention-based aggregator (Qiu et al., 2023).
 449 NICER is similar to MIL in that the patch features in each WSI (represented as a bag) is aggregated
 450 and condensed in different ways to produce a slide-level embedding. Nevertheless, NICER performs
 451 in an unsupervised manner, in contrast to supervised MIL approaches.

452 **Prototype Learning.** Prototypes, representative examples summarizing datasets, have been widely
 453 used in bioinformatics and NLP (dan Guo et al., 2022; Kim, 2022; Lee et al., 2019; Mialon et al.,
 454 2021), appear under related notions such as signatures (Lazebnik et al., 2005; Zhang et al., 2006;
 455 Caicedo et al., 2009) and bag-of-visual-words (Caicedo et al., 2009; Cruz-Roa et al., 2009; Sivic
 456 & Zisserman, 2003). In computational pathology, prototypical representations are natural since re-
 457 peating histology patterns often reflect shared morphology (Hou et al., 2016; Kalra et al., 2020; Pan
 458 et al., 2023; Wang et al., 2022b; Xu et al., 2012; Yu et al., 2023). Recent approaches (Vu et al.,
 459 2023; Claudio Quiros et al., 2024; Zaheer et al., 2017) build WSI prototypes using manual features
 460 and distance-based clustering (e.g., K-Means), with state-of-the-art variants adopting Gaussian Mix-
 461 ture Models (Mialon et al., 2021; Song et al., 2024). However, they all impose a fixed number of
 462 prototypes, ignoring the varying complexity of different WSIs. Adaptive clustering methods (Li &
 463 Nehorai, 2018; Vijayan & Aziz, 2023) offer partial flexibility but rely on rigid structures and in-
 464 cur high training costs, making them impractical for gigapixel slides. These limitations motivate
 465 NICER, a probabilistic nonparametric framework that is efficient and slide-adaptive.

466 **Dataset Condensation.** Dataset condensation, or distillation, compresses large datasets into small
 467 synthetic sets that preserve model performance (Wang et al., 2020). Unlike prototype learning,
 468 which selects subsets or analytic representations, it treats synthetic samples as learnable parameters
 469 and tries to balance between performance and efficiency via a bi-level learning approaches. This line
 470 of research typically involves in the nested optimization (Wang et al., 2020; Deng & Russakovsky,
 471 2022; Nguyen et al., 2021); or surrogate-objective approaches (Zhao et al., 2021; Wang et al., 2022a;
 472 Liu et al., 2023; Sajedi et al., 2023). In the context of WSIs, FedWSIDD (Jin et al., 2025) extends
 473 condensation to federated settings by synthesizing slides for efficient communication, but like con-
 474 ventional methods it relies on supervision signals, limiting its use in real-world scenarios where an-
 475 notations are scarce. In contrast, NICER introduces an unsupervised data condensation framework
 476 that eliminates the dependency on labels, enabling scalable, annotation-free WSI condensation.

477 5 CONCLUSIONS

478 In this paper, we introduce NICER, a novel framework for whole-slide image (WSI) condensation
 479 that addresses histological heterogeneity across slides. NICER reformulates prototype construc-
 480 tion as an unsupervised data condensation problem within a hierarchical probabilistic model, where
 481 prototypical information is distilled from raw features into patterns and then condensed into com-
 482 pact concepts nonparametrically. This adaptive process allows the prototype set to scale with WSI
 483 complexity, effectively handling variability across slides. We further derive a Bayesian inference
 484 algorithm to learn pattern-concept associations efficiently. Across datasets and tasks, NICER out-
 485 performs prior methods with up to 90% gains and strong efficiency, demonstrating practicality for
 486 real-world pathology under limited resources.

486 REFERENCES
487

488 Abdallah Alabdallah, Mattias Ohlsson, Sepideh Pashami, and Thorsteinn Rögnvaldsson. The con-
489 cordance index decomposition: A measure for a deeper understanding of survival prediction mod-
490 els. *Artif Intell Med*, 148:102781, January 2024.

491 Wouter Bulten, Hans Pinckaers, Hester van Boven, Robert Vink, Thomas de Bel, Bram van Gin-
492 neken, Jeroen van der Laak, Christina Hulsbergen-van de Kaa, and Geert Litjens. Automated
493 deep-learning system for gleason grading of prostate cancer using biopsies: a diagnostic study.
494 *Lancet Oncol*, 21(2):233–241, January 2020.

495 Wouter Bulten, Kimmo Kartasalo, Po-Hsuan Cameron Chen, Peter Ström, Hans Pinckaers, Kunal
496 Nagpal, Yuannan Cai, David F. Steiner, Hester van Boven, Robert Vink, Christina Hulsbergen-
497 van de Kaa, Jeroen van der Laak, Mahul B. Amin, Andrew J. Evans, Theodorus van der Kwast,
498 Robert Allan, Peter A. Humphrey, Henrik Grönberg, Hemamali Samaratunga, Brett Delahunt,
499 Toyonori Tsuzuki, Tomi Häkkinen, Lars Egevad, Maggie Demkin, Sohier Dane, Fraser Tan,
500 Masi Valkonen, Greg S. Corrado, Lily Peng, Craig H. Mermel, Pekka Ruusuvuori, Geert Lit-
501 jens, Martin Eklund, Américo Brilhante, Aslı Çakır, Xavier Farré, Katerina Geronatsiou, Vin-
502 cent Molinié, Guilherme Pereira, Paromita Roy, Günter Saile, Paulo G. O. Salles, Ewout Schaaf-
503 sma, Joëlle Tschui, Jorge Billoch-Lima, Emílio M. Pereira, Ming Zhou, Shujun He, Sejun Song,
504 Qing Sun, Hiroshi Yoshihara, Taiki Yamaguchi, Kosaku Ono, Tao Shen, Jianyi Ji, Arnaud Rous-
505 sel, Kairong Zhou, Tianrui Chai, Nina Weng, Dmitry Grechka, Maxim V. Shugaev, Raphael
506 Kiminya, Vassili Kovalev, Dmitry Voynov, Valery Malyshov, Elizabeth Lapo, Manuel Campos,
507 Noriaki Ota, Shinsuke Yamaoka, Yusuke Fujimoto, Kentaro Yoshioka, Joni Juvonen, Mikko Tuki-
508 ainen, Antti Karlsson, Rui Guo, Chia-Lun Hsieh, Igor Zubarev, Habib S. T. Bukhar, Wenyuan
509 Li, Jiayun Li, William Speier, Corey Arnold, Kyungdoc Kim, Byeonguk Bae, Yeong Won
510 Kim, Hong-Seok Lee, Jeonghyuk Park, and the PANDA challenge consortium. Artificial in-
511 telligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature
512 Medicine*, 28(1):154–163, Jan 2022. ISSN 1546-170X. doi: 10.1038/s41591-021-01620-2. URL
513 <https://doi.org/10.1038/s41591-021-01620-2>.

514 Dmitrii Bychkov, Nina Linder, Riku Turkki, Stig Nordling, Panu E Kovanen, Clare Verrill, Margarita
515 Walliander, Mikael Lundin, Caj Haglund, and Johan Lundin. Deep learning based tissue analysis
516 predicts outcome in colorectal cancer. *Scientific Reports*, 8(1):3395, February 2018.

517 Juan C. Caicedo, Angel Cruz, and Fabio A. Gonzalez. Histopathology image classification using
518 bag of features and kernel functions. In Carlo Combi, Yuval Shahar, and Ameen Abu-Hanna
519 (eds.), *Artificial Intelligence in Medicine*, pp. 126–135, Berlin, Heidelberg, 2009. Springer Berlin
520 Heidelberg. ISBN 978-3-642-02976-9.

521 Gabriele Campanella, Matthew G Hanna, Luke Geneslaw, Allen Miraflor, Vitor Werneck
522 Krauss Silva, Klaus J Busam, Edi Brogi, Victor E Reuter, David S Klimstra, and Thomas J Fuchs.
523 Clinical-grade computational pathology using weakly supervised deep learning on whole slide
524 images. *Nature Medicine*, 25(8):1301–1309, August 2019.

525 Richard J. Chen, Chengkuan Chen, Yicong Li, Tiffany Y. Chen, Andrew D. Trister, Rahul G. Kr-
526 ishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical
527 self-supervised learning. In 2022 IEEE/CVF Conference on Computer Vision and Pattern Recog-
528 nition (CVPR), pp. 16123–16134, 2022. doi: 10.1109/CVPR52688.2022.01567.

530 Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Bowen Chen,
531 Andrew Zhang, Daniel Shao, Andrew H Song, Muhammad Shaban, et al. Towards a general-
532 purpose foundation model for computational pathology. *Nature Medicine*, 2024.

533 Adalberto Claudio Quiros, Nicolas Coudray, Anna Yeaton, Xinyu Yang, Bojing Liu, Hortense Le,
534 Luis Chiriboga, Afreen Karimkhan, Navneet Narula, David A Moore, Christopher Y Park, Harvey
535 Pass, Andre L Moreira, John Le Quesne, Aristotelis Tsirigos, and Ke Yuan. Mapping the land-
536 scape of histomorphological cancer phenotypes using self-supervised learning on unannotated
537 pathology slides. *Nature Communications*, 15(1):4596, June 2024.

538 Angel Cruz-Roa, Juan C. Caicedo, and Fabio A. González. Visual pattern analysis in histopathol-
539 ogy images using bag of features. In *Proceedings of the 14th Iberoamerican Conference on*

540 *Pattern Recognition: Progress in Pattern Recognition, Image Analysis, Computer Vision, and*
 541 *Applications*, CIARP '09, pp. 521–528, Berlin, Heidelberg, 2009. Springer-Verlag. ISBN
 542 9783642102677. doi: 10.1007/978-3-642-10268-4_61. URL https://doi.org/10.1007/978-3-642-10268-4_61.

543

544 Dan dan Guo, Long Tian, Minghe Zhang, Mingyuan Zhou, and Hongyuan Zha. Learning prototype-
 545 oriented set representations for meta-learning. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=WH6u2Sv1Lp4>.

546

547 Zhiwei Deng and Olga Russakovsky. Remember the past: distilling datasets into addressable mem-
 548 ories for neural networks. In *Proceedings of the 36th International Conference on Neural Infor-
 549 mation Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN
 550 9781713871088.

551

552 Lei Fan, Arcot Sowmya, Erik Meijering, and Yang Song. Cancer survival prediction from whole
 553 slide images with self-supervised learning and slide consistency. *IEEE Transactions on Medical*
 554 *Imaging*, 42(5):1401–1412, 2023. doi: 10.1109/TMI.2022.3228275.

555

556 Simon Graham, Quoc Dang Vu, Mostafa Jahanifar, Shan E Ahmed Raza, Fayyaz Minhas, David
 557 Snead, and Nasir Rajpoot. One model is all you need: Multi-task learning enables simulta-
 558 neous histology image segmentation and classification. *Medical Image Analysis*, 83:102685,
 559 2023. ISSN 1361-8415. doi: <https://doi.org/10.1016/j.media.2022.102685>. URL <https://www.sciencedirect.com/science/article/pii/S1361841522003139>.

560

561 Ruoyu Guo, Kunzi Xie, Maurice Pagnucco, and Yang Song. Sac-net: Learning with weak and noisy
 562 labels in histopathology image segmentation. *Medical Image Analysis*, 86:102790, 2023.

563

564 Le Hou, Dimitris Samaras, Tahsin M. Kurc, Yi Gao, James E. Davis, and Joel H. Saltz. Patch-
 565 based convolutional neural network for whole slide tissue image classification. In *2016 IEEE*
 566 *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2424–2433, 2016. doi:
 567 10.1109/CVPR.2016.266.

568

569 Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learn-
 570 ing. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Con-
 571 ference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp.
 572 2127–2136. PMLR, 10–15 Jul 2018. URL <https://proceedings.mlr.press/v80/ilse18a.html>.

573

574 Syed Ashar Javed, Dinkar Juyal, Harshith Padigela, Amaro Taylor-Weiner, Limin Yu, and Aaditya
 575 Prakash. Additive mil: intrinsically interpretable multiple instance learning for pathology. In
 576 *Proceedings of the 36th International Conference on Neural Information Processing Systems*,
 577 NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

578

579 Haolong Jin, Shenglin Liu, Cong Cong, Qingmin Feng, Yongzhi Liu, Lina Huang, and Yingzi Hu.
 580 Fedwsidd: Federated whole slide image classification via dataset distillation, 2025. URL <https://arxiv.org/abs/2506.15365>.

581

582 Shivam Kalra, H R Tizhoosh, Charles Choi, Sultaan Shah, Phedias Diamandis, Clinton J V Camp-
 583 bell, and Liron Pantanowitz. Yottixel - an image search engine for large archives of histopathology
 584 whole slide images. *Med Image Anal*, 65:101757, June 2020.

585

586 Jakob Nikolas Kather, Johannes Krisam, Pornpimol Charoentong, Tom Luedde, Esther Herpel,
 587 Cleo-Aron Weis, Timo Gaiser, Alexander Marx, Nektarios A. Valous, Dyke Ferber, Lina Jansen,
 588 Constantino Carlos Reyes-Aldasoro, Inka Zörnig, Dirk Jäger, Hermann Brenner, Jenny Chang-
 589 Claude, Michael Hoffmeister, and Niels Halama. Predicting survival from colorectal cancer his-
 590 tology slides using deep learning: A retrospective multicenter study. *PLOS Medicine*, 16(1):
 591 1–22, 01 2019. doi: 10.1371/journal.pmed.1002730. URL <https://doi.org/10.1371/journal.pmed.1002730>.

592

593 Minyoung Kim. Differentiable expectation-maximization for set representation learning. In *Inter-
 594 national Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=MXdFBmHT4C>.

594 S. Lazebnik, C. Schmid, and J. Ponce. A sparse texture representation using local affine regions.
 595 *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1265–1278, 2005. doi:
 596 10.1109/TPAMI.2005.151.

597 Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set
 598 transformer: A framework for attention-based permutation-invariant neural networks. In Ka-
 600 malika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Con-
 601 ference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp.
 602 3744–3753. PMLR, 09–15 Jun 2019. URL <https://proceedings.mlr.press/v97/lee19d.html>.

603 Bin Li, Yin Li, and Kevin W Eliceiri. Dual-stream multiple instance learning network for whole slide
 604 image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF
 605 Conference on Computer Vision and Pattern Recognition*, pp. 14318–14328, 2021a.

606 Bin Li, Yin Li, and Kevin W Eliceiri. Dual-stream multiple instance learning network for whole slide
 607 image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF
 608 Conference on Computer Vision and Pattern Recognition*, pp. 14318–14328, 2021b.

609 Jichuan Li and Arye Nehorai. Gaussian mixture learning via adaptive hierarchical clustering. *Sig-
 610 nal Processing*, 150:116–121, 2018. ISSN 0165-1684. doi: <https://doi.org/10.1016/j.sigpro.2018.04.013>. URL <https://www.sciencedirect.com/science/article/pii/S0165168418301397>.

611 Yanqing Liu, Jianyang Gu, Kai Wang, Zheng Zhu, Wei Jiang, and Yang You. Dream: Efficient
 612 dataset distillation by representative matching. In *ICCV*, pp. 17268–17278, 2023. URL <https://doi.org/10.1109/ICCV51070.2023.01588>.

613 Ming Y. Lu, Bowen Chen, Andrew Zhang, Drew F. K. Williamson, Richard J. Chen, Tong Ding,
 614 Long Phi Le, Yung-Sung Chuang, and Faisal Mahmood. Visual language pretrained multiple in-
 615 stance zero-shot transfer for histopathology images. In *Proceedings of the IEEE/CVF Conference
 616 on Computer Vision and Pattern Recognition (CVPR)*, pp. 19764–19775, June 2023.

617 Grégoire Mialon, Dexiong Chen, Alexandre d’Aspremont, and Julien Mairal. A trainable optimal
 618 transport embedding for feature aggregation and its relationship to attention. In *International
 619 Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=ZK6vTvb84s>.

620 Pooya Mobadersany, Safoora Yousefi, Mohamed Amgad, David A. Gutman, Jill S. Barnholtz-Sloan,
 621 José E. Velázquez Vega, Daniel J. Brat, and Lee A. D. Cooper. Predicting cancer outcomes from
 622 histology and genomics using convolutional networks. *Proceedings of the National Academy of
 623 Sciences*, 115(13):E2970–E2979, 2018. doi: 10.1073/pnas.1717139115. URL <https://www.pnas.org/doi/abs/10.1073/pnas.1717139115>.

624 Manh Duong Nguyen, Nguyen Dang Huy Pham, Phi Le Nguyen, and Minh N. Do. A semi-
 625 supervised learning framework with cross-magnification attention for glioma mitosis classifica-
 626 tion. In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, pp. 1–4, 2025a.
 627 doi: 10.1109/ISBI60581.2025.10981240.

628 Manh Duong Nguyen, Dac Thai Nguyen, Trung Viet Nguyen, Homi Yamada, Huy Hieu Pham, and
 629 Phi Le Nguyen. Bridging classification and segmentation in osteosarcoma assessment via founda-
 630 tion and discrete diffusion models. In *2025 IEEE 22nd International Symposium on Biomedical
 631 Imaging (ISBI)*, pp. 1–5, 2025b. doi: 10.1109/ISBI60581.2025.10980711.

632 Timothy Nguyen, Zhourong Chen, and Jaehoon Lee. Dataset meta-learning from kernel ridge-
 633 regression. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=1-PrrQrK0QR>.

634 Wentao Pan, Jiangpeng Yan, Hanbo Chen, Jiawei Yang, Zhe Xu, Xiu Li, and Jianhua Yao. Human-
 635 machine interactive tissue prototype learning for label-efficient histopathology image segmenta-
 636 tion. 2023.

648 Peijie Qiu, Pan Xiao, Wenhui Zhu, Yalin Wang, and Aristeidis Sotiras. Sc-mil: Sparsely coded
 649 multiple instance learning for whole slide image classification. *arXiv preprint arXiv:2311.00048*,
 650 2023.

651 Alessio Sacco, Flavio Esposito, Guido Marchetto, Grant Kolar, and Kate Schwetye. On edge com-
 652 puting for remote pathology consultations and computations. *IEEE Journal of Biomedical and*
 653 *Health Informatics*, 24(9):2523–2534, 2020. doi: 10.1109/JBHI.2020.3007661.

654 Ahmad Sajedi, Samir Khaki, Ehsan Amjadian, Lucy Z. Liu, Yuri A. Lawryshyn, and Konstanti-
 655 nos N. Plataniotis. Datadam: Efficient dataset distillation with attention matching. In *Pro-
 656 ceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 17097–17107,
 657 October 2023.

658 Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, and Yongbing
 659 Zhang. Transmil: transformer based correlated multiple instance learning for whole slide im-
 660 age classification. In *Proceedings of the 35th International Conference on Neural Infor-
 661 mation Processing Systems*, NIPS ’21, Red Hook, NY, USA, 2021. Curran Associates Inc. ISBN
 662 9781713845393.

663 Sivic and Zisserman. Video google: a text retrieval approach to object matching in videos. In
 664 *Proceedings Ninth IEEE International Conference on Computer Vision*, pp. 1470–1477 vol.2,
 665 2003. doi: 10.1109/ICCV.2003.1238663.

666 Andrew H Song, Guillaume Jaume, Drew F K Williamson, Ming Y Lu, Anurag Vaidya, Tiffany R
 667 Miller, and Faisal Mahmood. Artificial intelligence for digital and computational pathology. *Na-
 668 ture Reviews Bioengineering*, 1(12):930–949, December 2023.

669 Andrew H. Song, Richard J. Chen, Tong Ding, Drew F.K. Williamson, Guillaume Jaume, and Faisal
 670 Mahmood. Morphological prototyping for unsupervised slide representation learning in compu-
 671 tational pathology. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition
 672 (CVPR)*, pp. 11566–11578, 2024. doi: 10.1109/CVPR52733.2024.01099.

673 Wenhao Tang, Sheng Huang, Xiaoxian Zhang, Fengtao Zhou, Yi Zhang, and Bo Liu. Multiple
 674 instance learning framework with masked hard instance mining for whole slide image classifica-
 675 tion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp.
 676 4078–4087, October 2023.

677 Kevin Thandiackal, Boqi Chen, Pushpak Pati, Guillaume Jaume, Drew F. K. Williamson,
 678 Maria Gabrani, and Orcun Goksel. Differentiable zooming fornbsp;multiple instance learning
 679 onnbsp;whole-slide images. In *Computer Vision – ECCV 2022: 17th European Conference,
 680 Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXI*, pp. 699–715, Berlin, Heidelberg,
 681 2022. Springer-Verlag. ISBN 978-3-031-19802-1. doi: 10.1007/978-3-031-19803-8_41. URL
 682 https://doi.org/10.1007/978-3-031-19803-8_41.

683 Susana M. Vieira, Uzay Kaymak, and João M. C. Sousa. Cohen’s kappa coefficient as a performance
 684 measure for feature selection. In *International Conference on Fuzzy Systems*, pp. 1–8, 2010. doi:
 685 10.1109/FUZZY.2010.5584447.

686 Darveen Vijayan and Izzatdin Aziz. Adaptive hierarchical density-based spatial clustering algo-
 687 rithm for streaming applications. *Telecom*, 4(1):1–14, 2023. ISSN 2673-4001. doi: 10.3390/
 688 telecom4010001. URL <https://www.mdpi.com/2673-4001/4/1/1>.

689 Quoc Dang Vu, Kashif Rajpoot, Shan E Ahmed Raza, and Nasir Rajpoot. Handcrafted histological
 690 transformer (h2t): Unsupervised representation of whole slide images. *Medical Image Analysis*,
 691 pp. 102743, 2023.

692 Kai Wang, Bo Zhao, Xiangyu Peng, Zheng Zhu, Shuo Yang, Shuo Wang, Guan Huang, Hakan
 693 Bilen, Xinchao Wang, and Yang You. Cafe: Learning to condense dataset by aligning features. In
 694 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*,
 695 pp. 12196–12205, June 2022a.

696 Tongzhou Wang, Jun-Yan Zhu, Antonio Torralba, and Alexei A. Efros. Dataset distillation, 2020.
 697 URL <https://arxiv.org/abs/1811.10959>.

702 Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and
 703 Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image clas-
 704 sification. Elsevier, 2022b.

705 Pei-Yau Weng, Minh Hoang, Lam M. Nguyen, My T. Thai, Tsui-Wei Weng, and Trong Nghia
 706 Hoang. Probabilistic federated prompt-tuning with non-IID and imbalanced data. In *The Thirty-
 707 eighth Annual Conference on Neural Information Processing Systems*, 2024. URL <https://openreview.net/forum?id=nw6ANsC66G>.

708 Jinxi Xiang and Jun Zhang. Exploring low-rank property in multiple instance learning for whole
 709 slide image classification. In *The Eleventh International Conference on Learning Representations*,
 710 2023. URL <https://openreview.net/forum?id=01KmhBsEPFO>.

711 Yan Xu, Jun-Yan Zhu, Eric Chang, and Zhuowen Tu. Multiple clustered instance learning for
 712 histopathology cancer image classification, segmentation and clustering. In *2012 IEEE Confer-
 713 ence on Computer Vision and Pattern Recognition*, pp. 964–971, 2012. doi: 10.1109/CVPR.2012.
 714 6247772.

715 Changmin Yu, Maneesh Sahani, and Máté Lengyel. Discovering temporally compositional neural
 716 manifolds with switching infinite GPFA. In *The Thirteenth International Conference on Learning
 717 Representations*, 2025. URL <https://openreview.net/forum?id=2iCIHgE8KG>.

718 Jin-Gang Yu, Zihao Wu, Yu Ming, Shule Deng, Yuanqing Li, Caifeng Ou, Chunjiang He, Baiye
 719 Wang, Pusheng Zhang, and Yu Wang. Prototypical multiple instance learning for predicting
 720 lymph node metastasis of breast cancer from whole-slide pathological images. *Med Image Anal*,
 721 85:102748, January 2023.

722 Shekoufeh Gorgi Zadeh and Matthias Schmid. Bias in cross-entropy-based training of deep survival
 723 networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(9):3126–3137,
 724 2021. doi: 10.1109/TPAMI.2020.2979450.

725 Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhut-
 726 dinov, and Alexander J Smola. Deep sets. In I. Guyon, U. Von Luxburg,
 727 S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Ad-
 728 vances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.,
 729 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/f22e4747da1aa27e363d86d40ff442fe-Paper.pdf.

730 Jianguo Zhang, M. Marszalek, S. Lazebnik, and C. Schmid. Local features and kernels for
 731 classification of texture and object categories: A comprehensive study. In *2006 Conference
 732 on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, pp. 13–13, 2006. doi:
 733 10.1109/CVPRW.2006.121.

734 Bo Zhao, Konda Reddy Mopuri, and Hakan Bilen. Dataset condensation with gradient matching. In
 735 *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=mSAKhLYLs1>.

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756 A BROADER STATEMENT OF IMPACT
757

758 This research develops an effective nonparametric compression and condensation algorithm for
759 whole-slide images (WSIs), enabling efficient learning from histological data with varying com-
760 plexity. The mathematical methods and insights presented in this work help bridge the gap between
761 large, gigabyte-scale images and their practical applications in healthcare. While the potential appli-
762 cation of our methods to real patient data may raise ethical considerations, such effects are indirect
763 and not the focus of this study. Our experiments rely solely on publicly available datasets, ensuring
764 that no ethical concerns are introduced in the evaluation of our algorithms.

765
766 B PSEUDOCODE FOR NICER
767768 **Algorithm 1** Nonparametric Unsupervised Data Condensation (NICER)
769

770 **input:** WSI feature bag \mathcal{H} , no. T of iterations, no. M of initial patterns
771 **output:** condensed concept prototype set Ω

772 1: initialize pattern set \mathcal{Z} of size M
773 2: **for** $t = 1$ to T **do**
774 3: **for** $i = 1$ to N **do**
775 4: query top- κ relevant patterns to h_i using Eq. 3
776 5: distill h_i to κ relevant z_m by maximizing Eq. 3
777 6: **end for**
778 7: $\Omega, b \leftarrow \text{condense}(\{z_m\}_{m=1}^M)$ // solving Eq. 8
779 8: $\Omega \leftarrow \left\{ \omega_k \in \Omega \mid \sum_{m=1}^M b_{mk} > 0 \right\}$ // remove redundant concepts
780 9: **end for**
781 10: **return** the set Ω of optimal concepts

782
783 C IMPLEMENTATION DETAILS
784785 C.1 DATASETS
786

787 We provide brief explanations for the datasets that were used for the evaluation of NICER for con-
788 densation ability and prototyping ability.

789 **PANDA.** (Bulten et al., 2022; 2020) For the ISUP grading task, we used prostate cancer core needle
790 biopsies (n=10,616) from the Prostate Cancer Grade Assessment (PANDA) challenge. Each biopsy
791 is given an ISUP grade, making this a 6-class classification task. These biopsies are collected from
792 Karolinska Institute (KRLS) and Radboud University Medical Center (RUMC). We label-stratify
793 the PANDA dataset into train/val/test of 80:10:10 and performance was evaluated using Cohen's
794 quadratic weighted Kappa κ^2 , accuracy and weighted F1 metrics.

795 **NSCLC.** For the non-small cell lung carcinoma (NSCLC) subtyping task, we use H&E WSIs from
796 TCGA for classifying lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC)
797 cases. The TCGA cohort contains a total of 1,4041 slides (LUAD: 529, LUSC: 512). We label-
798 stratify the TCGA cohort into train/val/test fold of 80:10:10 and evaluate performance using Cohen's
799 Kappa, balanced accuracy and F1 metrics.

800 **TCGA-BRCA.** The Breast Invasive Carcinoma (BRCA) cohort from The Cancer Genome Atlas
801 (TCGA), a joint effort of the NCI and NHGRI, provides one of the most comprehensive digital
802 pathology resources for breast cancer. The dataset contains 1,133 diagnostic WSIs spanning 1,062
803 patients, covering diverse histological and molecular subtypes. In addition to imaging, the cohort
804 supplies curated clinical outcomes, including Overall Survival (OS) and Progression-Free Interval
805 (PFI), which have been widely adopted as endpoints for developing and benchmarking survival
806 prediction models. We label-stratify the TCGA-BRCA dataset into train/val/test fold of 80:10:10
807 and evaluate performance C-Index for survival prediction

808 **TCGA-LUAD.** The Lung Adenocarcinoma (LUAD) cohort within TCGA offers a large-scale,
809 multi-institutional collection of pathology images and outcome data for a major subtype of non-

810 small cell lung cancer. It includes 529 diagnostic WSIs corresponding to 478 patients, with extensive clinical annotation. As with BRCA, the LUAD cohort provides OS and PFI as standardized survival endpoints, enabling robust prognostic modeling and cross-study comparison in computational pathology research. We label-stratify the TCGA-LUAD dataset into train/val/test fold of 80:10:10 and evaluate performance C-Index for survival prediction

817 C.2 EVALUATION SETTING DETAILS

819 We view unsupervised prototype construction not only as a condensation framework, whose primary
 820 role is to reduce redundancy in the WSI feature bag, but also as a form of representation learning,
 821 since the resulting prototypes are later used as inputs for downstream predictors. Accordingly, we
 822 evaluate all methods along two complementary axes:

824 **Condensation ability.** Here, the goal is to assess how well the condensed prototype set preserves
 825 information from the original WSI feature bag. Specifically, we apply NICER and baseline methods
 826 on the training set to obtain prototypes, and then evaluate them on the original uncompressed bag
 827 of features. This setup isolates the effectiveness of condensation by measuring how much predictive
 828 power is retained (or lost) after summarization, independent of downstream task complexity. It
 829 answers the key question: *Does condensation discard critical information or faithfully represent the*
 830 *original slide?*

831 **Morphological prototyping.** In this setting, we follow the prior evaluation protocol Song et al.
 832 (2024) to test whether condensed prototypes can generalize as useful, task-agnostic representations.
 833 Condensation methods are applied to both the training and test sets, and the resulting prototypes
 834 are directly used for downstream prediction. Unlike the condensation ability evaluation, this setup
 835 emphasizes the representation learning capacity of the prototypes, focusing on whether they capture
 836 robust morphological cues transferable across unseen WSIs rather than reconstruction fidelity. To
 837 ensure fairness, we use linear probing for all baselines, thereby isolating the quality of the learned
 838 prototypes. For PANTHER and NICER, we additionally evaluate with a transformer layer (with or
 839 without linearity), since NICER produces prototype sets whose capacity adapts to WSI complexity
 840 and thus cannot be fully exploited by a fixed linear layer alone.

842 C.3 TRAINING DETAILS AND COMPUTATIONAL RESOURCES

844 **Data preparation.** WSIs at $20\times$ magnification ($0.5, \mu\text{m}/\text{pixel}$) are divided into non-overlapping
 845 256×256 patches, and all patches are used without sampling. These patches are converted to
 846 representations using UNI Chen et al. (2024), a pretrained foundation encoder. We set $\kappa = 3$ and
 847 found $T = 50$ iterations sufficient for convergence across all datasets.

848 **Hyperparameter settings.** For training, we adopt the AdamW optimizer with weight decay set to
 849 1×10^{-5} and employ a cosine decay learning rate scheduler. In the *cancer subtyping* experiments,
 850 models are trained with cross-entropy loss for up to 50 epochs, with early stopping triggered if the
 851 validation loss fails to improve for 10 consecutive epochs. The initial learning rate is set to 1×10^{-4} .
 852 Since both the original feature bags and the NICER’s representations form variable-length WSI sets,
 853 we use a batch size of 1 combined with gradient accumulation over 32 steps across all methods.
 854 For the *survival prediction* task, we optimize using the negative log-likelihood (NLL) loss (Zadeh &
 855 Schmid, 2021), training over 50 epochs with a per-patient batch size of 1 and an initial learning rate
 856 of 1×10^{-5} . The training of NICER further involves an alternating optimization process over 20
 857 iterations, with $\kappa = 3$ and an initial pattern set size of $M_0 = 1000$, which decreases progressively
 858 through the condensation procedure. As a nonparametric random process, NICER allows us to
 859 directly regulate the number of generated prototypes via κ and M_0 . Following prior practices (Vu
 860 et al., 2023; Song et al., 2024), we set a number of prototypes generated by unsupervised baselines
 861 to 16 for all WSIs used in our experiments. The implementation details of predictor architectures,
 862 unsupervised baselines follow original papers and previous settings (Song et al., 2024).

863 **Computational considerations.** All experiments and data preprocessing are conducted on a
 864 NVIDIA RTX A6000 with 46GB of memory.

864 **D DERIVATION OF EQ. 2**
865866 We begin from the conditional likelihood in Eq. 2:
867

868
$$\log \mathbb{P}(\mathcal{H} \mid \mathcal{Z}, \theta) = \sum_{i=1}^N \log \mathbb{N}(h_i \mid z_{(i)}^*, \sigma^2 \mathbf{I}), \quad (10)$$

869

870 where each feature h_i is modeled as a Gaussian centered at its assigned pattern prototype $z_{(i)}^*$.
871 Expanding the Gaussian log-likelihood yields:
872

873
$$\log \mathbb{P}(\mathcal{H} \mid \mathcal{Z}, \theta) = -\frac{1}{2\sigma^2} \sum_{i=1}^N \|h_i - z_{(i)}^*\|^2 + C, \quad (11)$$

874

875 with C denoting terms independent of h_i or $z_{(i)}^*$. The squared distance can be written as
876

877
$$\|h_i - z_{(i)}^*\|^2 = \|h_i\|^2 + \|z_{(i)}^*\|^2 - 2\langle h_i, z_{(i)}^* \rangle. \quad (12)$$

878

879 Since both h_i and $z_{(i)}^*$ are ℓ_2 -normalized embeddings, we have $\|h_i\|^2 = \|z_{(i)}^*\|^2 = 1$. This reduces
880 the squared distance to
881

882
$$\|h_i - z_{(i)}^*\|^2 \approx 2 - 2\langle h_i, z_{(i)}^* \rangle. \quad (13)$$

883

884 Plugging this into the log-likelihood gives
885

886
$$\log \mathbb{P}(\mathcal{H} \mid \mathcal{Z}, \theta) \approx -\frac{1}{2\sigma^2} \sum_{i=1}^N (2 - 2\langle h_i, z_{(i)}^* \rangle) + C. \quad (14)$$

887

888 Dropping constants, we recover Eq. 2. This shows that maximizing the Gaussian likelihood is
889 approximately equivalent to maximizing feature - prototype similarity, providing a probabilistic
890 justification for our design.
891892 **E LEMMAS AND DERIVATIONS**
893894 **Lemma E.1.** (adapted from Weng et al. (2024)) For any scalar function $g(\mathbf{r})$ and a binary vector
895 $\xi = [\xi_1, \xi_2, \dots, \xi_n]$ such that $\xi_i \in \{0, 1\}$ and ξ has exactly one non-zero component, we have
896

897
$$g\left(\sum_{i=1}^n \xi_i \cdot \mathbf{r}_i\right) = \sum_{i=1}^n \left(\xi_i \cdot g(\mathbf{r}_i)\right) \quad (15)$$

898

899 with respect to any set $\{\mathbf{r}_i\}_{i=1}^n$ of valid inputs to $g(\mathbf{r})$.
900901 **Proof.** First, if there is no non-zero component, both sides of Eq. 23 evaluate to $g(0)$. Otherwise,
902 suppose the only non-zero component appears at position j , both sides of Eq. 23 will evaluate to
903 $g(\mathbf{r}_j)$. In both cases, Eq. 23 holds.
904905 **Lemma E.2.** Let $\mathbb{P}(z_m \mid b_m, \Omega)$ defined as in Eq. 5. Let $R_1(b) \triangleq \sum_{m=1}^M \log \mathbb{P}(z_m \mid b_m, \Omega)$,
906 considering (z_m, Ω) as constants. We have
907

908
$$R_1(b) = \sum_{i=1}^M \sum_{k=1}^K b_{mk} \cdot \log \mathbb{N}(z_m \mid \omega_k, \text{diag}(\delta(\omega_k; \gamma))), \quad (16)$$

909

910 which is linear in terms of the assignment parameter b .
911912 **Proof.** To derive results of Lemma E.2, note that Eq. 5 implies the following,
913

914
$$\log \mathbb{P}(z_m \mid b_m, \Omega) = \log \mathbb{N}\left(z_m \mid \sum_{k=1}^K b_{mk} \cdot \omega_k, \text{diag}\left(\delta\left(\sum_{k=1}^K b_{mk} \cdot \omega_k; \zeta\right)\right)\right) \quad (17)$$

915

916
$$= g\left(\sum_{k=1}^K b_{mk} \cdot \omega_k\right) \quad (18)$$

917

918 where we define

$$919 \quad 920 \quad 921 \quad 922 \quad 923 \quad 924 \quad 925 \quad 926 \quad 927 \quad 928 \quad 929 \quad 930 \quad 931 \quad 932 \quad 933 \quad 934 \quad 935 \quad 936 \quad 937 \quad 938 \quad 939 \quad 940 \quad 941 \quad 942 \quad 943 \quad 944 \quad 945 \quad 946 \quad 947 \quad 948 \quad 949 \quad 950 \quad 951 \quad 952 \quad 953 \quad 954 \quad 955 \quad 956 \quad 957 \quad 958 \quad 959 \quad 960 \quad 961 \quad 962 \quad 963 \quad 964 \quad 965 \quad 966 \quad 967 \quad 968 \quad 969 \quad 970 \quad 971$$

$$g(\mathbf{x}) \triangleq \log \mathbb{N}\left(z_m \mid \mathbf{x}, \text{diag}(\mathbf{x}; \zeta)\right) \quad (19)$$

In addition, since $\sum_k b_{mk} = 1$ with $b_{mk} \in \{0, 1\}$, Lemma E.1 implies that

$$g\left(\sum_{k=1}^K b_{mk} \cdot \omega_k\right) = \sum_{k=1}^K \left(b_{mk} \cdot g(\omega_k)\right) \quad (20)$$

We then plug Eq. 20 into Eq. 18 to have

$$\log \mathbb{P}(z_m \mid b_m, \Omega) = g\left(\sum_{k=1}^K b_{mk} \cdot \omega_k\right) = \sum_{k=1}^K \left(b_{mk} \cdot \omega_k\right) \quad (21)$$

$$= \sum_{k=1}^K b_{mk} \cdot \log \mathbb{N}\left(z_m \mid \omega_k, \text{diag}(\sigma(\delta_k; \zeta))\right) \quad (22)$$

Finally, taking summation over $m = 1, 2, \dots, M$ on both sides of Eq. 22, we arrive Lemma E.2.

Lemma E.3. *Let $\mathbb{P}(b_m \mid \zeta)$ defined as in Eq. 7. Let $R_2(b) \triangleq \sum_{m=1}^M \log \mathbb{P}(b_m \mid \zeta)$, considering (ζ) as constants. We have*

$$R_2(b) = \sum_{i=1}^M \sum_{k=1}^K b_{mk} \cdot \log \left(\frac{\exp(\alpha(\omega_k; \zeta))}{\sum_k \exp(\alpha(\omega_k; \zeta))} \right), \quad (23)$$

which is linear in terms of the assignment parameter b .

Proof. Plug Eq. 7 into the definition of $R_2(b)$, we have

$$R_2(b) = \sum_{m=1}^M \log \mathbb{P}(b_m \mid \zeta) = \sum_{m=1}^M \log \left(\prod_{k=1}^K \left(\frac{\exp(\alpha(\omega_k; \zeta))}{\sum_k \exp(\alpha(\omega_k; \zeta))} \right)^{b_{mk}} \right) \quad (24)$$

$$= \sum_{m=1}^M \sum_{k=1}^K b_{mk} \cdot \log \left(\frac{\exp(\alpha(\omega_k; \zeta))}{\sum_k \exp(\alpha(\omega_k; \zeta))} \right) \quad (25)$$

which naturally arrives Lemma E.3.

F ADDITIONAL RESULTS

F.1 PROTOTYPE ANALYSIS

Concept Prototype Diversity and Convergence. We extend the findings of Section 3.2 by examining additional WSI instances from TCGA-NSCLC. Figure 7 illustrates the trajectories of concept prototypes over 10 optimization iterations of NICER for three representative samples: *TCGA-93-A4JN-01Z-00-DX1.ED4C9365-6CCF-4AEE-B4C9-3CC5EC57286C*, *TCGA-50-6594-01Z-00-DX1.43b2005a-4245-4025-ad85-4a957f308a5c*, and *TCGA-49-4514-01Z-00-DX2.f1565a36-257d-432e-a84d-47c1d7a0185f*. The visualizations reveal that different WSIs exhibit distinct condensation dynamics, reflecting variation in morphological complexity and feature distribution. For example, the first sample (ending in “86C”) shows prototypes that initially cluster tightly, suggesting greater homogeneity, whereas the other two slides begin with more diffuse clusters, indicating higher heterogeneity.

Despite these sample-specific differences, a consistent pattern emerges across all trajectories: prototypes gradually diverge from their initialization with increasing variance, reflecting how condensation enforces specialization and reduces redundancy while preserving diversity. This behavior highlights NICER’s ability to uncover distinct and non-overlapping concept structures within each slide. Complementary evidence is shown in Figure 10, where the number of prototypes stabilizes after only a few iterations. This rapid convergence indicates that redundant concepts are pruned early, leaving a compact and stable set that continues to refine qualitatively rather than quantitatively. Together, these results emphasize NICER’s efficiency in learning diverse, non-redundant representations of WSIs with minimal optimization steps.

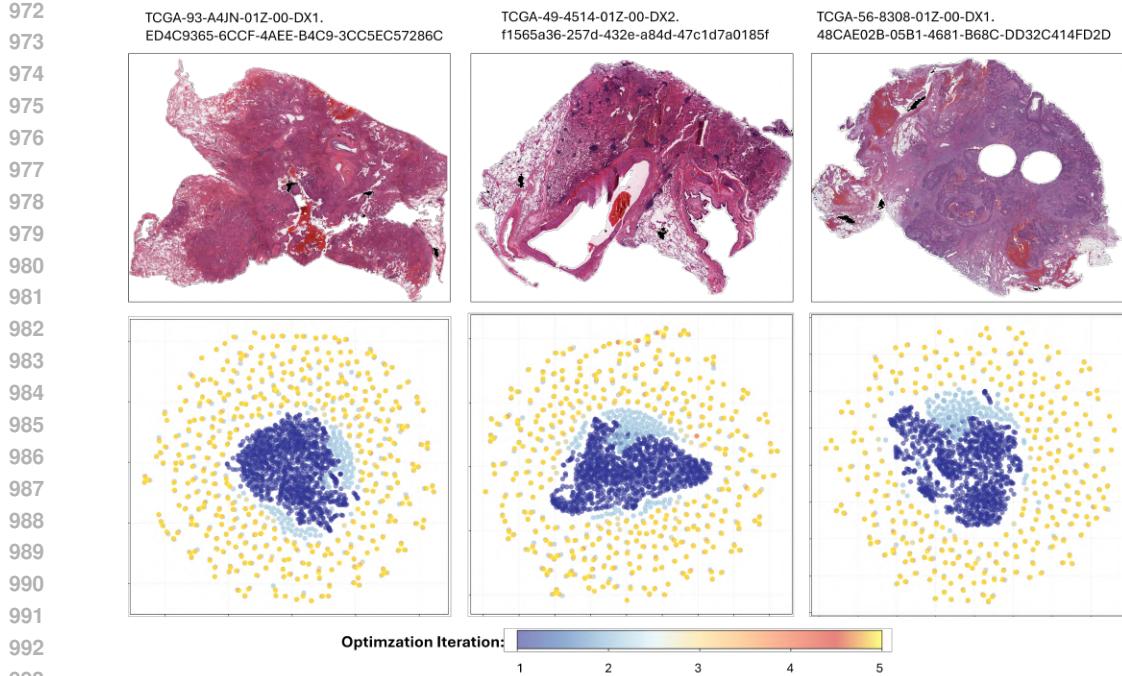


Figure 7: 2D t-SNE visualizations of NICER’s concept prototypes learned over five iterations using three representative TCGA slides. These plots highlight that NICER can capture different diverse information across different WSIs.

Table 5: Average number of final concept prototypes across different datasets

	PANDA	NSCLC	LUAD (Survival)	BRCA (Survival)
	62.80	154.67	157.40	158.60

Diversity Preservation. To evaluate NICER’s ability to preserve diversity across whole-slide images (WSIs), we compare 2D t-SNE visualizations of concept prototypes generated by NICER and PANTHER, alongside the original feature bags, on three TCGA-NSCLC samples: *TCGA-93-A4JN-01Z-00-DX1.ED4C9365-6CCF-4AEE-B4C9-3CC5EC57286C*, *TCGA-50-6594-01Z-00-DX1.43b2005a-4245-4025-ad85-4a957f308a5c*, and *TCGA-49-4514-01Z-00-DX2.f1565a36-257d-432e-a84d-47c1d7a0185f*. For fair visualization, NICER’s prototypes are clustered with K-Means to 16, matching the fixed prototype count used in PANTHER (Song et al., 2024). As shown in Figure 8, PANTHER fails to capture the inherent diversity of WSIs, collapsing heterogeneous regions into a limited number of clusters and discarding critical information required for downstream tasks, an effect consistent with its suboptimal performances in Table 1 and Table 2. This limitation arises because PANTHER enforces a rigid and heuristically small prototype budget, prioritizing efficiency at the expense of representational fidelity. In contrast, NICER employs an alternating optimization strategy that leverages pattern-based diversity preservation and condensation-driven efficiency, ensuring prototypes remain well-separated and encode distinct, slide-specific information. This balance enables NICER to adapt to the variability of individual WSIs while maintaining compact yet expressive concept sets.

Concept Prototype Capacity Analysis. To evaluate NICER’s ability to adapt its capacity to complexity of data, we report the average number of concept prototypes across different datasets. As can be seen from Table 5, for larger and more complex WSIs (e.g., NSCLC, LUAD, BRCA), NICER naturally allocates more capacity to preserve the underlying conceptual diversity in the slides. This aligns with our observation in Section 3.

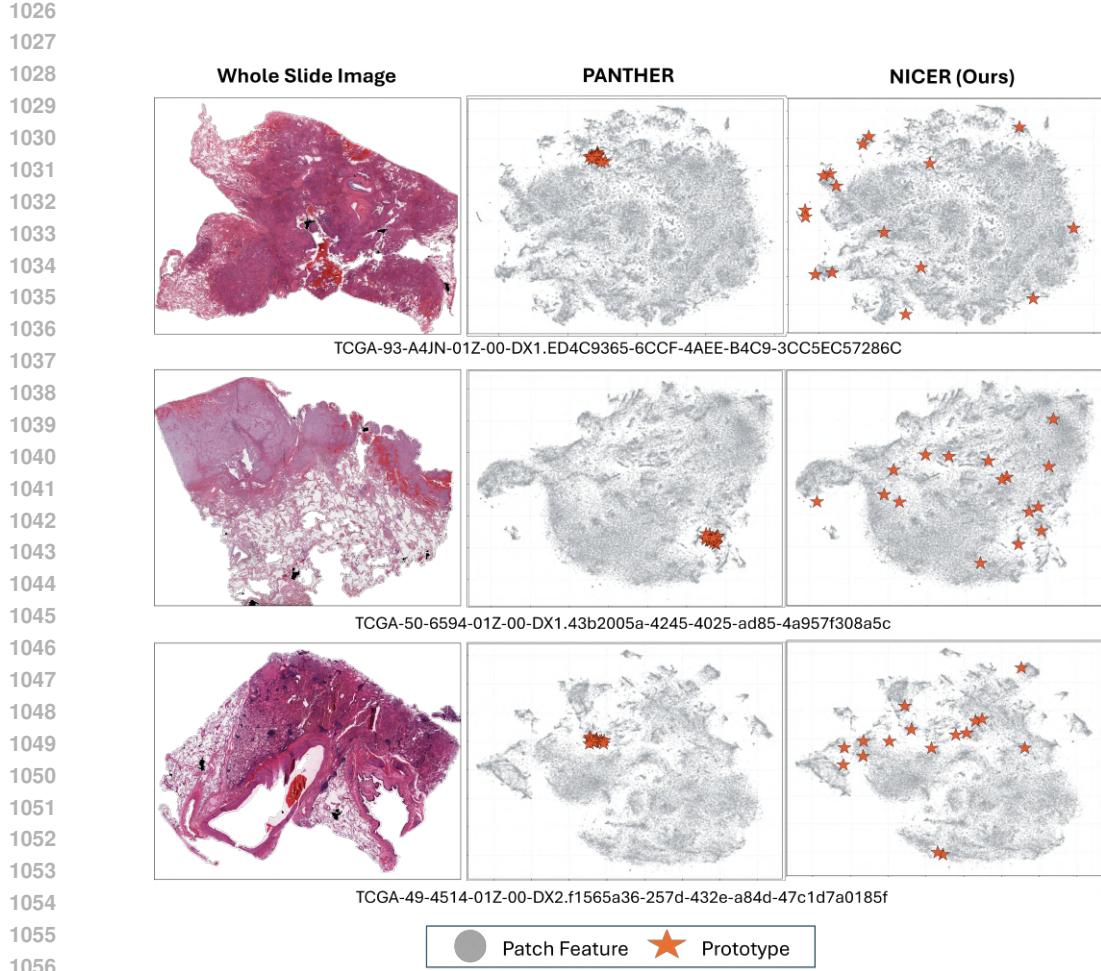


Figure 8: 2D t-SNE visualization of original WSI feature bags and learned concept prototypes, comparing our method (NICER) with a representative prior approach (PANTHER).

Table 6: Performance-Efficiency trade-offs comparison of NICER and the SOTA method. Conducted on PANDA using AB-MIL.

Predictor	M	Kappa	Accuracy	F1
ABMIL	50	85.48	67.56	67.64
	200	90.36	73.15	73.04
	500	91.85	74.50	74.46
	1000	91.48	74.94	75.00
DSMIL	50	82.75	60.18	61.03
	200	86.74	66.89	64.56
	500	88.59	67.34	68.03
	1000	88.73	68.46	67.52
ILRA	50	84.41	65.55	65.13
	200	89.67	66.62	67.34
	500	92.25	73.83	74.55
	1000	92.25	73.83	74.55

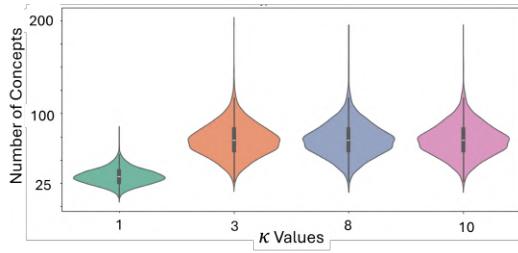


Figure 9: Impact of top- κ in number of final concept prototypes, conducted on PANDA. The plots show that number of concepts increases when we increase κ , allowing more preservation capacity. After a certain point, the condensation process saturates this count to a stable value range.

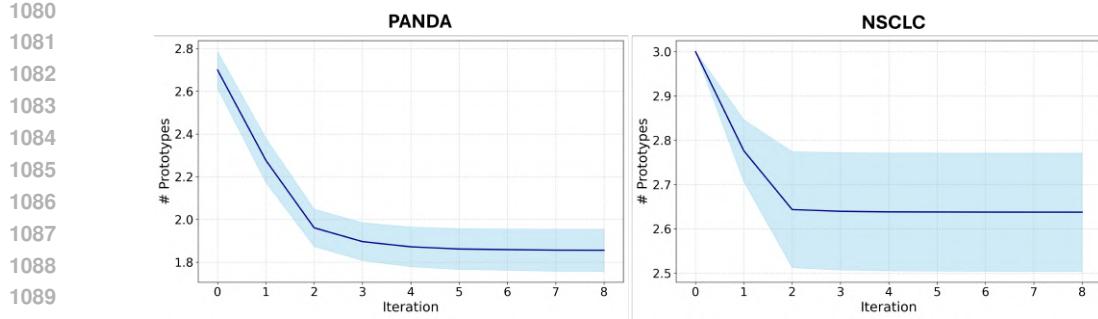


Figure 10: Number of prototypes tracked over the first 8 iterations on PANDA and NSCLC dataset. The values are reported at log10 scale

Table 7: Ablation study of top- κ on downstream tasks performance across different predictor architectures. The experiments are conducted on PANDA dataset.

Predictor	ABMIL			DSMIL			ILRA		
	Top-K	Kappa	Acc.	F1	Kappa	Acc.	F1	Kappa	Acc.
1	90.18	70.02	69.87	85.42	64.65	65.95	89.17	64.88	64.24
3	91.48	74.94	75.00	88.73	68.46	67.52	92.25	73.83	74.55
5	90.47	72.93	72.72	88.80	68.46	66.66	88.66	71.36	71.47
10	89.52	70.47	70.49	85.42	66.89	66.94	88.88	68.90	68.92

F.2 SENSITIVITY ANALYSIS

Extensive Results on Impacts of Number of Initial Prototypes (M). Table 6 complements the analysis in Figure 4 by evaluating how the initial number of patterns influences NICER’s performance across different predictor architectures. The results demonstrate that increasing the pattern set size consistently improves performance, as a larger pool enhances the model’s capacity to distill informative representations. However, these gains diminish once the pattern set grows beyond a certain threshold, with performance gradually reaching a plateau. This saturation reflects the role of the condensation process, which systematically eliminates redundancy and stabilizes the number of effective representations required to characterize a WSI, regardless of the starting pattern count, aligning our insights discussed from Figure 4.

Extensive Results on Impacts of Top- κ during Pattern Exploration. We further conduct a sensitivity analysis on the number of pattern assigned to each patch feature, controlled by the top- κ selection in the Condensation Ability tasks. The results, reported in Tab. 7, reveal a clear trend: model performance improves steadily across all evaluation metrics as κ increases, indicating that incorporating multiple patterns per patch allows richer information to be preserved. However, beyond a moderate value (e.g., $\kappa = 3$), this benefit begins to diminish, with performance gains plateauing or even slightly decreasing. The degradation at larger κ arises because information from each patch becomes overly dispersed across many patterns, weakening the sharpness of the distilled representation. These observations corroborate our earlier findings in Fig. 4 and Tab. 4, and reinforce the design principle outlined in Sec. 2.3 that pattern assignments must balance informativeness with compactness.

Extensive Results on Effectiveness of Condensation Stage. Fig. 9 shows the distribution of the number of final concepts produced by NICER across different κ values. As expected, the average number of prototypes increases with larger κ , since greater capacity enables the model to capture more diverse patterns from complex WSIs. Beyond a certain point, however, this growth plateaus, indicating that the condensation process has effectively merged overlapping patterns and compressed them into a compact, stable concept set. This stabilization highlights NICER’s ability to balance capacity with redundancy removal, yielding a consistent number of meaningful prototypes.

G LIMITATIONS

Despite NICER’s strengths in handling heterogeneous WSI complexity under limited-resource settings, there remain several avenues for improvement that we plan to explore in future work. First, NICER relies on a bag-of-features paradigm, where patches are treated independently and fine-grained spatial or multi-scale context is ignored. This prevents explicit modeling of tissue architecture and spatial priors—an important direction for future work, particularly in clinically critical settings. Second, prototype interpretability remains underexplored. While NICER’s nonparametric concepts improve efficiency and performance, their clinical meaning and uncertainty calibration have not been systematically assessed. Bridging these gaps is essential for making condensed representations both effective and trustworthy in medical applications.

1134
1135

1136

1137

1138

1139

1140

1141

1142

1143

1144

1145

1146

1147

1148

1149

1150

1151

1152

1153

1154

1155

1156

1157

1158

1159

1160

1161

1162

1163

1164

1165

1166

1167

1168

1169

1170

1171

1172

1173

1174

1175

1176

1177

1178

1179

1180

1181

1182

1183

1184

1185

1186

1187