Crystallization Signatures as Predictive Biomarkers in Histopathology

C. Dumitrescu, D. Hackel, I. Obeid and J. Picone

Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA
{claudia.dumitrescu, dmitry.hackel, iobeid, picone}@temple.edu

***Abstract*— Crystallization processes in soft tissues have long been analyzed in relation to pathology and have been proven to positively correlate with precursory pathological cell activity. Accurate and early detection of breast cancer remains a daunting challenge in artificial intelligence, in part, due to the prevalence of overdiagnosis and a lack of transparency in experimental results. Among histopathological specimens, tissue calcifications, ranging from dystrophic hydroxyapatite deposits to idiopathic oxalate crystals and psammoma bodies, have long held diagnostic promise but remain underexploited by visual models. In this paper, we have introduced a novel crystallization-focused approach to multi-classification tasks derived from the Fox Chase Cancer Center Breast Tissue Corpus (FCBR). We constructed an annotated subset comprising 439 patches categorized into Crystalline Non-Neoplastic (cnno: n = 51), Crystalline Ductal Carcinoma In Situ (cdcs: n = 168), and Crystalline Invasive Ductal Carcinoma (cidc: n = 220). Leveraging this subset, we conducted a baseline experiment comparing a simple Random Forest classifier trained on (1) a standard, non-crystallization dataset (1,850 patches) versus (2) an enriched dataset including the crystallization annotations (2,243 patches). Models were evaluated on held-out FCBR samples (18,224 patches) and externally validated on the Temple University Hospital Digital Pathology (TUDP) Corpus (46,666 patches). Incorporation of crystallization annotations more than doubled the overall accuracy: from 18.4% to 34.6% on FCBR and from 20.7% to 23.5% on TUBR. These improvements held under domain shift, indicating enhanced generalization and statistical significance. Our findings demonstrate that explicit modeling of microcalcification patterns provides biologically meaningful features that strengthen deep learning based breast cancer detection. We propose that further expansion of crystallization annotations, especially for underrepresented non-neoplastic cases, and integration into more complex architectures may drive additional gains in sensitivity and diagnostic granularity.**

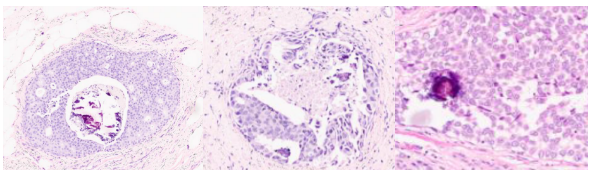
Keywords*— *breast cancer, calcification, histopathology, microcalcification, crystallization, biomarkers

I. Introduction

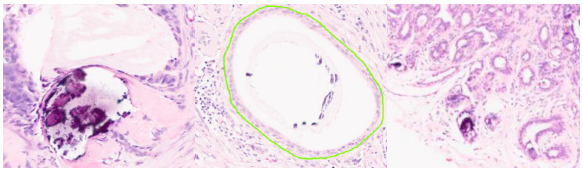
Crystallization in tissue refers to the formation and deposition of microscopic crystals within biological tissue [1]. The process of crystallization is naturally regulated by healthy tissues, such as in bone and teeth. Recent studies [2] highlighted that ectopic calcification is often an active, cell-driven process, contrary to the expectation of a passive precipitation, and that intracellular Ca^{2+} signaling is central to both physiological mineralization and pathological calcification. This shifts the view of calcification from being solely a product of aging or genetics to a process that is potentially preventable, monitorable, and treatable. A detailed understanding of its underlying mechanisms is therefore essential for clarifying its role in pathophysiology.

In breast cancer, calcifications arise through active processes promoted by the presence of cancer that mirror aspects of human physiology [3]. Some examples are shown in Figure 1. Cancer cells and adjacent stromal tissue can undergo osteogenic-like reprogramming, characterized by upregulation of transcription factors such as Runx2 and Osterix. This leads to the expression of bone-associated proteins including alkaline phosphatase and osteocalcin. These molecules promote the deposition of calcium phosphate, predominantly hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$), and calcium oxalate (CaC_2O_4), which serve as the fundamental components of calcification.

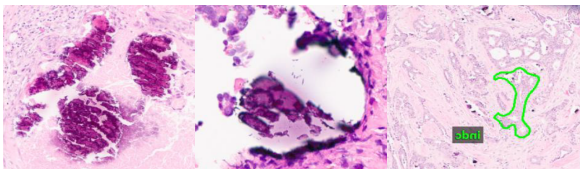
Concurrently, necrosis, apoptotic bodies, and matrix vesicles released by tumor cells generate localized niches enriched with calcium- and phosphate-binding proteins, while showing reduced concentrations of calcification inhibitors such as matrix Gla protein and fetuin-A. This



(a) samples of calcification associated with dcis



(b) samples of calcification associated with nneo



(c) samples of calcification associated with indc

Figure 1. Examples of calcification in a breast tissue image

breast cancer–derived microenvironment fosters hydroxyapatite and calcium oxalate crystal growth, rendering calcification a defining non-neoplastic hallmark of the disease. Nevertheless, mechanistic understanding remains incomplete [2], with particularly limited exploration in the context of computer vision-based analysis.

To this day, accurate and early detection of cancer remains a profound challenge in modern oncology [4]. Studies have shown that routine yearly imaging for people without symptoms usually does not lower the death rate when compared to routine check-ups. Instead, those scans often find numerous extra lesions that may not be carcinogenic. For example, in breast cancer screening, about 22% of detected cases are considered "over-diagnosed" [4]. One of the most dependable clues doctors look for in breast tissue is calcifications since these have long been recognized as an important diagnostic sign. But even though calcifications clearly play a key role in diagnosis, we still haven't fully utilized its potential with modern artificial intelligence (AI). In this study, we used the signatures of mineral deposits as morphological biomarkers to explore the influence of crystallization in cancerous tissue detection.

II. Pathological Crystallization

Pathological crystallization [5] in human tissues encompasses a variety of processes by which calcium salts and, in some cases, ionic structures are deposited in soft tissues under pathological conditions, such as necrosis, chronic inflammation, or apoptosis through apoptotic bodies [6]. Some examples are shown in Figure 1. Chemically, soft-tissue mineralization segregates into two primary crystal classes: Type I (calcium oxalate) and Type II (calcium phosphate/ hydroxyapatite).

Recent investigations [7] have increasingly suggested that specific mineral compounds may be linked to the development and progression of cancer. In the case of breast lesions, for example, work on ductal carcinoma in situ (dcis) has reported that hydroxyapatite deposits occur more frequently than in benign tissue, display lower levels of carbonate substitution, and contain higher concentrations of magnesium whitlockite, with these features aligning with both lesion grade and biological aggressiveness. Yet, the relationship is not entirely consistent. Shin et al. [3] observed findings that contradicted earlier associations of whitlockite with malignancy, while other studies [7] have more often connected this mineral with non-malignant processes. Nevertheless, it has been shown [2] that X-ray diffraction has demonstrated potential clinical utility, with measurements of carbonate substitution achieved alone a sensitivity of 85% and specificity of 88% in distinguishing benign and neoplastic cases using the average carbonate content alone.

In mammography, the most common type of crystallization is dystrophic calcification, independent of systemic calcium levels; as membrane integrity is lost, intracellular calcium floods the extracellular space precipitates with phosphate to form hydroxyapatite crystals within mitochondria and the surrounding matrix [9]. In contrast, metastatic calcification results from elevated serum calcium or phosphate, often due to hyperparathyroidism or chronic renal failure, leading to diffuse crystal deposition in otherwise healthy tissues. While metastatic calcifications typically affect the kidneys, lungs, and gastric mucosa, rare cases of breast parenchymal involvement have been reported in long-term hemodialysis patients and those with secondary hyperparathyroidism [10].

Secretory calcification, or idiopathic crystal deposition within glandular lumina, produces calcium oxalate (weddellite) crystals (Type I calcifications) most often in apocrine cysts of the breast. These birefringent, concentric crystals form within secretions and are typically benign. Their presence can occasionally coincide with proliferative lesions such as lobular carcinoma in situ [5].

Certain neoplasms generate psammoma bodies, concentric, lamellated calcium spherules thought to arise via dystrophic mechanisms within papillary tumor structures. In invasive micropapillary carcinoma of the breast, psammoma bodies have been observed in up to 64% of cases, reflecting localized cell death and mineral nucleation within micropapillary clusters [11].

At the extreme end of the spectrum, tumoral calcinosis and calcinosis cutis represent massive calcium phosphate accumulation, but the frequency is relatively low, especially in breast tissue. Another example of such rare formations is heterotopic ossification, representing true lamellar bone formation, with organized cortical and trabecular architecture, within soft tissues following trauma, surgery, or neurologic injury. This process, in which mesenchymal cells differentiate into osteoblasts outside the skeleton, can occur in muscle and periarticular tissues and is distinct from amorphous calcific deposits [12].

In this paper, we focus on the distinctive calcification patterns in breast tissue, exploring how these diverse pathological crystalline deposits can be quantitatively characterized and integrated into AI-driven models for improved cancer detection. We hypothesize that calcifications may serve as discriminative morphological markers, offering an opportunity for deep learning models to enhance the diagnosis and overall efficiency of cancer detection.

III. FCCC Breast Crystallization subset

The Neural Engineering Data Consortium (NEDC) has released two significant open-source annotated digital pathology datasets related to breast tissue [13]: the Temple University Hospital Digital Pathology Corpus Breast Tissue Subset (TUBR) and the Fox Chase Cancer Center (FCCC) Digital Pathology Corpus Breast Tissue Subsect (FCBR). A summary of the labels used in these corpora is given in Table 1. The process of annotating these corpora is described in detail in [13] and serves as a basis for this work.

Since FCCC specializes in cancer treatment, the FCBR data represents a large sample of the most common and most dangerous types of breast cancer – invasive ductal carcinoma. The FCBR contains 12,164 non-cancerous, 1,967 carcinogenic, and 5,954 cancerous identified structures. This corpus, heavily weighted towards malignant pathology, was selected as the basis for constructing a new subset focused specifically on tissue crystallization phenomena. To our knowledge, no other open-source subset exists that isolates and annotates crystallization patterns in breast histopathology.

We augmented the original FCCC dataset with an additional 440 detailed crystallization annotations. These annotations were developed to test our hypothesis that including calcification in our models would advance state of the art. We will refer to this subset as the Fox Chase Crystallization Corpus (FCCR). To facilitate structured analysis, we organized the crystallization annotations into three distinct classes, based on visual and spatial characteristics of the calcific deposits: Crystalline Non-Neoplastic (cnno: n = 51), Crystalline Ductal Carcinoma in Situ (cdcs) (n = 168), and Crystalline Invasive Ductal Carcinoma (cidc: n = 220). This categorization was informed by observed variations in morphology, localization, and spatial distribution of the calcific material, each correlating with distinct diagnostic.

Table 1. Labels used for annotation of TUBR and FCBR

Label	Description / Features
Normal (norm)	normal ducts and lobules
Ductal Carcinoma in Situ (dcis)	ductal carcinoma in situ, and lobular carcinoma in situ
Invasive Ductal Carcinoma (indc)	invasive ductal carcinoma, invasive lobular carcinoma, and invasive mammary carcinoma
Non-Neoplastic (nneo)	fibrosis, hyperplasia, intraductal papilloma, adenosis, ectasia, etc.
Inflammation (infl)	areas of inflammation
Artifact (artf)	grease pen marks, stitches, foreign bodies, etc.
Indistinguishable (null)	indistinguishable tissue, normally due to issues with the cut/stain
Suspected (susp)	regions that are at risk of developing into cancerous regions
Background (bckg)	stroma, no ducts or lobules

Importantly, we noted that the crystallization profiles differ subtly and slightly consistently across these pathological classes. For example, cnno structures tend to present with discrete, loosely clustered calcific fragments typically isolated within fibroglandular stroma, whereas cdcs annotations display larger, denser, and more irregular mineral formations situated within atypical ductal epithelial arrangements. In cidc cases, the calcifications are often embedded within chaotic, infiltrative patterns consistent with stromal invasion, some even having a scattered aspect. These findings point to potential microstructural markers that may assist in differentiating borderline or ambiguous lesions.

IV. Experimental Results

To evaluate the discriminative power of calcification patterns in histopathological classification, we conducted a baseline experiment to determine whether the inclusion of detailed annotations for crystalline structures improved the predictive accuracy of a deep learning model trained for multi-class classification of breast tissue pathology. The results of these experiments are shown in Table 2 through Table 5.

All experiments were conducted in a closed-set scenario using a fixed random seed. The data was extracted by converting the annotated image to a black and white image, resizing the image to a 256 x 256 px matrix, then flattening it and applying it to a balanced Random Forest (RNF) model. This RNF model was trained on two separate configurations: (1) a non-crystallization dataset consisting of 1,850 patches, and (2) a crystallization-enriched dataset containing 2,243 patches. The latter included an additional 440 manually annotated samples of cnno, cdcs, and cidc classes, mapped as nneo, dcis and indc. The models were evaluated on the remaining portion of FCBR (comprising 18,224 samples) and further validated on TUBR (comprising 46,666 samples). The latter constitutes true open set testing since no TUBR data was annotated.

We used an RNF model because of its simplicity, speed and stability. We have calibrated performance of several deep learning models on this data in [13]. Performance is generally consistent – improvements with one classifier tend to hold up across other classifiers.

Table 2. Confusion matrix and accuracy for FCBR

Fox Chase Cancer Center Breast Tissue Subset (FCBR)		
Metric	**Crystal**	**Non-Crystal**
Correct	6,308	3,345
Incorrect	11,916	14,879
Accuracy	34.6%	18.4%

Table 3. Comparison of the suspicious classes for FCBR

Fox Chase Cancer Center Breast Tissue Subset (FCBR)		
Class	**Accuracy (Crystal)**	**Accuracy (Non-Crystal)**
indc	19.1%	5.0%
dcis	17.2%	8.5%
nneo	7.5%	2.7%

Table 4. Confusion matrix and accuracy for TUBR

Temple University Digital Pathology Breast Tissue Subset (TUBR)		
Metric	**Crystal**	**Non-Crystal**
Correct	10,968	9,681
Incorrect	35,698	36,985
Accuracy	23.5%	20.7%

Table 5. Comparison of the suspicious classes for the TUBR

Temple University Digital Pathology Breast Tissue Subset (TUBR)		
Class	**Accuracy (Crystal)**	**Accuracy (Non-Crystal)**
indc	17.95%	4.97%
dcis	20.68%	6.10%
nneo	4.40%	2.72%

The inclusion of crystallization annotations improved the overall classification accuracy by more than doubling the no crystallization performance. In Table 2, we demonstrate significant gains in categories directly associated with calcification. These three classes are extremely difficult to classify using standard approaches and dominate the overall error rate. The most notable improvement was observed in the indc and dcis classes, indicating that annotated microcalcification structures enabled the model to better distinguish between early-stage and invasive carcinoma phenotypes.

In Table 3, we show the improvement in accuracy for each class. We see a significant across the board improvement in our ability to detect each of the three classes of interest. Each class experienced a three-fold improvement in accuracy.

The gain in accuracy is less pronounced for TUBR. This is not surprising since TUBR contains a much wider range of morphologies. While the overall accuracy in Table 4 increased by 20%, and the per class accuracies increased significantly as shown in Table 5, other challenges in TUBR masked the overall improvements. Given the domain variability in the TUH dataset, even modest improvements are indicative of a model better attuned to the signal.

Results for both experiments demonstrate that the crystallization model exhibits better calibrated behavior, with fewer high-magnitude misclassifications across dominant classes like indc and dcis. In contrast, the non-crystallization model shows heavy confusion between malignant and background classes, likely due to missing structural cues associated with mineralization.

V. Summary

Our results suggest that the inclusion of annotated crystalline structures enhances a model's ability to detect biologically significant pathology, particularly across classes that exhibit microcalcification as a morphological hallmark. In malignant conditions (dcis and indc), the localization, density, and pattern of calcification can serve as distinct morphological signatures that complement cellular and stromal features. Our findings support the hypothesis that explicitly modeling these patterns enables a more nuanced understanding of breast cancer pathology and strengthens the diagnostic of deep learning models. By capturing subtle calcific patterns linked to tumor biology, crystallization provides a biologically meaningful feature set that enhances both sensitivity and generalizability across datasets.

The substantial improvement in indc performance may be explained by the heterogeneous and often infiltrative distribution of calcifications in invasive cancers, which, when annotated, offer additional spatial cues to the model. Similarly, in pre-invasive lesions like dcis or nneo, smaller or more localized calcifications serve as early warning markers, something poorly captured in models trained on general-purpose datasets with coarse labels. Expanding the cnno annotation set may further improve performance on the nneo class, which remains comparatively underrepresented, due to its higher prevalence in the TUBR dataset.

The curated crystal subset, FCCR, provides a resource for future computational pathology studies exploring mineralization patterns as predictive or descriptive features of malignancy. The data is publicly available from our consortium website (*www.necdata.org*).

Acknowledgments

This material is based on work supported by several organizations over the years including the National Science Foundation (grants nos. 1925494 and 1726188), the Temple University Catalytic Collaborative Funding Initiative and most recently by the Pennsylvania Breast

Cancer Coalition Breast and Cervical Cancer Research Initiative. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these organizations.

REFERENCES

[1] D. Proudfoot, "Calcium Signaling and Tissue Calcification.," Cold Spring Harb Perspect Biol, vol. 11, no. 10, Oct. 2019, doi: *10.1101/cshperspect.a035303*.

[2] A. Ali, Y. W. Chiang, and R. M. Santos, "X-ray Diffraction Techniques for Mineral Characterization: A Review for Engineers of the Fundamentals, Applications, and Research Directions," *Minerals*, vol. 12, no. 2, 2022. doi: *10.3390/min12020205*.

[3] K. S. Shin, M. Laohajaratsang, S. Men, B. Figueroa, S. M. Dintzis, and D. Fu, "Quantitative chemical imaging of breast calcifications in association with neoplastic processes," *Theranostics*, vol. 10, pp. 5865–5878, 2020. doi: *10.7150/thno.43325*.

[4] A. B. Miller, C. Wall, C. J. Baines, P. Sun, T. To, and S. A. Narod, "Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial," *BMJ*, vol. 348, p. 10, 2014, doi: *10.1136/bmj.g366*.

[5] J. E. Gonzalez, R. Caldwell, and J. Valaitis, "Calcium Oxalate Crystals in the Breast: Pathology and Significance," *The American Journal of Surgical Pathology*, vol. 15, no. 6, 1991, [Online]. url: *https://journals.lww.com/ajsp/fulltext/1991/06000/calcium_oxalate_crystals_in_the_breast__pathology.7.aspx*.

[6] F. Boraldi, F. D. Lofaro, and D. Quaglino, "Apoptosis in the Extraosseous Calcification Process," *Cells*, vol. 10, no. 1, p. 29, 2021. doi: *10.3390/cells10010131.*

[7] C. Morasso *et al.*, "Whitlockite has a characteristic distribution in mammary microcalcifications and it is not associated with breast cancer," *Cancer Communications*, vol. 43, no. 10, pp. 1169–1173, 2023. doi: *https://doi.org/10.1002/cac2.12481*.

[8] N. Vasei, A. Shishegar, F. Ghalkhani, and M. Darvishi, "Fat necrosis in the Breast: A systematic review of clinical," *Lipids in Health and Disease*, vol. 18, no. 1, p. 139, Jun. 2019, doi: *10.1186/s12944-019-1078-4*.

[9] L. B. Resnikoff, E. B. Mendelson, C. E. Tobin, and T. M. Hendrix, "Breast imaging case of the day. Metastatic calcification in the breast from secondary hyperparathyroidism induced by chronic renal failure.," *RadioGraphics*, vol. 16, no. 6, pp. 1512–1513, 1996. doi: *10.1148/radiographics.16.6.8946553*.

[10] G. Pettinato, C. J. Manivel, L. Panico, L. Sparano, and G. Petrella, "Invasive Micropapillary Carcinoma of the Breast: Clinicopathologic Study of 62 Cases of a Poorly Recognized Variant With Highly Aggressive Behavior," *American Journal of Clinical Pathology*, vol. 121, no. 6, pp. 857–866, Jan. 2004, doi: *10.1309/XTJ7VHB49UD78X60*.

[11] B. Mujtaba *et al.*, "Heterotopic ossification: radiological and pathological review," *Radiology and Oncology*, vol. 53, no. 3, pp. 275–284, 2019. doi: *10.2478/raon-2019-0039*.

[12] D. Hackel et al., "Enabling Microsegmentation: Digital Pathology Corpora for Advanced Model Development," in *Signal Processing in Medicine and Biology: Applications of Artificial Intelligence in Medicine and Biology*, vol. 1, New York City, New York, USA: Springer, 2026, p. 50. [Online]. Available: *https://isip.piconepress.com/publications/book_sections/2026/springer/dpath/* (in publication).

[13] S. S. Shalamzari *et al.*, "Big Data Resources for Digital Pathology," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, Philadelphia, Pennsylvania, USA: IEEE, 2023, pp. 1–19. doi: *10.1109/SPMB59478.2023.10372721*.