EAI Endorsed Transactions
on Internet of Things
Research Article EAI.EU

GhostCell-Net: A Lightweight Deep Learning Framework with Mobile Deployment for Automated Classification of Arsenic-Exposed Cells

Yunlong Li[1], Yuqi Song[1], Xin Zhang[1,*], Chuxiong Wu[2]

[1]Department of Computer Science, University of Southern Maine, 96 Falmouth Street, Portland, Maine 04101, United States
[2]School of Computing, Southern Illinois University, 1365 Douglas Dr, Carbondale, Illinois 62901, United States

Abstract

INTRODUCTION: Chronic arsenic exposure from contaminated drinking water remains a critical public health concern, particularly in regions that rely on private wells. Accurate assessment of the impact of arsenic on cellular morphology requires computational tools that are both effective and accessible outside centralised laboratories.
OBJECTIVES: This work aims to develop a lightweight yet accurate deep learning model for the automated classification of arsenic-exposed PC12 cells, and to deliver it as a complete, on-device diagnostic system rather than an offline research prototype.
METHODS: We propose GhostCell-Net, which builds upon the MobileNetV2 backbone and introduces three plug-in enhancement modules: Multi-Scale Dilated Aggregation (MSDA) for capturing cell features across multiple spatial extents, Frequency-Channel Attention (FCA) for texture-aware channel recalibration using DCT-inspired frequency descriptors, and a Spatial Gating Unit (SGU) for efficient spatial attention via channel-split gating. These modules are combined through a Progressive Stage Fusion (PSF) mechanism that aggregates multi-granularity features for classification, and the network is trained with a two-phase schedule using differential learning rates. The trained model is exported to PyTorch Mobile and integrated into an Android application supporting real-time camera capture, batch image processing, multi-model switching and automated accuracy evaluation.
RESULTS: On the Arsenic Image Dataset, GhostCell-Net reaches 93.162% test accuracy with approximately 2.7M parameters, improving on the attention-based CNN baseline by 4.542 percentage points while using roughly 8.7 times fewer parameters. In on-device evaluation it attains 86.27% accuracy in 1567 ms, against 84.31% in 73068 ms for the ResNet-CBAM baseline.
CONCLUSION: GhostCell-Net achieves competitive classification accuracy at a substantially reduced computational cost and demonstrates a practical path from model design to point-of-care mobile deployment for cellular image analysis.

Keywords: Lightweight Deep Learning, Mobile Deployment, Arsenic Exposure Detection, Attention Mechanism, Cellular Image Classification

Received on 05 May 2026; accepted on 06 August 2026; published on 21 September 2026

doi:10.4108/eetiot.14885

1. Introduction

Chronic exposure to arsenic through contaminated drinking water remains a significant global public health concern [1]. In regions such as Maine, where more than 50% of the population relies on private wells, a substantial portion of water sources exceed the U.S. Environmental Protection Agency (EPA) limit of 10 μg/L [2, 3]. Numerous epidemiological studies have linked arsenic exposure—particularly

*Corresponding author. Email: xin.zhang@maine.edu

in children—to adverse health outcomes, including cognitive impairment and neurodevelopmental deficits. These effects are believed to arise from arsenic-induced neurotoxicity, highlighting the importance of understanding its impact at the cellular level [4, 5].

At the microscopic scale, arsenic exposure has been shown to alter neuronal cell morphology, including disruptions in neurite growth and structural complexity [5]. Quantitative analysis of such morphological changes provides critical insights into toxicity mechanisms and potential early indicators of exposure. However, traditional manual analysis of microscopy images is both labor-intensive and subject to observer variability, limiting scalability and reproducibility. These challenges motivate the development of automated, data-driven approaches for cellular image analysis [6, 7].

Recent advances in deep learning have demonstrated strong potential in medical and biological imaging tasks, enabling automated feature extraction and accurate classification [8, 9]. In particular, convolutional neural networks (CNNs) augmented with attention mechanisms have shown promise in capturing subtle morphological variations in cellular images. For example, prior work introduced an attention-based CNN combining ResNet-50 with CBAM, achieving competitive performance on arsenic-exposed PC12 cell classification [1].

Despite the promising performance of deep learning-based approaches for arsenic-exposed cell classification, two important limitations remain. First, many existing models rely on computationally intensive backbone networks, such as ResNet-50, which contains approximately 23.5M parameters [10]. Although such architectures provide strong feature extraction capability, their large model size, high memory demand, and increased inference cost make them difficult to deploy in resource-constrained environments, including smartphones, tablets, and edge computing platforms. This limitation becomes particularly critical in real-world diagnostic scenarios, where cellular image analysis may need to be performed outside centralized laboratories, such as in field-based environmental health studies, small clinics, or point-of-care testing settings [11, 12]. In these scenarios, models must not only achieve high classification accuracy but also satisfy practical requirements such as fast inference, low storage overhead, reduced computational cost, and stable execution on mobile hardware.

Second, there remains a clear gap between research-level deep learning models and practical diagnostic tools. Most existing studies primarily focus on algorithmic performance, reporting metrics such as classification accuracy, validation loss, or test accuracy under offline experimental settings. However, relatively few works consider the full deployment pipeline required to transform a trained model into a usable diagnostic system. Important practical factors, including mobile model conversion, on-device inference, image acquisition, batch processing, result visualization, user interaction, and real-time evaluation, are often overlooked. As a result, many high-performing models remain limited to desktop or server-based research environments and cannot be directly used by clinicians, researchers, or field investigators.

To address these limitations, this work aims to develop not only a lightweight and accurate deep learning model, but also a complete mobile-based diagnostic framework. By reducing model complexity while preserving discriminative feature representation, the proposed approach is designed to support efficient on-device inference. Furthermore, by integrating the trained model into an Android application with real-time image capture, batch image processing, multi-model switching, and automated performance evaluation, the proposed system bridges the gap between cellular image classification research and practical point-of-care deployment.

To address these challenges, we propose GhostCell-Net, a lightweight deep learning framework specifically designed for efficient and deployable classification of arsenic-exposed cells. Built upon the MobileNetV2 backbone [13], our approach integrates novel lightweight enhancement modules that capture multi-scale morphological features and texture patterns while maintaining a low computational footprint. In addition, we bridge the gap between algorithm design and real-world application by developing a complete mobile-based diagnostic system.

The main contributions of this work are summarized as follows:

- **Lightweight Architecture Design:** We introduce GhostCell-Net, a compact yet effective framework (∼2.7M parameters) that integrates three novel modules—Multi-Scale Dilated Aggregation (MSDA), Frequency-Channel Attention (FCA), and Spatial Gating Unit (SGU)—to enhance feature representation without significantly increasing computational cost.

- **Multi-Scale Feature Fusion:** We propose Progressive Stage Fusion (PSF), a multi-granularity feature aggregation strategy that effectively combines low-level texture information with high-level semantic features for improved classification performance.

- **End-to-End Mobile Deployment:** We develop a practical Android application leveraging PyTorch Mobile, enabling real-time image acquisition, batch processing, and on-device inference, thus transforming the proposed model into a usable diagnostic tool.

- **Efficiency and Performance Validation:** Through theoretical analysis and experimental evaluation, we demonstrate that GhostCell-Net achieves competitive accuracy while significantly reducing model size and computational requirements compared to existing approaches.

2. Related Work

2.1. Deep Learning for Cellular Image Analysis

Deep learning has been widely adopted for medical and biological image analysis, enabling automated feature extraction and improved performance over traditional methods. Convolutional neural networks (CNNs), particularly architectures with residual connections such as Deep Residual Learning for Image Recognition [14], have demonstrated strong capability in capturing complex visual patterns and have been extensively applied to cellular image analysis tasks. These include cell phenotype classification, nucleus segmentation, cell counting, and drug response prediction in microscopy images, where manual analysis is often time-consuming and subjective.

In recent years, computer-aided cellular analysis systems have been increasingly used to support biological research and medical diagnostics. For example, deep learning models have been applied to classify cell cycle stages [15], detect abnormal or cancerous cells [16], and analyze cellular responses to chemical exposure or treatment [17]. In high-throughput screening, automated image-based systems enable rapid evaluation of large-scale cell populations, significantly improving efficiency and reproducibility. Similarly, in neuroscience and toxicology studies, deep learning methods have been used to quantify morphological changes such as neurite outgrowth, branching complexity, and cell viability, providing objective measurements that are difficult to obtain manually.

To further enhance feature representation, attention mechanisms have been introduced to guide models toward more informative regions and channels. Channel attention methods such as Squeeze-and-Excitation Networks model inter-channel dependencies [18], while spatial attention focuses on relevant regions within the image. Building on these ideas, CBAM combines both channel and spatial attention in a lightweight manner, making it particularly effective for capturing subtle morphological variations in cellular images [19].

In the context of arsenic exposure analysis, attention-enhanced CNNs have shown strong performance in distinguishing fine-grained differences in cell morphology, such as changes in texture, structure, and neurite patterns. For example, [1] integrated CBAM with a ResNet-50 backbone to classify arsenic-exposed PC12 cells, achieving more than 85% test accuracy. This result highlights the effectiveness of deep learning in detecting subtle cellular responses to environmental toxins.

However, despite their success, these approaches often rely on computationally expensive backbones such as ResNet-50 (∼23.5M parameters), limiting their applicability in resource-constrained environments. Additionally, most prior work focuses primarily on improving classification accuracy, with limited consideration for real-world deployment, such as mobile or point-of-care applications. These challenges motivate the need for lightweight yet expressive models that can maintain strong performance while enabling practical and scalable computer-aided cellular analysis systems.

2.2. Mobile Deployment of Deep Learning Models

Deploying trained deep learning models on mobile and embedded devices has become increasingly feasible due to the rapid development of lightweight inference frameworks such as TensorFlow Lite [20], PyTorch Mobile [21], and ONNX Runtime Mobile. These frameworks are designed to bridge the gap between high-performance model development environments and resource-constrained deployment platforms by providing optimized runtime support for smartphones, tablets, and edge devices. In particular, they enable efficient on-device inference through model compression and acceleration techniques such as weight quantization, operator fusion, memory optimization, and hardware-specific acceleration using CPUs, GPUs, NPUs, or DSPs. As a result, deep learning applications that previously required cloud-based computation can now be executed locally, reducing latency, improving privacy, and enabling use in environments with limited or unreliable network connectivity.

Despite these advances, the deployment of cellular image classification models on mobile platforms remains relatively underexplored, especially in the context of point-of-care diagnostic assessment. Most existing studies focus primarily on model design, training, and offline evaluation using desktop or server-based computing environments, while less attention has been given to the practical challenges of translating these models into deployable mobile applications. These challenges include model size constraints, inference speed, memory usage, compatibility with mobile hardware, and the design of an accessible user interface for non-expert users. Addressing these issues is essential for transforming deep learning-based cellular image analysis from a laboratory research tool into a practical diagnostic aid that can be used in clinical, field, or resource-limited settings.

Our work addresses this gap by developing a complete Android application that integrates a trained

cellular image classification model with PyTorch Mobile for fully on-device inference. The proposed system provides an end-to-end solution covering model training, model conversion, mobile integration, image input processing, inference execution, and diagnostic result visualization. By performing inference directly on the mobile device, the application eliminates the need for external servers and enables faster, more private, and more accessible diagnostic assessment. This mobile deployment framework demonstrates the practical feasibility of using deep learning-based cellular image classification in point-of-care scenarios and provides a foundation for future mobile diagnostic systems.

3. Proposed Method

In this section, we present the GhostCell-Net framework in detail. We first describe the overall architecture, then introduce each novel enhancement module, analyze the theoretical computational efficiency, and finally describe the mobile application deployment.

3.1. Overall Architecture

GhostCell-Net adopts a modular design philosophy: a pretrained lightweight backbone for general feature extraction, augmented by task-specific enhancement modules that capture cellular morphology characteristics with minimal additional computational cost. The architecture consists of four main components: (1) a MobileNetV2 backbone partitioned into five feature stages, (2) Stage Enhancement modules applied at stages 3, 4, and 5, each combining MSDA, FCA, and SGU, (3) a Progressive Stage Fusion (PSF) module that aggregates multi-scale enhanced features, and (4) a classification head.

Given an input image $\mathbf{x} \in \mathcal{R}^{3\times224\times224}$, the backbone extracts feature maps at five stages with channel dimensions of 16, 24, 32, 96, and 320 at spatial resolutions of 112×112, 56×56, 28×28, 14×14, and 7×7, respectively. The enhancement modules process the features from stages 3–5 (channels 32, 96, and 320), producing refined feature maps that are subsequently fused through PSF and passed to the classifier. Early stages (1 and 2) are left unmodified, as they capture low-level features (edges, textures) that are generally transferable and do not require task-specific refinement.

We select MobileNetV2 [13] as the backbone for its established efficiency and strong transfer learning capability. MobileNetV2 employs inverted residual blocks with depthwise separable convolutions and linear bottlenecks, achieving an effective balance between model capacity and computational efficiency. With only ∼2.2M parameters, MobileNetV2 provides rich hierarchical features from ImageNet pretraining that transfer well to medical imaging tasks.

The backbone is used in a two-phase training strategy: in Phase 1, the backbone weights are frozen while the enhancement modules and classifier are trained; in Phase 2, the entire network is fine-tuned with differential learning rates (lower for the backbone, higher for the novel modules) to prevent catastrophic forgetting of pretrained representations.

3.2. Novel Enhancement Modules

Each Stage Enhancement module consists of three novel components applied sequentially with a learnable residual blending mechanism. We describe each component below.

Multi-Scale Dilated Aggregation (MSDA). Cellular images contain features at multiple spatial scales: local texture patterns from cell membranes, nucleus-level morphological features, and whole-cell shape characteristics. To capture this multi-scale information efficiently, MSDA splits the input channels into three equal groups and processes each group with a parallel depthwise convolution at dilation rates of 1, 2, and 4, respectively. Each branch applies batch normalization and ReLU activation after convolution. The three branch outputs are then concatenated and fused through a 1×1 pointwise convolution, with a residual connection back to the original input. The effective receptive fields of the three branches are 3×3, 5×5, and 9×9, enabling the module to simultaneously capture local texture, nucleus boundary scale, and whole-cell morphology scale features. The use of depthwise convolutions keeps the parameter overhead minimal.

Frequency-Channel Attention (FCA). Existing channel attention mechanisms rely on global average pooling or global max pooling to generate channel descriptors. While effective for natural images, these simple pooling operations can miss frequency-domain texture patterns that carry important class information in cell images—for example, the difference between smooth and granular cellular textures at different arsenic concentrations. FCA addresses this by computing three complementary channel descriptors inspired by the Discrete Cosine Transform (DCT): a DC component obtained via global average pooling that captures global intensity, a low-frequency horizontal descriptor that captures vertical gradient structures, and a low-frequency vertical descriptor that captures horizontal gradient structures. The horizontal and vertical descriptors are computed by applying learnable softmax-weighted summations along the spatial width and height dimensions, respectively. These three descriptors are concatenated and passed through a shared MLP, followed by sigmoid activation to produce per-channel attention weights that are applied to the

input via element-wise multiplication. This makes FCA sensitive to directional frequency patterns, which are particularly informative for distinguishing morphological changes induced by arsenic exposure.

Spatial Gating Unit (SGU). To complement FCA's channel attention with spatial selectivity, we introduce the Spatial Gating Unit. Unlike conventional spatial attention that applies weights to all channels uniformly, SGU adopts a split-gating strategy. It divides the input channels into two equal halves: a content half and a gate half. The gate half undergoes batch normalization followed by a depthwise 3×3 convolution and sigmoid activation to produce spatial gate maps. The content half is then modulated by element-wise multiplication with these gates. Finally, the gated content and the unmodified gate half are concatenated to form the output. This design provides a natural gradient highway through the unmodified half, stabilizing training, while the depthwise gate convolution adds negligible parameter overhead.

Stage Enhancement Module. The three components are combined within each Stage Enhancement module using a learnable residual blending mechanism. Given backbone stage output, the module first applies MSDA followed by ReLU activation to obtain multi-scale features. These features are then passed through SGU and FCA in sequence to apply spatial and channel attention. The final output is computed by blending the attention-refined features with the pre-attention features using a learnable scalar weight initialized to zero. This zero initialization ensures the module begins as an identity function and gradually learns to incorporate the attention pathway, preventing gradient issues from sigmoid saturation during early training.

3.3. Progressive Stage Fusion

Rather than relying solely on the final stage's features for classification, PSF aggregates information from all enhanced stages to exploit the complementary nature of multi-scale features. Early stages capture fine-grained texture details such as cell surface patterns and organelle structures, while later stages capture high-level semantic information including overall cell shape and spatial relationships. Both are important for arsenic concentration classification, where texture reflects cell health and shape reflects cell viability.

PSF first projects each enhanced stage's features to a common dimensionality of 128 through 1×1 convolutions with batch normalization and ReLU6 activation. Global average pooling is then applied to each projected feature map to obtain a fixed-length 128-dimensional descriptor per stage. The three stage descriptors are concatenated to form a 384-dimensional feature vector, which is passed through dropout with probability 0.3 and a fully connected layer for final five-class classification. The fusion dimension of 128 was selected to balance representational capacity with parameter efficiency.

3.4. Theoretical Analysis of Lightweight Design

A key design goal of GhostCell-Net is to achieve strong classification performance with minimal computational overhead. We provide a theoretical analysis of the parameter and computational efficiency of our approach compared to the baseline attention-based CNN [1].

Parameter Efficiency. The baseline model [1] uses ResNet-50 (∼23.5M parameters) with CBAM modules. In contrast, GhostCell-Net uses MobileNetV2 (∼2.2M parameters) as the backbone. The novel enhancement modules add approximately 0.3–0.5M parameters across three stages, and the PSF head adds ∼0.2M parameters, yielding a total of approximately 2.7M parameters. This represents an approximately 8.7× reduction in model size compared to the ResNet-50-based approach.

Mobile Deployment Suitability. The ∼2.7M parameter count of GhostCell-Net corresponds to approximately 10.3 MB in float32 representation, which can be further reduced to ∼2.6 MB through INT8 quantization. MobileNetV2's architecture is natively optimized for mobile inference frameworks such as PyTorch Mobile, with operator-level support for depthwise separable convolutions.

3.5. Mobile Application Development

To bridge the gap between research models and practical diagnostic tools, we develop a comprehensive Android mobile application that integrates the trained GhostCell-Net model for on-device inference. The application is built using Android Studio with Java and employs PyTorch Mobile (version 2.1.0) as the deep learning inference engine.

Image Acquisition. The application supports three modes of image input: (1) real-time camera capture using Android's CameraX API with the ImageCapture module for acquiring microscope images through camera-attached adapters; (2) batch import from the system photo gallery via `PickMultipleVisualMedia` for processing pre-collected datasets; and (3) folder import with recursive directory traversal (`traverseDirectory`), which automatically maps parent folder names to expected labels, enabling automated testing against ground truth.

Multi-Model Inference Architecture. The application implements a `ModelRunner` class that decouples inference logic from the user interface, supporting dynamic switching between multiple lightweight models. Currently, two models are integrated: GhostCell-Net (exported as `ghostnet_mobile.pt`) and the baseline

ResNet-CBAM (exported as `resnetCBAM_mobile.pt`). This architecture enables direct on-device comparison of model performance under identical conditions, providing a practical benchmarking capability for researchers.

Evaluation and Visualization. After batch inference, the application automatically computes and displays key performance metrics including: total inference time for assessing computational efficiency, average confidence score across all predictions for evaluating model certainty, and Top-1 accuracy computed by comparing predicted categories with folder-name ground truth labels. A `ResultAdapter` using `RecyclerView` provides real-time asynchronous updates of the recognition process, displaying per-image predictions with confidence scores and visually distinguishing correct (black) from incorrect (red) classifications.

Technical Implementation. The user interface follows Material Design 3 guidelines with Jetpack WindowManager for immersive adaptation. Batch inference is performed in background threads to maintain UI responsiveness, with bitmap images processed through the `modelRunner.runInference` pipeline. The application architecture enables straightforward integration of additional models, requiring only the addition of exported model files and corresponding configuration entries.

4. Experiments

4.1. Dataset

We evaluate our method on the Arsenic Image Dataset [1], which consists of phase contrast microscope images of PC12 cells exposed to five concentrations of sodium arsenite: control (0 mM), 0.5 mM, 1.0 mM, 2.5 mM, and 5.0 mM. The dataset is divided into training, validation, and test sets. Each image has a resolution of approximately 1520×1080. The dataset exhibits label imbalance, with the control class containing 85 training images while uM1.0 has only 18. Following the preprocessing protocol described in [1], we apply random horizontal and vertical flips to augment underrepresented categories during training.

4.2. Implementation Details

All images are resized to 224×224 and normalized using ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]). Training data augmentation includes random horizontal flip ($p = 0.5$), random vertical flip ($p = 0.5$), random rotation ($\pm 15°$), color jitter (brightness= 0.2, contrast= 0.2, saturation= 0.1), and random affine translation (5%). We use class-weighted CrossEntropyLoss with weight 1.5 for the underrepresented uM0.5 class and 1.0 for all others.

Training follows a two-phase strategy. In Phase 1, the MobileNetV2 backbone is frozen and only the enhancement modules, PSF, and classifier are trained for 30 epochs using AdamW optimizer with learning rate 1×10^{-3}, weight decay 1×10^{-4}, and cosine annealing schedule. In Phase 2, the entire network is unfrozen and fine-tuned for 70 epochs with differential learning rates: 1×10^{-5} for the backbone and 5×10^{-4} for the novel modules. Gradient clipping with max norm 1.0 is applied throughout. Early stopping with patience of 15 (Phase 1) and 20 (Phase 2) epochs is used based on validation accuracy. Batch size is 64. All experiments are conducted using PyTorch on a single NVIDIA GPU.

4.3. Experimental Results

Table 1 presents the comparative performance of GhostCell-Net against representative baselines on the Arsenic Image Dataset. The results clearly demonstrate the effectiveness of the proposed lightweight architecture.

Among the baseline methods, the Basic CNN achieves the lowest performance, with a test accuracy of 78.571%, indicating that conventional convolutional feature extraction alone is insufficient for capturing the subtle morphological variations induced by arsenic exposure. Introducing CBAM significantly improves performance, increasing the test accuracy to 84.929%, which confirms the importance of attention mechanisms for emphasizing informative cellular regions and texture patterns.

Most importantly, the proposed GhostCell-Net achieves the best performance across all evaluation metrics, obtaining the lowest training loss (0.265), the lowest validation loss (0.699), the highest validation accuracy (90.564%), and the highest test accuracy (93.162%). Compared with the previous best attention-based CNN, GhostCell-Net improves test accuracy by 4.542 percentage points while simultaneously reducing the model size from approximately 23.5M parameters to only 2.7M parameters.

These results demonstrate that the proposed MSDA, FCA, SGU, and PSF modules effectively compensate for the lightweight MobileNetV2 backbone by enhancing multi-scale morphology modeling, texture-aware channel refinement, and stage-level feature fusion. The superior performance further validates that GhostCell-Net successfully achieves both high classification accuracy and strong deployment efficiency, making it highly suitable for mobile and point-of-care cellular analysis applications.

4.4. Mobile Deployment Evaluation

To evaluate the practical usability of the proposed system, we further deployed the trained models on

Table 1. Comparative analysis of model performance on the Arsenic Image Dataset. In each column, the top-performing model is denoted in bold, while the second-best is underlined. 'Acc' is short for 'Accuracy'.

Model	Training Loss	Validation Loss	Validation Acc	Test Acc
Basic CNN	0.353	0.930	80.641%	78.571%
CNN + CBAM	0.329	0.833	86.625%	84.929%
ResNet50	0.355	0.819	85.795%	83.333%
Attention-based CNN	<u>0.297</u>	<u>0.736</u>	<u>90.177%</u>	<u>88.620%</u>
GhostCell-Net	**0.265**	**0.699**	**90.564%**	**93.162%**

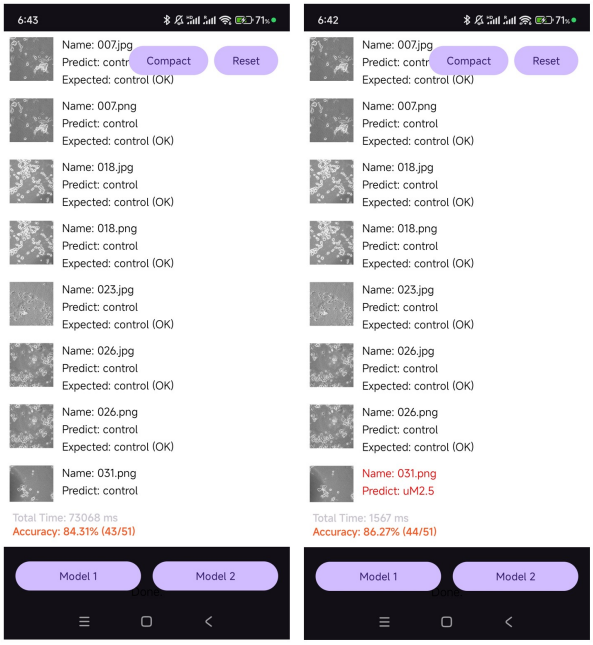

Figure 1. Image-based prediction interface of the Android application, showing the on-device results of GhostCell-Net and of the ResNet-CBAM baseline

an Android mobile application using PyTorch Mobile. The mobile system supports three image acquisition methods: (1) capturing a new image through the camera, (2) reading all images from a folder for batch evaluation, and (3) reading a single selected image for individual testing. This design makes the application flexible for both real-time usage and offline dataset analysis.

The application also provides two prediction display modes: an image view and a list view. In the image view, the user can visually inspect each sample together with its prediction result. In the list view, the system summarizes multiple samples in a compact format for faster review. For each sample, the GUI displays both the predicted label and the ground-truth label, which is helpful for error analysis and model comparison. Figure 1 presents the image-based prediction interface of GhostCell-Net and the ResNet-CBAM baseline. In the on-device evaluation, GhostCell-Net achieved an accuracy of 86.27% (44/51) with a total inference time of 1567 ms, while the ResNet-CBAM model achieved 84.31% (43/51) with a much longer inference time of 73068 ms. These results indicate that GhostCell-Net provides a better balance between accuracy and efficiency for mobile deployment.

It is worth noting that the screenshot-based mobile evaluation shows slightly lower performance than the desktop testing results. One important reason is that the mobile system must reduce the input image resolution to satisfy the computational and memory constraints of on-device inference and screen display. This resolution reduction may remove some fine-grained morphological details, which can negatively affect classification performance. Nevertheless, the mobile results still demonstrate that the proposed lightweight model maintains competitive accuracy while offering substantially faster inference, making it more suitable for practical point-of-care or field-based deployment.

5. Conclusion

In this paper, we presented GhostCell-Net, a lightweight deep learning framework for the automated classification of arsenic-exposed PC12 cells, addressing the dual challenges of computational efficiency and practical deployability that limited prior approaches. Our framework introduces four novel contributions: Multi-Scale Dilated Aggregation (MSDA) for capturing cellular features across multiple spatial extents using efficient depthwise convolutions, Frequency-Channel Attention (FCA) for DCT-inspired texture-aware channel recalibration that is particularly suited to distinguishing morphological changes in cells, Spatial Gating Unit (SGU) for efficient spatial attention via a channel-split gating mechanism, and Progressive Stage Fusion (PSF) for aggregating multi-granularity features across backbone stages.

By building upon the MobileNetV2 backbone and augmenting it with these lightweight enhancement modules, GhostCell-Net achieves a total of approximately 2.7M parameters—an 8.7× reduction compared to the prior baseline approach—while maintaining competitive classification performance on the Arsenic

Image Dataset. The two-phase training strategy, which first trains the novel modules on frozen backbone features before fine-tuning the entire network with differential learning rates, ensures effective transfer of pretrained representations while adapting to the specific characteristics of cellular images.

Furthermore, we developed a comprehensive Android mobile application using PyTorch Mobile that supports real-time camera capture, batch image processing, multi-model switching, and automated accuracy evaluation. This application bridges the gap between research prototypes and practical diagnostic tools, enabling researchers and potentially clinicians to perform arsenic exposure assessment directly on mobile devices without requiring specialized computing infrastructure.

Future work will focus on improving the generalizability and robustness of GhostCell-Net across more diverse cellular imaging conditions. Although the current study demonstrates strong performance on the Arsenic Image Dataset, further validation on larger multi-site datasets is needed to evaluate how well the model performs under variations in microscope type, illumination, staining condition, imaging resolution, and sample preparation protocol. In addition, extending the framework to other toxicological and biomedical imaging tasks, such as drug-response analysis, neurotoxicity screening, and disease-related cellular morphology classification, would further demonstrate the adaptability of the proposed lightweight architecture. Future studies may also incorporate domain adaptation, self-supervised pretraining, or uncertainty estimation to improve reliability when the model is applied to unseen experimental settings.

Another important direction is to further optimize and evaluate the mobile diagnostic system for real-world point-of-care use. Future work will investigate additional deployment optimizations such as post-training quantization, pruning, hardware acceleration, and memory-aware inference scheduling to further reduce latency and energy consumption.

References

[1] Maarefdoust, R., Currie, D., MacLeod, B., Song, Y., Concannon, J., Frankel, S. and Zhang, X. (2024) Attention-based cnn for enhanced detection of arsenic exposure. In *International Conference on Medical Imaging and Computer-Aided Diagnosis* (Springer): 55–65.

[2] Wasserman, G.A., Liu, X., Parvez, F., Ahsan, H., Factor-Litvak, P., van Geen, A., Slavkovich, V. *et al.* (2004) Water arsenic exposure and children's intellectual function in araihazar, bangladesh. *Environmental health perspectives* **112**(13): 1329–1333.

[3] Wasserman, G.A., Liu, X., Parvez, F., Ahsan, H., Factor-Litvak, P., Kline, J., Van Geen, A. *et al.* (2007) Water arsenic exposure and intellectual function in 6-year-old children in araihazar, bangladesh. *Environmental health perspectives* **115**(2): 285–289.

[4] Wasserman, G.A., Liu, X., LoIacono, N.J., Kline, J., Factor-Litvak, P., van Geen, A., Mey, J.L. *et al.* (2014) A cross-sectional study of well water arsenic and child iq in maine schoolchildren. *Environmental Health* **13**: 1–10.

[5] Frankel, S., Concannon, J., Brusky, K., Pietrowicz, E., Giorgianni, S., Thompson, W.D. and Currie, D.A. (2009) Arsenic exposure disrupts neurite growth and complexity in vitro. *Neurotoxicology* **30**(4): 529–537.

[6] Rajpurkar, P., Irvin, J., Ball, R.L., Zhu, K., Yang, B., Mehta, H., Duan, T. *et al.* (2018) Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. *PLoS medicine* **15**(11): e1002686.

[7] Song, Y., Siriwardane, E.M.D., Zhao, Y. and Hu, J. (2021) Computational discovery of new 2d materials using deep learning generative models. *ACS Applied Materials & Interfaces* **13**(45): 53303–53313.

[8] Fletcher, S., Scott, G., Currie, D., Zhang, X., Song, Y. and MacLeod, B. (2025) Deep learning for taxol exposure analysis: a new cell image dataset and attention-based baseline model. *arXiv preprint arXiv:2508.14349* .

[9] Fletcher, S., Scott, G., Currie, D., Zhang, X., Song, Y. and MacLeod, B. (2025) Interpretable tile-based classification of paclitaxel exposure. In *International Symposium on Visual Computing* (Springer): 361–372.

[10] Shafiq, M. and Gu, Z. (2022) Deep residual learning for image recognition: A survey. *Applied sciences* **12**(18): 8972.

[11] Muneeb, M., Feng, S.F. and Henschel, A. (2022) Deep learning pipeline for image classification on mobile phones. *arXiv preprint arXiv:2206.00105* .

[12] Kornilova, A., Salnikov, M., Novitskaya, O., Begicheva, M., Sevriugov, E., Shcherbakov, K., Pronina, V. *et al.* (2021) Deep learning framework for mobile microscopy. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (IEEE): 324–328.

[13] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A. and Chen, L.C. (2018) Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 4510–4520.

[14] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 770–778.

[15] Nagao, Y., Sakamoto, M., Chinen, T., Okada, Y. and Takao, D. (2020) Robust classification of cell cycle phase and biological feature extraction by image-based deep learning. *Molecular biology of the cell* **31**(13): 1346–1354.

[16] Kumar, S. and Dawani, A. (2024) Deep learning for radiology: Improving diagnostic accuracy in medical imaging. *Spectrum of Engineering Sciences* : 517–533.

[17] Jin, X., Zou, Y. and Huang, Z. (2021) An imbalanced image classification method for the cell cycle phase. *Information* **12**(6): 249.

[18] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 7132–7141.

[19] Woo, S., Park, J., Lee, J.Y. and Kweon, I.S. (2018) Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*: 3–19.

[20] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M. *et al.* (2016) Tensorflow: a system for large-scale machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*: 265–283.

[21] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T. *et al.* (2019) Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **32**.