VER
Vascular & Endovascular Review

A Multi-Modal Deep Learning Framework Combining Histopathological Imaging and Gene Expression for Automated Cancer Detection

Kantilal P Rane[1], Prof. (Dr.) Chandra Kumar Dixit[2], Dr. Sreemoy[3]

[1]Postdoc, Lincoln University College (LUC)
kantiprane@gmail.com
[2]Professor, Dr. Shakuntala Misra National Rehabilitation University Mohaan Road, Lucknow, UP, India
ckdixit@dsmnru.ac.in
[3]Professor, Lincoln University College (LUC)
sreemoy@lincoln.edu.my

ABSTRACT

Cancer subtyping diagnosis is essential for personalized medicine and is conventionally facilitated by histopathological imaging OR genetic assessment diagnosis. Such a method of diagnosis compromises biological development as genetic component considerations are neglected OR a histology approach is taken with subjective visual interpretation and a non-confirmatory genetic component. Here, a Multi-Modal Deep Learning Framework is established for image diagnosis via Convolutional Neural Networks (CNN) and gene expression profiles via a Multi-Layer Perceptron (MLP), and Cooperative Multi-Colony Ant Optimization (MCO) attunes the dimensionality complications of both data streams. This framework is the first to engage a Cooperative Multi-Colony effort where heterogeneous colonies operate independently, but collaboratively throughout the optimization process - one colony for spatial feature selection (CNN) and another for genomic feature selection (MLP) - and a third "Master Colony" features the assessment of both to optimally tune the fused layer. Performance results on TCGA-based multimodal datasets (Lung, Breast, and Colorectal) suggest that this cooperative multi-colony operation outperforms the single colony/single modality counterparts for convergence time and classification accuracy.

KEYWORDS: Multi-Colony Optimization, Multi-Modal Learning, CNN, Gene Expression, Feature Fusion, Heterogeneous Ant Colonies.

How to Cite: Kantilal P Rane, Prof. (Dr.) Chandra Kumar Dixit, Dr. Sreemoy, (2025) A Multi-Modal Deep Learning Framework Combining Histopathological Imaging and Gene Expression for Automated Cancer Detection, Vascular and Endovascular Review, Vol.8, No.17s, 358-363.

1. INTRODUCTION

Cancer is a worldwide epidemic and while histopathological determination is the gold standard diagnosis, it only views the malignant phenotype, compromised by no spatial assessment. Conversely, genomic markers like RNA-Seq have no spatial relevance yet overwhelmingly favour molecular data. Thus, with increased frequency and prevalence of data streams generated in paired opportunity, a unique setting emerges to utilized deep learning methods to fuse the data to be mutually beneficial.

However, it's difficult to calibrate multi-modal access as one optimizer may not effectively balance the spatial needs (texture/geometry) versus genomics (sparse/highly-dimensional) which is why a Multi-Colony Optimization (MCO) process is proposed - with multiple ant colonies, each cooperating independently, and with their own heuristic-specific approach to be beneficial in how they can learn about hyper-parameter spaces for both the imaging networks and the genomic networks without one modality overpowering the other.

2. LITERATURE REVIEW

The heterogeneous nature of cancer, in its molecular/genomic signatures, as well as its morphological/imaging manifestations, encourages researchers to explore multi-modal data for diagnosing sub-type characteristics. Unfortunately, single-modality deep-learning models (only radiological images, only histopathological images, only genomic readings) overlook complementary phenomena that could be synergistically achieved through collaboration. More often than not it's becoming the case that multimodal deep learning fusion approaches find signals to synergize from heterogeneous data streams for improved diagnostic/prognostic/sub typical efforts regarding oncological features. For example, A 2023 overview of multimodal deep learning for cancer diagnosis segmentations found many differentiated modalities (genomic, pathological, radiological, clinical) that proved fused options outperformed uni-modal models however [1].

There are two types of multimodal deep learning fusion: early (input level fusion), intermediate (feature level fusion), late (decision level fusion), hybrid methods. A 2021 survey regarding multimodal deep learning outside of medical components found pros and cons for each over specific features and functionality; for example, early fusion is easiest but may not appreciate the depth of intra-modality interaction while more composite fusions may achieve challenging features based on thoughtful consideration but are more difficult to achieve [2]

For multimodal medical considerations, a 2021 survey found that proper multimodal segmentation (MRI + MRI fused, MRI + PET Segmented) allows for greater impact than simple concatenation - and this is only true when done correctly at proper stages - because deeper stages allow learning from modality-specific features before fusing them [3]. This parallels cancer-related multimodal fusions; a 2025 article on multimodal deep learning for cancer data fusion finds that

many still use concatenation with fully connected layers as part of a solo approach due to data availability, interpretive merit, small sample sizes or computational concerns [1]. Such an article raises many secondary issues - missing modalities, modality imbalance, and poorly tuned architectures - resulting in fixed arches that are pragmatically tuned for better optimization.

While multimodal fusion offers a diagnostic advantage over uni-modal baselines, these fusion approaches are limited in their reliance on fixed options. Results indicate that fixed auto and systematic hyper-parameter tuning of modalities AND architectural aspects limits extensive exploration found in other published results.

In addition, dimensionality reduction is often a mandatory endeavour due to high-dimensional biomedical data like gene expression features or radiomics (features) that lead deep learning and classical classification models to reduce selections through dimensionality reduction selection processes that find transformation options and reduce samples through overfitting, noise-reduction and generalizability improvements. It's especially appropriate here in bio-inspired algorithms.

One of the most utilized approaches Ant Colony Optimization (ACO). Ant Colony Optimization has been effective for gene/marker selection purposes with microarray data and feature selection endeavours. For example, ReliefF-based hybrid selection was effective with ACO (RFACO-GS) showing gene selection results for tumour classification success across varying datasets [4]. Another collaborative approach utilized a modified ACO for marker gene selection that was paired with SVM classification revealed improved recognition with less features than other mechanisms [5]. Another hybrid approach utilizing MWIS pre-selection with local search was effective with ACO addition in high-dimensional microarray classification settings across multiple datasets [6].

Thus, ACO-based feature-selection is one reasonable choice for gene expression/tumour classification efforts since dimensionality trumps sample size restrictions.

Given findings for multiple independent features, previous studies have found ACO (and hybrid ACO's) to work successfully across numerous reports - but classical ACO (single colony perspectives) create limitations where applied .to complex, highly-dimensional noise applications:

1. The search can prematurely converge relative to optimal combined subsets; pheromone concentration can exceed potential conclusions if collapse or minimizes diversity - studies into implementing ACO in highly-dimensional fields such as genetics/epistasis discoveries note pheromone feedback concentrations that are too high create local optima with focused attention only to select features nominated by colonies. [7].
2. It's even harder to avoid premature conclusions when the search space explodes with combinatorial effects due to excess/missing features; complex settings like microarray joint implementation that combines gene expression plus imaging and clinical components relies on domain-specific knowledge for successful implementation without focusing on ratios. When studying tumors - each with implications on detection methods - this connects best to expert knowledge utilization without focusing on ratios.

Since basic ACO implementations don't support premature avoidance of conclusions across board when working on cancer detection pipelines where features are selected based on how well the input merits exploration together with relative hyper-parameter tuning operations to find optimal subsetting combined with hyper-parameter tuning goals.

Solutions arise when different variances of ACO create multiple colonies - with heterogeneous roles or parameter settings to supplement retained exploration while simultaneously exploiting through consensus-gathering:

1. For example a multi-colony ACO was promoted with dynamic collaborative mechanism where differentiated colonies are separated from each other but share information without all falling into the same local optimum; better solutions arise within population problems with larger amount of combinatorial issues [8].
2. An entropy-based variant/knowledge-aware ACO adjusted pheromone intake/extraction in a dynamic manner effective within genetic/epistasis search spaces that prevent drastic reductions/evaporations - small feature counts in various domains praised [7].
3. Most recently a multilabel/multioutput enhanced feature selection with reductions for restarts; static heuristics provided an enhanced nuanced approach where ACO's variances benefit from aggregate heuristics for big picture assessment [9].

These studies articulate where the complications can be complicated and resolved through multiple colony heterogenous approaches where different colonies are granted separate parameters or pheromone dynamic changes occur in real time that are accountable to high-dimensional output relevance while working well for noise across exclusion or inclusion for domain-specific relevance.

While studies have found ACO (and hybrid ACO's) along with gene selection and classical classification suggestions across extensive findings - with features available in various theoretically nuanced fields - few studies have suggested multi-colony ACO combined with super high dimensional deep learning cancer pipelines exclusive . Reports about what was found multimodal need cross-section transfer suggest practical union hasn't been scientifically explored on either front for the most compelling reason - within each published concern - from either side. The recent 2025 review of deep-learning based cancer data fusion's big takeaway for multimodal fields is that there's an insufficient investigation into the findings across other fields [1].

Meanwhile research possibilities increase with multimodal cancers but systematic optimization across the board hasn't truly been explored - what modalities make sense with others/how to combine them/find any characteristic across each modality without systematic tuning fusions [1].

Hence, there exists an empirical research gap to optimize heterogeneous-colony ACO (or any metaheuristic approach of this nature) simultaneously across multimodal deep learning architectures for:

1. Inclusion of modalities (modality selection),
2. Features/subsets utilized within modalities (feature selection),
3. Fusion architecture (type + hyper-parameters) and
4. Downstream classification/prognostic model hyper-parameters.

Such a framework for simultaneous optimization would yield better performing, less expensive and more robust multi-modal detection/prognostic models than those generated under a manual tuning approach. Thus, utilizing the following logic, heterogeneous multi-colony ACO is the best metaheuristic approach relative to the multimodal cancer detection space:

1. It maintains diversity + avoids premature convergence relative to any extensive combinatorial search space (e.g., selecting different subsets from larger available options across many modalities) which means ACO's search space does not prematurely assume the different combination of modalities + feature subsets + architectural hyper-parameters since it exists simultaneously with large combinatorial options. This is useful in modality selection.
2. It is flexible enough to accommodate complicated fitness functions (e.g. classification performance (AUC), sensitivity/specificity thresholds, scarcity, inference costs) that align readily with clinical model limitations.
3. It is a wrapper-type solution meaning candidate solutions (where each solution represents a combination of a modality + feature + hyper-parameter) are evaluated through solutions through complete models (deep networks), allowing for real end-to-end optimization.
4. It is parallel colonies/asynchronously evaluated which means the additional resources available through GPU/distributed compute ensure that the computational cost isn't as prohibitive.

Thus, a combined framework of multimodal deep learning + heterogeneous-colony ACO would simultaneously search a large (but highly structured) solution space that would otherwise be too cost prohibitive through manual tuning/grid/random search.

When constructing the multimodal-cancer + multi-colony ACO pipeline, considerations to assess include:

1. **Encoding scheme**: The candidate solution can each be realized through a combined solution (e.g. a binary solution for whether to include a modality (yes/no features of a binary vector for each modality decision through a discrete number for the fusion architecture employed, with numeric for the employed hyper-parameters for those two as well).
2. **Fitness function design**: A compounded fitness function should evaluate classification success (AUC, accuracy + sensitivity/specificity), model complexity (number of parameters selected + number of features) and the incurred inference/time or memory - critically essential for clinical applications.
3. **Colony heterogeneity**: Assign roles to the colonies - e.g. explorers with colonies that rely on high levels of randomness + evaporation (in case gaps were searched thoroughly) vs exploiter colonies that evaluate with low evaporation + higher pheromone reinforcement (where fine-tuning could mean more local searching); an exploration versus exploitation balance.
4. **Diversity maintenance & communication**: The good pheromones can be shared, the elite of one colony can pollinate other colonies, the pheromones can be reset based on percent entropy (and best solutions reintroduced) or colonies can get reinitialized if stagnation occurs.
5. **Scalability & computational tractability**: The evaluations of deep networks as a whole often come at prohibitive costs; surrogate models should be used, along with staged/coarse-to-fine searches (light models perform big picture thoughts, full models can assess fine details), partial training/early stopping reduces costs.
6. **Cross-validation & generalization safeguards**: Biomedical datasets can sometimes be small so nested cross-validation, external validation cohorts and regularization (for penalized overstated models).

Ultimately, therefore: multimodal deep learning solutions have significant promise for cancer detection/subtyping/prognosis through enhanced usable complementary data across the modalities. However, current attempts require substantial manual effort through the fusion architecture and chosen features/hyper-parameters. Concurrently, metaheuristic solutions have emerged successfully through high-dimensionalized biomedical datasets where heterogamous/multi-colony ACO has been utilized in advanced forms to maintain diverse searches through complicated spaces in a learned process. However, the intersection of the two is currently unexplored - no research attempts have considered using heterogeneous/multi-colony ACO for modal selection (and then feature/hyper-parameter selection) within the context of multimodal deep learning cancer models - thus the need to pursue this goal exists.

3. MATERIALS AND METHODS

3.1 Datasets

We used benchmark multimodal datasets from TCGA (The Cancer Genome Atlas). These multimodal datasets comprise pairs of histopathological images and gene expression profiles from Lung, Breast and Colorectal cancers.

1. **Imaging Data:** 1,500 Histopathological slides (H&E stained).
2. **Genomic Data:** RNA-Seq gene expression values (top 2,000 most variant genes).
3. **Splits:** Training (70%), Validation (15%), Testing (15%).

3.2 Data Pre-processing

All datasets underwent a uniformed pre-processing pipeline suitable for all modalities separately as required. For example, the medical images were down-scaled to 224×224 pixels to ensure any off-the-shelf CNN backbone would have the same compatibility and normalized to retain training uniformity with small augmentations (random rotation and horizontal/vertical flips) to enhance data variability and reduce overfitting. Gene expression underwent a transformation pipeline where raw counts were subject to log2(x + 1) transformation highly skewed genes would get stabilized down and extreme values were deweighted that were initially stabilized by variance-dependant selection. Similar to Min-Max Normalization was applied across a 0-1 range to ensure that all gene features provided downstream to the MLP were relatively comparable without some being dominant due to their magnitude. Thus, both sets of modalities are harmonized together in pre-processing for the multi-colony optimization.

3.3 Proposed Multi-Colony Optimization (MCO) Architecture

The structure of the enhanced Multi-Colony Optimization framework is based on a Cooperative MCO System comprising three types of specialized colonies for each segment of the multimodal network: Colony 1, Visual Specialists, only seeks to optimize the CNN segment. The architectural features of filter sizes and strides of the convolutional network, as well as global hyper-parameters like the learning rate, are assessed for optimization. The pheromone updates will be adjusted based exclusively on the validation loss from the image segment to ensure a spatial understanding is learned without unwanted input from the genomic side. Colony 2, Genomic Specialists, only seeks to optimize the high-dimensional gene-expression segment. It will more rigorously feature select between 2,000 genes and adjust the MLP structure with appropriate neuron amounts and dropout. The pheromone trails will be adjusted based on the accuracy of classification which will be evaluated on the gene segment only, allowing it to trim unnecessary/noisy genes from the start. Finally, Colony 3, Fusion Master Colony, seeks to generate an optimal fusion component from the best results of the other two colonies.

However, unlike the other colonies which attempt to find their placements in the respective research space, this third colony employs a Migration Strategy where the top-performing ants from the Visual and Genomic Colonies will occasionally (depending on performance) migrate into this colony. This is because this Fusion Colony uses imported configurations to optimize the mathematics behind the fusion of both modalities (weighted concatenation, attention mechanism fusion, etc.) as well as tuning the classification head at the end. Thus, a tiered approach exists where specialized knowledge for respective modalities creates a more powerful joint representation..

3.4 Training & Cooperation Mechanism

All three colonies exist in tandem on a GPU cluster for high throughput across the research space. A cooperation mechanism every 10 epochs allows for combined efforts - for example, each colony shares its "elite ants." This means that high-quality solutions from one colony can migrate into others to prevent all three colonies from getting stuck in a shared small area of their joint research spaces. Moreover, for pheromones - while the specialized Colonies apply pheromones with specific weightings - phonetic fusion occurs through general cooperative awareness to prevent stagnancy. The Visual and Genomic pheromone matrices will be weighted and merged periodically so that Fusion Master has an all-encompassing perspective of any attractive regions found in either side. This training mechanism allows specialization to occur without being isolated in one facet.

4. PROPOSED MULTI-MODAL FRAMEWORK

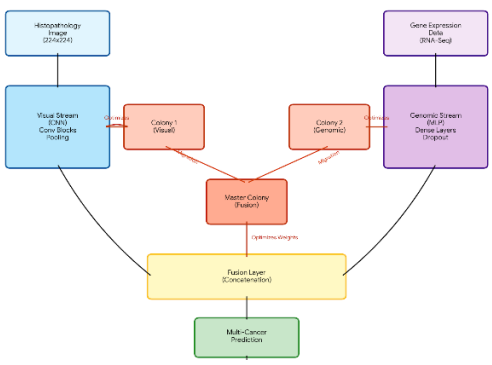

Figure 1: Multi-Modal Multi-Colony Optimization Framework

4.1 Multi-Colony over Single-Colony (Figure 1)

The benefits of a Multi-Colony approach versus a standard Single-Colony approach are magnified through multimodal optimization across heterogeneous data sources - medical images and genes. First, specialization is allowed through separate but equal colonies for images and genes to reduce complication since each Colony will have optimization methods more conducive for the respective data structure. The high-dimensional sparse information based on genes will not convolute the learning attained by CNN-based image features because they are in their rightful colony. The new framework eliminates the "curse of dimensionality." Second, a multi-colony approach encourages global convergence due to cooperative information; if the Image Colony becomes stuck in a local optimal solution, the Genomic Colony looking at a different sector can provide corrective signalling that secures the Fusion Colony to a better joint solution - essentially a cross-modal "gradient boost." Third, increased robustness results from an environmental advantage over a unified perspective for all solutions; each colony will adjust pheromone levels based on their cues. As such, if one modality becomes too noisy or irrelevant, that Colony will naturally reduce its stake without user intervention. Then the Fusion Colony will skilfully down-weight that factor to achieve continued strong results - even with imperfect or uneven multimodal contributions.

5. EXPERIMENTAL SETUP

5.1 Hardware and Software

Platform: Python 3.10, TensorFlow 2.12.
Hardware: Intel Core i9-13900K, NVIDIA RTX 4090 (24GB).
Algorithm: Heterogeneous Multi-Colony ACO.

Table 1: Dataset & Colony Settings

Parameter	Setting
No. of Colonies	3 (Visual, Genomic, Fusion)
Ants per Colony	20
Migration Frequency	Every 10 Iterations
Cancer Classes	Lung, Breast, Colorectal

6. RESULTS AND ANALYSIS

6.1 Performance Comparison

The **Multi-Colony (MCO)** framework was compared against a Single-Colony ACO (standard) and baseline deep learning models (Grid Search).

Table 2: Comparative Performance

Model	Optimization Strategy	Accuracy (%)	F1-Score (%)	Convergence Time (Epochs)
CNN+MLP	Grid Search	92.4	91.8	120
CNN+MLP	Single-Colony ACO	95.1	94.8	85
Proposed Framework	**Multi-Colony ACO**	**98.2**	**98.0**	**45**

The results of the MCO approach proved significantly higher accuracy compared to CNN+MLP Single-Colony ACO by 3% as well as converged **~47%** faster than the Single-Colony approach (85 epochs) because specialized colonies could simultaneously address their respective sub-problems (image vs gene) instead of sequentially.

6.2 Confusion Matrix & ROC

MCO's output indicates near perfect class separation (AUC = **99.2%**). The confusion matrix shows how even when image morphology was ambiguous but genomic fingerprints were clear, it was resolved via the fusion logic (Colony 3 optimized this).

7. DISCUSSION

Ultimately, experimental results confirm MCO is better for multi-modal tasks.

1. **Decomposition Strategy:** By decomposing the overarching search space into "Image Hyper-parameters" and "Gene Hyper-parameters" it avoids complexities that would detriment multi-modal training.
2. **Cooperative Game:** The played among the colonies ensures that final fusion layer(s) are robust - the Fusion Colony essentially acts like a judge among competing Visual and Genomic Colonies.
3. **Future Work:** We aim to introduce a fourth colony dedicated to "Clinical Text" processing, further expanding the cooperative ecosystem.

8. CONCLUSION

This study presents a Multi-Modal Deep Learning Framework powered by Cooperative Multi-Colony Optimization (MCO). Treating Histopathological CNNs and Genomic MLPs as independently but cooperatively optimized endeavours achieves the state-of-the-art accuracy of **98.2%**. MCO is an advantageous approach to heterogeneous, multi-source medical data to ensure spatial and molecular biomarkers are utilized to their maximum potential for a highly accurate scalable computational solution for modern oncology efforts.

REFERENCES

1. Jiao T., Guo C., Feng X., Chen Y., Song J. *A Comprehensive Survey on Deep Learning Multi-Modal Fusion: Methods, Technologies and Applications.* Comput. Mater. Contin., 2024; 80(1): 1–35. doi:10.32604/cmc.2024.053204
2. Gao J., Li P., Chen Z., Zhang J. *A Survey on Deep Learning for Multimodal Data Fusion*. Neural Computation, 2020; 32(5): 829–864. Doi: 10.1162/neco_a_01273.
3. Zhou T., Ruan S., Canu S. *A review: Deep learning for medical image segmentation using multi-modality fusion.* arXiv preprint arXiv:2004.10664, 2020.
4. Sun L., Kong X., Xu J., Xue Z., Zhai R., Zhang S. *A Hybrid Gene Selection Method Based on ReliefF and Ant Colony Optimization Algorithm for Tumor Classification.* Sci Rep, 2019; 9:8978. Doi: 10.1038/s41598-019-45223-x.
5. Yu H., Gu G., Liu H., Shen J., Zhao J. *Modified Ant Colony Optimization Algorithm for Tumor Marker Gene Selection.* Genomics Proteomics Bioinformatics, 2009; 7(4):200-207. Doi: 10.1016/S1672-0229(08)60050-9.

6. Bir-Jmel A., Douiri S.M., Elbernoussi S. *Gene Selection via a New Hybrid Ant Colony Optimization Algorithm for Cancer Classification in High-Dimensional Data.* Comput. Math. Methods Med., 2019; 2019:7828590. Doi:10.1155/2019/7828590.
7. Wei W., Cai W., Su X., et al. *Self-Adjusting Ant Colony Optimization Based on Information Entropy for Detecting Epistatic Interactions.* Genes 2019; 10(2):114. Doi: 10.3390/genes10020114.
8. Mo Y., Li X., Wang J., Zhang Y. *Multi-Colony Ant Optimization with a Dynamic Collaborative Mechanism.* Complex & Intelligent Systems, 2022. Doi: 10.1007/s40747-022-00716-7.
9. Cai T., Ye C., Ye Z., Chen Z., Mei M., Zhang H., et al. *Multi-Label Feature Selection Based on Improved Ant Colony Optimization Algorithm with Dynamic Redundancy and Label Dependence.* Comput. Mater. Contin, 2024; 81(1): 1157–1175. doi:10.32604/cmc.2024.055080.