# Representation of Features as Images with Neighborhood Dependencies for Compatibility with Convolutional Neural Networks

**Omid Bazgir** [1] **Ruibo Zhang** [1] **Saugato Rahman Dhruba** [1] **Raziur Rahman** [1] **Souparno Ghosh** [2 3] **Ranadip Pal** [1]

## Abstract

Deep learning with Convolutional Neural Networks has shown great promise in image-based classification and enhancement but is often unsuitable for predictive modeling using features without spatial correlations. We present a feature representation approach termed REFINED (REpresentation of Features as Images with NEighborhood Dependencies) to arrange high-dimensional vectors in a compact image form conducible for CNN-based deep learning. We consider the similarities between features to generate a concise feature map in the form of a two-dimensional image by minimizing the pairwise distance values following a Bayesian Metric Multidimensional Scaling Approach. We hypothesize that this approach enables embedded feature extraction and, integrated with CNN-based deep learning, can boost the predictive accuracy. We illustrate the superior predictive capabilities of the proposed framework as compared to state-of-the-art methodologies in drug sensitivity prediction scenarios using synthetic datasets, drug chemical descriptors as predictors from NCI60, and both transcriptomic information and drug descriptors as predictors from GDSC. Code: https://github.com/omidbazgirTTU/REFINED Full Paper: https://www.nature.com/articles/s41467-020-18197-y

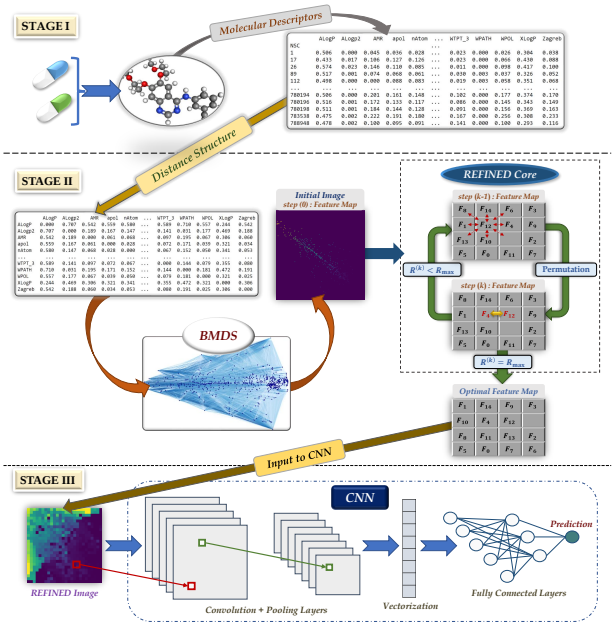


*Figure 1.* Overview of REFINED-CNN methodology for a representative application of drug sensitivity prediction using high-dimensional input features, such as molecular descriptors of drugs or genomic profiles of cell lines. STAGE I, calculate the pair wise dissimilarity matrix for the input features (672 × 672 Euclidean distance matrix for PaDEL descriptors of 52,000 unique drugs in NCI60 here). STAGE II,apply BMDS on this distance matrix to generate an initial image (of size 26 × 26 here) and apply hill climbing to arrive at an optimal configuration, i.e., the REFINED image, by maximizing the similarity between initial and final dissimilarity matrices. STAGE III, train a suitable CNN architecture with the REFINED images and predict sensitivity for a new sample (a given drug here)

## 1. Paper Summary

## References

Bazgir, O., Zhang, R., Dhruba, S. R., Rahman, R., Ghosh, S., and Pal, R. Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. *Nature communications*, 11(1):1–13, 2020.

*Equal contribution [1]Department of Electrical and Computer Engineering, Texas Tech University, USA [2]Department of Mathematics and Statistics, Texas Tech University, USA [3]Department of Statistics, University of Nebraska-Lincoln, USA. Correspondence to: Ranadip Pal <ranadip.pal@ttu.edu>.

The 2021 ICML Workshop on Computational Biology. 