# Help Conquer Cancer November 2012 Update

Since our last update, we have fully launched a GPU version of HCC, and below we highlight some useful improvements on image classification. Thank you all for your enthusiasm and important support.

Classifier improvements

Image features computed on the Grid are used to build an image classification system. We measure the accuracy of our classifier by comparing its output against scores assigned by human experts. Some images are difficult to interpret, causing experts to score these images differently.

While not yet fully finished and optimized, we have prepared an improved 11-way logistic regression classifier, which exceeds the HPCS[1] Random Forest model on precision, recall, or both for all image classes. Recall is improved on the most common classes, clear (98.8%) and precipitate (96.0). Precision is improved on crystal (74.9) and other less-common classes.

---
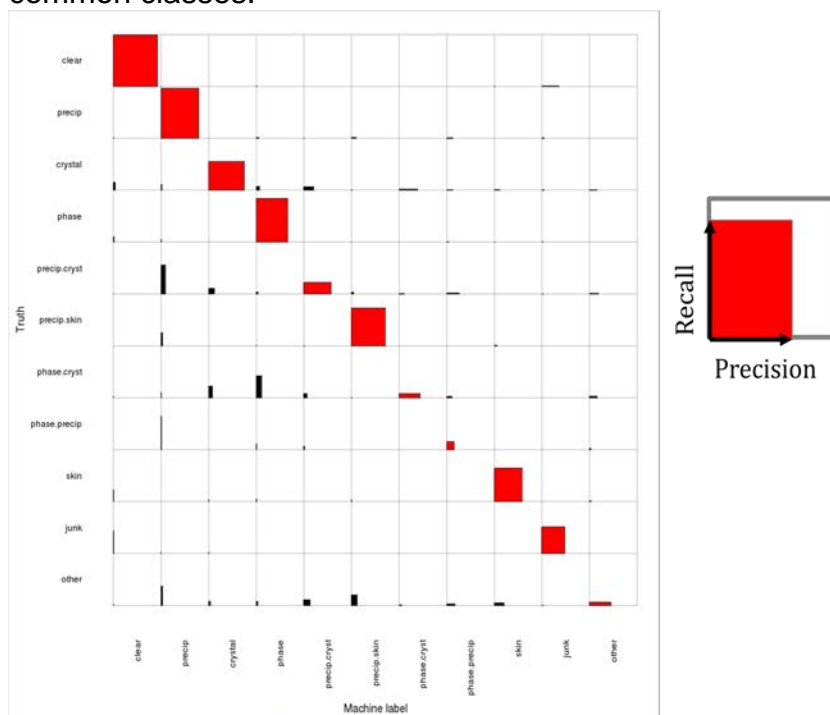
[1] Kotseruba, Y., Cumbaa, C.A. and Jurisica, I. (2012) High-throughput protein crystallization on the World Community Grid and the GPU. *Journal of Physics: Conference Series* **341** 012027 doi:10.1088/1742-6596/341/1/012027

When the 11 classes are aggregated into three super-classes (clear, has-crystal, other), the results are:

Machine label

	clear	cryst	other	Total	Recall
clear	5706	0	68	5774	99%
cryst	240	1522	1374	3136	49%
other	222	134	8358	8714	96%
Total	6168	1656	9800	17624	
Precision	93%	92%	85%		

(row labels under "Truth": clear, cryst, other, Total, Precision)

Note that this model is more precise about images it identifies as crystal, at a cost of poor recall (<50%). We continue to work on a neural network-based model that should provide a substantial improvement.

Other progress

In this direction, we have created SCRIPDB[2], which has been accessed by 691 unique users (since January 2012). Importantly, SCRIPDB is now deposited at http://pubchem.ncbi.nlm.nih.gov, and is the 7th largest source of deposited structures (out of 207; and the second largest academic deposition) - 2.5 times larger than the deposit from IBM last Fall.

While not directly linked to the HCC project, we have also co-authored several related papers:

Fortney, K., Xie, W., Kotlyar, M., Griesman, J., Kotseruba, J., Jurisica, I. NetwoRx: Connecting drugs to networks and phenotypes in S. Cerevisiae, *Nucl Acids Res*, In press.

Hammerman, P. S., et al., The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature*. 2012. doi:10.1038/nature11404

McKee, C.M., Xu, D., Cao, Y., Kabraji, S., Allen, D., Kearsmans, V., Beech, J., Smart, S., Hamdy, F., Ishkanian, A., Sykes, J., Pintile, M., Milosevic, M., Kwast, T. van der, Zafarana, G., Ramnarine, R., Jurisica, I., Mallof, C., Lam, W., Bristow, R.G., Muschel,

---

[2] Heifets, A. and Jurisica, I. SCRIPDB: A portal for easy access to syntheses, chemicals, and reactions in patents. *Nucl Acids Res*, **40, D**428-33, 2012. (http://dcv.uhnres.utoronto.ca/SCRIPDB)

R.J. Protease Nexin 1 modulates prostate adenocarcinoma by regulating the Hedgehog pathway. *J. Clin Invest.,* 2012. In press. Oct 8. pii: 59348, doi: 10.1172/JCI59348

Kotlyar, M., Fortney, F. and Jurisica, I. Network-based characterization of drug-regulated genes, drug targets, and toxicity. *Methods*, 57(4): 477-485, 2012, 2012.

Heifets, A., Jurisica, I. Construction of new medicines via game proof search. 26th *American Association for Artificial Intelligence Conference on Artificial Intelligence (AAAI-12)*, AAAI Press, Menlo Park, 1564-1570, 2012.

Wong, S., Cercone,N., Jurisica, I. Characterizing healthy and disease states by systematically comparing differential correlation networks in lung. *Advances in Health Informatics Conference*, Toronto, ON, 2012.

Otasek, D., Pastrello, C., Jurisica, I. Scalable, integrative analysis and visualization of protein interactions, *Protein-Protein Interactions - Computational and Experimental Tools*, Weibo Cai and Hao Hong (Ed.), ISBN: 978-953-51-0397-4, InTech, pp 457-472, 2012

Orchard, S., Kerrien, S., Abbani, S., Aranda, B., Bhate, J., Bidwell, S., Bridge, A., Briganti, L., Brinkmann, F.S.L., Cesareni, G., Chatr-aryamontri, A., Chautard, E., Chen, C., Dumousseau, M., Eisenberg, D., Goll, J., Hancock, R.E.W., Hannick, L.I., Jurisica, I., Khadake, J., Lynn, D.J., Mahadevan, U., Perfetto, L., Raghunath, A., Ricard-Blum, S., Roechert, B., Salwinski, L., Stümpflen, V., Tyers, M., Uetz, P., Xenarios, I., Hermjakob, H. Protein Interaction Data Curation - The International Molecular Exchange Consortium (IMEx), *Nat Methods*, 9(4): 345-350, 2012

Hai, J., Zhu, C. Q., Bandarchi-Chamkhaleh, B., Wang, Y. H., Navab, R., Shepherd, F. A., Jurisica, I., Tsao, M. S., L1 Cell Adhesion Molecule promotes tumorigenicity and metastatic potential in non-small-cell lung cancer, *Clin Cancer Res*, **18**(7):1914-1924, 2012.