_____

# GEDAS – Gene expression data analysis suite

**Tangirala Venkateswara Prasad [1]\*, Ravindra Pentela Babu[2], Syed Ismail Ahson[3]**

[1]Department of Computer Science, Jamia Millia Islamia University, New Delhi 110 025, India;
[2]Trisoft Systems Pvt. Ltd., 55 Community Centre, East of Kailash, New Delhi 110 065, India; [3]Department of Computer Science, Jamia Millia Islamia University, New Delhi 110 025, India;
Tangirala Venkateswara Prasad\* **-** E-mail: tvprasad2002@yahoo.com;
\* Corresponding author

## Abstract:
Currently available micro-array gene expression data analysis tools lack standardization at various levels. We developed GEDAS (gene expression data analysis suite) to bring various tools and techniques in one system. It also provides a number of other features such as a large collection of distance measures and pre-processing techniques. The software is an extension of Cluster 3.0 (developed based on Eisen Lab's Cluster and Tree View software). GEDAS allows the usage of different datasets with algorithms such as k-means, HC, SVD/PCA and SVM, in addition to Kohonen's SOM and LVQ.

**Availability:** http://gedas.bizhat.com/gedas.htm

**Keywords:** gene expression; standardization; GEDAS; cluster; software

## Background:
This work attempts to integrate different tools and techniques for gene expression analysis with an aim to standardize them for efficient usage. In this context, a number of tools such as Cluster/ Tree View **[1]**, SNOMAD **[2]**, Cluster 3.0 software **[3]**, GEDA suite **[4]**, GEPAS **[5]**, J-Express **[6]**, Cleaver 1.0 **[7]** and Expression Profiler **[8]** have been extensively studied and significantly improved in recent years. Here, we describe a software called GEDAS (gene expression data analysis suite) developed by integrating techniques such as OM, LVQ, k-means, hierarchical clustering, SVM **[9]** and PCA. The software supports a number of visualization techniques/gene expression data preprocessing algorithms **[1, 4]** and it contains over 10 visualizations and 19 distance measures.

## Methodology:
The GEDAS software has been developed as stand-alone software for analysis of microarray gene expression data using Visual Basic and Visual C++ programming languages. Microarray datasets can be loaded in plain text file, MS Excel or MS Access formats. The software uses Crystal Reports for generating outputs. A snapshot of GEDAS is shown in Figure 1.

## Utility:
The software facilitates various levels of data manipulation during pre-processing. GEDAS generates at least 6 different outputs for any analysis unlike other many tools producing just one output. The whole genome visualization tool is introduced in this development. **[10]** In addition to the traditional plots/graphs such as scatter plot and histograms, the temporal (or wave) graph, tree view, tree map, and whole genome view were standardized, developed and integrated into the software. We evaluated the tools using breast cancer, mouse (*Mus musculus*), *Arabidopsis thaliana*, *Homo sapiens* and sugarcane datasets. Another most important inclusion was the representation of hierarchical clustering output in the form of temporal (or wave) graph. In GEDAS, results are presented in a number of ways described elsewhere. **[4, 11, 12, 13, 14, 15, 16]** The techniques implemented in GEDAS are given in Table 1. The software facilitates sorting of data in rows, columns or both. The output can be exported in PDF, BMP, GIF and JIF formats.

## Further work:
In future development, we plan to incorporate other visualization tools **[4], [17]** including 2D and 3D score plots, profile plots, scatter plots (3D scatter plots, PCA visualization, ISOMAP visualization, and multi-dimensional scaling), Venn diagrams for visualizing similar elements in micro-arrays and SOM visualization for clustering result. We also plan to provide the software using a web interface. Our other plans include addition of robust distance measures and data mining tools (fuzzy c-means and agglomerative).
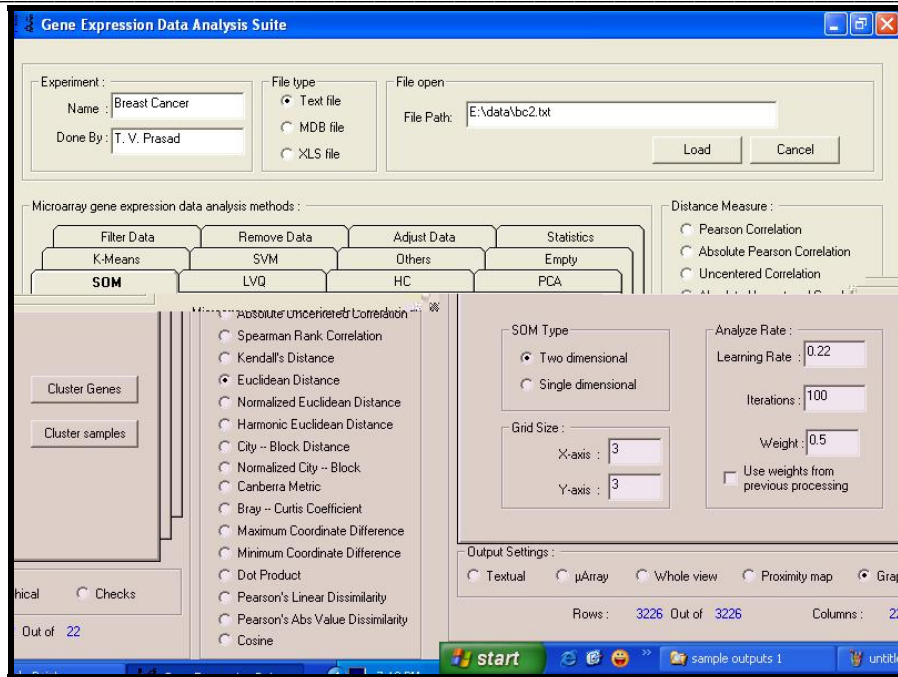
**Figure 1:** A snapshot of GEDAS is shown

| Visualization/Algorithm | Raw data | Pre-processed data | SOM | K-Means | LVQ | HC | PCA (gene) | SVM |
|---|---|---|---|---|---|---|---|---|
| Histogram | | | | | | | | |
| Checks view | | | | | | | | |
| Microarray | | | | | | | | |
| Whole sample | | | | | | | | |
| Proximity map | | | | | | | | |
| Temporal (incl. zoomed cluster view) | | | | | | | | |
| Textual | | | | | | | | |
| PC view | | | | | | | | |
| Eigen graph | | | | | | | | |
| Tree view | | | | | | | | |
| Scatter plot & M vs. A plot | | | | | | | | |
| Box-Whisker plot | | | | | | | | |
| Gene Ontology | | | | | | | | |

**Table 1:** The application of various visualization techniques included in GEDAS is listed.

**References:**
[1]    http://rana.lbl.gov/EisenSoftware.htm
[2]    http://pevsnerlab.kennedykrieger.org/snomad.htm
[3]    http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster
[4]    http://bioinformatics.upmc.edu/GE2/GEDA.html
[5]    http://gepas.bioinfo.cnio.es
[6]    http://www.molmine.com
[7]    http://classify.stanford.edu/
[8]    http://www.ebi.ac.uk/expressionprofiler
[9]    http://www.csie.ntu.edu.tw/~cjlin/libsvm
[10]   H. Caron, *et al., Science,* 291:1289 (2001) [PMID: 11181992]
[11]   C. H. Chen, *Statistica Sinica,* 12:7 (2002) [Link]
[12]   F. Luo, *et al., BIBE*, 328 (2003) [Link]
[13]   S. Tavazoie, *et al., Nature Genetics*, 22:281 (1999) [PMID: 10391217]
[14]   P. Toronen, *et al.,* FEBS Letters, 451:142 (1999)
[15]   http://bioinfo.cnio.es/docus/SOTA/#Software
[16]   http://cs.hut.fi
[17]   http://www.silicocyte.com