

**Statement of research achievements, if any, on which any award has already been received by the applicant. Please also upload brief citation(s) on the research work(s) for which the applicant has already received the award(s) (not to exceed 2000 words)**

**Statement of Research Achievements for which Prof. Bandyopadhyay was conferred the Infosys Prize in 2017**

Prof. Sanghamitra Bandyopadhyay is an outstanding scientist with top-class contributions in evolutionary computing based clustering, classifier design, many-objective optimization, and algorithms for biological data analysis.

Prof. Bandyopadhyay is internationally acclaimed for her contributions in clustering, a well-known NP-hard grouping problem. She developed single as well as multi-objective genetic algorithm based clustering methods that could come out of local optima and converged faster because of her pioneering strategy of cluster center based real-encoding of chromosomes integrated with domain-specific operators and local search. Hers was the first theoretically sound evolutionary clustering approach that directly optimized novel validity indices, whose uniqueness and global optimality for the correct number of clusters were theoretically proved. The impact of her work is evident from the fact that her fundamental paper in this regard is highly cited (more than 1000 times in Google Scholar), and is now considered state-of-the-art in evolutionary clustering.

Her evolutionary clustering methods are used extensively in domains like analog modulation classification, lip segmentation, biomedical and satellite image segmentation, modelling adiabatic temperature rise during concrete hydration, biometric analysis, reconstructing gene regulatory networks, identifying gene modules and detecting gene markers. Her cluster validity index has been utilized in the CANARY software of the US Environmental Protection Agency for detecting anomalous water quality triggered by contaminants. Her book "Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence", Springer 2007 has been critically reviewed and lauded in the IAPR newsletter, 30(4), page 12, 2008 (<http://www.iapr.org/docs/newsletter-2008-04.pdf>). Another book "Multiobjective Genetic Algorithms for Clustering: Applications in Data Mining and Bioinformatics" is the first of its kind on multiobjective clustering.

Although backed by strong theory, an effective version of multiobjective simulated annealing (SA) was lacking because of its point-based-search nature. By developing Archived Multiobjective SA (AMOS), Bandyopadhyay removed this gap in the literature. Its ability to handle four or more objectives, a major problem in existing methods, has been instrumental in its use for feature selection, gene selection, microarray data clustering, finding protein modules, segmenting remote sensing images, process plan generation in reconfigurable manufacturing systems and in selection of views in data warehousing.

Over the past decade Sanghamitra has developed new methods in computational biology, demonstrating clearly that innovative computational methods are essential for making

biological discoveries impacting therapeutic efforts. She developed TargetMiner, a machine learning based algorithm for predicting microRNA targets that provided the best specificity-sensitivity trade-off because of its systematic identification of tissue specific negative samples. The paper featured in BioTech Week, Nov. 4, 2009, and its web-based tool, made online in November 2009, is popular in the community. Genome-wide TargetMiner predictions are indexed by miRBase, a heavily used microRNA repository. Building upon these predictions, Sanghamitra revealed miR-155 as a novel genetic marker for breast cancer [Molecular BioSystems, 9(6), 2013], which has recently been independently established through experiments [Martin [et.al](#), Genes and Cancer, 5(9-10), 2014]. In another important computational study on host pathogen interaction networks, Sanghamitra provided for the first time an insight into the co-occurrence of HIV-1 infection and cancer. Her study on Alzheimer's patients has revealed the involvement of white matter of the brain in disease progression.

One of her remarkable recent contributions with great potential is the development of a fast, optimal global sequence alignment algorithm. This algorithm, integrated with the concepts of locality and context sensitive hashing, has already proved its mettle in big data scenario not only in biology, but also in web-mining and video retrieval. It opens the possibility of performing global sequence alignments at the genome level that was hitherto limited by high time complexity. Her algorithm will be particularly useful in big data scenario with next-generation sequencing machines churning out massive amounts of sequence data.

Sanghamitra has a brilliant academic career. To her credit, she is the second woman scientist to receive the coveted Bhatnagar award in Engineering Sciences. Her contributions have been acknowledged early on with several Young Scientist/Engineer awards bestowed on her from national academies and international bodies. She received the Swarnajayanti Fellowship from the Department of Science and Technology as well as young scientist/engineer awards from Indian National Science Academy (INSA), Indian National Academy of Engineers (INAE) and Department of Biotechnology (DBT). She has received the Humboldt Fellowship (Experienced Researchers) and Senior Associate position of ICTP, Italy. She is a Fellow of IEEE, INSA, INAE, National Academy of Sciences India (NASI) and West Bengal Academy of Sciences and Technology. Prof. Bandyopadhyay leads a team of dedicated students with whom she publishes in most of the top journals in her areas of research. She has published 160 journal papers, 20 book chapters and many conference papers, besides 3 authored and 3 edited books. Since 2015 she has taken charge as the Director of Indian Statistical Institute, a premier research institute in India spread over multiple cities, the first woman to occupy this position since its inception in 1931.

#### *Citation for the Infosys Prize*

The Infosys Prize 2017 in Engineering and Computer Science is awarded to Prof. Sanghamitra Bandyopadhyay for her scholarly record in algorithmic optimization and for its significant impact on biological data analysis. Her discoveries include a genetic marker for breast cancer, determination of co-occurrence of HIV and cancers and the role of white matter in Alzheimer's disease.

### **Statement of research achievements for which Prof. Bandyopadhyay was conferred the Bhatnagar award in 2010**

The major scientific contributions of the nominee are in Bioinformatics, Soft Computing (SC) and Pattern Recognition (PR) where she has conducted both theoretical development and methodology formulation. In the field of SC and PR, she developed genetic algorithm and simulated annealing based fuzzy clustering techniques using, for the first time, a cluster center based encoding and domain-specific operators [34,35]. The clustering methods could come out of local optima and evolve the number of clusters automatically, while the compact encoding led to faster convergence. Incorporation of measures of point symmetry, whose properties are mathematically analyzed, allows the identification of clusters of different, not necessarily convex, shapes [47,54]; while integration with the concepts of cluster stability and multi-seeds provide the ability to evolve more generalized clusters [81]. New cluster validity indices enable the automatic identification of the model and the model order [38,53,64]. The uniqueness and global optimality of the partitioning thus obtained is theoretically established through a relationship with the well-known Dunn's index [70]. Her pioneering approach of clustering by simultaneously optimizing multiple, often conflicting, objectives [43,49,56,68] enables the identification of clusters that provides a superior trade-off between the different objectives. In multiobjective particle swarm optimization, her time variant inertia and acceleration coefficients have eliminated the need of fixing them *a priori* [48]. She has, through the development of a novel multiobjective simulated annealing method AMOSA, provided a superior optimization technique especially for many-objective test functions [52].

Some of her recent significant contributions in Bioinformatics are as follows:

- Development of novel algorithms for finding functionally enriched gene modules through clustering of gene expression data [49,56].
- Integration of multiple data sources using adaptive weighting for prediction of the functions of yeast genes [57]. This paper is included as a curated paper in Saccharomyces Genome Database.
- Leading a pioneering effort in studying the involvement of microRNAs (miRNAs), small non-coding RNAs believed to be intricately involved in many diseases, in different types of human cancer. This has led to the construction of a cancer-miRNA network, an invaluable resource for studying the different factors involved in oncogenesis [74]. The network indicates that neighboring miRNAs are differentially co-expressed. It is used to identify hub miRNAs that could be used to detect miRNAs with strong onco/tumor suppressor characteristics. The paper appeared as a featured article in BiomedCentral website, life-sciences newsletter eBioNews (<http://www.ebioNews.com/news-center/research-frontiers/rnai-a-microrna/14853-development-of-the-human-cancer-microrna-network.html>), and will appear in the annual cancer issue of Genome Technology.
- Algorithm for the identification of densely associated modules in scale-free miRNA co-expression graphs, based on a theoretical analysis of the graph properties [67], has led to new discoveries in miRNA regulation [58]. The paper is "highly accessed" in BiomedCentral.
- Aggregating the rankings provided by several miRNA target prediction algorithms in order to arrive at the consensus ordering through an unsupervised measure of

rank importance. This work has been selected for patenting by Intellectual Ventures.

- Development of TargetMiner, a web based tool for the prediction of miRNA targets. The major contribution in this regard is the systematic identification of tissue specific negative samples, a crucial issue in machine learning. The earlier target prediction methods that relied on randomly selected or artificially generated negative samples suffered from high false positive rates and varied widely in their predictions. TargetMiner provided improved accuracy and superior specificity-sensitivity trade-off as compared to almost all the existing methods. The paper [60] was featured in BioTech Week, Nov. 4, 2009, and its web-based tool made online in November 2009 ([http://www.isical.ac.in/~bioinfo\\_miu](http://www.isical.ac.in/~bioinfo_miu)) has already received more than 1000 hits.

#### *Citation for the Bhatnagar Prize*

Made pioneering contributions in computational analysis of microRNAs through the development of a microRNA-cancer network, and discovery of miRNA modules and their putative regulatory elements by graph theoretic study of co-expression patterns. Developed a novel approach for biologically relevant tissue-specific negative example selection for microRNA target prediction resulting in a method providing the best specificity-sensitivity trade-off till date.