

Paper:

Importance of Computational Intelligent in Proteomics

Kabir Mamun and Alok Sharma

School of Engineering & Physics, The University of the South Pacific, Fiji

Laucala Campus, Suva, Fiji

E-mail: {mamun_k, sharma_al}@usp.ac.fj

[Received December 16, 2013; accepted February 15, 2014]

Computational Intelligent (CI) techniques have become an apparent need in many bioinformatics applications. In this article, we make the interested reader aware of the necessity of CI, providing a basic taxonomy of proteomics, and discussing their use, variety and potential in a number of both common as well as upcoming proteomics application.

Keywords: computational intelligence, proteomics

1. Introduction

Computational intelligence (CI) is a branch of computer science studying problems for which there are no effective computational algorithms and plays a very essential role in 21st century's world. A good part of CI research is concerned with low-level cognitive functions such as perception, object recognition, signal analysis, discovery of structures in data and so on. It has several emerging applications in biological sciences, forensic science, pattern recognition, banking and much more. In the biological sciences, deciphering of a three-dimensional structure of a protein sequence is considered to be a vital and challenging task. The identification of the three dimensional structure of a protein sequence provides objective information about the function of a protein. The structure of a protein can be determined by experimental methods which is not only a slow and time consuming process, but in many cases it's almost impossible. The abundance of protein data requires the advancement of computational techniques to predict protein structures in a reasonable amount of time.

2. Computational Intelligent in Proteomics

The knowledge of protein three-dimensional (3D) structures is vitally important for biomedical research, such as protein function analysis, mutagenesis experiments and rational drug design. Although the X-ray crystallography technique can determine protein 3D structures with high resolution, they are still time consuming, expensive and cannot be readily applied to the proteins that cannot be successfully crystallized, including most mem-

brane proteins. The nuclear magnetic resonance (NMR) is a powerful tool that can determine the 3D structures of membrane proteins of small and medium size in solutions [1–3], but it is also time-consuming and costly. In order to acquire the protein structural information at a large scale and in a timely manner, high throughput fast computational protein structure prediction methods, such as homology modelling [4, 5], could be used. Since the accuracy of predicted protein structures depend on the relatedness of homologous structural templates and the correctness of sequence alignment [4], assessing the quality of protein structural models is important for controlling and analysing the quality of the predicted models.

Protein model quality assessment plays a profound role in protein structure prediction and related applications [6]. Accurate quality assessment of protein models can help rank a pool of candidate models predicted for a given query protein. A number of model quality assessment methods and tools, such as Model Evaluator [7], APOLLO [8], and QMEAN [9] have been developed by the blessing of CI. These methods evaluate the quality of models based on the structural information extracted from protein models, without considering the source information (e.g., sequence alignment, homologous template structure), used to generate the models.

3. Protein Fold

In general terms fold assignment is one of the basic moments in any effort towards understanding protein structure and function. Protein fold assignment will often disclose evolutionary relationships, which sometimes are difficult to detect at a sequence level. This may help in a better understanding of protein function, its biological activity and role in living organisms.

The prediction of the fold of a query protein from its primary sequence has become very challenging. Fold is a three-dimensional pattern that according to Structural Classification of Protein (SCOP) [10] is characterised by a set of major secondary structural elements (e.g., α -helices and β -sheets) with certain arrangement and topological connections. Each fold can belong to one out of four structural classes, namely α , β , $\alpha + \beta$, α/β , which are used to characterised a protein in an upper level of structural organisation depending on the major-

ity of secondary structural elements in a protein structure and the succession of these elements in the structure. It has been estimated that the number of folds is $\sim 1,000$ [11], a very small number compared to the number of proteins. Fold recognition (FR) or threading is a cousin to sequence homology. Instead of searching for significantly similar sequences and deducing the structure of the protein, Fold recognition methods try to recognize the structural fold of a protein by using a structure template library and the protein's sequence information then generate an alignment between the query and the recognized template from which the structure of query protein can be predicted [12]. One family of methods that are used to assign a fold to a protein sequence uses alignment of the query sequence in order to assign to it the fold of a protein which has an adequate level of similarity with it. Profile-profile [13, 14] alignment methods are the most sensitive at detecting distant homologues with low sequence similarity ($<20\%$) among the sequence-based alignment methods, while sequence-structure alignment performance depends on the data derived from known structures which are still, though, very small in number compared to protein sequences. Another strategy for fold prediction is to use sequence-derived features combined with classification techniques (e.g., train and test decision trees, k -nearest neighbor (KNN) and neural network (NN) classifiers). This family of methods has gained great interest as the work described in [15]. Fold prediction gets more challenging when more classes appear and the classification problem becomes more complex. Work in this field has included the application of one-versus-others [16], unique one-versus-others and all-versus-all methods [17], all of which use NNs or support vector machines (SVMs) as classifiers in multiple binary classification tasks. In [18], authors used Bayesian classifiers, decision trees and SVMs as building blocks in multi-level classification scheme in order to classify proteins firstly at class (structural) level and then at fold level, while in [19], authors combined evolutionary algorithms for the selection of the most informative sequence-derived features and SVMs as classification modules in the fold recognition problem. There are two aspects of protein fold recognition problem: first is the computational difficulty and second is that the current energy functions are still not accurate enough to calculate the free energy of a given conformation. Computational difficulty can be solved by parallelization of one of the evolutionary methods so it can give a high performance [12].

4. Protein Structure

Prediction of protein structural classes is a very important and challenging problem. The protein structure determination for several proteins, e.g., transmembrane proteins or some large proteins may not be possible with X-ray crystallography and NMR techniques. It is worth noting that UniProtKB/TrEMBL database [20] contains 28,395,832 protein sequence entries, while the number of

stored protein structures in Protein Data Bank (PDB) [21] is 65,643. Thus, the need of extracting structural information through computational analysis of protein sequences has become very important and a lot of research has been conducted towards this goal in the late years [12]. Protein structure can be analysed using the following techniques; (a) Structural Analysis, (b) Physiochemical, and (c) Evolution approach (i.e. Position-Specific Scoring Matrix (PSSM)). Over the last three decades many attempts, with varying degrees of success and novelty and CI techniques are pioneering in this field.

4.1. Protein Sequence

Over the last twenty years, the number of known protein structures has significantly increased. As a result, numerous computational methods have been developed to predict protein structural class based on the primary amino acid sequence, beginning in the 1980s [22, 23], some advancements made in 1990s [18, 24] and the most methods [25–29] developed recently. The early methods were tested on very limited protein sets, which resulted in very low performance. More recently, performance results have ranged from relatively low to high. In addition, Primary protein structural class (i.e., Protein sequence) prediction results are quite poor compared to other protein secondary structure prediction methods. In [30] rectify these shortcomings, by applying an ensemble of classification algorithms and developing a compact feature representation of protein sequences. The method applies a custom designed feature-based representation of the sequences and an ensemble of four complementary classifiers to improve prediction accuracy for sequences of varying homology.

The SCOP classification is performed manually, using structural elements located in individual domains within the protein. Researchers claim that the SCOP classification is more natural and provides more reliable information to study protein structural classes [27, 31, 32]. Structural class prediction is usually performed in two steps. First, the primary AA sequence is transformed into a fixed-length feature vector. Next, the feature vectors are fed to a classification algorithm to perform the prediction. The early computational prediction methods represented the primary sequence using only the composition vector, threshold-based class definitions, and applied discriminant analysis with simple distance definitions as the classification algorithm (i.e., Euclidean distance [23], the Hamming distance [22] and the Mahalanobis distance [24]). Later prediction methods used more complex classification algorithms, and the same composition vector-based representation (i.e., the maximum component coefficient principle algorithm [33], fuzzy clustering [34], artificial neural networks [35], vector decomposition [36], Bayesian classification [20, 37], and most recently SVM [21, 38] and so on). The most noticeable progress among these algorithms is the inclusion of the coupling effect among different AAs [31]. Recent works also improve structural class prediction by using

alternative sequence representations. Examples include auto-correlation functions based on non-bonded residue energy [39] and polypeptide composition [40]. However, these algorithms are often only tested on very small datasets, with uncontrolled (often high) sequence homology. This tends to result in overestimated prediction accuracy. In addition, they do not perform reliable comparison with other algorithms on common datasets and some incorrectly perform out-of-sample tests [26].

4.2. Secondary Structure

The prediction of protein secondary structure is an important step in the prediction of protein tertiary structure. From the point of view of pattern recognition, protein secondary structure prediction can be seen as a 3-class discrimination task, which consists in assigning to each residue of a sequence its conformational state, either α -helix, β -strand or aperiodic (coil). Researchers have started working on this problem as early as in the late sixties. Since then, almost all the main families of machine learning methods have been assessed on it. Some of the highly accurate prediction methods are connectionist architectures [41], SPINE-X & PSIPRED [42].

Though prediction of protein secondary structures has been an active research issue in bioinformatics for quite a few years and many approaches have been proposed, a new challenge emerges as the sizes of contemporary protein structure databases such as the Protein Data Bank (PDB) continue to grow exponentially.

Although kernel methods have already found many applications in bioinformatics, but they have seldom been applied to protein secondary structure prediction. As an example in [43], the authors implemented different combinations of bi-class SVMs to perform the prediction from alignment process generated by BLAST. In [44], different multi-class SVMs (M-SVMs) were used to combine several prediction methods.

The new challenge concerns how to effectively exploit the huge amount of structural information deposited in large protein structure databases and deliver ever-improving accuracy in secondary structure prediction as the sizes of the databases continue to grow. This new challenge is addressed in [45] where the authors use divergent PSI-BLAST profiles to train and test neural networks. The particular way of averaging over many networks, as well as the amazing number of networks averaged (up to 800) is the contribution of this work where prediction accuracy is estimated to be higher than 77%.

4.3. Feature Selection (FS) Techniques

In many case, both the protein sequence and secondary structure prediction are not good enough to identify the protein structure. Hence applying a feature selection (FS) technique can improve the prediction rate. During the last decade, the motivation for applying FS techniques in bioinformatics has shifted from being an illustrative example to becoming a real prerequisite for model building.

Table 1. General purpose feature selection (FS) software.

WEKA	Java
Fast Correlation Based Filter	Java
Feature Selection Book	Ansi C
MLC++	C++
Spider	Matlab
SVM and Kernel Methods Matlab Toolbox	Matlab

In particular, the high dimensional nature of many modelling tasks in bioinformatics, going from sequence analysis over microarray analysis to spectral analyses and literature mining has given rise to a wealth of feature selection techniques being presented in the field [50–52]. In the context of classification, feature selection techniques can be organized into three categories, depending on how they combine the feature selection search with the construction of the classification model: filter methods, wrapper methods and embedded methods, reported in [30]. Filter techniques assess the relevance of features by looking only at the intrinsic properties of the data [46, 53]. The filter techniques treat the problem of finding a good feature subset independently of the model selection step, whereas wrapper methods embed the model hypothesis search within the feature subset search. In a third class of feature selection techniques, termed embedded techniques, the search for an optimal subset of features is built into the classifier construction, and can be seen as a search in the combined space of feature subsets and hypotheses.

In order to provide the interested reader with some pointers to existing software packages, **Table 1** shows an overview of existing software implementing general purpose feature selection methods. All software packages mentioned are free for academic use.

5. Structural Motifs: Connectivity Between Secondary Structure Elements

Structural motif level, sometimes also called super-secondary structure, which is dependent on the connectivity between secondary structure elements. In protein structures helices and strands are connected to each other and combined in many different ways. Although, from known protein three-dimensional structures, there is a limited number of possible ways in which secondary structure elements are combined in nature. Someone should be able to distinguish such motifs by using a graphics display program, like SwissPDB viewer (Deep View) [47].

6. Function Detection

Functional annotation of proteins is a fundamental problem in the proteomics. The recent availability of protein interaction networks for many model species has

spurred on the development of computational methods for interpreting such data in order to elucidate protein function. In [46] authors have reported a review of the current computational approaches for the task, including direct methods, which propagate functional information through the network, and module-assisted methods, which infer functional modules within the network and use those for the annotation task. Despite these caveats, analysis of interaction networks is a young, promising and very active research area. The utilization of such networks for function prediction is just one of a plethora of possible ways by which this rich source of information can be exploited. Although techniques for function prediction using CI have been continuously improving, there is still a lot of room for improvement, both in terms of the methodologies and in terms of their evaluation.

7. Protein Databases: Short Overview

The primary protein database for protein structure information is the PDB [14], created sometime in the beginning of the 1970s. OCA [49], a browser-database for protein structure/function integrates information from KEGG,¹ OMIM,² Pfam,³ PubMed,⁴ SCOP⁵ [13], SwissProt,⁶ and others. Proteomics Identifications Database (PRIDE⁷) is a public repository for proteomics data, containing protein and peptide identifications and their associated supporting evidence as well as details of post-translational modifications.

8. Conclusions

In this article, we reviewed the main importance and contributions of computational intelligence (CI) research in a set of well-known Proteomics applications. Two main issues emerge as common problems in the proteomics domain: the large input dimensionality, and the small sample sizes. To deal with these problems, a wealth of techniques has been designed by researchers in bioinformatics, machine learning and data mining and implementation of CI techniques significantly increased the performance [54, 55].

To conclude, we would like to note that, in order to maintain an appropriate size of the article, we had to limit the number of referenced studies. We therefore apologize to the authors of papers that were not cited in this work.

Acknowledgements

The authors would like to thank SEP, FSTE, USP for the support.

1. <http://www.genome.jp/kegg/> [Accessed November 29, 2013]
2. <http://www.omim.org/> [Accessed November 29, 2013]
3. <http://pfam.sanger.ac.uk/> [Accessed November 29, 2013]
4. <http://www.ncbi.nlm.nih.gov/pubmed/> [Accessed November 29, 2013]
5. <http://scop.mrc-lmb.cam.ac.uk/scop/> [Accessed November 29, 2013]
6. <http://www.uniprot.org/> [Accessed November 29, 2013]
7. <http://www.ebi.ac.uk/pride> [Accessed November 29, 2013]

References:

- [1] J. R. Schnell and J. J. Chou, "Structure and mechanism of the M2 proton channel of influenza A virus," *Nature* No.451, pp. 591-595, 2008.
- [2] M. J. Berardi, W. M. Shih, S. C. Harrison, and J. J. Chou, "Mitochondrial uncoupling protein 2 structure determined by NMR molecular fragment searching," *Nature* No.476, pp. 109-113, 2011.
- [3] B. OuYang, S. Xie, M. J. Berardi, X. Zhao, J. Dev, et al., "Unusual architecture of the p7 channel from hepatitis C virus," *Nature* Vol.498, pp. 521-525, 2013.
- [4] K. C. Chou, "Structural bioinformatics and its impact to biomedical science," *Curr Med Chem* Vol.11, pp. 2105-2134, 2004.
- [5] K. C. Chou, "Coupling interaction between thromboxane A2 receptor and alpha-13 subunit of guanine nucleotide-binding protein," *J Proteome Res* Vol.4, pp. 1681-1686, 2005.
- [6] J. Lundström, L. Rychlewski, J. Bujnicki, and A. Elofsson, "Pcons: A neuralnetwork-based consensus predictor that improves fold recognition," *Protein Sci* Vol.10, pp. 2354-2362, 2001.
- [7] Z. Wang, A. N. Tegge, and J. Cheng, "Evaluating the absolute quality of a single protein model using structural features and support vector machines," *Proteins* Vol.75, pp. 638-647, 2009.
- [8] Z. Wang, J. Eickholt, and J. Cheng, "APOLLO: A quality assessment service for single and multiple protein models," *Bioinformatics* Vol.27, pp. 1715-1716, 2011.
- [9] P. Benkert, S. C. Tosatto, and D. Schomburg, "QMEAN: A comprehensive scoring function for model quality assessment," *Proteins* Vol.71, pp. 261-277, 2008.
- [10] T. J. P. Hubbard, A. Bart, S. E. Brenner, A. G. Murzin, and C. Chothia, "SCOP: a structural classification of proteins database," *Nucleic Acids Research*, Vol.27, pp. 254-256, 1999.
- [11] C. Chothia, "One thousand families for the molecular biologist," *Nature*, Vol.357, pp. 543-544, 1992.
- [12] A. S. Eldin, T. H. A. Soliman, M. E. Marie, and M. M. M. Ghareeb, "A Deep Glimpse into Protein Fold Recognition," *Int. J. of Sciences, Research Article*, ISSN:2305-3925, Vol.2, June 2013.
- [13] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice," *Nucleic Acids Research*, Vol.22, pp. 4673-4680, 1994.
- [14] A. Marti-Renom, M. S. Madhusudhan, and S. Andrej, "Alignment of protein sequences by their profiles," *Protein Science*, Vol.13, pp. 1071-1087, 2004.
- [15] M. W. Craven, R. J. Mural, L. J. Hauser, and E. C. Uberbacher, "Predicting protein folding classes without overly relying on homology," *Proc. of Intelligent Systems in Molecular Biology (ISMB)*, Vol.3, pp. 98-106, 1995.
- [16] I. Dubchak, I. Muchnik, C. Mayor, I. Dralyuk, and S. H. Kim, "Recognition of a protein fold in the context of the structural classification of proteins (SCOP) classification," *Proteins*, Vol.35, pp. 401-407, 1999.
- [17] C. H. Ding and I. Dubchak, "Multi-class protein fold recognition using support vector machines and neural networks," *Bioinformatics*, Vol.7, pp. 349-358, 2001.
- [18] K. Marsolo, S. Parthasarathy, and C. Ding, "A multi-level approach to SCOP fold recognition," *Proc. of the Fifth IEEE Symp. on Bioinformatics and Bioengineering*, pp. 57-64, 2005.
- [19] S. Y. M. Sanganthan and P. N. Kalyanmoy, "Multi-class protein fold recognition using multi-objective evolutionary algorithms," *Proc. of the 2004 IEEE Symp. on Computational Intelligence in Bioinformatics and Computational Biology*, pp. 61-66, 2004.
- [20] R. Apweiler, A. Bairoch, C. H. Wu, W. C. Barker, B. Boeckmann, S. Ferro, E. Gasteiger, H. Huang, R. Lopez, M. Magrane, M. J. Martin, D. A. Natale, C. O'Donovan, N. Redaschi, and L. L. Yeh, "UniProt: the universal protein knowledge base," *Nucleic Acids Research*, Vol.32, pp. D115-D119, 2004.
- [21] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The Protein Data Bank," *Nucleic Acids Research*, Vol.28, pp. 235-242, 2000.
- [22] P. Klein and C. Delisi, "Prediction of protein structural class from the amino-acid sequence," *Biopolymers* Vol.25, pp. 1659-1672, 1986.
- [23] H. Nakashima, K. Nishikawa, and T. Ooi, "The folding type of a protein is relevant to the amino acid composition," *J. of Biochemistry*, Vol.99, pp. 153-162, 1986.
- [24] K. C. Chou and C. T. Zhang, "Predicting protein-folding types by distance functions that make allowances for amino-acid interactions," *J. of Biological Chemistry* Vol.269, pp. 22014-22020, 1994.
- [25] Y. Cai, "Is it a paradox or misinterpretation?," *Proteins* Vol.43, pp. 336-338, 2001.

- [26] K. D. Kedariseti, L. A. Kurgan, and S. Dick, "Classifier ensembles for protein structural class prediction with varying homology," *Biochemical and Biophysical Research Communication*, Vol.348, pp. 981-988, 2006.
- [27] Z.-X. Wang and Z. Yuan, "How good is the prediction of protein structural class by the component-coupled method?," *Proteins* Vol.38, pp. 165-175, 2000.
- [28] A. Sharma, J. Lyons, A. Dehzangi, and K. K. Paliwal, "A feature extraction technique using bi-gram probabilities of position specific scoring matrix for protein fold recognition," *J. of Theoretical Biology*, Vol.320, No.7, pp. 41-46, 2013.
- [29] A. Sharma, K. K. Paliwal, A. Dehzangi, J. Lyons, S. Imoto, and S. Miyano, "A Strategy to Select Suitable Physicochemical Attributes of Amino Acids for Protein Fold Recognition," *BMC Bioinformatics*, Vol.14, No.233, pp. 1-11, 2013.
- [30] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, Vol.23, No.19, pp. 2507-2517, DOI: 10.1093/bioinformatics/btm344, 2007.
- [31] K. C. Chou and G. M. Maggiora, "Domain structural class prediction," *Protein Engineering* Vol.11, pp. 523-538, 1998.
- [32] A. Murzin, S. Brenner, T. Hubbard, and C. Chothia, "SCOP: a structural classification of protein database for the investigation of sequence and structures," *J. of Molecular Biology* Vol.247, pp. 536-540, 1995.
- [33] C. T. Zhang and K. C. Chou, "An optimization approach to predicting protein structural class from amino-acid composition," *Protein Science* Vol.1, pp. 401-408, 1992.
- [34] H. B. Shen, J. Yang, X.-J. Liu, and K. C. Chou, "Using supervised fuzzy clustering to predict protein structural classes," *Biochemical and Biophysical Research Communications*, Vol.334, pp. 577-581, 2005.
- [35] J. Cornette, K. B. Cease, H. Margalit, J. L. Spouge, J. A. Berzofsky, and C. DeLisi, "Hydrophobicity scales and computational techniques for detecting amphipathic structures in protein," *J. of Molecular Biology* Vol.195, pp. 659-685, 1987.
- [36] F. Eisenhaber, C. Frömmel, and P. Argos, "Prediction of secondary structural content of proteins from their amino acid composition alone. II. The paradox with secondary structural class," *Proteins* Vol.25, pp. 169-179, 1996.
- [37] J. Yu and X. Chen, "Bayesian neural network approaches to ovarian cancer identification from high-resolution mass spectrometry data," *Bioinformatics*, Vol.21, pp. i487-i494, 2005.
- [38] Y. Cai, X. J. Liu, X. B. Xu, and K. C. Chou, "Support vector machines for prediction of protein domain structural class," *J. of Theoretical Biology* Vol.221, pp. 115-120, 2003.
- [39] W.-S. Bu, Z.-P. Feng, Z. Zhang, and C.-T. Zhang, "Prediction of protein (domain) structural classes based on amino-acid index," *European J. of Biochemistry* Vol.266, pp. 1043-1049, 1999.
- [40] L. Jin, W. Fang, and H. Tang, "Prediction of protein structural classes by a new measure of information discrepancy," *Computational Biology and Chemistry* Vol.27, pp. 373-380, 2003.
- [41] B. Rost and S. O'Donoghue, "Sisyphus and prediction of protein structure," *CABIOS*, Vol.13, pp. 345-356, 1997.
- [42] D. T. Jones, "Protein secondary structure prediction based on position-specific scoring matrices," *J. Mol. Biol.* Vol.292, pp. 195-202, 1999.
- [43] S. Hua and Z. Sun, "Support vector machine approach for protein subcellular localization prediction," *Bioinform.*, Vol17, pp. 721-728, 2001.
- [44] Y. Guermeur, "Combining discriminant models with new multi-class SVMs," *Pattern Anal. and Applications*, Vol.5, No.2, pp. 168-179, 2002.
- [45] S. Altschul, T. Madden, A. Shaffer, J. Zhang, Z. Zhang, W. Miller, and D. Lipman, "Gapped Blast and PSIBlast: A new generation of protein database search programs," *Nucleic Acids Res.*, Vol.25, pp. 3389-3402, 1997.
- [46] T. Szecsi and K. A. Mamun, "Situation Recognition Module for Hospital Robots," in R. Scheidl and B. Jakoby (eds.), *Proc. of the 13th Mechatronics Forum Biennial Int. Conf. (MECHATRONICS 2012)*, pp. 47-54, 2012.
- [47] N. Guex and M. C. Peitsch, "SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling," *Electrophoresis*, Vol.18, pp. 2714-2723, 1997.
- [48] R. Sharan, I. Ulitsky, and R. Shamir, "Network-based prediction of protein function," *Mol Syst Biol* Vol.3, No.88, DOI: 10.1038/msb4100129, 2007.
- [49] J. Prilusky, "OCA, a browser-database for protein structure/function," 1996.
- [50] A. Sharma, S. Imoto, S. Miyano, and V. Sharma, "Null space based feature selection method for gene expression data," *Int. J. of Mach. Learning and Cybernetics* Vol.3, No.4, pp. 269-276, 2012.
- [51] A. Sharma, S. Imoto, and S. Miyano, "A top-r feature selection algorithm for microarray gene expression data," *IEEE/ACM Trans. on Comp. Biol. and Bioinform.*, Vol.9, No.3, pp. 754-764, 2012.
- [52] A. Sharma, C. H. Koh, S. Imoto, and S. Miyano, "Strategy of finding optimal number of features on gene expression data," *Electronics Letters*, Vol.47, No.8, pp. 480-482, 2011.
- [53] A. Sharma, S. Imoto, and S. Miyano, "A between-class overlapping filter-based method for transcriptome data analysis," *J. of Bioinform. and Comp. Biol.*, Vol.10, No.05, 2012.
- [54] A. Sharma, K. K. Paliwal, S. Imoto, and S. Miyano, "Principal component analysis using QR decomposition," *Int. J. of Mach. Learn. and Cyber.*, Vol.4, No.6, pp. 679-683, 2013.
- [55] A. Sharma and K. K. Paliwal, "A new perspective to null linear discriminant analysis method and its fast implementation using random matrix multiplication with scatter matrices," *Pattern Recognition*, Vol.45, No.6, pp. 2205-2213, 2012.

**Name:**

Kabir Mamun

Affiliation:

Lecturer, School of Engineering & Physics, The University of the South Pacific, Fiji

Address:

SEP, FSTE, Laucala Campus, Suva, Fiji

Brief Biographical History:

2007- Research Assistant at Dublin City University

2012- Mechanical & Instrumentation Design Engineer at Sonex Metrology, Ireland

2013- Lecturer at the University of the South Pacific

Main Works:

- "Situation Recognition Module for Hospital Robots," in R. Scheidl and B. Jakoby (eds.), *Proc. of the 13th Mechatronics Forum Biennial Int. Conf. (MECHATRONICS 2012)*, August 2012.
- "Hospital Robot Module Development in the iWARD Project," *VIMation J.*, No.1, pp. 6-12, 2010.

Membership in Academic Societies:

- The Institute of Electrical and Electronics Engineers (IEEE)

**Name:**

Alok Sharma

Affiliation:

Associate Professor, School of Engineering & Physics, The University of the South Pacific, Fiji

Address:

SEP, FSTE, Laucala Campus, Suva, Fiji

Brief Biographical History:

2010-2012 Research Fellow, the University of Tokyo

2013- Adjunct Associate Professor at the Institute for Integrated and Intelligent Systems (IIIS), Griffith University

2014- Associate Professor at the University of the South Pacific

Main Works:

- Various projects carried out in conjunction with Motorola (Sydney), Auslog Pty., Ltd. (Brisbane), CRC Micro Technology (Brisbane), the French Embassy (Suva), and JSPS (Japan). Research interests include pattern recognition, computer security, human cancer classification and protein fold and structural class prediction problems.

Membership in Academic Societies:

- The Institute of Electrical and Electronics Engineers (IEEE)