PhD Application Cover Letter

I am writing to express my interest to work as a Ph.D. student under your kind supervision in your research lab in the area of ***"Bioinformatics".*** I have completed my Master of Science (MSc) in Statistics from the research group (one year, 2014-2015) in the Department of Statistics, University of Rajshahi, Bangladesh. In my master's thesis title was **"Robustification of Naïve Bayes Classifier for Gene Expression Data and Protein-Protein Interaction (PPI) Sites Prediction".** I have also completed the 4-years B.Sc. Honors degree (2009-2013) from the same Department. I have studied two courses related to bioinformatics with entitle **(i) Statistical Genomics & (ii) Advanced Bioinformatics** at master's level. Which includes different types statistical algorithms for genetic linkage analysis, QTL, eQTL, GWAS with SNP analysis, DNA/protein sequence analysis, Gene-expression and NGS RNA-Seq data analysis. After completing my master's I joined as a **"Research Assistant"** in the Health Research Group at Department of Statistics, University of Rajshahi, Bangladesh. During this period (March, 2015 to March, 2018), I am jointly developed some statistical algorithms for bioinformatics. I have been participated more than 10 national and international workshops on mathematics, Machine Learning, Robust Statistics and Bioinformatics. Also I have attendant several computer programming training on R, MATLAB, C, Perl and Python for software development in Bioinformatics organized by Department of Statistics, University of Rajshahi, Bangladesh. I have presented more than 10 scientific papers at international conferences in the field of Bioinformatics and Genomics. After that, I worked as a **"Statistical Officer"** from April, 2018 to ongoing at International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) under the infectious diseases divisions. As a responsibility I worked in public health, epidemiology and vaccine clinical trials research domain mainly data management, cleaning, statistical analysis, manuscript writing and reporting. And I have received a lot of effective knowledge like programming and software development in- R, Visual Studio 2010 & C# programming, SPSS and STATA, MS SQL Server 2007 & 2010 and Mysql. I have good hands on experiences in Bioinformatics and Genomics as well as in machine learning techniques and basic knowledge for developing of few bioinformatics software tools. I have also published 14 journal articles with good ISI impact factor, 9 full length conference proceedings, and 19 abstract conference proceedings at international conferences in the field of Bioinformatics and Environmental science. My academic curriculum provided me a good knowledge on different theoretical subjects implied in the research projects, and the capacity to integrate different disciplines of Statistics, Machine Learning, Programming, Bioinformatics, and Genomics. Further, I have completed my undergraduate and postgraduate studies in English medium, thus I believe I have well command over my language and won't be facing any problems in communicating in your lab. I am very enthusiastic and keen to work under your guidance and equally confident that I will be able to provide valuable contribution to your laboratory's future prospective and initiatives in the **Bioinformatics** project. Thank you very much and I appreciate your time. Please find my CV with all the details. I am looking forward for your kind communication. Thanking You.

Sincerely,

Md. Shakil Ahmed
Email: shakil.statru@gmail.com
https://www.google.com/amp/s/www.researchgate.net/profile/Md_Ahmed28/amp

CURRICULUM VITAE
OF
MD. SHAKIL AHMED

CONTACT ADDRESS:

Md. Shakil Ahmed

JANATA HOUSING, Road#06,
House#09, AGARGON,
Dhaka, Bangladesh
Cell Phone: +880-1737-725636,
+880-1867-117428
E-mail: shakil.statru@gmail.com

Academic Qualification (From the Top Label)

Name of School/ University	Location	Years and Label of Graduation	Result	Years Required
University of Rajshahi	Rajshahi, Bangladesh	2015 (Masters of Science in Statistics)*	3.794 (out of 4) [4th position]	1
University of Rajshahi	Rajshahi, Bangladesh	2013 (Bachelors of Science in Statistics)*	3.586 (out of 4) [3rd position]	4
Mochmoil Degree College	Rajshahi, Bangladesh	2008 (Higher Secondary)	4.00 (out of 5)	2
Mochmoil High School	Rajshahi, Bangladesh	2006 (Secondary)	4.75 (out of 5)	5
Mochmoil Govt. Primary School	Rajshahi, Bangladesh	2000 (Primary)	Pass	5
			Total years	17

*Medium of Instruction at B.Sc. (Honors) and M.Sc. label was in English.

Personal Details

Name	:	Md. Shakil Ahmed
Father's Name	:	Md. Shamsul Haque
Mother's Name	:	Rabeya Bagum
Spouse's Name	:	Suraiya Khanam
Date of Birth	:	November 24, 1989
Nationality	:	Bangladeshi
Sex	:	Male
Marital Status	:	Married
Religion	:	Islam
Email Address	:	shakil.statru@gmail.com

EMPLOYMENT HISTORY (6.3 YEARS)

Position	Duration	Years	Organization/Institute
Statistical Officer	April, 2018-January 22, 2021	2 Years 9 months	ERI, IDD, icddr,b
Research Assistant	March, 2015-March, 2018	3 Years	Health Research Group, Dept. Statistics, University of Rajshahi
Lecturer (Statistics)	December, 2016-May, 2017	6 Months	Rajshahi Engineering Science and Technology College (Affiliated with University of Rajshahi), Rajshahi-6206

As Statistical Officer have successfully completed the following responsibilities

I am working as a **"Statistical Officer"** at ICDDR,B, Dhaka, Bangladesh. I am working in public health, epidemiology and vaccine clinical trials research monitoring and evaluation domain mainly M & E, data management, cleaning, statistical analysis, manuscript writing and reporting from April, 2018 to till now. And I have received a lot of effective Knowledge like Programming in- R, Python, Perl, C, Visual Studio 2010 & C# Programming, SPSS and STATA, MS SQL Server and Mysql. I also have basic computer skills like MS office Suite and internet communication. I am very much confident about Monitoring and Evaluation and strong analytical ability in the social research field.

As Research Assistant have performed following responsibilities

Provided technical support for Survey, Making Questionnaires, Training and Monitoring of Data collection process, Making Data entry template/table, Supervision of Data Entry Team, Data cleaning, Data analysis in SPSS and Stata. Report writing. Provided assistance to write project proposal and program fund raising.

As Lecturer (Statistics) have performed following responsibilities

Teaching, research, and supervised students, student training for statistical analysis. Lecturer preparation for class room presentation. I thought two subjects with entitle "Statistics for Engineers" and "Matrix Algebra and Differential Equations" at B. Sc. in Engineering label students.

SURVEILLANCE AND VACCINE TRAIL EXPERIENCE

Sl	Project Name	Study design	Sponsors and Collaborators	Responsibilities
1)	Immunogenicity study of an inactivated bivalent whole-cell oral cholera vaccine among	Double-Blinded, Individually Randomized Placebo-Controlled	icddr,b, Govt. of Bangladesh (EPI), WHO and UNICEF	Study design, Sample selection, Data management, Vaccine coverage,

	Unregistered Myanmar Nationals (URMN)	Trial		Statistical analysis and research report writing
2)	Food Borne Illness Surveillance in Bangladesh (FBIS)	Observational case-control study of 10 sentinel sites of different districts in Bangladesh	icddr,b, Govt. of Bangladesh (IEDCR) and FAO	Study design, Data management, Systematic reviews and meta-analyses, Graphically presentation
3)	Endemic Cholera Surveillance in Bangladesh (ECSB)	Observational case-control study of 22 sentinel sites of different districts in Bangladesh	icddr,b, Govt. of Bangladesh (IEDCR) and Gates Foundation	Study design, Data management, Systematic reviews and meta-analysis, Graphically presentation
4)	ETEC ETVAX Vaccine trial in Bangladesh.	Dose-escalating, Double-Blinded, Individually Randomized Placebo-Controlled Trial	icddr,b and PATH	Statistical analysis, research report writing
5)	Study of field application of oral cholera vaccine, Shanchol for use in developing country settings	Open-label Trial	icddr,b , International Vaccine Institute (IVI), EPI	Statistical analysis and research report writing

Master's Thesis (Bioinformatics & Machine Learning)

Master's Thesis Title is ***"Robustification of Naïve Bayes Classifier for Gene Expression Analysis and Protein-Protein Interaction (PPI) Sites Prediction".*** Department of Statistics, University of Rajshahi, Bangladesh.

Bachelor's Thesis

Bachelor's Thesis Title is ***"Factor Associated with Betel Cultivation in Bangladesh: A Cohort Study".*** Department of Statistics, University of Rajshahi, Bangladesh.

Awards

"Gold Medal for B. Sc (Hons) Academic Achievement (30 Dec 2015)". Given by Bangabandhu Sheikh Mujibur Rahman Hall, University of Rajshahi, Rajshahi-6205, Bangladesh.

Studied Courses

Part-1:	**Part-2:**
1. Probability Theory 2. Principle of Statistics-I 3. Principle of Statistics-II 4. Algebra and Numerical Methods 5. Analytical Geometry and Calculus 6. Matrix Algebra **7.** Introduction to Computers with Task-Oriented Software	1. Statistical Methods 2. Regression and Modeling 3. Analysis of Variance 4. Mathematical Economics 5. Real Analysis 6. Differential Equations and Fourier Series 7. Computer Programming (C & C++)
Part-3:	**Part-4:**
1. Multinormal Distribution and Order Statistics 2. Estimation 3. Hypothesis Testing 4. Regression and Diagnostics 5. Stochastic Process 6. Survey Methods and Sampling 7. Complex Variable and Matrix Analysis **8.** Simulation and Modeling	1. Multivariate Analysis 2. Measure and Modern Probability Theory 3. Demography 4. Biometrics 5. Economic Statistics and Econometrics 6. Sampling and Methodology 7. Operations Research and Quality Control 8. Social and Occupational Statistics
Part-5:	7. Advanced Stochastic Modeling 8. Advanced Biostatistics
1. Advance Statistical Inference 2. Generalized Linear Model 3. Advanced Multivariate Analysis 4. Statistical Genomics 5. Advanced Bioinformatics 6. Health and Epidemiology	

Bioinformatics Study

I have studied two courses in Bioinformatics at master's level, the details contents as follows:

M-STAT 506: Statistical Genomics	**M-STAT 507: Advanced Bioinformatics**
Basic Genomics, Marker Analysis of Phenotypes, The Structure of QTL Mapping, Interval Mapping Approaches for QTL Analysis, Gene Expression QTL (eQTL) Analysis, Genome-wide SNPs and Haplotype Analysis.	Introduction, Analysis of DNA Sequence, Sequence Alignment, Biological Databases, Statistical Phylogenetics, Protein Classification and Structure Prediction, Statistical Analysis of Microarray Gene Expression Data, Drug Discovery Informatics.

Publications List:

Book Publication (Bioinformatics & Machine Learning)

1. **M. Shakil Ahmed** and M. Kamruzzaman (January, 2018): *"Robust Naïve Bayes Algorithm for Gene Expression and PPI Data"*. LAMBERT Academic Publishing, ISBN: 978-613-3-99594-9; OmniScriptum GmbH & Co. KGHaroldstraße, 14, 40213 Düsseldorf, Germany.
2. **M. Shakil Ahmed,** Md. Selim Reza and M. Kamruzzaman (January, 2018): *"Ubiquitination PTM Sites Prediction via Random Forest Algorithm"*. LAMBERT Academic Publishing, ISBN: 978-613-4-92320-0; OmniScriptum GmbH & Co. KGHaroldstraße, 14, 40213 Düsseldorf, Germany.

Journal Publications (Bioinformatics & Machine Learning)

1. **M. Shakil Ahmed**, Md Shahjaman, Md Rana, Md Mollah, and Nurul Haque. "Robustification of Naïve Bayes Classifier and Its Application for Microarray Gene Expression Data Analysis." *BioMed research international* 2017 (2017).[IF=2.476]
2. **M. Shakil Ahmed**, Md Shahjaman, Enamul Kabir, and Md Kamruzzaman. "Structure modeling to function prediction of Uncharacterized Human Protein C15orf41." *Bioinformation* 14, no. 5 (2018): 206-212.
3. **M. Shakil Ahmed**, Md Shahjaman, Enamul Kabir, and Md Kamruzzaman. "Prediction of Protein Acetylation Sites using Kernel Naïve Bayes Classifier Based on Protein Sequences Profiling." *Bioinformation* 14, no. 5 (2018): 213-218.
4. **M. Shakil Ahmed**, M. Kamruzzaman, M. M. Rana, Z. Akond, and M. N. H. Mollah. "In silico analyses of human collagen protein function prediction." *Journal of Bio-Science* 24 (2016): 55-65.
5. Shahjaman, Md, Nishith Kumar, **M. Shakil Ahmed**, Anjuman Ara Begum, SM Shahinul Islam, and Md Nurul Haque Mollah. "Robust Feature Selection Approach for Patient Classification using Gene Expression Data." *Bioinformation* 13, no. 10 (2017): 327.
6. Shahjaman, Md, Nishith Kumar, Md Mollah, Manir Hossain, **M. Shakil Ahmed**, Anjuman Ara Begum, S. M. Shahinul Islam, Md Mollah, and Nurul Haque. "Robust Significance Analysis of Microarrays by Minimum β-Divergence Method." *BioMed research international* 2017 (2017). [IF=2.476]
7. M. Hadiul Kabir, **M. Shakil Ahmed** et al., (December, 2016): *"Gene Selection for Patients Clustering by Gaussian Mixture Model"*. International Journal of Biometrics and Bioinformatics (IJBB); ISSN 1985-2347, Volume - 10, Issue – 3, Pages - 34 – 45, CSC-Open Access Journals, Kuala Lumpur, Malaysia.
8. Akond, Z., M. Alam, **M. Shakil Ahmed**, and M. N. H. Mollah. "Multivariate statistical techniques for metagenomic analysis of microbial community recovered from environmental samples." *Journal of Bio-Science* 24 (2016): 45-53.

9. Rana, M. M., M. N. Hasan, **M. Shakil Ahmed**, and M. N. H. Mollah. "A novel computational approach for toxicogenomics biomarker discovery in drug development pipeline." *Journal of Bio-Science* 25 (2017): 57-66.
10. Md Zakiul Hassan, **M. Shakil Ahmed**, Md Marufuzzaman Khan, Mohammad Ahsan Uddin, Fahmida Chowdhury, and Md Kamruzzaman. "Genomic profiling of Nipah virus using NGS driven RNA-Seq expression data." *Bioinformation* 15, no. 12 (2019): 853.

Journal Publications (Environment and Hydrology)

11. ATM Rahman, **M. Shakil Ahmed**, Hasnat Mohammad Adnan, Mohammad Kamruzzaman, M. Abdul Khalek, Quamrul Hasan Mazumder, and Chowdhury Sarwar Jahan. "Modeling the changes in water balance components of the highly irrigated western part of Bangladesh." *Hydrology and Earth System Sciences* 22, no. 8 (2018): 4213-4228.[IF=4.01]
12. Md Kamruzzaman, ATM Sakiur Rahman, **M. Shakil Ahmed**, Md Enamul Kabir, Quamrul Hasan Mazumder, M. Sayedur Rahman, and Chowdhury Sarwar Jahan. "Spatio-temporal analysis of climatic variables in the western part of Bangladesh." *Environment, Development and Sustainability* 20, no. 1 (2018): 89-108.

Journal Publications (Public Health and Epidemiology)

13. Khan, Ashraful Islam, Md Taufiqul Islam, Shah Alam Siddique, **M. Shakil Ahmed**, Nurnabi Sheikh, Ashraf Uddin Siddik, Muhammad Shariful Islam, and Firdausi Qadri. "Post-vaccination campaign coverage evaluation of oral cholera vaccine, oral polio vaccine and measles–rubella vaccine among Forcibly Displaced Myanmar Nationals in Bangladesh." *Human vaccines & immune therapeutics* (2019): 1-5. [IF=3.643]
14. Uddin, M. A. and **M. Shakil Ahmed, "**Modified naive Bayes classifier for classification of protein-protein interaction sites". *Journal of Bioscience and Agriculture Research*, 2020, 26(02), 2177- 2184.

Proceedings Publications (Full Length Paper)

1. **M. Shakil Ahmed**, Supria Saha et al., (January, 2017):*"Statistical Computation for Detection of Extracellular Collagen Protein Influencing in Human Skin".* Paper ID: P111, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, p:485-491, Department of Statistics, Rajshahi University, Bangladesh.
2. **M. Shakil Ahmed**, Supria Saha et al., (January, 2017*):"Microarray Gene Expression Data Analysis by Robust Naïve Bayes Classifier".* Paper ID: P104, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, p:421-427, Department of Statistics, Rajshahi University, Bangladesh.
3. **M. Shakil Ahmed,** Atul Chandra Singha et al., (2015): "*Bayesian Approach for Prediction of Interface and Non-interface Residues from Protein Sequence".* Paper ID:123,

International Conference on Materials, Electronics & Information Engineering(ICMEIE)-2015,ISBN 978-984-33-8940--4,Faculty of Engineering, University of Rajshahi, Bangladesh.

4. M. Hadiul Kabir, **M. Shakil Ahmed** et al.,(January, 2017): *"Feature Selection for Two-way Clustering of Gene Expression Data by Gaussian Mixture Model"*, Paper ID: P099, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, p:390-396, Department of Statistics, Rajshahi University, Bangladesh.
5. Fee Faysal Ahmed, M. Parvez Mosharaf, Adiba Sultana, M. Selim Reza, **M. Shakil Ahmed** et al., (January, 2017*):"Identification of protein S-Nitrosylation sites using composition of amino acid frequency".* Paper ID: P126, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, p:577-585, Department of Statistics, Rajshahi University, Bangladesh.
6. Selim Reza, Adiba Sultana, Mst. Shamima Khatun, Fee Faysal Ahmed, **M. Shakil Ahmed** et al., (January, 2017*):"An Accurate Prediction of Protein Ubiquitination Sites Using the Composition of Amino Acid Pairs".* Paper ID: P091, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, p:357-364, Department of Statistics, Rajshahi University, Bangladesh.
7. Adiba Sultana, Selim Reza, Mst. Shamima Khatun, M. Parvez Mosharaf, **M. Shakil Ahmed** et al., (January, 2017*):"Identification of Protein Phosphorylation Sites Using a Non-parametric Feature Selection Approach".* Paper ID: P102, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, p:411-417, Department of Statistics, Rajshahi University, Bangladesh.
8. M. Parvez Mosharaf, Fee Faysal Ahmed, Abida Sultana, M. Selim Reza, **M. Shakil Ahmed** et al., (January, 2017*):"In silico prediction of protein Ubiquitination sites mapping on Arabidopsis Thaliana".* Paper ID: P127, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, p:586-593, Department of Statistics, Rajshahi University, Bangladesh.
9. M. Masud Rana, Mohammad Nazmol Hasan, **M. Shakil Ahmed** et al., (January, 2017*):"Toxicogenomics Biomarker Detection for Liver Toxicity using Multilevel Hierarchical Modeling".* Paper ID: P114, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, p:508-515, Department of Statistics, Rajshahi University, Bangladesh.

Proceedings Publications (Poster & Abstract Paper)

1. **M. Shakil Ahmed**, Supria Saha et al., (September, 2016):*"Functional Analysis of Extracellular Matrix Collagen Protein Sequences in Human Skin Using Statistical Approach".* Paper ID:38, International Conference on Analysis of Repeated Measures

Data, August, 2016, Department of Applied Statistics under Higher Education Quality Enhancement Sub-Project, CP-3293,EastWest University, Bangladesh.

2. **M. Shakil Ahmed**, Atul Chandra Singh et al., (March 12-13, 2016): *"Statistical Analysis of Protein Sequences for PTM Sites Prediction*". Celebration of Glorious 55-Years 2nd Reunion of RUSA, Page No.105. Jointly organized by Department of Statistics and Rajshahi University Statistics Alumni, University of Rajshahi, Bangladesh.
3. **M. Shakil Ahmed,** Atul Chandra Singha et al., (27-29 Dec, 2015):*"Feature Selection for Acetylation Sites Prediction Based on Naïve Bayes Classifier".* Paper ID: 069, The 2nd International Conference on Theory and Application of Statistics, Dhaka University Statistics Department Alumni Association (DUSDAA), University of Dhaka, Bangladesh.
4. **M. Shakil Ahmed,** Atul Chandra Singha et al., (2015): "*Biomarker Identification and Cancer Classification Using Robust Naïve Bayes Classifier Based on Microarray Gene Expression Data with Mean Shrinkage".* Poster ID: 47, International Conference on Biological Sciences: Food, Health and Environmental Perspectives (ICBSc-2015), Institute of Biological Sciences, University of Rajshahi, Bangladesh.
5. M. Abu Horaira, M. Selim Reza, Adiba Sultana, **M. Shakil Ahmed** et al., (January, 2017*):"Feature Selection for Multiclass Linear Bayes Classifier and Its Application for Gene Expression Data Analysis".* Paper ID: 70, International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment (20-23 January) 2017, ISBN: 978-984-34-0996-6, Department of Statistics, Rajshahi University, Bangladesh.
6. M. Masud Rana, M. Nazmol Hasan, **M. Shakil Ahmed** et al., (September, 2016):*"Identification of Chemically-induced Liver Toxicity Genomic Biomarkers using Factorial ANOVA Approach"*. Paper ID:36, International Conference on Analysis of Repeated Measures Data, August, 2016, Department of Applied Statistics under Higher Education Quality Enhancement Sub-Project, CP-3293,EastWest University, Bangladesh.
7. M. MasudRana, Atul Chandra Singha, **M. Shakil Ahmed** et al., (2015*):"Identification of the toxicogenomic biomarker using the multistage hierarchical ANOVA approach".*Journal of Drug Metabolism & Toxicology, 6:3, ISSN: 2157-7609 JDMT,DOI: 10.4172/2157-7609.S1.002. (IF=1.00)
8. Adiba Sultana, M. Selim Reza, M. Mahabubur Rahman, **M. Shakil Ahmed** et al., (27-29 Dec, 2015):*"Statistical Computation for Protein PTM Sites Prediction".* Paper ID: 076, The 2nd International Conference on Theory and Application of Statistics, Dhaka University Statistics Department Alumni Association (DUSDAA), University of Dhaka, Bangladesh.
9. Fahima Farhana Anni, M. Mamunur Rashid, Nishith Kumar, **M. Shakil Ahmed** et al., (27-29 Dec, 2015):*"Differential Analysis of Metabolite Profiles".* Paper ID: 078, The 2nd International Conference on Theory and Application of Statistics, Dhaka University Statistics Department Alumni Association (DUSDAA), University of Dhaka, Bangladesh.
10. M. Mamunur Rashid, Fahima Farhana Anni, Nishith Kumar, **M. Shakil Ahmed** et al., (27-29 Dec, 2015):*"Statistical Computation for Metabolomics Data Analysis".* Paper ID:

0187, The 2nd International Conference on Theory and Application of Statistics, Dhaka University Statistics Department Alumni Association (DUSDAA), University of Dhaka, Bangladesh.

11. M. Masud Rana, Mohammad Nazmul Hasan, Atul Chandra Singh, **M. Shakil Ahmed** et al., (27-29 Dec, 2015): *"Nested ANOVA-Simultaneous Component Analysis (NASCA): An Extended Tool of ANOVA-SCA for Toxicogenomic Biomarker Discovery".* Paper ID: 083, The 2nd International Conference on Theory and Application of Statistics, Dhaka University Statistics Department Alumni Association (DUSDAA), University of Dhaka, Bangladesh.

12. M. Selim Reza, Adiba Sultana, M. Mahabubur Rahman, **M. Shakil Ahmed** et al.,(27-29 Dec, 2015): *"Prediction of Ubiquitination Sites from Protein Sequences Using Random Forest".* Paper ID: 085, The 2nd International Conference on Theory and Application of Statistics, Dhaka University Statistics Department Alumni Association (DUSDAA), University of Dhaka, Bangladesh.

13. Atul Chandra Singh, M. Jahangir Alam, **M. Shakil Ahmed** et al., (27-29 Dec, 2015): *"Robust Canonical Correlation Analysis for Genome-Wide Association Studies".* Paper ID: 077, The 2nd International Conference on Theory and Application of Statistics, Dhaka University Statistics Department Alumni Association (DUSDAA), University of Dhaka, Bangladesh.

14. Atul Chandra Singha, **M. Shakil Ahmed** et al., (2015): *"Robust Phylogenetic Analysis for Morphological Evolution".* Poster ID: 16, International Conference on Biological Sciences: Food, Health and Environmental Perspectives (ICBSc-2015), Institute of Biological Sciences, University of Rajshahi, Bangladesh.

15. M. Masud Rana, **M. Shakil Ahmed** et al., (2015): *"Identification of Genomic Biomarker of Detoxification of Organ Toxicity Using Hierarchical ANOVA Approach".* Poster ID: 69, International Conference on Biological Sciences: Food, Health and Environmental Perspectives (ICBSc-2015), Institute of Biological Sciences, University of Rajshahi, Bangladesh.

16. Atul Chandra Singha, **M. Shakil Ahmed** et al., (2014): *"Robust Phylogenetic Canonical Correlation Analysis for Bioinformatics"* Paper ID:19, International Conference on Applied Statistics 2014, Institute of Statistical Research and Training(ISRT), University of Dhaka, Bangladesh.

17. M. Masud Rana, Atul Chandra Singha, **M. Shakil Ahmed** et al., (2014): *"Discovery of the Toxicogenomic Biomarker using the Multistage Experimental Design".* Paper ID: 92, International Conference on Applied Statistics 2014, Institute of Statistical Research and Training (ISRT), University of Dhaka, Bangladesh.

18. Nusrat Jahan, M. Bahadur Badsha, **M. Shakil Ahmed** et al., (2014): *"A Comparison of Protein-Protein Interaction Networks".* Paper ID: 122, International Conference on

Applied Statistics 2014, Institute of Statistical Research and Training (ISRT), University of Dhaka, Bangladesh.

19. M. Mehedi Hasan, **M. Shakil Ahmed** et al., (2014): *"Prediction of Protein Post-Translational Modification Sites Using the Bayes Classifier".* Paper ID:23, International Conference on Applied Statistics 2014, Institute of Statistical Research and Training(ISRT), University of Dhaka, Bangladesh.

Workshop and Training

1. *"R programming for Making Shiny Apps and R Markdown for Bioinformatics"* (1 September, 2016), Bioinformatics Lab, Statistics Department, University of Rajshahi, Bangladesh.
2. *"Python Programming for Bioinformatics"* (5-6 January, 2016), Bioinformatics Lab, Statistics Department, University of Rajshahi, Bangladesh.
3. *"Perl Programming and Online Software Development for Bioinformatics"* (31 July to 01 August, 2015), Bioinformatics Lab, Statistics Department, University of Rajshahi, Bangladesh.
4. *"R-Programming and Applications of R-Packages for Bioinformatics"* (02-05 June, 2015), Bioinformatics Lab, Statistics Department, University of Rajshahi, Bangladesh.
5. *"R-Packages/Software Development for Bioinformatics"* (20th June to 6th July, 2015), Bioinformatics Lab, Statistics Department, University of Rajshahi, Bangladesh.
6. **"***International Workshop on Large Scale National Surveys***"** (18-19 October, 2012), Department of Statistics, University of Rajshahi, Bangladesh.
7. "Optimization" (21-22 September, 2012) Department of Statistics, University of Rajshahi, Bangladesh.
8. *"Environmental Analysis under Climate Change"* (27-28 July, 2012) Department of Statistics, University of Rajshahi, Bangladesh.
9. **"***Digital Signal and Image Processing***" (**25-26 May, 2012**)** Department of Statistics, University of Rajshahi, Bangladesh.

Paper Presentation in International Conferences

1. *"International Conference on Bioinformatics and Biostatistics for Agriculture, Health and Environment"* (20-23 January, 2017), Department of Statistics, Rajshahi University, Bangladesh.
2. *"The 2nd International Conference on Theory and Application of Statistics, Dhaka University Statistics Department Alumni Association (DUSDAA)"* (27-29 December, 2015), University of Dhaka, Bangladesh.
3. *"International Conference on Biological Sciences: Food, Health and Environmental Perspectives (ICBSc) 2015" (8 August 2015)*. Institute of Biological Sciences, University of Rajshahi, Bangladesh.

4. *"International Conference on Materials, Electronics & Information Engineering (ICMEIE) 2015" (05-06 June, 2015),* Faculty of Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh.
5. *"International Conference on Applied Statistics 2014" (27-29 December, 2014),* Institute of Statistical Research and Training (ISRT), University of Dhaka, Bangladesh.
6. *"International Conference on Statistical Data Mining for Bioinformatics Health Agriculture and Environment"* **(**21-24 December, 2012**)** Department of Statistics, University of Rajshahi, Bangladesh.

Computer Literacy

Operating System	:	Microsoft XP Professional, Windows-7, Windows-8 and UBUNTU Linux.
Application Software	:	Certificate in Computer Application (MS-Word, MS-Excel, MS-Power Point), Latex.
Statistical Packages	:	SPSS (with syntax), STATA, Weka (Data Mining Software).
Programming Languages	:	R, Perl, Python, C, MATLAB, HTML, Simulation and Modeling.
Graphic Software	:	Adobe Photoshop
Internet	:	Browsing, E-mail &Web Design.
Other	:	Printing and Scanning
Typing	:	Good Typing speed in Bangla and English

References

Mohammad Azizur Rahman
Associate Professor
Dept. of Management Studies
Begum Rokeya University,
Rangpur, Bangladesh
HP: +88-01716-554463
E-mail: azizru07@gmail.com
Web: http://www.brur.ac.bd/

Dr. M. Sayedur Rahman
Professor
Faculty of Science,
Department of Statistics,
University of Rajshahi, Rajshahi-6205,
Bangladesh
HP: +88-01715360318
E-mail: say_rahman@yahoo.co.uk
Homepage: http://www.ru.ac.bd/stat/

STATEMENT OF INTEGRITY

☼ I declare that all information furnished in these curriculum vitae are correct and complete.

Signature ----------
Date: 14-02-2021

UNIVERSITY OF RAJSHAHI

Rajshahi, Bangladesh

GRADING SYSTEM		
Marks	LG	GP
80% and above	A+	4.00
75% to less than80%	A	3.75
70% to less than75%	A-	3.50
65% to less than70%	B+	3.25
60% to less than65%	B	3.00
55% to less than60%	B-	2.75
50% to less than55%	C+	2.50
45% to less than50%	C	2.25
40% to less than45%	D	2.00
Less than40%	F	0.00
Incomplete	I	--

TRANSCRIPT OF ACADEMIC RECORDS

Name: Md. Shakil Ahmed
Father's Name: Md. Shamsul Haque
Mother's Name: Rabeya Bagum
Hall/College: Bangabandhu Sheikh Mujibur Rahman
Degree Conferred: M.Sc. (Thesis)
Entrance Qualification: B.Sc. (Honours)

Examination Year: 2013
Roll No. 09114718
Faculty: Science
Department: Statistics
Course Duration: One (01) Year
Date of Birth: 24.11.1989

Reg. No. 3335
Session: 2008-2009

Course Code	Course Title	Credit	Letter Grade
M-Stat 501	Advanced Statistical Inference	3	B+
M-Stat 502	Generalized Linear Models	3	A-
M-Stat 503	Advanced Multivariate Analysis	3	A-
M-Stat 506	Statistical Genomics	3	A+
M-Stat 507	Advanced Bioinformatics	3	A+
M-Stat 511	Health and Epidemiology	3	A
M-Stat 512	Advanced Stochastic Modeling	3	A-
M-Stat 516	Advanced Biostatistics	3	A
M-Stat 519	Tutorial/Terminal and Class Attendance	4	A+
M-Stat 520	Viva-Voce	4	A+
M-Stat 522	Written Thesis	4	A+
M-Stat 523	Inplant Training	2	A+
M-Stat 524	Thesis Defense	2	A+

Total Credit : 40	Earned Credit : 40	GPA: 3.794	LG: A	Remarks: Passed

Sl. No: 1151

Issued on: **24.07.2016**

Composed By.....

Verified By.....

Controller of Examinations
University of Rajshahi

Asstt. Controller of Examinations
University of Rajshahi

*** Not valid without seal**

UNIVERSITY OF RAJSHAHI

Rajshahi, Bangladesh

GRADING SYSTEM		
Marks	LG	GP
80% and above	A+	4.00
75% to less than80%	A	3.75
70% to less than75%	A-	3.50
65% to less than70%	B+	3.25
60% to less than65%	B	3.00
55% to less than60%	B-	2.75
50% to less than55%	C+	2.50
45% to less than50%	C	2.25
40% to less than45%	D	2.00
Less than40%	F	0.00
Incomplete	I	--

Sl. No: **723**

Issued Date: **20.07.2016**

TRANSCRIPT OF ACADEMIC RECORDS

Name: Md. Shakil Ahmed **Exam. Year:** 2012 **Reg. No.** 3335 **Session:** 2008-2009

Father's Name: Md. Shamsul Haque **Roll No.** 09114718 **Date of Birth:** 24.11.1989

Mother's Name: Rabeya Bagum **Faculty:** Science

Hall/College: Bangabandhu Sheikh Mujibur Rahman **Department:** Statistics

Degree Conferred: B.Sc. (Honours) **Course Duration:** Four (04) Years **Entrance Qualification:** Higher Secondary Certificate

B.Sc. (Honours) Part-I Exam-2009 (*Course Improved in:)

Course Code	Course Title	Credit	Letter Grade	Year of GPA
B.Stat. 101	Probability Theory	4	A-	
B.Stat. 102	Principles of Statistics-I	4	A+	
B.Stat. 103	Principles of Statistics-II	4	B+	
B.Stat. 104	Algebra and Numerical Methods	4	A-	
B.Stat. 105	Analytical Geometry and Calculus	4	B+	
B.Stat. 106	Matrix Algebra	4	B	3.45
B.Stat. 107	Introduction to Computers with Task-Oriented Software	4	A	
B.Stat. 108	English for Statistics	0	S	
B.Stat. 109	Statistics Practical	6	A-	
B.Stat. 110	Viva-Voce	2	C+	
B.Stat. 111	Tutorial/Terminal & Class Attendance	2	A+	

B.Sc. (Honours) Part-II Exam-2010 (*Course Improved in:)

Course Code	Course Title	Credit	Letter Grade	Year of GPA
B.Stat. 201	Statistical Methods	4	A-	
B.Stat. 202	Regression and Modeling	4	B+	
B.Stat. 203	Analysis of Variance	4	A-	
B.Stat. 204	Mathematical Economics	4	A	
B.Stat. 205	Real Analysis	4	B+	3.49
B.Stat. 206	Differential Equations and Fourier Series	4	A+	
B.Stat. 207	Computer Programming	4	B+	
B.Stat. 208	Statistics Practical	6	A-	
B.Stat. 209	Viva-Voce	2	A-	
B.Stat. 210	Tutorial/Terminal & Class Attendance	2	B+	

Office of the Controller of Examinations ★ University of Rajshahi ★

- *Not valid without seal*

(Continued)

(Continued Page-2)

B.Sc. (Honours) Part-III Exam-2011 (*Course Improved in:)				
Course Code	Course Title	Credit	Letter Grade	Year of GPA
B.Stat. 301	Multinormal Distribution and Order Statistics	4	B+	
B.Stat. 302	Estimation	4	A+	
B.Stat. 303	Hypothesis Testing	4	A	
B.Stat. 304	Regression and Diagnostics	4	B+	
B.Stat. 305	Stochastic Processes	4	A	
B.Stat. 306	Survey Methods and Sampling	4	A-	3.66
B.Stat. 307	Complex Variable and Matrix Analysis	2	B+	
B.Stat. 308	Simulation and Modeling	2	A-	
B.Stat. 309	Statistics Practical	10	A+	
B.Stat. 310	Viva-Voce	2	B	
B.Stat. 311	Tutorial/Terminal & Class Attendance	2	A+	
B.Sc. (Honours) Part-IV Exam-2012 (*Course Improved in:)				
Course Code	Course Title	Credit	Letter Grade	Year of GPA
B.Stat. 401	Multivariate Analysis	4	B+	
B.Stat. 402	Measure and Modern Probability Theory	4	A+	
B.Stat. 403	Demography	4	B+	
B.Stat. 404	Biometrics	4	A+	
B.Stat. 405	Economic Statistics and Econometrics	4	B+	
B.Stat. 406	Sampling and Methodology	4	A	
B.Stat. 407	Operations Research and Quality Control	2	A	3.72
B.Stat. 408	Social and Occupational Statistics	2	A-	
B.Stat. 409	Statistics Practical	8	A+	
B.Stat. 410	Research Project and Field Studies	2	A	
B.Stat. 411	Viva-Voce	2	A+	
B.Stat. 412	Tutorial/Terminal & Class Attendance	2	A+	

Total Credit : 160	Earned Credit :160	CGPA: 3.586	Letter Grade: A-	Remarks: Passed

Composed By.....

Verified By.....

Controller of Examinations
University of Rajshahi

Asstt.Controller of Examinations
University of Rajshahi

Office of the Controller of Examinations ★ University of Rajshahi ★

* Not valid without seal

University of Rajshahi
Department of Statistics

TO WHOM IT MAY CONCERN

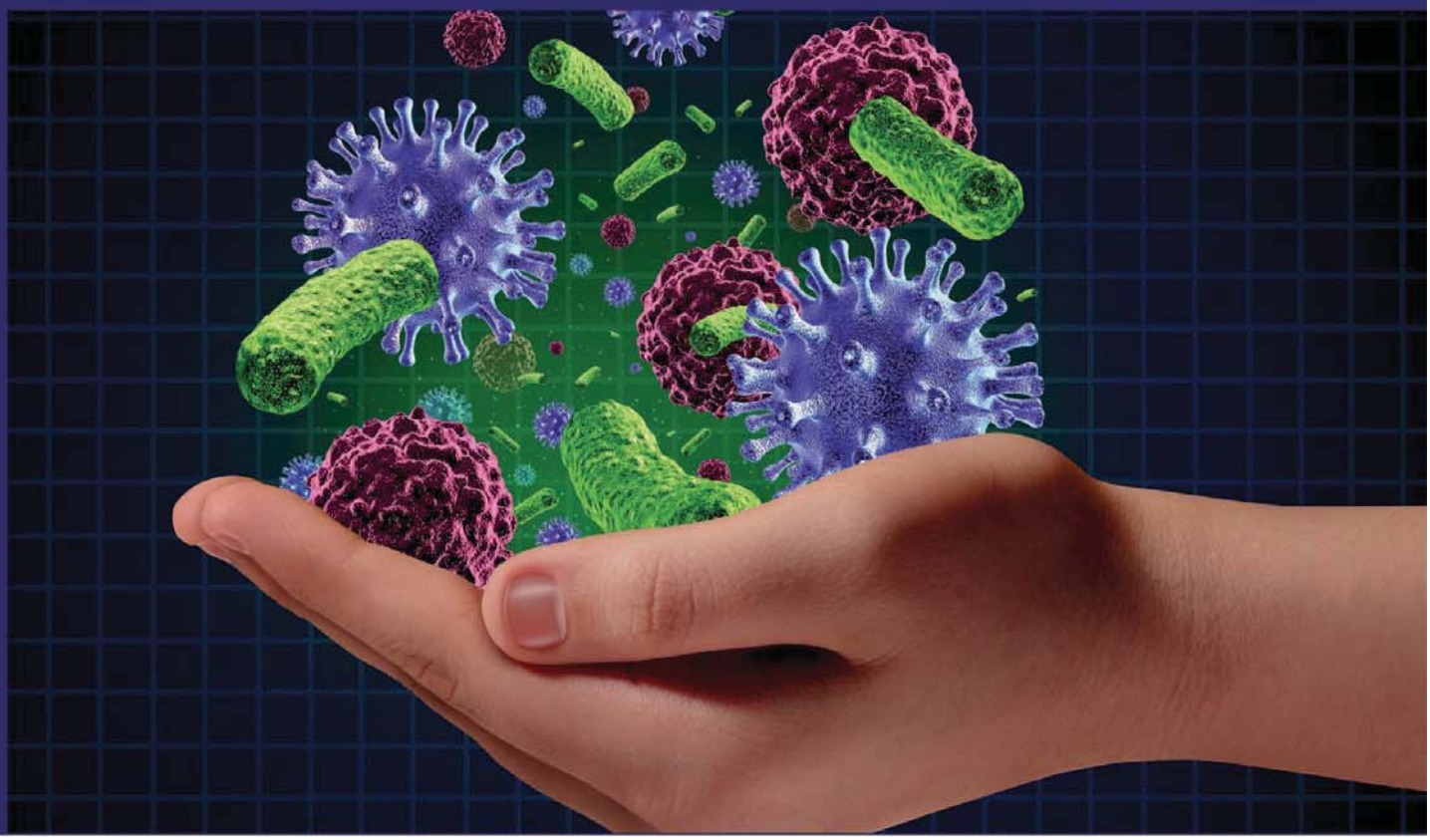
It is my pleaser that ***Md. Shakil Ahmed*** son of ***Rabeya Bagum*** (Mother) ***Md. Shamsul Haque*** (Father) bearing Examination Roll No. 09114718 of ***Bangabandhu S. M. Hall*** Registration No. 3335 Session 2008-2009 has successfully completed the Bachelor of Science (B.Sc) in Statistics in 2012 (held in April 2013) and Masters of Science (M.Sc) from thesis group in Statistics in 2013 (held in October 2014) from the Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh. The medium of instructions for all of our academic activities is English.

I wish him every success in life.

31. 3. 16

Chairman
Department of Statistics
University of Rajshahi
Rajshahi-6205, Bangladesh

Chairman
Department of Statistics
Rajshahi University.



Shakil Ahmed
Selim Reza
Kamruzzaman

Ubiquitination PTM Sites Prediction via Random Forest Algorithm

LAP LAMBERT Academic Publishing

Md. Shakil Ahmed
Md. Kamruzzaman

Robust Naïve Bayes Algorithm for Gene Expression and PPI Data

LAP LAMBERT Academic Publishing

HUMAN VACCINES & IMMUNOTHERAPEUTICS
https://doi.org/10.1080/21645515.2019.1616502

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS Check for updates

Post-vaccination campaign coverage evaluation of oral cholera vaccine, oral polio vaccine and measles–rubella vaccine among Forcibly Displaced Myanmar Nationals in Bangladesh

Ashraful Islam Khan*, Md. Taufiqul Islam*, Shah Alam Siddique, Shakil Ahmed, Nurnabi Sheikh, Ashraf Uddin Siddik, Muhammad Shariful Islam, and Firdausi Qadri

Infectious Diseases Division, icddr,b (International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh

ABSTRACT

Background: The new influx of Forcibly Displaced Myanmar Nationals (FDMNs) into Bangladesh started in August 2017 through different entry points of Bangladesh. Considering the imminent threat of infectious diseases outbreaks, the Government of Bangladesh (GoB) decided to vaccinate children against three deadly diseases (measles, rubella and poliomyelitis) and oral cholera vaccine (OCV) for all except <1 year children. After completion of the campaigns, post-vaccination campaign evaluation was carried out to assess the coverage of OCV, OPV and MR vaccines during campaigns.

Methods: Post-vaccination campaign evaluation was conducted after completion of the 2nd dose of oral cholera vaccine (OCV2) and oral polio vaccine (OPV2) through a cross-sectional survey. The evaluation was conducted in the Balukhali camps under Ukhiya upazilla. Precision-based sample size was calculated to estimate the vaccine coverage. Ninety-two trained interviewers were involved to collect data from the target of approximately 40000 FDMNs between 18 and 25 November 2017.

Results: Data were collected from 39,438 FDMNs during the survey period. The highest coverage was observed for OCVs (94% for OCV1 and 92% for OCV2). On the other hand, lower coverage was observed for the other vaccines; the coverage for OPV1, OPV2 and MR were 75%, 88% and 38%, respectively. Unawareness (30.7% did not know about the campaign) was the most notable cause of lowering down MR vaccine coverage.

Conclusion: The experience in Bangladesh demonstrates that vaccine campaigns can be successfully implemented as part of a comprehensive response toward disease outbreak among high-risk populations in humanitarian crisis.

ARTICLE HISTORY
Received 29 December 2018
Revised 24 March 2019
Accepted 2 May 2019

KEYWORDS
Vaccine; coverage; OCV; OPV; MR; FDMN; Bangladesh

Introduction

Vaccination against infectious diseases is one of the most important public health tools that can be implemented quickly during the inter-epidemic period and which results in significant reduction of mortality and morbidity.[1,2] The risk of communicable disease epidemics is increased due to the increasingly large numbers of displaced populations residing in camps, informal settlements, inadequate water and sanitation facilities or temporary placement sites.[3] The most recent massive exodus of Forcibly Displaced Myanmar Nationals (FDMNs) from northern parts of Myanmar's Rakhine State started on 25 August 2017. The people, globally known as Rohingya, entered through different entry points of Cox's Bazar, the southern border district of Bangladesh. Till November 2018, there has been an influx of more than 700,000 FDMNs, mostly women and children, who joined an estimated 300,000 Rohingyas from past migrations, making a total of more than a million people living in 32 camps in Ukhiya and Teknaf upazilas (sub-districts) of Cox's Bazar.[4] Due to the lack of resources, they were sheltered in makeshift settlements with unhygienic living conditions, poor access to safe water, and lack of a proper sanitation system. Observing risk factors in this humanitarian crisis and initial rapid risk assessment for disease (such as cholera) outbreak, was done by a team which consisted of GoB, WHO, UNICEF, icddr,b and IOM. Given the risk factors, they took into consideration that cholera might occur in refugee settlements especially in cholera endemic settings like Bangladesh.

After the influx, people were being affected by infectious diseases including diarrheal diseases, respiratory tract infections, diphtheria, and measles among others. Since 8 November 2017, diphtheria case was reported and this outbreak continued among the FDMNs camps lead to 44 deaths (Bangladesh Rohingya Emergency Response, Epidemiological Bulletin, Week 47). In the recent past, for example in Yemen, South Sudan, Haiti and other countries, the lack of WaSH and public health facilities have led to large epidemics with high numbers of cholera cases and death.[5–7] Bangladesh is an endemic country with one of the world's highest burdens of cholera, with an estimated 109,052 cholera cases annually while ~66 million population are at risk with an annual incidence rate of 1.64/1,000 along with 3% case fatality.[8] After analyzing the risk, International Co-ordination Group

CONTACT Firdausi Qadri fqadri@icddrb.org Infectious Diseases Division, icddr,b (International Centre for Diarrhoeal Disease Research, 68 Shaheed Tajuddin Ahmed Sharani, Mohakhali, Dhaka 1212, Bangladesh

*Contributed as first author

Environ Dev Sustain
DOI 10.1007/s10668-016-9872-x

Spatio-temporal analysis of climatic variables in the western part of Bangladesh

Md. Kamruzzaman[1] · A. T. M. Sakiur Rahman[2] · Md. Shakil Ahmed[3] · Md. Enamul Kabir[4] · Quamrul Hasan Mazumder[5] · M. Sayedur Rahman[3] · Chowdhury Sarwar Jahan[5]

Received: 26 January 2016 / Accepted: 17 October 2016

Abstract Monitoring and detecting trends of climatic variables like rainfall and temperature are essential for agricultural developments in the context of climate change. The present study has detected trends in annual and cropping seasonal rainfall and temperature data for the period of 1961–2011 using Mann–Kendall (MK) test, Spearman's rho (SR) test and modified Mann–Kendall test that has been applied to the significant lag-1 serial correlated time series data, and slope has been estimated using Sen's Slope estimator for twelve meteorological stations located in the western part of Bangladesh covering about 41 % of the country. Almost 71 % trends explored by MK test in annual rainfall are

✉ Md. Kamruzzaman
mkzaman@ru.ac.bd; mkzaman.ru@gmail.com

A. T. M. Sakiur Rahman
shakigeo@gmail.com

Md. Shakil Ahmed
shakil.statru@gmail.com

Md. Enamul Kabir
Enamul.Kabir@usq.edu.au

Quamrul Hasan Mazumder
qhm27@yahoo.com

M. Sayedur Rahman
say_rahman@yahoo.co.uk

Chowdhury Sarwar Jahan
sarwar_geology@yahoo.com

[1] Institute of Bangladesh Studies, University of Rajshahi, Rajshahi 6205, Bangladesh

[2] Institute of Environmental Science, University of Rajshahi, Rajshahi 6205, Bangladesh

[3] Department of Statistics, University of Rajshahi, Rajshahi 6205, Bangladesh

[4] School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Toowoomba Campus, Toowoomba, Australia

[5] Department of Geology and Mining, University of Rajshahi, Rajshahi 6205, Bangladesh

Published online: 22 October 2016

Hydrol. Earth Syst. Sci., 22, 4213–4228, 2018
https://doi.org/10.5194/hess-22-4213-2018

Hydrology and Earth System Sciences
Open Access
EGU

Modeling the changes in water balance components of the highly irrigated western part of Bangladesh

A. T. M. Sakiur Rahman[1], M. Shakil Ahmed[2], Hasnat Mohammad Adnan[3], Mohammad Kamruzzaman[4], M. Abdul Khalek[2], Quamrul Hasan Mazumder[3], and Chowdhury Sarwar Jahan[3]

[1]Hydrology Lab, Department of Earth and Environmental Sciences, Graduate School of Science and Technology, Kumamoto University, 2-40-1 Kurokami, Kumamoto, Japan
[2]Department of Statistics, University of Rajshahi, Rajshahi 6205, Bangladesh
[3]Department of Geology and Mining, University of Rajshahi, Rajshahi 6205, Bangladesh
[4]Institute of Bangladesh Studies, University of Rajshahi, Rajshahi 6205, Bangladesh

Correspondence: A. T. M. Sakiur Rahman (shakigeo@gmail.com)

Received: 23 August 2017 – Discussion started: 10 October 2017
Revised: 12 June 2018 – Accepted: 9 July 2018 – Published: 9 August 2018

Abstract. The objectives of the present study were to explore the changes in the water balance components (WBCs) by co-utilizing the discrete wavelet transform (DWT) and different forms of the Mann–Kendall (MK) test and develop a wavelet denoise autoregressive integrated moving average (WD-ARIMA) model for forecasting the WBCs. The results revealed that most of the potential evapotranspiration (P_{ET}) trends (approximately 73 %) had a decreasing tendency from 1981–1982 to 2012–2013 in the western part of Bangladesh. However, most of the trends (approximately 82 %) were not statistically significant at a 5 % significance level. The actual evapotranspiration (A_{ET}), annual deficit, and annual surplus also exhibited a similar tendency. The rainfall and temperature exhibited increasing trends. However, the WBCs exhibited an inverse trend, which suggested that the P_{ET} changes associated with temperature changes could not explain the change in the WBCs. Moreover, the 8-year (D3) and 16-year (D4) periodic components were generally responsible for the trends found in the original WBC data for the study area. The actual data was affected by noise, which resulted in the ARIMA model exhibiting an unsatisfactory performance. Therefore, wavelet denoising of the WBC time series was conducted to improve the performance of the ARIMA model. The quality of the denoising time series data was ensured using relevant statistical analysis. The performance of the WD-ARIMA model was assessed using the Nash–Sutcliffe efficiency (NSE) coefficient and coefficient of determination (R^2). The WD-ARIMA model exhibited very good performance, which clearly demonstrated the advantages of denoising the time series data for forecasting the WBCs. The validation results of the model revealed that the forecasted values were very close to actual values, with an acceptable mean percentage error. The residuals also followed a normal distribution. The performance and validation results indicated that models can be used for the short-term forecasting of WBCs. Further studies on different combinations of wavelet analysis are required to develop a superior model for the hydrological forecasting in the context of climate change. The findings of this study can be used to improve water resource management in the highly irrigated western part of Bangladesh.

1 Introduction

The water balance model is considerably important for water resource management, irrigation scheduling, and crop pattern designing (Kang et al., 2003; Valipour, 2012). The model can also be used for the reconstruction of catchment hydrology, climate change impact assessment, and streamflow forecasting (Alley, 1985; Arnall, 1992; Xu and Halldin, 1996; Molden and Sakthivadivel, 1999; Boughton, 2004; Anderson et al., 2006; Healy et al., 2007; Moriarty et al., 2007; Karimi et al., 2013). Therefore, accurately forecasting the water balance components (WBCs) and detecting the changes in them is important for achieving sustainable water resource management. However, hydrometeorological time series are con-

BIOINFORMATION
Discovery at the interface of physical and biological sciences

OPEN ACCESS GOLD

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 15(12)

Research Article

Genomic profiling of Nipah virus using NGS driven RNA-Seq expression data

Md. Zakiul Hassan[1†], Md. Shakil Ahmed[1†*], Md. Marufuzzaman Khan[2], Mohammad Ahsan Uddin[3], Fahmida Chowdhury[1], Md. Kamruzzaman[4]

[1]Infectious Diseases Division, International Centre for Diarrheal Disease Research, Bangladesh, Dhaka, Bangladesh; [2]Department of Public Health, The University of Tennessee, Knoxville, Tennessee, USA; [3]Department of Statistics, University of Dhaka, Dhaka, Bangladesh; [4]Institute of Bangladesh Studies, University of Rajshahi, Rajshahi, Bangladesh; Md. Shakil Ahmed - Email: md.shakil@icddrb.org; †Contributed equally; *Corresponding author

Received December 28, 2019; Revised December 31, 2019; Accepted December 31, 2019; Published December 31, 2019

DOI: 10.6026/97320630015853

Abstract:

Nipah virus (NiV) is an ssRNA, enveloped paramyxovirus in the genus Henipaveridae with a case fatality rate >70%. We analyzed the NGS RNA-Seq gene expression data of NiV to detect differentially expressed genes (DEGs) using the statistical R package limma. We used the Cytoscape, Ensembl, and STRING tools to construct the gene-gene interaction tree, phylogenetic gene tree and protein-protein interaction networks towards functional annotation. We identified 2707 DEGs (p-value <0.05) among 54359 NiV genes. The top-up and down-regulated DEGs were EPST1, MX1, IFIT3, RSAD2, OAS1, OASL, CMPK2 and SLFN13, SPAC977.17 using log2FC criteria with optimum threshold 1.0. The top 20 up-regulated gene-gene interaction trees showed no significant association between Nipah and Tularemia virus. Similarly, the top 20 down-regulated genes of neither Ebola nor Tularemia virus showed an association with the Nipah virus. Hence, we document the top-up and down-regulated DEGs for further consideration as biomarkers and candidates for vaccine or drug design against Nipah virus to combat infection.

Keywords: Nipah virus, NGS RNA-Seq, limma, Phylogenetic gene tree, Protein-protein interaction network

Background:

Nipah virus (NiV) is a stage III zoonotic pathogen from the family of Paramyxoviridae and a new genus from the Henipavirus **[1]**. Nipah virus was first discovered in a large encephalitis outbreak in Malaysia in 1998 **[2-4]**. Nipah virus outbreak has been recognized nearly every year in Bangladesh since 2001 and occasionally in neighboring India **[5-9]**. With the capacity of person-to-person transmission, high case fatality rate (>70%) and no availability of treatment or vaccine, the World Health Organization included the Nipah virus among the 7 Blueprint list of priority diseases and effort for Nipah vaccine development is underway **[10-12]**. Genes are strongly involved in NiV infection in interferon response in endothelial cells. The chemokine CXCL10 (interferon-induced protein 10, IP-10) gene was identified among the top 10 up-regulated genes. The cellular functionality of CXCL10 is a generation of inflammatory immune response and neurotoxicity **[13]**. Arankalle, V. A *et. al.* performed the NiV whole-genome sequencing (18,252 nucleotides) from the lung tissue samples **[14]**. Detection of DEGs is an important branch of transcriptomics research in bioinformatics. RNA-sequencing (RNA-seq) is the modern Next Generation Sequencing (NGS) technology for genomic profiling of any bacteria, virus or pathogens and other causes of diseases. Identification of DEGs or transcripts associated with the specific trait of interest from the high dimension of transcriptomic data based on NGS RNA-Seq gene expression technique. Previously microarray technology had been used by biological and biomedical researchers for discovering the candidate genes and differentially expressed markers between two or more groups of interest. Additionally, this approach includes the identification of disease biomarkers that may be important in the diagnosis of the different types and subtypes of diseases, with several implications in terms of prognosis and therapy **[15].** This sequence-based technology has created significant scope for studying the transcriptome and enabling a wide range of novel

Gene Selection for Patient Clustering by Gaussian Mixture Model

Md. Hadiul Kabir *hadi_ruo7@yahoo.com*
Laboratory of Bioinformatics,
Department of Statistics
University of Rajshahi
Rajshahi-6205, Bangladesh

Md. Shakil Ahmed *shakil.statru@gmail.com*
Laboratory of Bioinformatics,
Department of Statistics
University of Rajshahi
Rajshahi-6205, Bangladesh

Md. Nurul Haque Mollah *mollah.stat.bio@ru.ac.bd*
Laboratory of Bioinformatics,
Department of Statistics
University of Rajshahi
Rajshahi-6205, Bangladesh

Abstract

Clustering is the basic composition of data analysis, which also plays a significant role in microarray analysis. Gaussian mixture model (GMM) based clustering is very popular approach for clustering. However, GMM approach is not so popular for patients/samples clustering based on gene expression data, because gene expression datasets usually contains the large number (m) of genes (variables) in presence of a few (n) samples observations, and consequently the estimates of GMM parameters are not possible for patient clustering, because there does not exists the inverse of its covariance matrix due to $m>n$. To conquer these problems, we propose to apply a few 'q' top DE (differentially expressed) genes (i.e., $q<n/2<m$) between two or more patient classes, which are selected proportionally from all DE gene's groups. Here, the fact behind our proposal that the EE (equally expressed) genes between two or more classes have no significant contribution to the minimization of misclassification error rate (MER). For selecting few top DE genes, at first, we clustering genes (instead of patients/samples) by GMM approach. Then we detect DE and EE gene clusters (groups) by our proposed rule. Then we select q (few) top DE genes from different DE gene clusters by the rule of proportional to cluster size. Application of such a few 'q' number of top DE genes overcomes the inverse problem of covariance matrix in the estimation process of GMM's parameters, and ultimately for gene expression data (patient/sample) clustering. The performance of the proposed method is investigated using both simulated and real gene expression data analysis. It is observed that the proposed method improves the performance over the traditional GMM approaches in both situations.

Keywords: Gene Expression, Patient Clustering, Gaussian Mixture Model, Inverse Problem of Covariance Matrix, Top DE genes Selection for Patient Clustering.

1. INTRODUCTION

Clustering is a useful exploratory technique for gene expression data analysis. In fact, clustering is usually performed when no information is available concerning the membership of data items to predefined classes. That's why; clustering is traditionally seen as part of unsupervised learning. However, several heuristic clustering algorithms have been proposed in microarray data analysis. Model (especially, GMM) based clustering offer a principled alternative to heuristic algorithms.

Hindawi
BioMed Research International
Volume 2017, Article ID 5310198, 18 pages
https://doi.org/10.1155/2017/5310198

Research Article

Robust Significance Analysis of Microarrays by Minimum β-Divergence Method

Md. Shahjaman,[1,2] Nishith Kumar,[1,3] Md. Manir Hossain Mollah,[4] Md. Shakil Ahmed,[1] Anjuman Ara Begum,[1] S. M. Shahinul Islam,[5] and Md. Nurul Haque Mollah[1]

[1]*Bioinformatics Lab, Department of Statistics, University of Rajshahi, Rajshahi 6205, Bangladesh*
[2]*Department of Statistics, Begum Rokeya University, Rangpur, Rangpur 5400, Bangladesh*
[3]*Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh*
[4]*Department of Biostatistics, BUHS, Dhaka, Bangladesh*
[5]*Institute of Biological Science (IBSc), University of Rajshahi, Rajshahi 6205, Bangladesh*

Correspondence should be addressed to Md. Shahjaman; shahjaman_brur@yahoo.com

Received 29 March 2017; Accepted 28 May 2017; Published 27 July 2017

Academic Editor: Hesham H. Ali

Identification of differentially expressed (DE) genes with two or more conditions is an important task for discovery of few biomarker genes. Significance Analysis of Microarrays (SAM) is a popular statistical approach for identification of DE genes for both small- and large-sample cases. However, it is sensitive to outlying gene expressions and produces low power in presence of outliers. Therefore, in this paper, an attempt is made to robustify the SAM approach using the minimum β-divergence estimators instead of the maximum likelihood estimators of the parameters. We demonstrated the performance of the proposed method in a comparison of some other popular statistical methods such as ANOVA, SAM, LIMMA, KW, EBarrays, GaGa, and BRIDGE using both simulated and real gene expression datasets. We observe that all methods show good and almost equal performance in absence of outliers for the large-sample cases, while in the small-sample cases only three methods (SAM, LIMMA, and proposed) show almost equal and better performance than others with two or more conditions. However, in the presence of outliers, on an average, only the proposed method performs better than others for both small- and large-sample cases with each condition.

1. Introduction

Microarray experiments are usually conducted with expressions of huge number of genes (G) and a small number of experimental samples (n). This unique data structure has been discovered as a completely new promising area for the researchers. At the same time, it provides a challenge to the researchers because of high dimensionality and its complexity with small sample size. Among this huge number of genes, discovery of few biomarker genes those are differentially expressed (DE) between two or more experimental conditions with multiple patterns is one of the main objectives of this experiments. These biomarker genes are important in the diagnosis of different types and subtypes of diseases for patient prognosis and treatment [1–3]. Nowadays, researchers are also interested in exploring the gene coexpression network or interaction of DE genes to predict the hub genes that are associated with different types and subtypes of cancer [4]. The most commonly used statistical tests for the discovery of DE genes between two or more conditions are t-test or ANOVA (F-test). However, both testing procedures sometimes produce misleading results to discover few biomarker genes, because both of them suffer from small-sample sizes and normality assumptions, and they do not share the information of all genes [5]. Therefore, a gene-specific t-statistic or F-statistic becomes large even for low differential expressions of genes between two or more conditions. Thus, the false discovery rate (FDR) may increase. Tusher et al. [6] introduced a popular statistical technique to detect the DE genes by assimilating a set of gene-specific t-tests. This approach is known as Significance Analysis of Microarrays (SAM). It controls the FDR by

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

BIOMEDICAL INFORMATICS

www.bioinformation.net
Volume 13(10)

Hypothesis

Robust Feature Selection Approach for Patient Classification using Gene Expression Data

Md. Shahjaman[1, 2*], Nishith Kumar[1, 3], Md. Shakil Ahmed[1], AnjumanAra Begum[1], S. M. Shahinul Islam[4], Md. Nurul Haque Mollah[1]

[1]Bioinformatics Lab, Department of Statistics, University of Rajshahi-6205, Bangladesh; [2]Department of Statistics, Begum Rokeya University, Rangpur-5400, Bangladesh; [3]Department of Statistics, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh; [4]Institutitute of Biological Science (IBSc), University of Rajshahi, Rajshahi-6205, Bangladesh; Md. Shahjaman - Email: shahjaman_brur@yahoo.com; *Corresponding Author

Received August 26, 2017; Revised September 11, 2017; Accepted September 12, 2017; Published October 31, 2017

Abstract:
Patient classification through feature selection (FS) based on gene expression data (GED) has already become popular to the research communities. T-test is the well-known statistical FS method in GED analysis. However, it produces higher false positives and lower accuracies for small sample sizes or in presence of outliers. To get rid from the shortcomings of t-test with small sample sizes, SAM has been applied in GED. But, it is highly sensitive to outliers. Recently, robust SAM using the minimum β-divergence estimators has overcome all the problems of classical *t*-test & SAM and it has been successfully applied for identification of differentially expressed (DE) genes. But, it was not applied in classification. Therefore, in this paper, we employ robust SAM as a feature selection approach along with classifiers for patient classification. We demonstrate the performance of the robust SAM in a comparison of classical t-test and SAM along with four popular classifiers (LDA, KNN, SVM and naive Bayes) using both simulated and real gene expression datasets. The results obtained from simulation and real data analysis confirm that the performance of the four classifiers improve with robust SAM than the classical t-test and SAM. From a real Colon cancer dataset we identified 21 additional DE genes using robust SAM that were not identified by the classical *t*-test or SAM. To reveal the biological functions and pathways of these 21 genes, we perform KEGG pathway enrichment analysis and found that these genes are involved in some important pathways related to cancer disease.

Keywords: Feature selection, classification, robust SAM, β-divergence estimators.

Background:
Nowadays the big biological data is one of the hottest topics for the researchers. Gene expression datasets is the high-dimensional big datasets because it contains ten thousands of genes/features with very few patients/samples **[1].** This behavior of gene expression data often refers to the curse of dimensionality **[2-3].** Thus analyzing of these types of datasets has become complicated and challenging for the researchers. The goal of classification is to allocate/classify the new objects into one of two or more population of the training dataset whose categories are known in advance. Cancer classification based on gene expression dataset is important for subsequent diagnosis and treatment. Without correct classification of different cancer types of the patient, it is very difficult to provide proper treatment and therapies **[4].** The conventional classification methods are largely dependent on different morphological parameters to classify cancer. Thus their applications become limited with low prediction accuracies. To get rid from the curse of dimensionality of GED, classification through informative gene identification or feature selection (FS) has already attracted to the research communities **[5].** FS can boost the performance of the classifiers by selecting smaller number of features. It also reduces the computational time and provides more reliable estimates to train the classifiers. There are three types of FS methods for GED analysis; (a) wrapper method, (b) embedded method and (b) filter based method **[6-7].** Wrapper method searches the features until a certain accuracy of the classifier was achieved. Embedded methods embed feature selection within classifier construction. Filter based method first select few informative features (DE genes) using the labeled samples of training dataset and based on these pre selected features, researchers perform the further classification task. Filter based methods are easily understandable and computationally faster than the wrapper and embedded methods, thus they are better suited to high dimensional datasets **[8].** Among the filter-based methods, t-test is one of the popular and widely used methods in gene expression data analysis **[9].**

J. Bio-Sci. 24: 55-65, 2016 ISSN 1023-8654
http://www.banglajol.info/index.php/JBS/index

IN SILICO ANALYSES OF HUMAN COLLAGEN PROTEIN FUNCTION PREDICTION

MS Ahmed[1*], M Kamruzzaman[2], MM Rana[1], Z Akond[1] and MNH Mollah[1]

[1]*Bioinformatics Lab., Department of Statistics, University of Rajshahi, Bangladesh*
[2]*Institute of Bangladesh Studies, University of Rajshahi, Bangladesh*

Abstract

Collagen is the extracellular matrix protein in the several connective tissues in the human body. It is an important component for mediating cell-cell interactions and pathological conditions in human body. In this study we perform the analysis of physiochemical properties and investigate the functional characteristics of human collagen proteins. Also investigate the functional protein groups by the statistical analysis. The collagen protein family consisting 28 members in human which are involving in the complex structure of protein. The protein function, protein sequence properties, domain composition, phylogenetic and protein-protein interaction (PPI) networks analysis of human collagen alpha-1 protein sequences are implemented by the online bioinformatics tools which are currently available. Based on the PCA analysis amino acid composition, features of collagen protein sequences are divided into two supreme influential functional groups such as collagen 12, 14, 20 formed one group and the rest of others formed another group. The protein-protein interaction network study using STRING showed that top interacting score of functional group proteins 0.952, 0.939 and 0.929. The most common functional domain of collagen proteins are VWC, C4, LamG, VWA, KU, C1Q, TSPN and FN3. Physicochemical, functional and phylogenetic classification can give extensive information of protein's structure and function. The depiction of alpha-1 chains of collagen protein family in human collagen 12, 14 and 20 as a prospective protein cluster. These three proteins are possess, low glycine and proline, very high aliphatic index and a close evolutionary relation in the human skin.

Key words: Collagen protein sequences, k-means clustering, PCA, phylogenetic tree, PPI network, protein domain structure

Introduction

The extracellular matrix (ECM) is consisting of collagens, proteoglycans, glycoproteins and proteases. The extracellular matrix of connective tissues represents a complex alloy of variable members of diverse protein families defining structural integrity and various physiological functions. It is the main component of connective tissue and makes up from 25% to 35% of the whole-body protein content. Collagen Type I protein found in bone, skin, muscles and walls of blood vessels in human body (Järveläinen et al. 2009). Neighboring a substantial volume of cells the ECM is an intricate network of macromolecules. For the multiple processes such as cell migration, cell-cell interaction and cell proliferation these components play vital role (Bowers et al. 2010). The collagen protein is a triple helical structure of polypeptide chains, commonly known as the alpha chains. The common sequence pattern of triple helix is "Gly-X-Y" (Kadler et al. 1996). The stability of the helical structure depends on the presence of glycine as every third residue and being other property of the smallest amino acid. Any amino acid can be taken instead of X and Y but frequently occupied by the proline residue. Every mature active collagen protein molecules were shown that the peptidases form of pro-peptides present at the N and C terminal. The genetically distinct 28 members

*Author for correspondence: shakil.statru@gmail.com

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

BIOMEDICAL INFORMATICS

www.bioinformation.net

Volume 14(5)

Hypothesis

Prediction of Protein Acetylation Sites using Kernel Naive Bayes Classifier Based on Protein Sequences Profiling

Md. Shakil Ahmed[1*], Md. Shahjaman[2], Enamul Kabir[3], Md. Kamruzzaman[4]

[1]Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh; [2]Department of Statistics, Begum Rokeya University, Rangpur-5400, Bangladesh; [3]School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Australia; [4]Data Science for Knowledge Creation Research Center, Seoul National University, Korea; Md. Shakil Ahmed – E-mail: shakil.statru@gmail.com; *Corresponding Author:

Received March 17, 2018; Revised April 29, 2018; Accepted April 29, 2018; Published May 31, 2018

doi: 10.6026/97320630014213

Abstract

Lysine acetylation is one of the decisive categories of protein post-translational modification (PTM), it is convoluted in many significant cellular developments and severe diseases in the biological system. The experimental identification of protein-acetylated sites is painstaking, time-consuming and expensive. Hence, there is significant interest in the development of computational approaches for consistent prediction of acetylation sites using protein sequences. Features selection from protein sequences plays a significant role for acetylation sites prediction. We describe an improved feature selection approach for acetylation sites prediction based on kernel naïve Bayes classifier (KNBC). We have shown that KNBC generated from selected features by a new feature selection method outperforms than the existing methods for identification of acetylation sites. The sensitivity, specificity, ACC (Accuracy), MCC (Matthews Correlation Coefficient) and AUC (Area under Curve of ROC) in our proposed method are as follows 80.71%, 93.39%, 76.73%, 41.37% and 83.0% with the optimum window size is 47. Thus the kernel naïve Bayes classifier finds application in acetylation site prediction.

Keywords: Acetylation, Protein Sequences, Kernel Naive Bayes Classifier, Binary Encoding, CKSAAP Encoding and Kruskal-Wallis test.

Background:

The lysine residues in a protein are acetylation for exist the acetyle group in the N terminus. The lysine acetylation is one of the most vital for a lot of cellular progressions **[1-5].** For example, the dynamic interaction between lysine acetyl transferases (KATs) and lysine deacetylases (KDACs) is used to maintain the appropriate levels of histone acetylation for normal cell growth, proliferation and differentiation **[6].** Acetylation has been shown to regulate of protein expression, complex steadiness, localization and fusion **[7-12].** Lysine acetylation is intricate in the thoughtful diseases comparable with the cancer for the abnormality of KAT/KDAC function of impacting the cell division **[13-15].** The significant aims of the biological research are to describe the genome perspectives and recognize the function of genetic material in the post-genomic period **[16].** For understanding the genome backgrounds the significant information can be provided by the proteomics and transcriptomics data **[17-18].** The acetylation is one of the most significant protein modifications with an important impact on the protein functions based on the proteomic data. In the amino acid it is frequently catalyzed through acetyl transferase that transmissions acetyl group of the acetyl coenzyme (Acetyl-CoA). A large scale of the mammalian acetylated proteins has been notorious via the proteomics techniques and which are suggesting that the acetylation may be as the ubiquitous as phosphorylation **[6, 19].** The human proteins of 85% and yeast proteins of 68% were acetylated at N-terminus is described by Van Damme **[20],** The two forms of Acetylation occur in cellular methods such as Na-acetylation and Ne-acetylation. Na-acetylation is the irreversible modification happens during the translation of protein at N-terminus and the posttranslational practice it arises only for the chloroplast proteins **[21-22].** On the other hand the Ne-acetylation is the

ISSN 0973-2063 (online) 0973-8894 (print)

BIOINFORMATION
Discovery at the interface of physical and biological sciences

Open access

BIOMEDICAL INFORMATICS

www.bioinformation.net

Volume 14(5)

Hypothesis

Structure modeling to function prediction of Uncharacterized Human Protein C15orf41

Md. Shakil Ahmed[1*], Md. Shahjaman[2], Enamul Kabir[3], Md. Kamruzzaman[4]

[1]Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh; [2]Department of Statistics, Begum Rokeya University, Rangpur-5400, Bangladesh; [3]School of Agricultural, Computational and Environmental Sciences, University of Southern Queensland, Australia; [4]Data Science for Knowledge Creation Research Center, Seoul National University, Korea; Md. Shakil Ahmed – E-mail: shakil.statru@gmail.com; *Corresponding author

Received March 17, 2018; Revised April 29, 2018; Accepted April 29, 2018; Published May 31, 2018

doi:10.6026/97320630014206

Abstract:
The dyserythropoietic anemia disease is a genetic disorder of erythropoiesis characterized by morphological abnormalities of erythroblasts. This is caused by human gene C15orf41 mutation. The uncharacterized C15orf41 protein is involved in the formation of a functional complex structure. The uncharacterized C15orf41 protein is thermostable, unstable and acidic. This is associated with TPD (Treponema Pallidum) domain (135 to 265 residue position) and three PTM sites such as K50 (Acetylation), T114 (Phosphorylation) and K176 (Ubiquitination). C15orf41 is paralogous to isoform-1 (gi|194018542|) and open reading frame isoform-CRA_c (gi|119612744|) of Homo sapiens located at chromosome 15. It interacts with the human ATP (Adenosine Triphosphate) binding domain 4 (ATPBD4) having similarity score 0.725 as per protein-protein interaction (PPI) network analysis. This data provides valuable insights towards the functional characterization of human gene C15orf41.

Keywords: Uncharacterized human protein C15orf41, Phylogenetic analysis, Protein domain, PTM sites and PPI networks.

Abbreviations:
TPD - Treponema Pallidum; PTM - post translational modification; 3D - three-dimensional; MSA - multiple sequence alignment; PDB - Protein Data Bank; pI - isoelectric point; GRAVY - Grand Average of Hydropathicity; NJ - Neighbor Joining; PPI - protein-protein interaction; ATPBD4 - ATP (Adenosine Triphosphate) binding domain 4.

Background

The human uncharacterized gene C15orf41 is located at chromosome 15 encodes a protein with two predicted helix-turn-helix domains. Mutations of this gene are found in the family of congenital dyserythropoietic anemia type-I **[1].** This anemia disease description is an autosomal recessive blood disorder characterized by morphological abnormalities of erythroblasts, macrocytic anemia, secondary hemochromatosis and unproductive erythropoiesis. It is ccasionally associated with bone abnormalities, especially of the hands and feet (acrodysostosis), nail hypoplasia and scoliosis. Ultrastructural features include inter-nuclear chromatin bridges connected with some nearly erythroblasts. It is completely separated and an abnormal appearance (spongy or Swiss-cheese entrance) of the heterochromatin in a high proportion of the erythroblasts.

The structural and functional characteristics of proteins play the significant role in drug design and discovery. Investigations of these proteins characteristics experimentally in the wet lab are much laborious, time consuming and costly. The computational/statistical tools of bioinformatics reduce this cost and time significantly to characterize the uncharacterized proteins. These tools are widely used for homology modeling of sequence profiles and predicting the three-dimensional (3D) structure of the targeted protein. The homology modeling is utilized when the experimentally obtained structure is unavailable. It can provide a useful 3D model for the protein of interest that is related to at least one known protein structure. It is also used to predict the 3D structure of one or more proteins of known structure for a given protein sequence based on the primarily sequence alignment. The inclusive municipal sequences are increasing in some databases like SwissProt **[2]** and NCBI

Hindawi
BioMed Research International
Volume 2017, Article ID 3020627, 17 pages
https://doi.org/10.1155/2017/3020627

Research Article

Robustification of Naïve Bayes Classifier and Its Application for Microarray Gene Expression Data Analysis

Md. Shakil Ahmed,[1] Md. Shahjaman,[1,2] Md. Masud Rana,[1] and Md. Nurul Haque Mollah[1]

[1]*Lab of Bioinformatics, Department of Statistics, University of Rajshahi, Rajshahi 6205, Bangladesh*
[2]*Department of Statistics, Begum Rokeya University, Rangpur, Rangpur 5400, Bangladesh*

Correspondence should be addressed to Md. Shakil Ahmed; shakil.statru@gmail.com

Received 18 March 2017; Revised 10 June 2017; Accepted 14 June 2017; Published 7 August 2017

Academic Editor: Federico Ambrogi

The naïve Bayes classifier (NBC) is one of the most popular classifiers for class prediction or pattern recognition from microarray gene expression data (MGED). However, it is very much sensitive to outliers with the classical estimates of the location and scale parameters. It is one of the most important drawbacks for gene expression data analysis by the classical NBC. The gene expression dataset is often contaminated by outliers due to several steps involved in the data generating process from hybridization of DNA samples to image analysis. Therefore, in this paper, an attempt is made to robustify the Gaussian NBC by the minimum β-divergence method. The role of minimum β-divergence method in this article is to produce the robust estimators for the location and scale parameters based on the training dataset and outlier detection and modification in test dataset. The performance of the proposed method depends on the tuning parameter β. It reduces to the traditional naïve Bayes classifier when $\beta \to 0$. We investigated the performance of the proposed beta naïve Bayes classifier (β-NBC) in a comparison with some popular existing classifiers (NBC, KNN, SVM, and AdaBoost) using both simulated and real gene expression datasets. We observed that the proposed method improved the performance over the others in presence of outliers. Otherwise, it keeps almost equal performance.

1. Introduction

Classification is a supervised learning approach for separation of multivariate data into various sources of populations. It has been playing significant roles in bioinformatics by class prediction or pattern recognition from molecular OMICS datasets. Microarray gene expression data analysis is one of the most important OMICS research wings for bioinformatics [1]. There are several classification and clustering approaches that have been addressed previously for analyzing MGED [2–11]. The Gaussian linear Bayes classifier (LBC) is one of the most popular classifiers for class prediction or pattern recognition. However, it is not so popular for microarray gene expression data analysis, since it suffers from the inverse problem of its covariance matrix in presence of large number of genes (p) with small number of patients/samples (n) in the training dataset. The Gaussian naïve Bayes classifier (NBC) overcomes this difficulty of Gaussian LBC by taking the normality and independence assumptions on the variables. If these two assumptions are violated, then the nonparametric version of NBC is suggested in [12]. In this case the nonparametric classification methods work well but they produce poor performance for small sample sizes or in presence of outliers. In MGED the small samples are conducted because of cost and limited specimen availability [13]. There are some other versions of NBC also [14, 15]. However, none of them are so robust against outliers. It is one of the most important drawbacks for gene expression data analysis by the existing NBC. The gene expression dataset is often contaminated by outliers due to several steps involved in the data generating process from hybridization of DNA samples to image analysis. Therefore, in this paper, an attempt is made to robustify the Gaussian NBC by the minimum β-divergence method within two steps. At step-1, the minimum β-divergence method [16–18] attempts to estimate the parameters for the Gaussian NBC based on the

antibiotics

MDPI

Article

Antibiotics Use and Its Knowledge in the Community: A Mobile Phone Survey during the COVID-19 Pandemic in Bangladesh

Zubair Akhtar [1,*], Syeda Mah-E-Muneer [1], Md. Mahbubur Rashid [1], Md. Shakil Ahmed [1], Md. Ariful Islam [1], Sukanta Chowdhury [1], Zobaid Khan [2], Md. Zakiul Hassan [1,3], Khaleda Islam [2], Shahana Parveen [1], Nitish Debnath [2], Mahmudur Rahman [1,4] and Fahmida Chowdhury [1]

1 Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh (icddr,b), Dhaka 1212, Bangladesh; mahe@icddrb.org (S.M.-E.-M.); mahbubur.rashid@icddrb.org (M.M.R.); shakil.statru@gmail.com (M.S.A.); arif@icddrb.org (M.A.I.); sukanta@icddrb.org (S.C.); zhassan@icddrb.org (M.Z.H.); shahana@icddrb.org (S.P.); mrahman@globalhealthdev.org (M.R.); fahmida_chow@icddrb.org (F.C.)

2 Fleming Fund Country Grant to Bangladesh, DAI Global, LLC, House 3, First Floor, Road 23B, Gulshan 1, Dhaka 1212, Bangladesh; Zobaid_Khan@dai.com (Z.K.); Khaleda_Islam@dai.com (K.I.); Nitish_Debnath@dai.com (N.D.)

3 Nuffield Department of Medicine, University of Oxford, Oxford OX1 2JD, UK

4 Global Health Development, EMPHNET, 69 Mohakhali, Dhaka 1212, Bangladesh

* Correspondence: zakhtar@icddrb.org

Citation: Akhtar, Z.; Mah-E-Muneer, S.; Rashid, M.M.; Ahmed, M.S.; Islam, M.A.; Chowdhury, S.; Khan, Z.; Hassan, M.Z.; Islam, K.; Parveen, S.; et al. Antibiotics Use and Its Knowledge in the Community: A Mobile Phone Survey during the COVID-19 Pandemic in Bangladesh. *Antibiotics* **2021**, *10*, 1052. https://doi.org/10.3390/antibiotics10091052

Academic Editor: Seok Hoon Jeong

Received: 5 August 2021
Accepted: 27 August 2021
Published: 29 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: The general population has been excessively using antibiotics during the COVID-19 pandemic. Therefore, the use of antibiotics for any reported illnesses in the preceding four weeks and knowledge of antibiotics among the general population in the community were assessed for possible interventions. A mobile phone survey among a general population across eight administrative divisions of Bangladesh was conducted during January–March 2021. Reported illness episodes irrespective of COVID-19 in the preceding four weeks of the interview, use of antibiotics for the illnesses, and knowledge on antibiotics among the general population were recorded. Descriptive analyses were performed. We randomly interviewed 1854 participants, with a mean age of 28.5 years (range: 18–75 years); 60.6% were male. Among all participants, 86.3% (95% CI: 84.7–87.8) heard names of antibiotics, but only 12.1% reported unspecified harmful effects, and 3.5% reported antimicrobial resistance when antibiotics were taken without a physician's prescription. Among 257 (13.9%) participants, who consumed medicines for their recent illness episode, 32.7% (95% CI: 27.2–38.6) reported using antibiotics. Of those who could recall the names of antibiotics prescribed (n = 36), the most frequently used was azithromycin (22.2%) followed by cefixime (11.1%) and ciprofloxacin (5.6%). Our findings show an increased antibiotic use for illnesses reported in the preceding four weeks and an elevated knowledge at the community level during the COVID-19 pandemic compared with the pre-pandemic period.

Keywords: antibiotic resistance; COVID-19; pandemic; antibiotic awareness; antibiotic use

1. Introduction

Modern healthcare is predominantly reliant on antibiotic treatment [1], but there has been a phenomenal and imprudent use of antibiotics leading to the advent of resistant strains of bacteria [2,3]. To address the antimicrobial resistance (AMR) as a whole, with more emphasis on antibiotic resistance, the World Health Organization (WHO) initiated a range of AMR-related activities, including the development of the Global Action Plan on Antimicrobial Resistance (GAP-AMR) by the 68th World Health Assembly in May 2015 [4]. Published literatures exhibit a high proportion of inappropriate use of antimicrobials,