RGC Ref. No.:
UGC/FDS11/E04/16
(please insert ref. above)

RESEARCH GRANTS COUNCIL
COMPETITIVE RESEARCH FUNDING SCHEMES FOR THE LOCAL SELF-FINANCING DEGREE SECTOR

FACULTY DEVELOPMENT SCHEME (FDS)

Completion Report
(for completed projects only)

Submission Deadlines:
1. *Auditor's report with unspent balance, if any: within **six** months of the approved project completion date.*
2. *Completion report: within **12** months of the approved project completion date.*

Part A: The Project and Investigator(s)

1. Project Title

Structural Modeling, Characterization and Analysis of New Receptor-Ligand Systems: Applying Computational Methods in Molecular Binding Affinity Analysis

2. Investigator(s) and Academic Department(s) / Unit(s) Involved

Research Team	Name / Post	Unit / Department / Institution
Principal Investigator	HO Wai-shing / Assistant Professor	School of Computing and Information Sciences / Caritas Institute of Higher Education
Co-Investigator(s)		
Others		

3. Project Duration

	Original	Revised	Date of RGC / Institution Approval *(must be quoted)*
Project Start Date	1/10/2016		
Project Completion Date	30/9/2019	31/3/2020	Institution Approval granted on 16/7/2019
Duration *(in month)*	36	42	Institution Approval granted on 16/7/2019

Deadline for Submission of Completion Report	30/9/2020	31/3/2021	

Part B: The Final Report

5. Project Objectives

5.1 Objectives as per original application

1. *To structurally model an irreversible or hybrid receptor-ligand system based on efficient computational techniques.*
2. *To collect effective characterization methods for classic reversible binding systems, and to develop new/adapted methods for irreversible/hybrid ones.*
3. *To reasonably estimate the binding affinity of a new system based on different characterizations, and to further predict the affinity for a variant system.*
4. *To integrate the involved algorithms into a comprehensive software/program suite that comprises structural modeling, characterization, evaluation and post-analysis of various receptor-ligand binding systems, and to further develop compatible interfaces to classic molecular-structural-analysis tools.*
5. *To provide valuable experiences for teachers and undergraduates to develop useful skills in molecular structural analysis, programming, and algorithm design, and to further improve their computer application skills in a specialized field.*

5.2 Revised objectives

Date of approval from the RGC: N/A

Reasons for the change:

1.

2.

3. ....

5.3 Realisation of the objectives

(Maximum 1 page; please state how and to what extent the project objectives have been achieved; give reasons for under-achievements and outline attempts to overcome problems, if any)

Objective 1. For this objective, two methods for visualising gigantic medical structures were proposed. For spatial medical data collected from computed tomography (CT) or 3D scanners, two learning based methods to improve the quality of the visualisation were proposed. First, a unilateral filtered facet normal descriptor (uFND) for measuring the geometry features around each facet of a given mesh was designed and such features would be learned through regression. Second, a novel algorithm based on Iterative Closest Points (ICP) were proposed to optimise the display of different parts together through feature selection and suitable training. These two methods significantly improved the visualisation of medical images.

Objective 2. Protein-protein interaction is an important problem for locating potentially useful characteristics of proteins by studying its structure. Good prediction algorithms can save us inefficient and costly experimental work for studying PPI. In the project, a boosting based algorithm was proposed for the imbalanced prediction problem of protein-ligand sites. With the use of stochastic sensitivity tree boosting algorithm (SSTBoost), the performances of PPI site prediction improved significantly against other state-of-the-art prediction algorithms. This helped us in discovering previously not known PPI sites in a more efficient way.

Objective 3. During the analysis of the receptor-ligand systems, graphical models have been widely used to learn the conditional dependence structures among random variables. Such graphs may be constructed from data collected in multiple controlled experiments. However, most existing models are developed for estimating a single graph only. Thus, a new joint differential network analysis (JDNA) model for jointly estimate multiple differential networks with latent variables from multiple datasets was developed in this project. Experiments showed that JDNA model was effective in identifying differential networks under different conditions.

In addition, the architecture selection problem of artificial neural network was studied. The study focused at the process of evaluating, ranking and making choices form a set of network structures. Choosing the correct structure to be used in our future learning would be essential as that would be critical to the final quality of our solutions. We applied this method to the PPI site prediction problem above.

These two had significant potential in estimating the binding affinity of ligand-receptor systems.

Objective 4. All the programs developed above could be used in the process of identifying potentially useful PPI and analysing graphs. Those programs were integral parts for the integrated software suite.

Objective 5. Our work on PPI site prediction was summarised in lecture notes. This is a good example for our students on the application of boosting techniques. Our School has started a new Bachelor of Science degree in Artificial Intelligence in 2020. Our work will be part of the efforts in showing the application of machine learning techniques to different areas.

5.4 Summary of objectives addressed to date

Objectives *(as per 5.1/5.2 above)*	**Addressed** *(please tick)*	**Percentage Achieved** *(please estimate)*
1. *To structurally model an irreversible or hybrid receptor-ligand system based on efficient computational techniques.*	✓	100%
2. *To collect effective characterization methods for classic reversible binding systems, and to develop new/adapted methods for irreversible/hybrid ones.*	✓	100%
3. *To reasonably estimate the binding affinity of a new system based on different characterizations, and to further predict the affinity for a variant system.*	✓	100%
4. *To integrate the involved algorithms into a comprehensive software/program suite that comprises structural modeling, characterization, evaluation and post-analysis of various receptor-ligand binding systems, and to further develop compatible interfaces to classic molecular-structural-analysis tools.*	✓	100%
5. *To provide valuable experiences for teachers and undergraduates to develop useful skills in molecular structural analysis, programming, and algorithm design, and to further improve their computer application skills in a specialized field.*	✓	100%

6. Research Outcome

6.1 Major findings and research outcome
(Maximum 1 page; please make reference to Part C where necessary)

Our research outcome can be divided into three main parts. First, Dr. M Wei and our team mainly studied a number of methods in visualising gigantic biological structures. The work were presented in first two journal articles stated in Part C. The second part was about the analysis of features in protein-ligand interactions. The application of stochastic sensitivity tree boosting to protein-protein interaction and protein-ligand site prediction was recorded in the fourth journal article stated in Part C. The application of joint differential network analysis on networked data, including the PPI sites, were presented in the fifth journal article in Part C. The third part was on the application of multiple-criteria based optimisation to the evaluating, ranking and making choices from a set of artificial neural network structures. This work is presented in the third journal article in Part C.

6.2 Potential for further development of the research and the proposed course of action *(Maximum half a page)*

We applied stochastic sensitivity tree boosting to refine the data for PPI site prediction. The result was promising. One natural extension is to try using other existing boosting strategies to study the improvements to the prediction result. This should help us further reduce the work wastes in experiments on fruitless PPI.

Another potential further development was on the visualisation techniques for gigantic structures of receptor-ligand systems. The methods we developed worked well on other medical images whilst the experience gained in our project could lead to other ways for visualising the structures and collect important characteristics from the visualisation.

7. Layman's Summary

(Describe in layman's language the nature, significance and value of the research project, in no more than 200 words)

Through this research project, two novel algorithms for visualising gigantic 3D medical images were developed. They can improve the visualisation quality by applying machine learning based training. They helped in visualising the gigantic protein structures. Structural visualisation is an important aspect in ligand modelling.

Protein-protein interactions (PPIs) are the molecular basis for many biological processes. Identifying the locations of the interactions is an important step in understanding those processes. Experimental methods to solve PPI sites are expensive and time-consuming, thus different kinds of prediction algorithms were proposed. The research team proposed a stochastic sensitivity tree boosting algorithm for refining data for PPI site prediction. The method has significant better results than other state-of-an-art methods in terms of recall and other quality measures.

The research team also studied the machine learning algorithms applicable to protein analysis. We proposed a new method for evaluating, ranking and making choices from a set of network structures for artificial neural networks (ANNs). Moreover, we also proposed the joint differential network analysis (JDNA) model for analysing graphical models of condition dependencies. It helped the analysis of relationships between variables over different controlled experiments.

Part C: Research Output

8. Peer-Reviewed Journal Publication(s) Arising Directly From This Research Project
(Please attach a copy of the publication and/or the letter of acceptance if not yet submitted in the previous progress report(s). All listed publications must acknowledge RGC's funding support by quoting the specific grant reference.)

The Latest Status of Publications				Author(s) *(denote the corresponding author with an asterisk*)*	Title and Journal / Book *(with the volume, pages and other necessary publishing details specified)*	Submitted to RGC *(indicate the year ending of the relevant progress report)*	Attached to this Report *(Yes or No)*	Acknowledged the Support of RGC *(Yes or No)*	Accessible from the Institutional Repository *(Yes or No)*
Year of Publication	Year of Acceptance *(For paper accepted but not yet published)*	Under Review	Under Preparation *(optional)*						
2017				WANG Ran*, XIE Haoran, FENG Jianqiang, WANG Fu Lee, XU Chen	Multi-criteria decision making based architecture selection for single-hidden layer feedforward neural networks, *International Journal of Machine Learning and Cybernetics*, vol. 10, pp 655-666, https://doi.org/10.1007/s13042-017-0746-9, Springer	Yes	No	Yes	
2017				LIANG Luming, WEI Mingqiang*, SZYMCZAK Andrzej, PETRELLA Anthony, XIE Haoran, QIN Jing, WANG Jun,	Nonrigid iterative closest points for registration of 3D biomedical surfaces, *Optics and Lasers in Engineering*, vol. 100, pp 141-154,	Yes	No	Yes	

				WANG Fu-lee	http://dx.doi.org/10.1016/j.optlaseng.2017.08.005, Elsevier				
2017				WEI Mingqiang, WANG Jun, GUO Xianglin, WU Huisi, XIE Haoran*, WANG Fu-lee, QIN Jing	Learning-based 3D surface optimization from medical image reconstruction, *Optics and Lasers in Engineering*, vol. 103, pp 110-118, https://doi.org/10.1016/j.optlaseng.2017.11.014, Elsevier	Yes	No	Yes	
2019				NG Wing W. Y., WANG Debby D., ZHANG Jianjun*, WANG Fu Lee, ZHANG Yuda	Stochastic Sensitivity Tree Boosting for Imbalanced Prediction Problems of Protein-ligand Interaction Sites, *IEEE Transactions on Emerging Topics in Computational Intelligence*, https://doi.org/10.1109/TETCI.2019.2922340, IEEE	No	Yes (Attachment 1)	Yes	
2019				OU-YANG Le, ZHANG Xiao-Fei*, ZHAO Xing-Ming,	Joint Learning of Multiple Differential Networks with Latent Variables,	No	Yes (Attachment 2)	Yes	

				WANG Debby D., WANG Fu Lee, LEI Baiying, YAN Hong	***IEEE Transactions on Cybernetics*, vol. 49, no. 9, pp 3494-3506**				

9. Recognized International Conference(s) In Which Paper(s) Related To This Research Project Was / Were Delivered
(Please attach a copy of each conference abstract)

Month / Year / Place	**Title**	**Conference Name**	**Submitted to RGC** *(indicate the year ending of the relevant progress report)*	**Attached to this Report** *(Yes or No)*	**Acknowledged the Support of RGC** *(Yes or No)*	**Accessible from the Institutional Repository** *(Yes or No)*
NIL						

10. Whether Research Experience And New Knowledge Has Been Transferred / Has Contributed To Teaching And Learning
(Please elaborate)

The research work has been summarised in various presentations for students to understand the skills and techniques used in the development of the research. Moreover, the parts for boosting on PPI site prediction will be included the lecture notes as an example application of boosting to different types of biological problems.

11. Student(s) Trained

(Please attach a copy of the title page of the thesis)

Name	Degree Registered for	Date of Registration	Date of Thesis Submission / Graduation
N/A			

12. Other Impact

(e.g. award of patents or prizes, collaboration with other research institutions, technology transfer, teaching enhancement, etc.)

NIL

13. Statistics on Research Outputs

	Peer-reviewed Journal Publications	Conference Papers	Scholarly Books, Monographs and Chapters	Patents Awarded	Other Research Outputs (please specify)	
No. of outputs arising directly from this research project	5	0	0	0	Type	No.

14. Public Access Of Completion Report

(Please specify the information, if any, that cannot be provided for public access and give the reasons.)

Information that Cannot Be Provided for Public Access	Reasons
NIL	