

# Contents

|   | <i>page</i> |
|---|-------------|
| Abstract  | i           |
| Acknowledgements  | lii         |
| Contents  | v           |
| List of Figures   | vii         |
| List of Tables  | ix          |
| List of Symbols   | xi          |
| List of Abbreviations   | xiii        |
| <b>Chapter 1: Introduction</b>  | <b>1</b>    |
| 1.1 Motivation  | 1           |
| 1.3 Objectives  | 1           |
| 1.4 Contribution  | 2           |
| 1.5 Outline of the thesis   | 2           |
| <b>Chapter 2: Literature Review</b>   | <b>3</b>    |
| 2.1 Assessing adverse drug reactions  | 3           |
| 2.2 Computational Strategies used in Drug Discovery   | 4           |
| 2.3 Computational Models for Side Effects Prediction  | 5           |
| 2.4 Systems modeling  | 7           |
| 2.5 Graph theoretical modeling  | 7           |
| 2.6 Biological space  | 8           |
| 2.7 Feature selection   | 9           |
| 2.8 Prediction of absent interaction  | 12          |
| <b>Chapter 3: Data Collection and Curation</b>  | <b>13</b>   |
| 3.1 Integration of DrugBank and SIDER2  | 13          |
| 3.2 Statistics of DrugBank3.0 and SIDER   | 15          |
| 3.3 Chemical Descriptors  | 17          |
| 3.3.1 Computational representation  | 17          |
| 3.3.2 Chemical properties   | 17          |
| <b>Chapter 4: Mathematical and Computational techniques</b>   | <b>23</b>   |
| 4.1 Canonical Correlation Analysis  | 23          |
| 4.1.1 Unconstrained optimization  | 23          |
| 4.1.2 Constrained optimization  | 24          |
| 4.2 Generalized Canonical Correlation analysis  | 26          |
| 4.2.1 Unconstrained optimization  | 26          |
| 4.2.2 Constrained optimization  | 27          |
| 4.3 Generalized Partial Canonical Correlation analysis  | 28          |
| <b>Chapter 5: Phenotypic side effects prediction by optimizing correlation with chemical and target profiles of drugs</b> | <b>31</b>   |
| 5.1 Simplified GCCA   | 32          |
| 5.2 Results   | 33          |
| 5.2.1 Drug-target profile   | 34          |
| 5.2.2 Drug-chemical profile   | 35          |
| 5.2.3 Effects of increasing number of eigenvectors  | 40          |
| 5.3 Performance evaluation  | 40          |
| 5.4 Conclusions   | 40          |
| <b>Chapter 6: Role of secondary neighbors of targets in specifying phenotypic side effects of drugs</b>                   | <b>43</b>   |
| 6.1 Genomic Space   | 44          |
| 6.2 Measuring secondary interactions using graph theoretical metrics  | 45          |
| 6.3 Representation of features of drug profile  | 47          |
| 6.4 Results   | 47          |

|   |    |
|---|----|
| <b>Chapter 7: Predicting hidden drug reactions using partial information of side effect classes</b>                         | 49 |
| 7.1 Organ/system-based classes of side effects  | 50 |
| 7.2 Identification of set of crucial 'known side effects'   | 53 |
| 7.3 Predicting hidden side effects  | 55 |
| 7.2.1 Blood   | 56 |
| 7.2.2 Eye   | 58 |
| 7.2.3 Abdominal   | 59 |
| 7.2.4 Hepatic   | 61 |
| 7.2.5 Gastrointestinal  | 63 |
| 7.2.6 Kidney  | 65 |
| 7.2.7 Nose  | 66 |
| 7.4 Discussion  | 67 |
| <b>Chapter 8: A partial canonical correlation model for identification of drug features specific to side effect classes</b> | 69 |
| 8.1 Materials   | 69 |
| 8.2 Methods   | 71 |
| 8.3 Performance evaluation  | 73 |
| 8.4 Result  | 75 |
| 8.4.1 Strategy for identification of 2D features specific to organ/class  | 76 |
| 8.4.2 Chemical features contributing to Eye related side effects  | 77 |
| 8.4.3 Chemical features contributing to Nose related side effects   | 78 |
| 8.4.4 Chemical features contributing to Abdomen related side effects  | 79 |
| 8.5 Discussion  | 78 |
| <b>Chapter 9: Summary</b>   | 81 |
| <b>Annexure A: Formulae for numerator layout derivation</b>   | 83 |
| <b>Annexure B: Applications of PCCA for finding class-specific features</b>   | 85 |
| <i>References</i>   | 93 |