**TABLE OF CONTENT**

| **Chapter No.** | **Title** | | **Page No.** | |
| --- | --- | --- | --- | --- |
|  | | **INTRODUCTION** | | 1 |
|  | | **PROBLEM STATEMENT** | | 4 |
|  | | **LITERATURE REVIEW** | | 6 |
|  | | 3.1 Background on Embeddings and Drug Repurposing | | 6 |
|  | | 3.1.1 Seq2seq Fingerprint: An Unsupervised Deep Molecular Embedding for Drug Discovery | | 6 |
|  | | 3.1.2 Predicting New Molecular Targets for Known Drug | | 7 |
|  | | 3.1.3 Drug Target Identification Using Side-Effect Similarity | | 7 |
|  | | 3.2 Generating Embeddings for SMILES | | 8 |
|  | | 3.2.1 SMILES2Vec: An Interpretable General-Purpose Deep Neural Network for Predicting Chemical Properties | | 8 |
|  | | 3.2.2 SPVec: A Word2vec-Inspired Feature Drug-Target Interaction Prediction | | 8 |
|  | | 3.2.3 Mol2vec: Unsupervised Machine Learning Approach with Chemical Intuition | | 9 |
|  | | 3.2.4 SWeeP: Representing Large Biological Sequences Datasets in Compact Vectors | | 9 |
|  | | 3.3 Generating Embeddings for Gene Expression | | 10 |
|  | | 3.3.1 Drug Repurposing Using Deep Embeddings of Gene Expression Profiles | | 10 |
|  | | 3.3.2 A Large-Scale Gene Expression Intensity-Based Similarity Metric for Drug Repositioning | | 11 |
|  | | 3.3.3 Gene2vec: Distributed Representation of Genes Based on Co-Expression | | 11 |
|  | | 3.3.4 Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data | | 11 |
|  | | 3.3.5 Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics | | 12 |
|  | | 3.3.6 Drug Repositioning: A Machine-Learning Approach Through Data Integration | | 12 |
|  | | 3.4 Adverse Drug Reactions | | 13 |
|  | | 3.4.1 Systematic Drug Repositioning Based on Clinical Side- Effects | | 13 |
|  | | 3.4.2 Predicting ADR of Combined Medication from Heterogeneous Pharmacologic Databases | | 13 |
|  | | 3.4.3 Predicting Adverse Drug Reactions Through Interpretable Deep Learning Framework | | 14 |
|  | | 3.4.4 Detecting Potential Adverse Drug Reactions Using a Deep Neural Network Model | | 15 |
|  | | **DATA** | | 17 |
|  | | 4.1 Overview | | 17 |
|  | | 4.2 SMILES: Structural Indication of a Drug | | 17 |
|  | | 4.3 LINCS: Functional Indication of a Drug | | 18 |
|  | | 4.4 Combined Feature Set for ADR Detection | | 19 |
|  | | 4.5 Classification Systems | | 21 |
|  | | 4.5.1 ATC: Anatomical Therapeutic Chemical Classification | | 21 |
|  | | 4.5.2 SIDER: Side Effects Resource | | 23 |
|  | | 4.5.3 Adverse effects combined dataset | | 24 |
|  | | **SYSTEM REQUIREMENTS SPECIFICATION** | |  |
|  | | **SYSTEM DESIGN (detailed)** | |  |
|  | | **IMPLEMENTATION AND PSEUDOCODE (if applicable)** | |  |
|  | | **CONCLUSION OF CAPSTONE PROJECT PHASE-1** | |  |
|  | **PLAN OF WORK FOR CAPSTONE PROJECT PHASE-2** | | |  |
| **REFERENCE/ BIBLIOGRAPHY** | | | |  |
| [**APPENDIX A DEFINITIONS, ACRONYMS AND**](https://docs.google.com/document/d/1sxLzRaxGh0fTgzlXiH-5RgoxpmOZgtX9RD5OH3Dfigw/edit#heading%3Dh.qzpp4n2gssp1) **ABBREVIATIONS** | | | |  |
| [**APPENDIX B**](https://docs.google.com/document/d/1sxLzRaxGh0fTgzlXiH-5RgoxpmOZgtX9RD5OH3Dfigw/edit#heading%3Dh.mywen4vn97v4) **USER MANUAL (OPTIONAL)** | | | |  |

**LIST OF FIGURES**

| **Figure No.** | | **Title** | **Page No.** | |
| --- | --- | --- | --- | --- |
| Figure 1 | ADRs caused by drugs when used separately as compared to used together | | | 8 |
| Figure 2 | Representation of the L1000 Assay | | | 25 |
| Figure 3 | Combined feature set for ADR prediction | | | 27 |
| Figure 4 | Representation of the combine feature set | | | 27 |
| Figure 44 | Graph of testing and training losses | | | 85 |

**LIST OF TABLES**

| **Table No.** | | **Title** | **Page No.** | |
| --- | --- | --- | --- | --- |
| Table 1 | [Statistics of the L100 Assay](#_1fob9te) | | | 23 |