UTM SCHOOL OF COMPUTING Faculty of Engineering	**FACULTY OF ENGINEERING** **UNIVERSITI TEKNOLOGI MALAYSIA**

TECHNOLOGY & INFORMATION SYSTEM
(SECP1513)

PROJECT LOW FIDELITY PROTOTYPE: PART 1

Submitted To:

DR. AZURAH BINTI ABU SAMAH

From:

GROUP 3			
No.	Name	Matric no.	Section
1	AINA NAJWA BINTI MOHD ROZI	A21EC0010	1
2	LOKE RUI KEE	A20ET0329	
3	MUHAMMAD AZAMUDDIN NASUTION BIN RADUAN	A21EC0069	
4	NUR AISYAH FATIHAH BINTI MOHAMED ROZI	A21EC0107	
5	SOFIA ANAK HENRY	A21EC0133	

ACKNOWLEDGEMENT

First and foremost, we would like to express our gratitude to Allah SWT as He gives us a chance to finish this assignment and manage to give the best on this assignment till the end. We also would like to thank our beloved TIS lecturer, Dr Azurah binti A Samah for the valuable guidance and advice. We also would like to express our gratitude to Dr Azurah for giving us guidance in this assignment despite her hectic schedule. We appreciate it a lot.

Besides, we would like to express our gratitude to our family members, classmates and friends who give their support for us directly or indirectly in this assignment until it is fully complete.

Finally, an honourable mention goes to all the team members in the group for their commitment and hard work in completing this assignment. Without the great teamwork among us, we would face many difficulties while doing this assignment. Thankfully with the great teamwork in the group, the process in completing this assignment went smoothly. As the saying goes, many hands make light work.

Once again, a million thanks to everyone who contributes in this assignment directly and indirectly. With the support from everyone, we finally completed this assignment.

Table of Contents

1. Introduction

Bioinformatics is a subfield of biology and computer science. It is often involved in data analysis, modelling and software development. Most often, these approaches are used for various applications in predicting shapes of proteins, determining gene and protein functions by storing, analysing and disseminating biological data from the source, particularly amino acid and deoxyribonucleic acid (DNA), ribonucleic acid (RNA), and protein sequences. (National Human Genome Research Institute, n.d.)

This is where artificial intelligence (AI) techniques come into place. According to Microsoft, AI is a computer system that makes predictions or takes actions based on patterns in existing data and can then learn from its errors to increase its accuracy (Microsoft Azure, n.d.). Underneath AI, there is machine learning, which refers to the algorithm that an AI program possesses that makes it have the ability to learn things on its own and make predictions. As Industry 4.0 continues with the focus in smart manufacturing, the role of machine learning has been increasingly getting more attention and become significantly important.

Just like in any other fields, machine learning is able to give a rapid change to the development of bioinformatics, including the discovery of drugs. Drug discovery and development is the process whereby new drugs are discovered or developed to be used in treating diseases and are safe and effective to use (Deore, 2019). According to the article Applications of Machine Learning in Drug Discovery and Development (Vamathevan, 2019), it is said that machine learning is able to improve the discovery and decision making in drug discovery for well-specified questions, provided that the machine learning model is trained with high-quality data. With that being said, the application of machine learning can indeed reduce failure rates and speed up the process with its data-driven nature.

2. Report Content

2.0 Problem statement

As technology advanced, drug data grew larger, and drug discovery was shifted to machine learning. However, as time passed and data became larger and more complex, using machine learning produced some issues, and the detection of protein structures became less precise (Zhang, Lu & Tan, Jianjun & Han, Dan & Zhu, Hao. (2017)). And, in general, machine learning is not suitable for solving a complex quarry with a large amount of data because it can only handle structured data. However, as technology drastically rose, Artificial intelligence (AI) is also referred to as machine intelligence because machines can be trained or customized to perform tasks similar to those performed by the human brain (Poole et al. 1998; Vinod and Anand 2021; Gopal 2018). Protein structure-activity is thought to be useful in drug design. As a result of protein malfunctions, the human body has developed a plethora of impurities. The small different molecules of proteins are used to develop drug-design strategies. As a result, more people will die because of taking medications. As the fatality rate rises, the human population will decline, and our society will not grow. Because most protein structures are not structured, the goal of this study is to identify how deep learning techniques can improve the accuracy of the 3D structure of proteins (Miles J., Walker A.).

2.1 Suggested Solution

Artificial intelligence has been used to anticipate drug-protein interactions, identify drug efficacy, and ensure biomarkers for safety. By utilising deep learning techniques, we can transform our simple data into complex data. And one advantage of deep learning is that you make less errors if you stereotype excessively. A robot called Adam was used to code the gene sequence and, surprisingly, nine were new and accurate, while only one was inaccurate. Other than that, the german biotechnology company Evotec have also announced a phase 1 clinical trial on a new anticancer molecule, a result of collaboration with Exscientia, a 2012 spinoff-company of the University of Dundee that applies AI techniques to small molecule drug discovery. The discovery of the new anticancer molecule has been achieved through the use of Exscientia's "Centaur Chemist" AI design

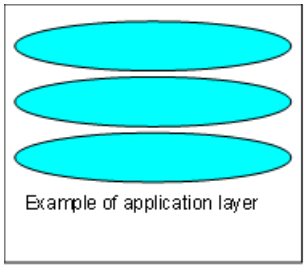
platform where millions of potential small molecules are computationally sorted and their various properties compared thoroughly, allowing them to select the potential molecules deemed most suitable to be synthesized, tested, and optimized before the eventual molecule is chosen. The anticancer molecule, which is an A2-receptor antagonist used to suppress tumor growth, would take 4-5 years in order to be discovered, had it being done through the traditional discovery processes, while it only took Exscientia 8 months to achieve the same amount of progress. In short, these examples serves as evidence of the immense potential of AI in the domain of drug discovery.

2.2 Potential Client

1. Pharmaceutical industry

It can automate the data in less than a second. Drug development will speed up, and data will be processed quickly. Because Al techniques have a lot of different skills, like categorization, analysis, and prediction, they can save you a lot of time and energy. Extending this into the pharmaceutical industry opens up a lot of possibilities. One of those possibilities is that data can be translated in a second. Because the machine's memory can read text, biometrics, and other types of data, we don't have to figure out what kind of data it is and the information is also precise.

3. Architecture Planning and Design



These layers are frequently referred to as UI, BLL (Business Logic Layer), and DAL (Data Access Layer) (Data Access Layer). Users submit requests via the UI layer, which only interacts with the BLL in this architecture. For data access requests, the BLL can then contact the DAL. The UI layer should not make any direct requests to the DAL or interact with persistence in any way. In the same way, the BLL should only interact with persistence via the DAL. As a result, each layer is assigned a well-defined responsibility.

In drug discovery, deep learning algorithms' numerous models are used such as Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Autoencoder.

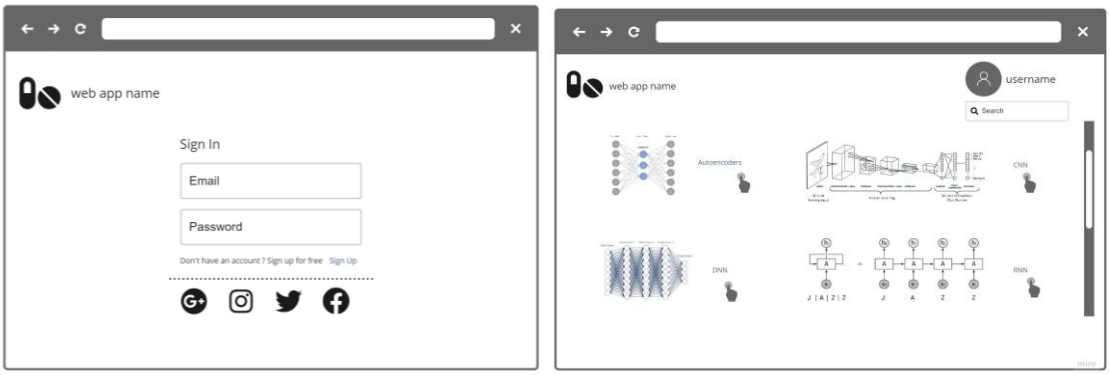
The DNN architecture arose from an extension of the Artificial Neural Network (ANN), which included multiple layers between input and output nodes. The DNN architecture tracks the outcomes in a mathematical model, which can be either non-linear or linear.

CNN is a subclass of DNN that is commonly used to analyse visual images. CNN is also known as shift-invariant ANN due to its reliance on weights. CNN is a multi-layer perceptron that has been regularised. Fully connected networks are characterised by the concept of multi-layer perceptron, in which each neuron in the first layer is associated with the next layer.

The RNN algorithm is a subset of the artificial neural network in which connections can form between the input and output nodes. In this manner, a directed graph and a temporal sequence can be created in the network. Similarly, the RNN network makes use of its internal memory to perform grouping in input variables.

An autoencoder is used to represent the encoding data format in dimensionality reduction in

order to keep a strategic distance from the network's 'noise' signal. Autoencoders are a type of artificial neural network that retrieves data through unsupervised learning. The autoencoder must explore the input data before copying it to the output layer. Autoencoders are made up of two parts: encoders and decoders, as well as a hidden layer



Example of web app wireframe design

4. Conclusion

Machine learning in drug discovery is not a new thing in the industry. There are lots of open-source models and tools in the field and the players – the developers and researchers that have contributed to the open-source, are not stopping any time soon. They are continuing their journey to develop more advanced machine learning assets for innovative drug discovery. TorchDrug by Mila and China research team for rapid prototyping (Mila, 2021), AlphaFold by Google's DeepMind to solve protein folding problem by predicting the protein's 3D structure (Pati, 2021), Open Drug Discovery Toolkit (ODDT) by researchers in Poland as extensible computer aided drug discovery package (Wójcikowski, 2015), PaDELPY library to calculate molecular fingerprints for building a machine learning model (Nantasenama, 2021), are all powerful drug discovery tools that has created an impact in life sciences. For instance, AlphaFold is used to predict the protein structure of the causative agent of COVID-19, SARS-CoV-2 and was pending experimental detection in early 2020 which is then later proven accurate. Hence, machine learning techniques are proven to be able to accelerate real-world problem solving with high accuracy of data-driven prediction, and have left several impactful contributions in the drug discovery industry.

References

Deore, A. &. (2019, December). The Stages of Drug Discovery and Development Process. *Asian Journal of Pharmaceutical Research and Development.* , 62-67.

Microsoft Azure. (n.d.). *What is artificial intelligence?* Retrieved from Microsoft Azure: https://azure.microsoft.com/en-au/overview/what-is-artificial-intelligence/

National Human Genome Research Institute. (n.d.). *Bioinformatics*. Retrieved from National Human Genome Research Institute: https://www.genome.gov/genetics-glossary/Bioinformatics

Vamathevan, J. C. (2019). Applications of machine learning in drug discovery and development. *Nat Rev Drug Discov 18*, 463-477.

Zhang, Lu & Tan, Jianjun & Han, Dan & Zhu, Hao. (2017). From machine learning to deep learning: Progress in machine intelligence for rational drug discovery.

Machine learning methods for property prediction in chemoinformatics: Quo Vadis?

J. Chem. Inf. Model., 52 (2012), pp. 1413-1437

Miles J., Walker A. The potential application of artificial intelligence in transport. *IEE Proc.-Intell. Transport Syst.* 2006;153:183–198.

Mila. (2021, August 11). A team led by Mila researcher Jian Tang launches TorchDrug, an open-source platform for drug discovery. Retrieved from Mila: https://mila.quebec/en/a-team-led-by-mila-researcher-jian-tang-launches-torchdrug-an-open-source-platform-for-drug-discovery/

Nantasenama, C. (2021, July 5). How to build machine learning models for drug discovery using PaDELPy. Retrieved from Data Professor: https://youtu.be/rEmDyZHz5U8

Pati, S. (2021, April 30). Top 6 AI-Powered Drug Discovery Tools In 2021. Retrieved from Analytics India Magazine: https://analyticsindiamag.com/top-6-ai-powered-drug-discovery-tools-in-2021/

Wójcikowski, M. Z. (2015). Open Drug Discovery Toolkit (ODDT): a new open-source player in the drug discovery field. . J Cheminform, 7-26.

Ardalis. (n.d.). Common web application architectures. Retrieved from https://docs.microsoft.com/en-us/dotnet/architecture/modern-web-apps-azure/common-web-application-architectures

Dara, S., Dhamercherla, S., Jadav, S. S., Babu, C. H. M., & Ahsan, M. J. (n.d.). Machine Learning in Drug Discovery: A Review. Artificial Intelligence Review, 1. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8356896/

Savage, N. (2021, May 27). *Tapping into the drug discovery potential of AI.* Biopharma Dealmakers.

https://www.nature.com/articles/d43747-021-00045-7