**AI-Driven Drug Screening for solving substance abuse**

Adhvik Chakraborty

Semester 1, Block 1

January 17, 2025

David Thompson Secondary School

# Abstract

Opioid use disorder (OUD) crisis involves unregulated use of synthetic opioids due to their psychotropic properties and is a well-established factor for drug-related deaths in British Columbia. In Vancouver, the crisis continues to worsen affecting individuals, families and communities resulting in an alarming number of deaths and putting enormous burden to the public health care system and safety of the city. The BC Coroners report recommended creating alternative safer drugs as an immediate means of combating this public health emergency.

The primarily target of opioids are the MOR, DOR and KOR protein receptors in the brain. When drugs or opioids are used, potential side-effects occur as the drug targets other critical proteins in the body. Safer alternatives to existing opioids can be developed by systematically unveiling its interaction with the opioid protein-protein interaction (PPI) network. This project addresses the OUD problem by screening and identifying known drugs (drug-repurposing) that can serve as a safer alternative to existing opioids by lowering side-effects and overdose risks. The central idea is to elucidate the complex effect of a drug by developing reliable machine learning models of the opioid receptors and their PPI network to predict the binding of a drug to various interacting proteins in the body. The most optimal ML model is developed by performing a thorough evaluation and tuning based on datasets available on chEMBL. Further, in this work an innovative drug screening technique has been devised by leveraging the system of ML models of MOR and its PPI network. The strategy has been validated by accurately characterizing existing Health Canada based opioids and their side effects. When applied to screen 11,915 drugs, a number of safer opioid substitutes were revealed and are also customized based on vulnerable heath conditions. The result of this project would provide clinicians with a variety of options, strategies, and flexibility to offer care to treat individuals with OUD. Therefore, the novel approach in this project aims to directly take a step towards the recommendation outlined in the BC coroners report.

Table of Contents

[Introduction 4](#_Toc187963764)

[Background Information 5](#_Toc187963765)

[Procedure 8](#_Toc187963766)

[Part 1: Understanding binding of proteins with ligands 8](#_Toc187963767)

[Part 2: Developing ML Models to predict binding of proteins and ligands 10](#_Toc187963768)

[Part 3: Understanding the PPI Network related to MOR 12](#_Toc187963769)

[Part 4: Drug Screening 15](#_Toc187963770)

[Lab Question 17](#_Toc187963771)

[Hypothesis 17](#_Toc187963772)

[Purpose 17](#_Toc187963773)

[Materials 17](#_Toc187963774)

[Data and Observations 18](#_Toc187963775)

[Part 2: ML Models to predict binding of proteins and ligands 20](#_Toc187963776)

[Part 3: Understanding the PPI Network related to MOR 25](#_Toc187963777)

[Part 4: Drug Screening 27](#_Toc187963778)

[Analysis 30](#_Toc187963779)

[Part 1: Understanding binding of proteins with ligands 30](#_Toc187963780)

[Part 2: ML Models to predict binding of proteins and ligands 30](#_Toc187963781)

[Part 3: Understanding the PPI Network related to MOR 32](#_Toc187963782)

[Part 4: Drug Screening 33](#_Toc187963783)

[Sources of Error 34](#_Toc187963784)

[Future Improvements 34](#_Toc187963788)

[Conclusion 35](#_Toc187963789)

[Acknowledgments 35](#_Toc187963791)

[References 35](#_Toc187963792)

[Appendix 37](#_Toc187963793)

# Introduction

Substance use is a significant global issue with far-reaching consequences, impacting individuals, families, and communities through health complications, social challenges, and economic burdens. In Vancouver, this crisis is particularly dire, as the city remains in a state of public health emergency. The severity of the problem has drawn national attention, becoming a central topic of debate during BC’s recent provincial election.

 The drug use crisis is rapidly worsening, with increasing rates of addiction and overdose. In 2021, approximately 40 million people globally, suffered from drug use disorders, nearly 50% more than a decade ago, with cannabis being the most frequently used substance. Over the past seven years, more than 13,000 lives have been lost to drug-related causes, and in 2023, an average of 6.9 deaths occurred daily in British Columbia.

This issue continues to influence public health priorities and drives effort toward effective interventions. As per the 2023 report by “BC Coroners Service Death Review Panel” it is estimated that as many as 225,000 people in BC remain at risk of unregulated drug injury or death. Their top recommendation and call to action mentions that “The immediate priority for action must be on elements of that system that can be rolled out quickly in order to save lives now. As the primary cause of the current crisis is the unregulated toxic drug supply, our urgent attention must be on creating access to alternatives to the unregulated drug supply for people who use drugs." The panel reaffirmed and recognized that in short term the solution of providing quality-controlled alternatives of sufficient quantity and potency will immediately reduce the risk to people who would otherwise access the substances they use through the unregulated drug supply. In addition to expanding what is available in the medical model, these substances could be responsibly provided without a prescription in a manner that also includes a robust system of oversight and evaluation to respond to concerns about individual risk, public health and public safety.

The report also mentions that the primary driver of death is the increased toxicity, volatility, and unpredictability of the unregulated drug supply. Various synthetic opioids poisoning continues to be the leading cause of unnatural death. At least 32,632 Canadians died from an opioid overdose between January 2016 and June 2022. In Vancouver’s Downtown Eastside, unregulated drug toxicity has led to a death rate over 12 times the provincial average, with fentanyl found in more than 85% of overdoses in 2018–2019.

Opioid Use Disorder (OUD) refers to long-term uncontrolled and continued use of various synthetic opioids despite significant consequences. Although there are existing solutions to help with opioid use disorder (OUD), they have too many side effects and often don't work as patients don't adhere to their treatment plan.

The objective of this project is to address the OUD problem by screening and identifying known drugs that can serve as a safer alternative to the existing opioids. A favorable outcome of this project would provide clinicians with a variety of options, strategies, and flexibility to offer care to treat individuals with OUD. Therefore, this project aims to directly take a step towards the recommendation outlined in the “BC Coroners Service Death Review Panel” report by discovering a safer supply of non-toxic drugs available to people who are at the risk of death due to dependence on illegal drug supply. The main idea of the project is to build machine learning models that can predict the binding affinity between ligands and proteins for drug repurposing. By predicting the binding affinities of a drug to the various receptors in the human body a suitable drug screening technique can be developed. Additionally, repurposing drugs for OUD will reduce costs and time since, developing new drugs is very expensive and multi-year effort.

This project includes: the background information. Then described are the lab question, hypothesis, and purpose of the project. Next, listed are the materials and procedure. Then described are the Data & Observations. Next, is the data analysis where the results of the data and observation are discussed. Finally, the lab report ends with the sources of error faced in this project, future improvements, the conclusion, the references, and the appendix.

# Background Information

While there are existing remedies and medications available to assist with Opioid Use Disorder (OUD), their effectiveness remains limited. Machine learning can help with the process of developing new solutions by reusing drugs originally designed for a different purpose.

Protein-drug binding is a fundamental biological process where a drug molecule (ligand) interacts with a specific site on a protein, often referred to as the binding site. This interaction is driven by various non-covalent forces, including hydrogen bonds, ionic interactions, hydrophobic effects, and van der Waals forces. The binding can alter the protein’s function, either enhancing or inhibiting its activity, which is the basis for the therapeutic effects of many drugs. The specificity and strength of these interactions are influenced by the three-dimensional structure of both the protein and the drug, highlighting the importance of molecular compatibility. Understanding protein-drug binding is critical in drug discovery, as it provides insight into how drugs exert their effects at the molecular level, enabling the design of more effective and targeted therapies.

AutoDock Vina is a widely used molecular docking software designed to predict the binding affinity such as drug candidates, to their target proteins. It is an open-source tool that offers a balance between computational efficiency and accuracy, making it a popular choice for virtual screening and drug discovery projects. AutoDock Vina uses a scoring function to estimate the binding free energy, helping researchers identify promising ligands for further experimental validation. Its ability to handle flexible ligands and receptors enables more realistic docking simulations, facilitating the exploration of molecular interactions critical for drug design. AutoDock Vina is an accessible and powerful resource in computational biology and cheminformatics.

Opioid receptors are G protein-coupled receptors (GPCRs) mediate the human body’s response to most hormones, neurotransmitters, drugs, and are involved in perception of vision, and taste. The 4 main opioid receptors are Mu-Opioid Receptor (MOR), Kappa-Opioid Receptor (KOR), Delta-Opioid Receptor (DOR) and Nociceptin receptor (NOP). Each receptor has distinct functions and effects on the body, influencing pain management, mood, and various physiological responses.

MOR is one of the primary opioid receptors in the body, playing a critical role in pain management, reward, and addictive behaviors. It is widely distributed in the central and peripheral nervous systems, with high concentrations in brain regions associated with pain perception, stress, and emotions, such as the thalamus, brainstem, and limbic system. Activation of MOR by opioids like morphine or fentanyl results in potent analgesic effects, making it a key target for managing severe pain. However, MOR activation is also linked to several adverse effects, including respiratory depression, euphoria, physical dependence, and tolerance, which contribute to the high risk of opioid misuse and overdose. Because of its central role in both pain relief and opioid addiction, MOR remains a critical focus in the search for safer and more effective therapies to manage pain and treat Opioid Use Disorder (OUD).

An agonist is a substance that binds to a specific receptor and activates it to produce a biological response. An opioid agonist binds to opioid receptors in the brain and other parts of the body, triggering effects such as pain relief, euphoria, or sedation. These effects are typically the desired outcomes for medical treatments like pain management. An antagonist refers to a substance that binds to opioid receptors in the brain but does not activate them, thereby blocking or inhibiting the effects of opioid agonists

AI stands for artificial intelligence, and it is a field of study in computer science that develops machines with intelligence. Machine Learning (ML), a subset of AI, is a computer algorithm that focuses on analyzing and finding patterns in data to allow learning, decision–making and reasoning for that set of data. In contrast to an algorithm-based software design where every event and scenario is precisely mentioned, ML deals with designing a computer model that learns to perform its task on its own. Some of the most common learning methods are supervised learning, unsupervised learning, and reinforcement learning. Once the model learns then the model can be used to do the prediction for an input it has never seen before. This is called ML inference and is used by the general users.

Regression in ML is a type of supervised learning used to predict a continuous numerical target based on input features. It models the relationship between the independent variables the features and the dependent variable the target. Common types include Linear regression, Logistic Regression, and Polynomial Regression. The goal is to minimize the error between predicted and actual values, often measured by metrics like Mean Squared Error (MSE).

A decision tree is a type of regression algorithm in ML. It splits data into branches based on feature values, it resembles a tree structure with a root node, branches, and leaf nodes where each node represents a decision.

The Random Forest Regressor is a collection of decision trees to predict continuous target values. It works by creating a "forest" of decision trees, each trained on a random subset of the data and using a random subset of features. Random Forest handles non-linear relationships well, is robust to noise, and performs effectively on large datasets.

A particular function or a disease in an organism is governed by a group of proteins. A Protein-Protein Interaction (PPI) network is a representation of the interactions between proteins within a cell or organism. PPI networks are crucial for understanding cellular mechanisms, signaling pathways, and disease mechanisms. By analyzing these networks, researchers can identify key proteins involved in specific functions or diseases. Nodes represent proteins. Edges represent interactions between these proteins.

Human Ether-à-go-go-Related Gene (hERG) is a gene that makes a protein important for controlling the electrical activity in the heart. This protein forms part of a channel that helps the heart cells return to their normal state after each beat. If this channel doesn't work correctly, it can cause the heart to beat irregularly or too fast, which can be dangerous. A drug should not block or affect the hERG channel significantly, as blocking this channel can lead to serious heart problems, like irregular heartbeats or arrhythmias. Additionally, a suitable, and safe drug should have a low binding affinity with hERG.

The figure below shows a random forest machine learning model.

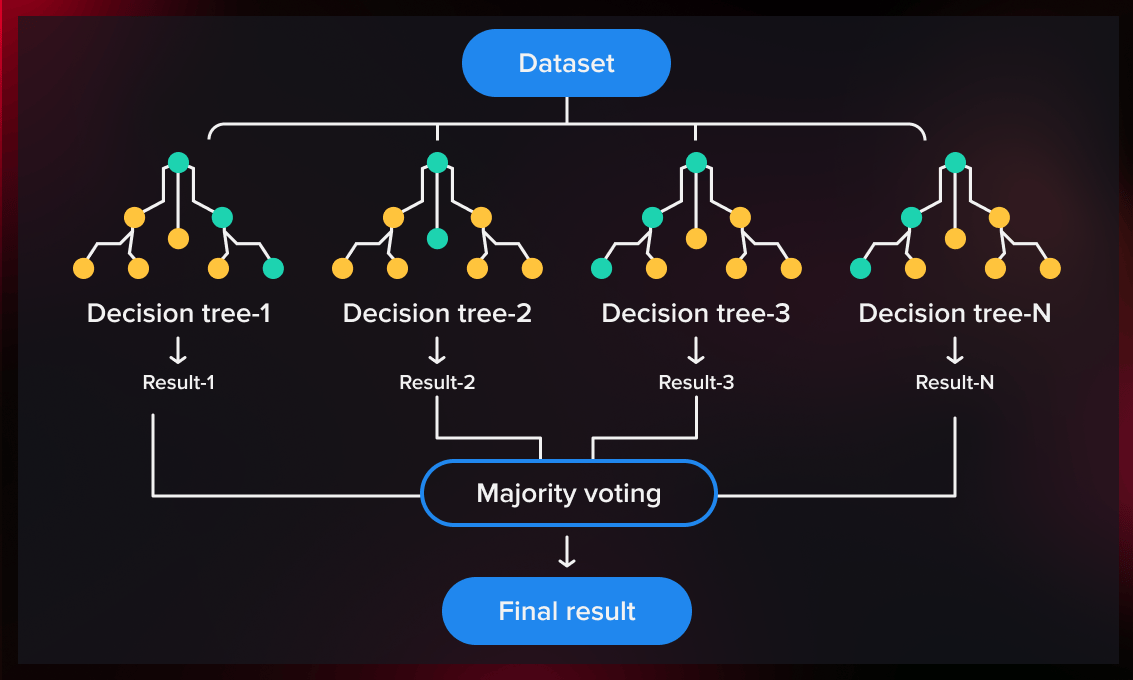


Figure 1: Random Forest Model, Taken From: https://serokell.io/blog/random-forest-classification

# Procedure

## Part 1: Understanding binding of proteins with ligands

Protein-ligand binding occurs when a small molecule (ligand) attaches to a protein, often inducing a change in the protein's shape or function to perform a biological task. To better understand this process, AutoDockTools (a graphical user interface for preparing and analyzing molecular docking simulations) and AutoDock Vina (a software program that performs docking simulations, predicting how ligands bind to proteins and assessing their binding affinities) are used to predict binding affinities between proteins and ligands. The steps involved in this process are as follows:

1. Install Necessary Software

* Install AutoDock Vina for molecular docking: Follow the instructions on the AutoDock Vina website.
* Install AutoDockTools: Download from [AutoDock Suite](https://autodocksuite.scripps.edu/adt/).
* Install PyMOL: Download from [PyMOL](https://www.pymol.org/).

2. Download Required Files

* Protein: Download the Mu opioid receptor (PDB ID: 5C1M) from the [PDB](https://www.rcsb.org/).
* Ligands: Download Fentanyl, Buprenorphine, and Aspirin from [PubChem](https://pubchem.ncbi.nlm.nih.gov/):
  + Aspirin: 2244
  + Fentanyl: 3345
  + Buprenorphine: 644073
* Create a folder to store these files, naming the protein as "MOR" and ligands as "Fentanyl," "Aspirin," and "Buprenorphine."

3. Prepare the Protein (MOR)

* Open AutoDockTools and load the protein file.
* Delete water molecules: Edit > Delete Water
* Add hydrogens: Edit > Hydrogens > Add (Polar only)
* Add Kollman charges: Edit > Add Kollman Charges
* Save the protein as a PDBQT file: Grid > Macromolecules > Choose > Select Protein > Save as PDBQT

4. Prepare the Ligands

* Open PyMOL and load the ligand.
* Export the ligand as a PDB file: File > Export Molecule > Save as PDB
* Load the ligand into AutoDockTools and convert it to PDBQT: Ligand > Input > Choose > Save as PDBQT

5. Create the Grid for Binding

* Drag the PDBQT files of the protein and ligand into AutoDockTools.
* Define the grid box: Grid > Macromolecules > Choose > Select Protein > Grid Box
* Set the grid box dimensions and center to focus on the binding site.
* Save the grid dimensions as "grid.txt".

6. Create the Configuration File

* Create a "config.txt" file with the following details:
  + Receptor and Ligand: file names of protein and ligand (e.g., MOR.pdbqt, Fentanyl.pdbqt).
  + Center (x, y, z): values from "grid.txt".
  + Size (x, y, z): values from "grid.txt".
  + Energy range: 4
  + Exhaustiveness: 8

7. Run AutoDock Vina

* Open Command Prompt (cmd).
* Navigate to the folder with the files: cd [folder path]
* Run AutoDock Vina with the following command: vina.exe --receptor MOR.pdbqt --ligand Fentanyl.pdbqt --config config.txt --log log.txt --out output.pdbqt
* Wait for the results: The log file ("log.txt") contains docking information, and the "output.pdbqt" file shows the 3D structure of the docking.

8. Repeat for Each Ligand

* Repeat steps 7-8 for each ligand (Fentanyl, Buprenorphine, and Aspirin) with the MOR protein.

## Part 2: Developing ML Models to predict binding of proteins and ligands

ML Training data collection

As described earlier, Machine Learning can be very fast and efficient for predicting the binding of ligands with proteins. However, finding a good machine learning model is a complex method and requires good representative training data.

ChEMBL, maintained by the European Bioinformatics Institute (EMBL-EBI) is a large, open-access database of bioactive molecules with drug-like properties. It contains data on the relationships between chemical compounds, their biological activities, and drug targets. ChEMBL includes information such as compound structures, biological assay results, target details, and binding affinities. ChEMBL is used to download the training data necessary for our model of the protein of interest as follows: (Example for MOR shown)

* Go to ChEMBL and search the protein name (e.g. “Mu opioid Receptor”)
* Select the protein with the type “SINGLE PROTEIN” and organism as “Homo sapiens” (e.g. CHEMBL233 for MOR)
* Scroll down to Approved Drugs and Clinical Candidates and download all compounds.
* The data contains the SMILES representation of the ligand and its binding affinity with the protein.

Previous research in prediction of binding affinities through machine learning has shown that molecular fingerprints are more efficient and accurate than SMILES. Previous work done by Dong Chen to convert SMILES string representation into Molecular fingerprints (512 length vector) is used in this project. It converts each SMILE into a fixed 512 length embedding vector which is used as the input to the ML model as shown in figure 2.



Figure 2: SMILES to Molecular Fingerprints

Machine Learning Model

The objective is to build a ML model for each protein to accurately predict the binding affinity of the ligand with the protein. To determine the most effective machine learning model, several regression-based algorithms were developed and trained on the dataset. The performance of these models was carefully evaluated, and the three best-performing models were highlighted in this report.

* Bayesian Ridge Regression Model
* Random Forest Regressor Model
* Linear Regression Model

Among them, the Random Forest Regressor was ultimately selected based on its superior score and predictive accuracy.

Figure 3 shows all the steps to build the machine learning models. The basic training algorithm takes the fingerprint of the ligand as the input and outputs a binding affinity of the ligand with the protein. The predicted binding affinity and the actual binding affinity is fed into the loss function to determine the prediction error which is used to update the weights of the model. After going through the complete dataset, the three final prediction scores R2, MAE, and MSE are calculated.

Hyperparameter tuning is used to improve the model’s accuracy, efficiency, and generalization by changing its parameters. For every combination of hyperparameters, the training process is repeated to find which set of parameters would produce the best model.

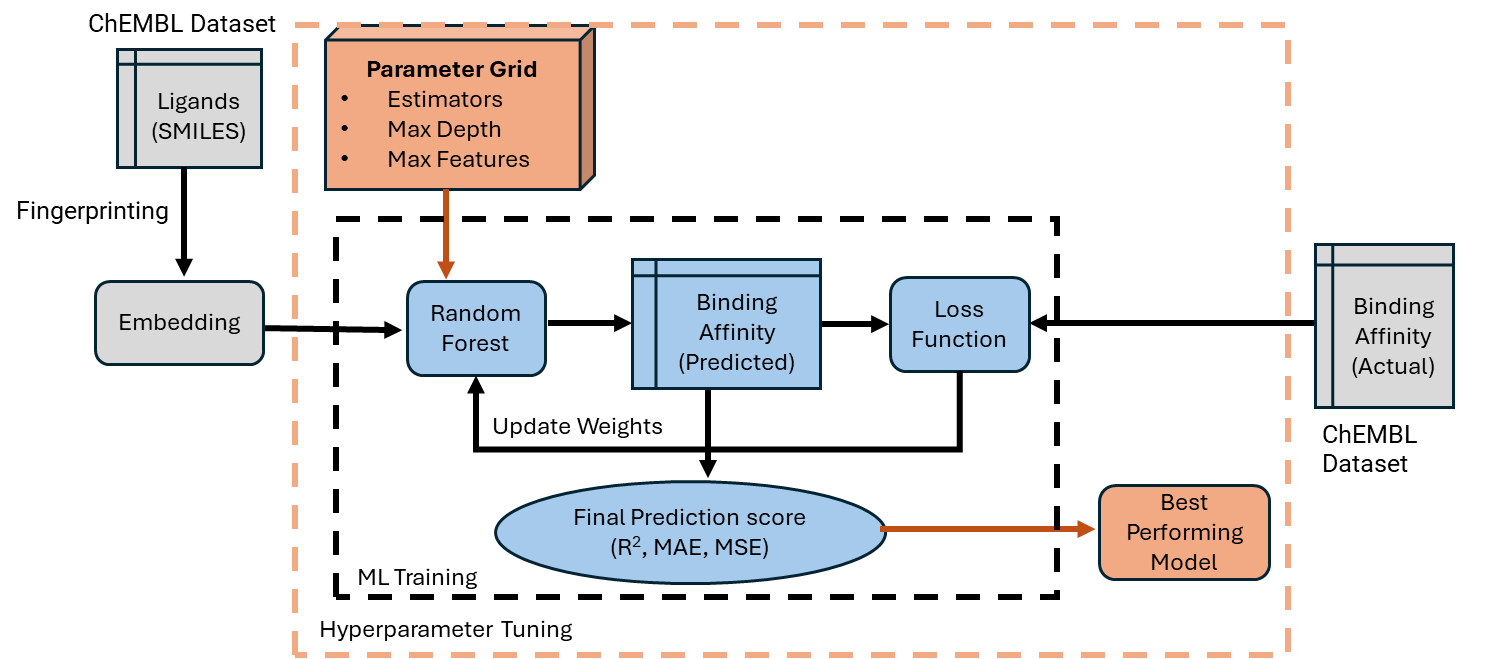


Figure 3: Flow chart showing how hyperparameter tuning works

## Part 3: Understanding the PPI Network related to MOR

A perfect drug or ideal drug should only bind to MOR and none of the other related proteins. To find the best candidates, the screening should identify drugs with a high binding affinity with MOR and low binding affinities for the rest of the proteins in MOR’s PPI network. To get the proteins in MOR’s PPI network follow the steps below:

* From the [STRING](https://string-db.org/cgi/network?taskId=b7XtEt6m40iZ&sessionId=bnlV1MS18z0K) website search up “Mu Opioid Receptor” and click continue. This will show all the proteins directly related to MOR.

Figure 4 below shows the MOR PPI network. OPRM1’s alias is MOR and the proteins directly connected are shown below.

A network of lines and dots

Description automatically generated

Figure 4: MOR PPI Network from STRING

However, the dataset required for ML training for all the proteins directly connected to OPRM1/MOR is not readily available. Therefore, the PPI network had to be expanded to find proteins that are available in the ML training dataset as shown in the graph below. The grey circles are proteins whose training data is not available, and for the remaining coloured circles an ML model was created for each of them as explained in part 2. The proteins collected are: AVPR2, ADRB2, AKT1, CNR1, CXCR1, CXCR2, CXCR4, MC1R, S1PR1, SMO, and SRC.



Figure 5: Expanded MOR PPI network to find available proteins.

In addition to the proteins mentioned above, an ML model was developed for hERG, a protein that plays a vital role in regulating the heart’s electrical signal. This regulation is essential for maintaining proper heart rhythm. If a drug blocks hERG, it can disrupt this balance, leading to prolonged QT intervals and increasing the risk of arrhythmias, which can cause heart attacks or even sudden cardiac death. As a result, it is critical for drugs and opioids to exhibit a "low" binding affinity with hERG to minimize these dangerous side effects. The table below shows all the proteins chosen and their functions.

|  |  |  |
| --- | --- | --- |
| Full Name | Name | Description |
| Arginine Vasopressin Receptor 2 | AVPR2 | Regulates water balance and blood pressure. |
| Adrenergic Beta-2 Receptor | ADRB2 | Influencing heart rate, smooth muscle relaxation. |
| Protein Kinase B | AKT1 | Promotes cell survival, growth, and metabolism. |
| Cannabinoid Receptor 1 | CNR1 | Regulating mood, memory, appetite, and pain perception. |
| C-X-C Chemokine Receptor Type 1 | CXCR1 | Involved in immune response, directing neutrophils to sites of infection or inflammation. |
| C-X-C Chemokine Receptor Type 2 | CXCR2 | Guides white blood cells to areas of inflammation or infection. |
| C-X-C Chemokine Receptor Type 4 | CXCR4 | Regulates immune cell signaling, stem cell migration, and HIV spread. |
| Melanocortin 1 Receptor | MC1R | Controls skin pigmentation and response to UV radiation. |
| Sphingosine-1-Phosphate Receptor 1 | S1PR1 | Regulates immune cell migration and vascular development in the immune system. |
| Smoothened, Frizzled Class Receptor | SMO | Involved in cell growth, development, and differentiation. |
| Non-Receptor Tyrosine Kinase | SRC | Regulates cellular processes like growth, differentiation, and, involved in cancer signaling. |
| Human Ether-à-go-go-Related Gene | hERG | The hERG protein controls the heart's electrical signals. If it’s blocked, it can cause heart rhythm problems. |

Table 6

Once the proteins are known, an ML model is created for each protein as described in part 2.

## Part 4: Drug Screening

As explained earlier, repurposing drugs that work as an alternative and/or substitute to existing recommended opioids involve selecting all drugs that primarily have high binding affinity with MOR. This is essential to ensure that the drug has similar euphoric effect if works as an agonist or has similar potential to work as antagonist on the person suffering with OUD. Next, to make it safer than current opioids it must have the least binding affinities with the rest of the proteins.

1. From the DrugBank website download the csv format of the list of all known drugs. These drugs will be screened to see if they are safer drugs than existing recommended opioids.
2. For each drug extract the SMILES strings and convert them into molecular fingerprints as explained in part 2.
3. Using each drug’s molecular fingerprint, the ML model predicts the corresponding binding affinities.
4. For screening the drugs, it may be noted that there is a very less probability to find perfect or ideal drug alternatives and hence a practical screening and selection technique is used where the proteins are prioritized from most to least important for screening as given in the table below
   * Filter all drugs that have a “High” MOR binding.
   * Further filter the drugs that have “Low” hERG binding as it interferes with the heart’s activity.
   * Next, filter the drugs that have “Low” CXCR4 binding since, it is involved in cancer metastasis, HIV infection, and immune response.
5. For a more customized selection of drugs a person’s health condition must be accounted for. The table below lists the critical proteins a drug should have a low binding with based on common health conditions.

|  |  |
| --- | --- |
| **Health Condition of person with OUD** | **Proteins with “Low” binding affinity with the drugs** |
| Diabetes | **CNR1:** May exacerbate weight gain and appetite dysregulation, worsening metabolic control. |
| Cardiovascular Diseases | **CNR1:** Can alter central nervous system activity, impacting cardiovascular regulation. |
| Cancer | **AVPR2:** Can lead to water retention issues, complicating chemotherapy side effects.  **ADRB2:** May promote cancer cell migration in specific tumor types. |
| Neurological Disorders | **CXCR1/CXCR2:** Excessive immune modulation can exacerbate neuroinflammation. |
| Autoimmune Diseases | **MC1R:** Pigmentation effects are irrelevant and could cause unnecessary side effects. |

Table 7

The process of finding suitable safe drugs is shown below in Figure 8.

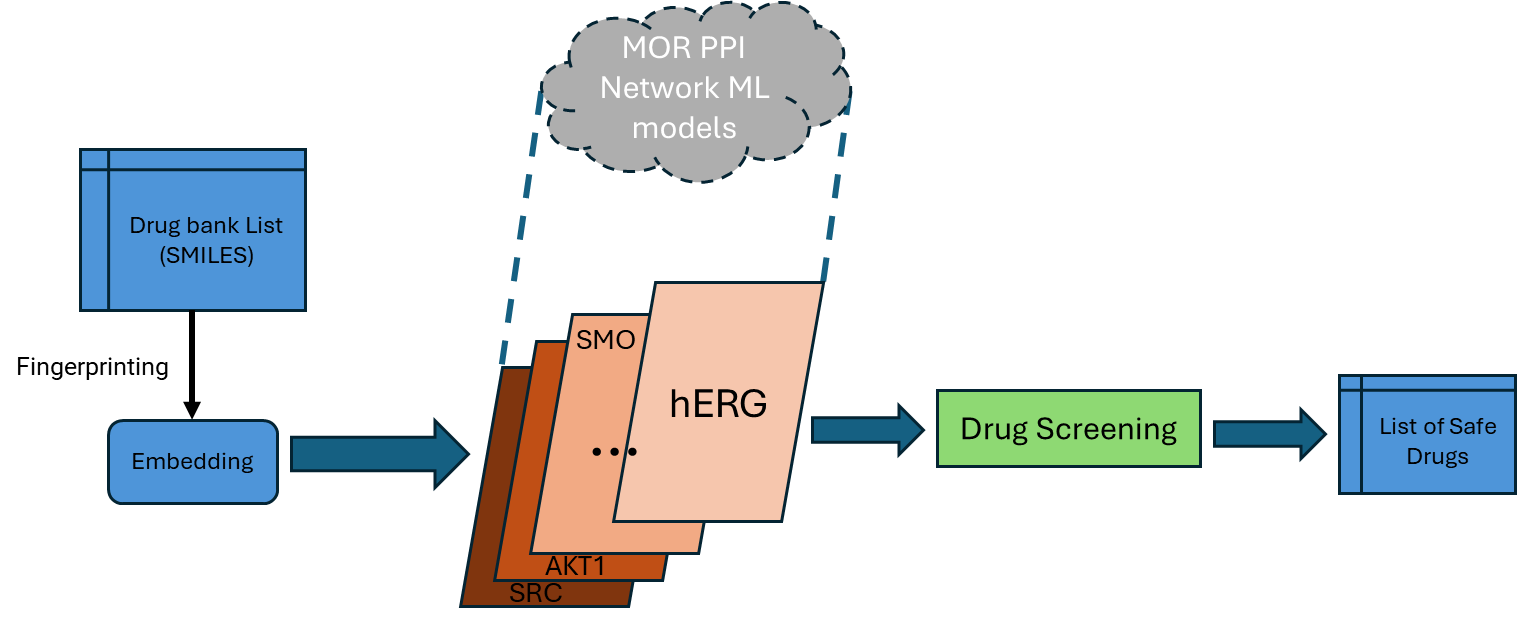


Figure 8: Steps to find safe drugs

# Lab Question

Can an efficient machine learning technique be developed to identify safer alternatives to current opioids that have reduced side effects, and overdose risk for individuals with Opioid Use Disorder (OUD)?

# Hypothesis

If reliable machine learning models are built that can predict the binding between drugs and different proteins associated with the opioid receptor in the human body then, it is possible to find existing drugs that are safer substitutes to the existing opioids.

# Purpose

The purpose of this project is to leverage machine learning to repurpose existing drugs, offering healthcare providers a broader range of cost-effective treatment options tailored to individual medication needs, thereby taking a step towards the implementation of the BC Coroners' recommendation.

# Materials

Hardware

* PC/Laptop
* Internet Connection

Software

* AutoDock Vina - (<https://vina.scripps.edu/>)
* Molecular Fingerprinting (<https://github.com/ChenDdon/AGBTcode>)
* Python Packages
  + Scikit Learn
  + Numpy
  + Pandas
  + Joblib

Data

* chEMBL
* STRING
* Protein Data Bank (PDB)
* DRUGBANK Online

# Data and Observations

Part 1: Understanding binding of proteins with ligands

AutoDock Vina is used to predict the binding affinity after preparing the protein and ligands. Three different drugs were selected to interact with the Mu opioid receptor (MOR): Aspirin, which is expected to have a weak binding affinity; Fentanyl, which is known to bind strongly; and Buprenorphine, which also binds well with MOR. Figures 9, 10, 11, and 12 show the MOR protein alongside the ligands Aspirin, Fentanyl, and Buprenorphine, respectively, in AutoDockTools after preparation. The actual compound structures taken from chEMBL can be seen in the appendix.A purple and blue lines on a black background

Description automatically generated

Figure 9: MOR protein shown in Autodock tools

A hexagon shaped lines on a black background

Description automatically generatedA purple hexagon on a black background

Description automatically generated

Figure 10: Aspirin ligand shown in Autodock tools Figure 11: Buprenorphine shown in ADT

A line of lines on a black background

Description automatically generated

Figure 12: Fentanyl ligand shown in Autodock tools

Using these models the MOR and respective ligands were binded and there binding affinities were recorded. Figure 13 shows the picture of MOR bonded to fentanyl and table 14. shows the consisting of binding affinities between the protein and compound.

A purple and blue lines

Description automatically generated with medium confidence

MOR Protein

Fentanyl Ligand docked with the MOR protein

Figure 13: MOR and Fentanyl bonded together shown in Autodock Tools

When AutoDock Vina predicts the binding affinity between a protein and a ligand, it provides several potential binding affinity predictions. The minimum and maximum values of these predictions are shown below. The table below displays the results for MOR with Aspirin, Fentanyl, and Buprenorphine.

|  |  |  |
| --- | --- | --- |
| **Drugs** | **Binding Affinity using AutoDock Vina** | **Binding Affinity from chEMBL** |
| Aspirin | Maximum Binding Affinity: -5.5 Minimum Binding Affinity: -5.3 | N/A |
| Fentanyl | Maximum Binding Affinity: -7.1 Minimum Binding Affinity: -6.7 | -8.7 |
| Buprenorphine | Maximum Binding Affinity: -7.5 Minimum Binding Affinity: -5.7 | -12.5 |

Table 14: Binding affinities between ligands and MOR protein.

## Part 2: ML Models to predict binding of proteins and ligands

Several regression-based models were developed and trained on the MOR dataset to help identify the optimal model for the prediction of ligands binding to proteins. The performance of the ML models was measured using the following scoring methods:

* Mean Squared Error (MSE) - The average absolute difference between the predicted and actual values. Sensitive to the absolute magnitude of errors but less sensitive to outliers than MSE. Lower is better and a perfect score would be 0
* Mean Absolute Error (MAE) - The average of the squared differences between the predicted and actual values. Penalizes larger errors more heavily due to squaring, making it sensitive to outliers. Lower the better and a perfect score would be 0
* R2 - Indicates the proportion of the variance in the target variable explained by the model. Provides a relative measure of how well the model fits compared to a naive baseline. A perfect score would be 1

Figure 15 shows the performance of the top 3 regression-based ML models of MOR and measured by MSE, MAE, and R2. The first, second and third bar shows the scoring results of linear regression, Random Forest and Bayesian Ridge respectively.

Figure 15: Performance comparison of MOR ML models using a train-test split of 90:10

Next, to further improve the Random Forest ML model for MOR, Hyperparameter tuning was utilized to find the optimal hyperparameters that would further increase the accuracy and efficiency of the ML models. The table below shows the details of the hyperparameters used.

|  |  |  |
| --- | --- | --- |
| **Hyperparameters** | **Values Used** | **Description** |
| N estimators | [50, 100, 200] | The number of decision trees in the random forest regressor. Higher values usually lead to better performance but also increases computational time. |
| Min samples split | [2, 5, 10] | The minimum number of samples required to split a node. Used to prevent overfitting by controlling tree depth. |
| Max Features | ['sqrt', 'log2', None, 0.8, 0.9] | Maximum number of features considered for splitting a node. Helps to prevent overfitting by adding randomness. |
| Max Depth | [None, 10, 20, 30] | Maximum depth of the decision tree. Controls the growth of the tree. Deeper trees fit more complex patterns but may overfit. |

Table 16: Description and values of hyperparameters

To better understand the effect of each hyperparameter of on the MOR ML models, a spider chart is created in Figure 17 comparing the hyperparameters of the worst and best scoring models and a model using cross validation. The best model has hyperparameters of {N Estimators = 100, Max Depth = 30, Min Samples Split = 5, and Max Features = None}. Whereas the model with the worst scores has hyperparameters of {N Estimators = 50, Max Depth = 10, Min Samples Split =10, and Max Features = Log2}. The model created using cross validation had parameters of {N Estimators = 200, Max Depth = 30, Min Samples Split =2, and Max Features = 0.8}.

Figure 17: The spider chart shows the parameters of the best, worst, and the model using cross validation for MOR.

The bar graph below shows the performance of the random forest MOR model with various hyperparameters.

A graph of different colored bars

Description automatically generated

Figure 18: Performance comparison of MOR ML models with different parameters

Figure 19 below presents a scatter and line plot illustrating the error in prediction of the binding affinities of the MOR ML model and various ligands. The x-axis corresponds to the different ligands ordered from highest to lowest binding affinities. The blue scatter plot depicts the actual binding affinities (left y-axis) whereas the orange scatter plot represents the corresponding prediction errors (right y-axis), which range from 25% to -35%.

A graph with orange lines and blue lines

Description automatically generated

Figure 19: The data shows how the prediction error changes as the actual binding affinity between MOR and other ligands decrease.

The table below shows the binding affinity prediction distribution of the ligands categorized in bins of high, medium and low binding with MOR. All ligands with a Binding Affinity (BA) less than equal to -9 is considered high, BA less than -9 but more than -7 is considered medium and BA more than equal to -7 is considered as low. The column shows the predicted bin of the ligands. Each cell in the table represents its predicted BA bin as per its column that categorized as corresponding BA bin in the actual dataset as given by its row. For example, 46 ligands in cell (High, Medium) are predicted to be in the medium bin but actually belong to the high bin in the dataset.

|  |  |  |  |
| --- | --- | --- | --- |
| **Predicted Category →**  **Actual Category**  **↓** | High | Medium | Low |
| High | 62.9% | 1% | 0% |
| Medium | 5.4% | 27.5% | 0% |
| Low | 0.1% | 2.7% | 0.4% |

Table 20

## Part 3: Understanding the PPI Network related to MOR

Figure 21 below shows the evaluation results of the models of the proteins in MOR’s PPI network. Just like the MOR model the evaluation metrics used were MSE, MAE, and R2.

Figure 21: Data results of proteins in MOR’s PPI network

The list of all Health Canada approved opioids is shown in table 22 with their brief description.

|  |  |
| --- | --- |
| **Drug** | **Description** |
| **Buprenorphine** | A partial opioid agonist used to treat opioid addiction and manage pain, offering lower abuse potential. |
| **Butorphanol** | A mixed opioid agonist-antagonist primarily used for pain relief and as a pre-anesthetic medication. |
| **Codeine** | A mild opioid used to treat mild to moderate pain and as a cough suppressant. |
| **Fentanyl** | A synthetic opioid that is highly potent and used to manage severe pain, often in cancer or surgical settings. |
| **Hydrocodone** | A semi-synthetic opioid commonly prescribed for moderate to severe pain and as a cough suppressant. |
| **Hydromorphone** | A potent opioid derived from morphine, used to treat severe pain. |
| **Meperidine** | A synthetic opioid used for moderate to severe pain, though less commonly due to safety concerns. |
| **Methadone** | A long-acting opioid used in pain management and as a treatment for opioid use disorder. |
| **Morphine** | A natural opioid derived from opium, used as a benchmark drug for severe pain relief. |
| **Normethadone** | A synthetic opioid with analgesic properties, primarily used for pain management and cough suppression. |
| **Opium** | A natural extract from the poppy plant containing several active compounds, used historically for pain and sedation. |
| **Oxycodone** | A semi-synthetic opioid used for managing moderate to severe chronic pain. |
| **Oxymorphone** | A potent opioid analgesic used to manage severe pain, often in patients tolerant to other opioids. |
| **Pentazocine** | A synthetic opioid with mixed agonist-antagonist effects, used for moderate to severe pain. |
| **Tapentadol** | A dual-action pain reliever that combines opioid effects with norepinephrine reuptake inhibition. |
| **Tramadol** | A weak opioid analgesic with additional serotonin and norepinephrine reuptake inhibition, used for moderate pain relief. |

Table 22: Canadian approved opioids and their details: (<https://www.canada.ca/en/health-canada/services/drugs-health-products/reports-publications/medeffect-canada/list-opioids.html>)

Next, to evaluate the effects of the above opioids, all MOR PPI random forest models as mentioned were used to predict their binding affinities with these opioids. The predictions were categorized based on the same thresholds: High, Medium, and Low as shown below in figure 23.



Figure 23: Data of binding affinity of Canadian approved opioids and MOR and it’s PPI network.

## Part 4: Drug Screening

To determine a final list of safe drugs, the DrugBank drug list was screened by predicting the binding affinity of each drug to the MOR PPI proteins. The binding affinities were categorized as High, Medium, or Low. Figure 24 below presents a bar graph illustrating the percentage distribution of the binding affinity categories for each protein.

Figure 24: Percentage of binding affinities categorized as High, Medium, and Low for each protein with the entire DrugBank list.

After predicting and categorizing the binding affinities, the drugs were further screened for potential safety across five categories, including a general "no disease" group. For each disease, drugs with the highest priority were filtered to achieve the lowest possible binding affinity category, as illustrated in the figure below. The orange boxes highlight the key proteins relevant to each disease, while the arrows indicate comparisons with Canadian-approved opioids. The binding affinity categories were tallied, and the most frequent category was selected as the summary for each protein. This process was repeated for all repurposed drug categories, and the summaries of both were compared to generate the table shown below.

The table below compares the performance of the screened drugs in this project with the baseline (Health Canada approved opioids). The orange boxes highlight the key proteins relevant to each disease, while the arrows indicate comparisons with Canadian-approved opioids. The red downward arrows indicate higher binding affinities than the baseline, the yellow sideways arrows show no change from the baseline, and the green upward arrows represent lower binding affinities than the baseline.

A chart with arrows pointing to different directions

Description automatically generated

Figure 25: Comparison of Canadian approved opioids and repurposed drug categories.

The table below shows the list of drugs for each repurposed drug category.

|  |  |  |
| --- | --- | --- |
| **Repurposed Drug Category** | **Number of Drugs** | **Drugs** |
| Canadian Approved Opioids  (Baseline) | 17 | Buprenorphine, Butorphanol, Codeine, Fentanyl, Hydrocodone, Hydromorphone, Meperidine, Methadone, Morphine, Normethadone, Opium, Oxycodone, Oxymorphone, Pentazocine, Tapentadol, Tramadol |
| General Screened Opioids | 7 | Telithromycin, Clarithromycin, Bicifadine, Cethromycin, Y-27632, Solithromycin, Mericitabine |
| Diabetic Safe Opioids | 4 | Diphenhydramine, Ofloxacin, Bicifadine, Mericitabine |
| Cardiovascular Safe Opioids | 4 | Diphenhydramine, Ofloxacin, Bicifadine, Mericitabine |
| Cancer Safe Opioids | 7 | Telithromycin, Clarithromycin, Bicifadine, Cethromycin, Y-27632, Solithromycin, Mericitabine |
| Neurological safe Opioids | 5 | Fexofenadine, Diphenhydramine, Desipramine, Bicifadine, Y-27632 |
| Autoimmune Safe Opioids | 5 | Fexofenadine, Ofloxacin, Bicifadine, Y-27632, Mericitabine |

Table 26

# Analysis

## Part 1: Understanding binding of proteins with ligands

As mentioned earlier the purpose of the experiment in this part is to understand and assess the binding affinity values of various ligands with the MOR receptor protein. Therefore, three drugs that are known to bind strongly (Buprenorphine and Fentanyl), and weakly (Aspirin) are selected. AutoDock Vina, a widely used computational molecular docking tool is used for this purpose. Table 14 shows the binding of MOR with the ligands Aspirin, Fentanyl, and Buprenorphine using AutoDock Vina. Also, the binding affinity of these drugs as available from chEMBL is included to compare against AutoDock vina value. As can be seen, the range of binding affinity predictions for Aspirin and Fentanyl is relatively narrow, with a difference of 0.2 and 0.4, respectively. This indicates a higher level of confidence in AutoDock Vina's predictions for these ligands. In contrast, the range for Buprenorphine is significantly wider, with a difference of 1.8, suggesting lower confidence in its binding affinity prediction. Based on the results it can be deduced that:

* Predicting the binding of ligands with protein is a complex process and need a robust and efficient solution where AI can play a key role.
* The measure of binding affinity value varies based on the technique used that need to be accounted for screening the drugs to minimize both false positives and false negatives.

## Part 2: ML Models to predict binding of proteins and ligands

The comparison of the performance the top three performing regression-based models is shown in Figure 15. The Linear regression model has scores of 1.71 and 1.02 for MSE and MAE respectively. With the binding affinities of the ligands ranging from -15 to -5 this means that the range of error for the linear regressor model in correctly predicting the binding affinity is approximately 6.7% - 18.2%. Whereas for the Bayesian Ridge regressor with scores of 1.57 and 0.98 for MSE and MAE respectively, the range of error is approximately 6.4% - 17.5%. Lastly, the random forest regressor had MSE and MAE scores of 1.37 and 0.9 respectively with a range of error of approximately 5.9% - 16.1%. The smallest MSE and MAE scores are obtained from the random forest regressor which signifies that random forest has the most accuracy in predicting the MOR model’s binding affinity with ligands. A value greater than 0.5 for R2 is a good score for a model. Based on this data it is concluded that the random forest regressor is best suited to perform the prediction of binding of ligands and proteins for this project’s purposes.

Once the ML technique is selected the immediate next step is to evaluate whether its accuracy can be further improved using Hyperparameter tuning. Using figure 17 it can be observed that with increasing N Estimators which increases the number of decision trees the model’s performance improves up to a certain extent which is around 100 but beyond that it may not provide any further benefit. The max depth is the number of levels each decision tree has, and both of the best performing models have a max depth of 30 which signifies that the prediction of the binding affinity needs a lot of comparisons. Similarly, the best performance for the max features which impacts the splitting of the node when constructing the trees show that a high amount of features is necessary for an accurate prediction. It can be observed that for min sample split which is the smallest amount of data a node must have to be eligible for splitting, the lower the value the better the performance.

The accuracy of the random forest models' predictions is shown in Figure 18. The lowest-performing model is significantly worse than the others. The best-performing model and the cross-validation model have nearly identical results across all three-evaluation metrics. Using this, the final hyperparameter chosen for MOR is the best model with values of {N Estimators = 100, Max Depth = 30, Min Samples Split = 5, and Max Features = None}.

The scatter and line plot in Figure 19 illustrates the prediction error of the MOR ML model for various ligands' binding affinities. For the majority of the ligands (78%) the prediction error is within 5%. Another 19% of ligands have a prediction error of between 10 -15%. Very few ligands have a large prediction error with maximum of 68% which can be seen in the spikes in the scatter plot. Another key observation is that ligands with high binding affinities have positive prediction error percentages while low binding affinity have prediction error percentages that are negative. This means that the model is predicting a little lower for high binding ligands and a little higher for low binding ligands.

To make the screening technique simpler the ligands are further categorized based on their binding affinities as High (BA <= -9), Medium (-9 < BA <= -7), and Low (BA > -7). This can be seen in table 20 which shows the binding affinity prediction distribution of the ligands categorized in bins of high, medium and low binding with MOR. Using this categorization method, it is seen that the prediction accuracy of the categorized bin is quite accurate as seen by the diagonal cells in the table. Note that two observable prediction errors are seen where 5.4% of ligands that were actually Medium but were predicted High and another 2.7% were low but predicted as medium. Based on this part of the experiment the key takeaways are:

* The random forest model is the most suitable for this task.
* Hyperparameter tuning optimized the model, achieving a prediction error within 5% for 78% of ligands.
* Binding affinity categorization underpins the drug screening method developed in this project

## Part 3: Understanding the PPI Network related to MOR

The comparison of scores for the proteins in MOR's PPI network is shown in Figure 21. Most proteins exhibit relatively good scores, except for ADRB2, which received MSE, MAE, and R2 scores of 3, 1.4, and 0.48, respectively. The high MSE indicates that the model made large errors on a few predictions but was accurate for most of them. On the other hand, some of the best-performing protein models include SMO, S1PR1, AVPR2, and AKT1.

To evaluate the ML model's accuracy and analyze the characteristics and side effects of Health Canada-approved opioids, all the drugs were tested using the ML models. Figure 23 shows the categories of binding affinities, and it can be observed that all the opioids bind well to MOR, as they specifically target this receptor. They also show medium binding with hERG which is a critical protein. However, they also show high binding affinities with AVPR2, CXCR2, CXCR4, S1PR1, and SMO. Based on these high binding affinities, potential side effects include swelling, high blood pressure leading to dizziness, cancer growth, chronic inflammation, and organ damage. The clinically known harmful side effects of these drugs on the human body are quite similar to the predicted side effects listed above. This further illustrates that all the ML models are very reliable, robust, and effective for the drug screening purposes.

In addition, the proteins ADRB2, AKT1, CNR1, CXCR1, MC1R, and SRC have mostly fall in the medium binding category. This highlights that Health Canada-approved opioids are not entirely safe and serve as the project's baseline. The drugs selected after screening should have fewer side effects. The key takeaways are:

* The machine learning models effectively characterize the impact of opioids on individuals.
* Health Canada-approved opioids are associated with significant side effects and cannot be considered completely safe.
* The goal of this project is to identify and propose a list of alternative drugs with fewer side effects compared to existing opioids.

## Part 4: Drug Screening

The bar graph in figure 24 obtained by running each drug from DrugBank with MOR PPI ML models. Very few drugs fall under the low binding category across all the proteins. Specifically, except SRC and hERG with low binding percentages of 2% and 1% respectively, no other proteins have drugs with predicted low binding. This indicates that only 1% of the drugs have a low hERG binding value, which is essential for a drug to be considered safe and also will have an improved hERG value compared to the baseline. For MOR, 60% of the predictions indicated high binding affinity, with the remaining 40% falling in the medium range. For an opioid substitute the repurposed drug needs to have a high MOR binding and hence it was filtered based on that. CXCR4 a protein associated with cancer growth has 35% of the predictions indicating a medium binding and 65% of the remaining prediction falling in the high range. The screened drugs selected all have medium binding with CXCR4 something that was improved from the baseline. The other proteins CNR1, CXCR1, ADRB2, and MC1R have predicted medium binding percentages of 39%, 99%, 72%, and 72% respectively.

Figure 25 shows the comparison of the screened and repurposed drugs obtained using the technique developed in this project with the Health Canada approved opioids baseline. For MOR all the bindings stay the same (High), something that was being aimed for since, it is needed for an opioid substitute. Next, all of the hERG and CXCR4 values have improved (lower binding) compared to the baseline. Both of these proteins are critical for a drug to be considered safe. Additionally, for specific proteins targeted based on a health condition the binding has improved or stayed the same compared to the baseline. This ensures that the screened drugs are much safer than the currently existing opioids. Finally, the names of the opioids substitute as obtained in this project for various categories is shown in table 26.

The final outcomes of this project are:

* Seven drugs were identified as generally safe opioids.
* Four opioids are tailored for individuals with OUD and diabetes.
* Four opioids are customized for those with OUD and cardiovascular issues.
* Seven opioids are available for individuals with OUD and cancer.
* Five opioids are designed for people with OUD and neurological issues.
* Five opioids are suitable for those with OUD and autoimmune conditions.

# Sources of Error

# In this project several potential sources of error may have influenced the outcomes of the results. In terms of data, incomplete or inaccurate SMILES strings and PPI protein information from databases like DrugBank, PDB, and ChEMBL could have created errors for predicting the binding affinities. Additionally, errors in the data, such as incorrect molecular fingerprint conversion, might have also impacted the results.

# On the machine learning aspect, overfitting or underfitting the machine learning models could have led to reduced prediction accuracy which would invalidate the screening results. Suboptimal hyperparameter tuning may also have prevented the model from achieving its best performance. Additionally, assuming that Low binding affinity always leads to clinical effectiveness ignores important biological factors, like how drugs behave in the body or interact with proteins.

# Lastly, limitations, such as rounding errors or insufficient resources, could have disrupted the training process, while human errors during implementation or interpretation of results may have introduced additional inaccuracies in the results.

# Future Improvements

In the future this project can be enhanced in three key areas. First, prediction accuracy can be significantly improved by developing transformer-based machine learning models. Second, the scope of the PPI network will be expanded by including additional opioid receptors, such as DOR and KOR, to provide a more comprehensive understanding of drug interactions. Lastly, the screening process will be refined by analyzing and incorporating drug toxicity and absorption properties.

# Conclusion

# In this project, a novel approach leveraging machine learning models to understand a drug's effects on the human body through the PPI network was proposed and successfully implemented. The results supported the hypothesis, demonstrating that machine learning can effectively identify safer alternatives to existing opioids. By applying this method to screen known drugs, the project highlighted its potential to combat opioid use disorder (OUD) by providing a safer and more effective solution to this global health crisis.

# Acknowledgments

I would like to thank my parents and my family, who have supported and helped me throughout this project. I would also like to thank Ms. Leong, who has guided and encouraged me in this project.

# References

Benyamin, R., Trescot, A., Datta, S., Buenaventura, R., Adlaka, R., Sehgal, N., Glaser, S., & Vallejo, R. (2008, March 1). *Opioid Complications and Side Effects*. Pain Physician. https://pubmed.ncbi.nlm.nih.gov/18443635/

Bhatti, A. (2020, April 17). *Fentanyl Addiction In Vancouver | Andy Bhatti Interventions*. Andy Bhatti. https://andybhatti.com/addictions/fentanyl-addiction-in-vancouver/

Bioinformatics With BB. (2020, July 30). *Molecular Docking | Autodock VINA Virtual Screening | VINA Docking tutorial | Bioinformatics*. YouTube. https://www.youtube.com/watch?v=tFFxNTvvoJI

Dhaliwal, A., & Gupta, M. (2023, July 24). *Physiology, Opioid Receptor*. PubMed; StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK546642/

Feng, H., Jiang, J., & Wei, G.-W. (2023). Machine-learning repurposing of DrugBank compounds for opioid use disorder. *Computers in Biology and Medicine*, *160*, 106921–106921. https://doi.org/10.1016/j.compbiomed.2023.106921

Fox, L. (2023, February 13). *How AI Can Help Design Drugs to Treat Opioid Addiction*. The Biophysical Society . https://www.biophysics.org/news-room/how-ai-can-help-design-drugs-to-treat-opioid-addiction

Khushaktov, M. F. (2023, August 26). *Introduction Random Forest Classification By Example*. Medium. https://medium.com/@mrmaster907/introduction-random-forest-classification-by-example-6983d95c7b91

Liu, Y. S., Kiyang, L., Hayward, J., Zhang, Y., Metes, D., Wang, M., Svenson, L. W., Talarico, F., Chue, P., Li, X.-M., Greiner, R., Greenshaw, A. J., & Cao, B. (2022). Individualized Prospective Prediction of Opioid Use Disorder. *The Canadian Journal of Psychiatry*, *68*(1), 54–63. https://doi.org/10.1177/07067437221114094

mgl-admin. (n.d.-a). *Downloads*. AutoDock Vina. https://vina.scripps.edu/downloads/

mgl-admin. (n.d.-b). *Homepage*. Mgltools. http://mgltools.scripps.edu/

Nyuytiymbiy, K. (2021, April 5). *Parameters and Hyperparameters in Machine Learning and Deep Learning*. Medium. https://towardsdatascience.com/parameters-and-hyperparameters-aa609601a9ac

*OPRM1 protein (human) - STRING interaction network*. (2024). String-Db.org. https://string-db.org/cgi/network?taskId=bOq6tL3uVdpQ&sessionId=b5sqoi5NkBCu&allnodes=1

Rutherford, G. (2022). *Machine learning predicts risk of opioid use disorder for individual patients*. Ualberta.ca. https://www.ualberta.ca/folio/2022/12/machine-learning-predicts-risk-of-opioid-use-disorder.html

Schrodinger. (2019). *PyMOL | pymol.org*. Pymol.org. https://pymol.org/2/

UNODC. (2023). *World Drug Report 2023*. United Nations : Office on Drugs and Crime. https://www.unodc.org/unodc/en/data-and-analysis/world-drug-report-2023.html

V=Bapat, S. (2020, March 20). *Molecular Docking Tutorial: AUTODOCK VINA - PART 1*. Www.youtube.com. https://www.youtube.com/watch?v=Sux91FJ3Xe8

Warren, D. K., Marashi, A., Siddiqui, A., Asim Adnan Eijaz, Pradhan, P., Lim, D., Call, G., & Dras, M. (2022). Using machine learning to study the effect of medication adherence in Opioid Use Disorder. *PLOS ONE*, *17*(12), e0278988–e0278988. https://doi.org/10.1371/journal.pone.0278988

Chief Coroner. *An Urgent Response to a Continuing Crisis Report*. 1 Nov. 2023, www2.gov.bc.ca/assets/gov/birth-adoption-death-marriage-and-divorce/deaths/coroners-service/death-review-panel/an\_urgent\_response\_to\_a\_continuing\_crisis\_report.pdf.

Centre for Addiction and Mental Health. *What Is Opioid Agonist Therapy?* 2016.

GeeksforGeeks. “Random Forest Algorithm in Machine Learning.” *GeeksforGeeks*, 12 July 2024, www.geeksforgeeks.org/random-forest-algorithm-in-machine-learning/.

Government of Canada. “Opioid Use Disorder and Treatment.” *Www.canada.ca*, 25 July 2023, www.canada.ca/en/health-canada/services/opioids/opioids-use-disorder-treatment.html.

IBM. “Random Forest.” *Ibm.com*, 20 Oct. 2021, www.ibm.com/think/topics/random-forest.

National Cancer Institute. “NCI Dictionary of Cancer Terms.” *National Cancer Institute*, Cancer.gov, 2020, www.cancer.gov/publications/dictionaries/cancer-terms/def/agonist.

“Opioid Use Disorder.” *BCCSU*, www.bccsu.ca/opioid-use-disorder/.

OpenAI. “Introducing ChatGPT.” *Openai.com*, 30 Nov. 2022, openai.com/index/chatgpt/.

# Appendix

A screenshot of a computer

Description automatically generated

A structure of a chemical formula

Description automatically generated

Figure 28: Buprenorphine compound structure taken from chEMBL.

Figure 27: Aspirin compound structure taken from chEMBL.

A diagram of a molecule

Description automatically generated

Figure 29: Fentanyl compound structure taken from chEMBL.