in this study we obtained compared genertive diffusion model diffdock with a graph neural network regression model equibind which was layered with a random forest model (RFscore) as a scoring function and a xgboost classifer. The discrimative pipeline was intialized by curating and combining a dataset dopamine receptor actives from binding database and decoys (chemically similar inactives) from the Database of Useful Decoys Enhanced (DUD-E) server. The combined dataset consisting of 3027 actives and 7300 decoys was docked on the D1 holoreceptor from PDB - 7joz. The tensor output was featurized with RFscore, the featuized data was trained and vaildated with Xgboost. Using establised actives, we procceded to do a similarity screening from the enamine chemical library (>44 billion compounds space at the time of writing) with the tanimoto parameter set at 0.8, we intended to hunt for novel scaffold and chemotypes. We obtaned 10000 compounds that were docked with diffdock and Equibind-Rfscore-Xgboost pipeline. we did a min-max normalization of the confidence metric from diffdock and a hit probability prediction from the EFX pipeline. Overall, Diffdock 34% was more confident. We curated the 40 top scoring compound for both diffdock and the EFX model, with an emphasis for compound with unique scaffolds. Prediction accuracy will be tested with a novel real time in vitro validation system with genetically encoded biosensor dLight

Discrimination vs. Generation: The Machine Learning Dichotomy in Dopaminergic Hit Detection.

Certainly, here's a refined version of your abstract:

In recent times, Computer Aided Drug Design (CADD) has become the new paradigm in small molecule drug screening and discovery. These High-Throughput systems modify and eliminate the bottlenecks associated with the traditional radioligand binding assays to accelerate drug development. Although, these systems are not pristine and devoid of challenges. A major pitfall with CADD is the absence of conventional definition for the scoring function for predicted protein-ligand binding. Here, we attempt to provide context to this problem with Artificial Intelligence, by screening for dopamine D1 hits with a combination of ligand and structure-based approaches. In this study, we juxtaposed the generative diffusion model, DiffDock, with a graph neural network regression model, EquiBind, which was further complemented by the RFscore (a random forest scoring function) and an XGBoost classifier. Our discriminative pipeline commenced with the curation and amalgamation of a dataset of dopamine receptor actives sourced from the Binding Database, and decoys (chemically analogous inactives) extracted from the Database of Useful Decoys Enhanced (DUD-E) server. This composite dataset, comprising 3,027 actives and 7,300 decoys, was docked onto the D1 dopamine holoreceptor from PDB-7joz. The resulting tensor output, post-featurization with RFscore, underwent training and validation with XGBoost. Guided by known actives, we embarked on a similarity screening of the Enamine chemical library, which, at the time of this study, spanned over 44 billion compounds. Employing a Tanimoto coefficient threshold of 0.8, our objective was to discern novel scaffolds and chemotypes. This search yielded 10,000 compounds that were subsequently docked using both DiffDock and the EquiBind-RFscore-XGBoost pipeline. After applying min-max normalization to the confidence metric derived from DiffDock and procuring hit probability predictions from the EFX pipeline, DiffDock emerged as 34% more confident. We shortlisted the top 40 scoring compounds from both methodologies, placing a premium on compounds boasting unique scaffolds. The predictive accuracy of these selections will be assessed via a pioneering real-time in vitro validation system, harnessing the genetically encoded biosensor, dLight.

When I first think about the drug approval process, it seems a bit counterintuitive to me. Why should drug companies have to go through such rigorous testing to prove a drug's efficacy when we, as consumers, have a plethora of information at our fingertips? Platforms like Yelp, Consumer Reports, and Quora offer reviews on almost everything, from the best restaurants in town to which plumber you should hire. And generally, these reviews are pretty reliable. So why can't we just apply this same logic to the pharmaceutical industry?

But then I started to think deeper about the implications. Unlike choosing a bad restaurant and getting an upset stomach, choosing a wrong drug can have severe, if not fatal, consequences. The stakes are incomparably higher. We're dealing with people's health and lives, not just their Saturday night dinner plans.

Of course, there's the argument that the marketplace will force out bad products. It's a basic economic principle. If a product isn't good, it won't sell. Businesses that don't adapt to consumers' needs go under. And while this principle might hold true for industries where consumers can easily judge a product's quality after one use, the pharmaceutical world is a lot murkier. Just because a drug sells well doesn't necessarily mean it's effective. We've seen this in the past, before the 1962 regulations, where profit-driven motives didn't always align with the drug's clinical utility.

Now, don't get me wrong. I'm all for a bit of deregulation here and there. In some areas, bureaucracy can be more of a hindrance than a help. But when it comes to health and safety, I'd rather err on the side of caution. People can be desperate for cures and treatments, and they might not have the luxury to wait and see how the market reacts to a new drug. They need something that's proven to work, not just something that's proven to sell.

I also worry about the long-term impacts of a market-driven approach to drug approval. What happens to research and development? Will drug companies be more inclined to produce drugs they think will sell rather than drugs that are genuinely needed? This could stifle innovation and leave us without treatments for less common, but still serious, diseases.

I also find the comparison to review platforms a bit misplaced. Sure, we can go on Yelp and see if a restaurant is worth our time, but can you imagine deciding on a drug based on Quora answers? It just doesn't sit right with me. Plus, while these platforms offer a wealth of consumer perspectives, they can't replace the rigorous, scientific analysis that goes into proving a drug's efficacy.

To sum up, while I understand the frustrations of business leaders and politicians who view the current drug approval process as burdensome, I don't think the solution is as simple as letting the market decide. Health isn't something we can gamble with. So no, I wouldn't be comfortable with removing requirements for companies to prove efficacy during the drug approval process. Some things are just too important to leave up to chance or market forces.

Top of Form

Problem: The research question is whether there is a difference in the efficacy of drug A and drug B for the treatment of diabetes. The investigator plans to test the effect of each drug in a randomized trial after 4 weeks of treatment. If the mean fasting plasma glucose levels in treated type 2 diabetes patients was found from literature to be 110 mg/dL, with a standard deviation of 20 mg/dL.  If the investigator wants to detect a difference of 15% or more in mean fasting plasma glucose between the two treatment groups:

1. What would be your null hypothesis?

The null hypothesis (*H*0​) is a statement that there is no difference between the two groups

1. What would be your alternative hypothesis

The alternative hypothesis (*Ha*​) posits that there is a difference between the two groups. Given the scenario, it doesn't specify whether one drug is better or worse than the other, just that there's a difference.

1. What level of statistical significance would you suggest (type I error, alpha)? And why?

I would suggest an alpha level of 0.05 (or 5%), as this is the most commonly used threshold in biomedical research, balancing the risks of Type I and Type II errors. It provides a reasonable balance between being too strict (potentially missing true effects) and being too lenient (risking false positives).

1. What power do you suggest?

I would suggest a power of 0.80 (or 80%)

|  |  |  |
| --- | --- | --- |
| |  | | --- | | 1. Please help the investigator to select the appropriate sample size to detect this desired effect | | 53 |

To determine the appropriate sample size for this study, it's essential to consider our statistical hypotheses and the details of the study design. This includes focusing on the effect size, the minimum meaningful difference we're looking to detect (15% in this scenario), the chosen power of the study (80%), and the significance level (5%). Based on the data provided from previous literature, the variability in measurements is 110mg/dL. With a standard deviation given as 20mg/dL, the expected difference between the two group means is 15% of 110mg/dL, which translates to 16.5mg/dL. Given that we're operating with a two-sided t-test, calculations indicate that for this study design and to achieve the desired detectable difference, the required sample size would be approximately 44 participants in each group. Hence, a total of 88 participants are recommended for the entire study.

I hope this message finds you well. My name is Temi, and I am reaching out to express my deep interest in the Innovation Lead position at the Cambridge Innovation Center.  My background in Pharmacology and vast experience in both academic research at Harvard Medical School and the commercial sphere at Sanofi, I bring a unique blend of scientific expertise and business acumen to the table.

My doctoral research at Harvard Medical School, specifically in Dr. Dong Kong's Lab, has honed my skills in using cutting-edge technologies, especially within the realm of drug discovery and machine learning. A testament to my innovative acumen, I played a pivotal role in the design and optimization of the novel dLight biosensor for high throughput in vitro screening of dopaminergic hits. Furthermore, I've successfully leveraged virtual screening augmented with artificial intelligence to optimize dopaminergic hits for the modulation of neuronal circuit regulating appetite. Also, on my thesis project, I explored both discriminative and generative machine learning paradigms and develop a scoring system for high-potential compounds. These capabilities align seamlessly with the Innovation Lead's role in shaping product concepts or breakthrough technologies to fit Novo Nordisk's demands.

Furthermore, my tenure at Sanofi in Nigeria showcased my capacity to drive business growth and profitability. As a Key Account Manager, I spearheaded strategies that revived dwindling accounts and significantly improved revenue streams, generating $5.7M in value over 4 years. The experience of collaborating transversally to broker public-private agreements and mentoring new representatives adds depth to my understanding of the pharmaceutical industry's nuances.

I am excited about the prospect of leveraging this unique blend to drive Novo Nordisk’s solutions in the biotech arena, and I'd welcome the opportunity to discuss how I can add value to your team. I have my resume attached to this mail. I look forward to hearing from you.

I hope this message finds you well. I am Temi, I am reaching out to express my deep interest in the Innovation Lead position at the Cambridge Innovation Center.  My background in Pharmacology and my vast experience in both academic research at Harvard Medical School and the commercial realm at Sanofi, I bring a unique blend of scientific expertise and business acumen to the table.

Bottom of Form