In Silico Prediction of Drug Permeability through an Inflamed Blood-Brain Barrier using Molecular Feature Modeling

Tanish Jain

Abstract

The introduction of computational techniques to analyze chemical data has given rise to the analytical study of biological systems, known as “bioinformatics”. One facet of bioinformatics is using machine learning (ML) technology to detect multivariable trends in various cases. Among the most pressing cases is predicting blood-brain barrier (BBB) permeability. The development of new drugs to treat central nervous system disorders presents unique challenges due to poor penetration efficacy across the blood-brain barrier. This research aims to mitigate this problem through an ML model that analyzes chemical features and accounts for patient variance. To do so: (i) An overview into the relevant biological systems and processes as well as the use case is presented. (ii) Second, an in-depth literature review of existing computational techniques for detecting BBB permeability was undertaken. From there, inflammation is identified as a variable characteristic unexplored across current techniques and an initial solution is proposed. (iii) Lastly, a two-part *in silico* model to quantify the likelihood of permeability of drugs with defined features across an inflamed BBB through passive diffusion is developed, tested, and discussed. Testing and validation with the dataset determined the predictive logBB model’s mean squared error to be ~ 0.112 units. The currently used neuroinflammation model’s mean squared error was approximately 0.3 units. The developed model outperforms the currently used model to predict permeability into the BBB.

Introduction

Background

Drug development is a lengthy, complex, and costly process, accompanied by a high degree of risk surrounding eventual success. The development of a single prescription medicine that is approved for marketing is estimated to cost drugmakers > $2 billion. The challenge is amplified in the case of the development of central nervous system (CNS) drugs. CNS drugs typically take 20% longer to develop and 38% longer to get FDA approval than non-CNS drugs, with a failure rate of 85% (DiMasi, 2016). The larger rejection rate, stems in part, due to the poor penetration ability across the blood-brain barrier (BBB), hence limiting the growth of the neuro-therapeutics field (Liebner et. al., 2018; Pardridge, 2005; Muehlbacher et. al., 2011). The BBB comprises epithelial-like tight junctions that limit diffusion from the blood to the extracellular fluid of the CNS to molecules with a molecular mass < 400 Da and a hydrogen bond count < 8 bonds; the tight junctions act as a physical and biochemical barrier between the CNS and the bloodstream, maintaining the homeostasis of the CNS (Lochhead et. al., 2020; Daneman & Prat, 2015). While the BBB shields the brain from infectious and toxic substances, it also restricts the ability of drugs to target specific locations in the brain to treat neurological disorders.

A thorough understanding of the BBB is necessary in both the academic and industrial fields due to its importance for treating longstanding untractable diseases such as Alzheimer's and Parkinson's (Reichel, 2006). The highly selective nature of the BBB deters effort to develop solutions for neurological diseases and disorders; therefore, the neurotherapeutics front faces a dilemma in that there is a small number of molecules for the majority of CNS disorders. Without a method to gauge permeability before approaching clinical trials, decades could lapse before the successful treatment of neurological diseases, leading many to live with brain dysfunction.

Goals

The goal of this project was to create an applicable machine learning model (ML) to predict the permeation value of chemical compounds across the blood-brain barrier and serve as a preliminary step for drug discovery. This includes the design and development of an *in silico*

model to input a compound's molecular features to calculate logBB values, which is an established metric of permeability across the blood-brain barrier. The preceding step would apply for healthy barriers. The applicability of the proposed model was derived from the neuroinflammation sub-model, using an unprecedented component that factored a patient's C-reactive protein (CRP) level and adjusted the predicted logBB value for neuroinflammation. Such a model can predict the permeability of the BBB under diseased and/or inflamed conditions, which is more likely to be encountered in neurological diseases. Machine learning models can thus be applied in preclinical trials to discard drugs that cannot permeate the BBB before spending money and time in clinical trials.

Literature Review

The BBB holds the most prominent effects in determining the pharmacokinetic properties of bioactive drugs in the brain. It is composed of endothelial cells, pericytes, and astrocytes in direct contact with brain tissue and differs from the typical blood vessel because the endothelial cells form tight junctions, heightening the selectivity (Abbott et. al., 2010; Bechmann et. al., 2007). This allows it to restrict the time course of a compound's absorption into the extracellular brain space. The measure of a molecule's permeability is governed by logBB value, using the formula:

$$logBB = log(\frac{C_{brain}}{C_{blood}}) \quad \text{Eq. 1}$$

Certain physicochemical descriptors of drug and drug-like compounds that are indicative of molecule binding capacity have heavy influence in determining whether a molecule can diffuse, actively or passively, over the tight junctions of the BBB (Geldenhuys et. al., 2015). These drug properties cannot be linked, however, to BBB diffusion without intricate nonlinear computation methods because there are no present equations that can identify this correlation. Fortunately, computational techniques such as deep learning technology can provide this multidimensional simulation (Gerebtzoff & Seelig, 2006).

The literature review was categorized into three types of models. First, standard computational approaches without artificial intelligence (AI) techniques were explored to determine the need for machine learning implementation. Second, ML models that focused on structural

quantification as input data were analyzed. Third, similar ML models that either solely focused on chemical features or had a joint focus that prioritized chemical features were analyzed.

Conflicting literature surrounding the merit of ML approaches compared to traditional regression for analysis of clinical data has encouraged the research of both techniques for the prediction of BBB permeability (Christodoulou et. al., 2019; Huang et. al., 2020). A prominent feature of traditional analysis is taking 3D structures of molecules and quantifying the data into 1D descriptors, and methods such as VolSurf ® (Molecular discovery, UK) have been utilized for this purpose. Results in the form of correct logBB classifications have had a wide range of accuracy, from 79% to 90%, supposedly credited to variance in technique (Crivori et. al., 2000; Ooms et. al., 2002). Other approaches have varied both the input type and method of output. One study calculated logPS values to represent penetration through the BBB and used the values as a dataset to make predictions based on a drug's logD value, polar surface area, and van der Waals surface area of basic atoms. More unique attempts as such don't do classification accuracy but instead R^2 values yet have had scores tending to be less than 0.75 (Liu et. al., 2004; Nicolazzo et. al, 2006; Pucas et. al., 2019).

Machine learning is a subfield of AI that relies on a computer's ability to learn patterns on its own. Its application is explored in three methods: supervised, unsupervised, and reinforcement learning (Réda & Delhaye-duriez, 2020). In the field of drug discovery, supervised learning has proved to be the most common technique due to the need to verify information and inability to cluster data of >10 variables. Similar to standard computation techniques, machine learning approaches have also explored structural descriptors such as cross-sectional area to predict BBB permeability; such models have achieved accuracies as high as 88% but necessitate a larger dataset relative to standard approaches (Plisson & Piggott, 2019). A more recent approach quantified the structure through its molecular fingerprint and used that as its primary descriptor alongside supplementary chemical features. A molecular fingerprint converts a molecule's structure into a bit string which encodes the structure as a descriptor (Muegge & Mukherjee, 2015). This approach achieved an improved accuracy of 91.9%, boosting the reputability of ML approaches for classification of drug permeability through the BBB.

Furthermore, ML can also be leveraged for a much larger variable count that traditional regression techniques cannot handle. The ability to do so has led scientists to hand-select features believed to play a role in BBB permeability. One comprehensive model attempted logistic regression, linear discriminant analysis, k nearest neighbor, C4.5 decision tree, probabilistic neural network, and support vector machine techniques using their custom dataset but was unable to par computational techniques (Li et. al., 2005). Generating descriptors has become a common technique for these approaches, and programs such as CODES which organizes molecules from a topological point of view have been leveraged to do so (Dorronsoro et. al., 2004). This has also given rise to the use of deep neural networks, a technical mimic of the human brain through the use of nodes, to discover underlying relations in drug data (Wang, 2003; Alsenan et. al., 2020). Neural networks are a proven computing technology for identifying hidden patterns in raw data and generalizing nonlinear correlations to go beyond a given dataset (White, 1989). An important aspect of neural networks is the adaptable training mechanism. The shifting of weights for different types of data allows a model to account for incomplete datasets and varying importance in the molecules' descriptors. This allocates room for model improvement without rewriting source code because inputting new, diverse data can be seamlessly integrated (Guo et. al., 2017). Perhaps one of the strongest models in this literature review came from a multi-core SVM method that used drug side effects and indications as inputs for the prediction. This allowed the model to account for non-passive diffusion and achieve an accuracy of 97% with the limitation of being unable to determine how the drug entered (Miao et. al., 2019). The latter is critical in the field of drug discovery as chemical descriptors such as molecule size can point to the inability to permeate to a specific cause, and this is necessary for creating permeable drugs.

Despite accurate models having been developed to relate inputs such as structure or features to an output such as logBB, these forecasters are seldom employed because of assumptions of constants in the brain, limited validation, and in some cases, proprietary nature of the model (Broccatelli et. al., 2016).

To counter this issue, this research explored the causes behind the variation of BBB permeability in patients with neurological disorders (Casella et. al., 2014; Johansson et. al., 2009). The most

influential variance discovered was inflammation. Measured through a patient's acute phase CRP level, inflammation levels have been shown to have a direct correlation to BBB permeability by prompting leptin resistance across the BBB (Hsuchou et. al., 2008; Pan et. al., 2008). The protein is in the pentraxin family and produced primarily in the liver in response to cytokines interleukin-6 and interleukin-1β, both reactive to the inflammatory cycle and thus outlining the rationale behind using the CRP pathway for incorporation neuroinflammation (Dube et. al., 2008; Chen et. al., 2006). Wet lab research has determined a CRP threshold of alternance at 2.5 μg/ml; if higher than this, BBB impairment will factor into expected logBB value of chemical compounds (Hsuchou et. al., 2012). The model incorporates the aforementioned threshold as a basis for whether an adjustment is needed to the logBB output.

Methods

Data Acquisition

This project employed a custom-built, verified dataset of 281 molecules with varying permeability values. Names of molecular compounds and associated logBB values were pulled from a previous study that compiled data from over 100 source publications and verified each compound (Radchenko et. al., 2020). From here, the following 16 physicochemical descriptors for each molecule was extracted from the public PubChem database using Selenium tools ® (Thoughtworks Ltd., USA) and incorporated into the machine learning model: Molecular Weight, Mass, XLogP, Hydrogen Bond Acceptor Count, Hydrogen Bond Donor Count, Rotatable Bond Count, Monoisotopic Mass, Formal Charge, Topological Polar Surface Area, Heavy Atom Count, Isotope Atom Count, Atom Stereocenter Count, Bond Stereocenter Count, Covalently-Bonded Unit Count, Vapor Pressure, and Complexity. Molecules that had less than 50% of data unattained from the PubChem database were removed from the final set, with remaining missing values undergoing data correction.

Model Selection

This work defined a two-step process to determine the permeability of a compound on a patient-by-patient basis. 3D molecular modeling had been extensively researched over decades through deep learning techniques; therefore, a tabular data analysis approach was favored as it potentially

had unexplored avenues. With background from Alsenan et. al, a multi-layer perceptron regression (MPR) model was initially developed for the predictive logBB model. The MPR was proficient in determining patterns of high correlation. However, outlier accountancy was poor because the model was indeterminate for molecules that fell outside the concentrated features of the dataset. Repeated accuracy analysis failed to show improvement in accuracy in spite of changing hyperparameters; this was because the model was unable to learn the patterns from the least significant features, resulting in a mean absolute error of 0.179 and a mean squared error of 0.453. Since the goal of this project was to predict permeability of compounds that have not yet been synthesized or used, this posed a greater risk due to the inability to forecast the descriptors of future drugs in development. Ensemble modeling was employed next using the methodology from Plisson and Piggott. Bagging methods were used to train a series of weak models and combine them to create a stronger, more predictive model. Boosting methods were also used to sequentially train weak models so that multiple techniques could be used, each building off the previous. The techniques used in both bagging and boosting were Linear Regression, Ridge Regression, Lasso Regression, Bayesian Regression, and SVM modeling was less than that of existing models, Regression. The issue with ensemble modeling was the accuracy (error level) computed was less than that of existing research, although still stronger than the MPR model.

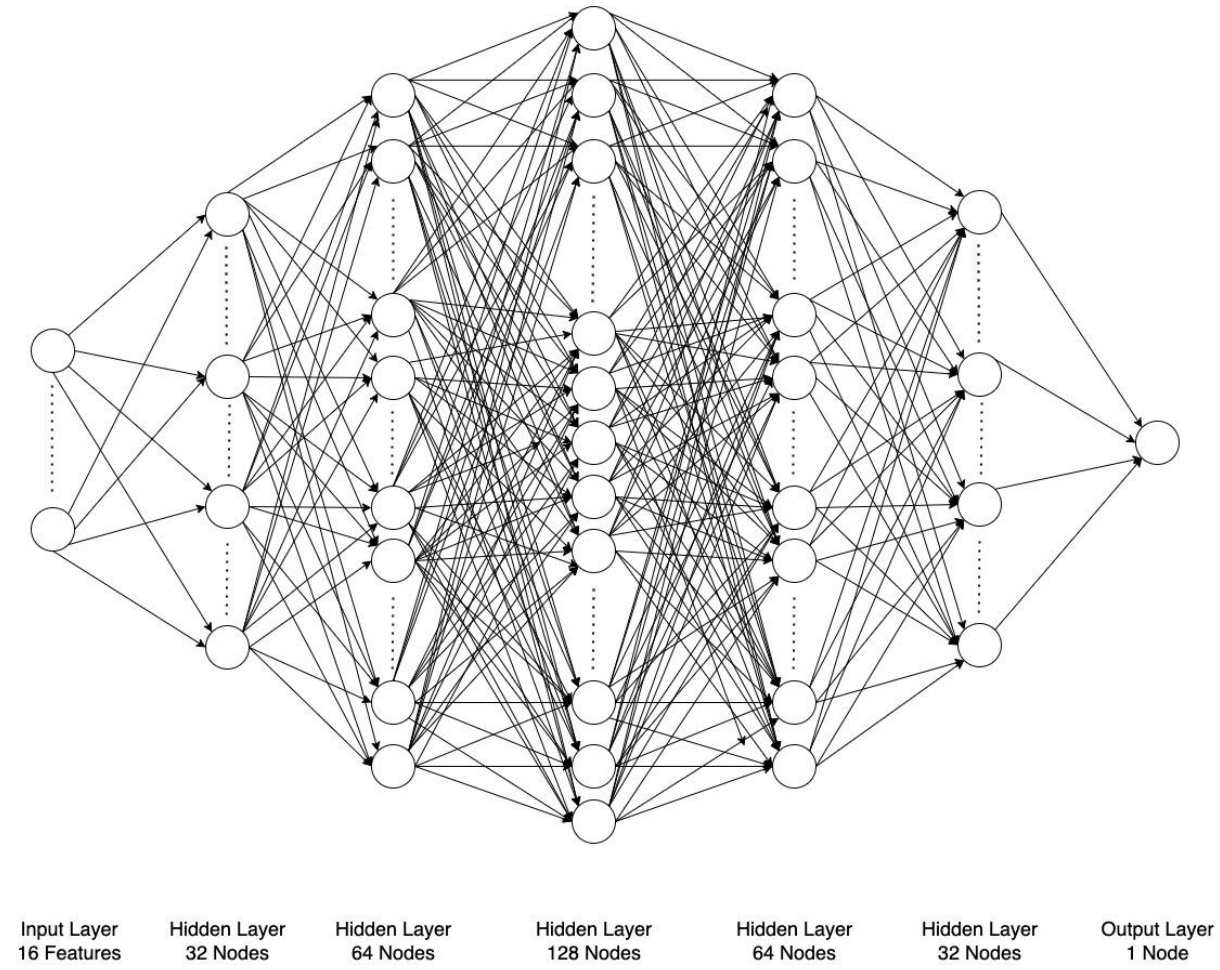

Figure 1. Predictive logBB Model Network Architecture (Original Image)

After experimenting with different model architectures and their performance, a decision was made to go with a fully connected neural network (FCNN). The FCNN, cited in this research as the predictive logBB model, takes in an assortment of preprocessed features. It is a neural network with an input layer of 16 nodes and an output layer with a single node depicting the logBB value. Between the input and the output layer there are 5 hidden layers with a node breakdown per layer of 64, 128, 256, 128, and 32 nodes, respectively, that are trained to learn the logBB value through multivariable analysis. This model was selected on two bases: it was able to learn the significance across the distribution, and it achieved an error level less than both models.

The neuroinflammation model works in conjunction with the logBB Model if there is additional information about the C-Reactive Protein levels in a patient. CRP levels on the surface did not

suggest a direct correlation with the logBB values; however, the second and third-order feature derivatives were successfully forecasted to have significant correlations with the logBB values. Hence, the decision was made to use a quadratic polynomial regression model for determining the validity of that correlation.

Solution

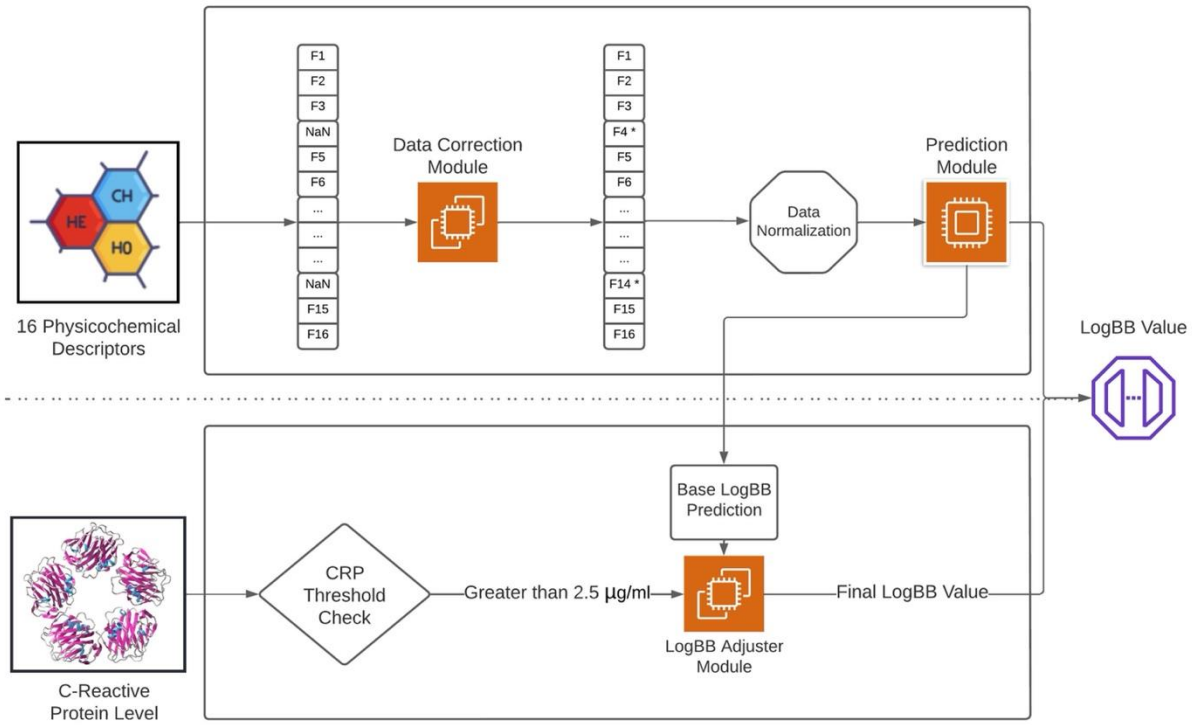

Figure 2. Synchronous Model Architecture (Original Image)

The software consists of three key components aside from data acquisition: preprocessing, the predictive logBB model and the neuroinflammation model.

The input receiver takes in the raw input of the 16 aforementioned physicochemical descriptors and, if desired, a patient's CRP level. Preprocessing is split into two consecutive steps: data correction and data normalization. To maximize compound inclusion in the model, missing features from the 16 descriptors were predicted using multiple quadratic polynomial regression models. Using an extended, manipulable equation generated with quadratic polynomial regression techniques, present features could be substituted in as constants to solve for missing

features. Yet since data correction was an educated prediction, the model associated lower weights with these values, so the neural network could include a variety of compounds without incorrect or exaggerated permeability output. The weights were determined based on the strength of the correlation between the missing feature and the features used for predicting it; a higher correlation indicated confidence in the prediction and therefore a higher weight. The strongest correlation was seen between formal charge and complexity and the weakest correlation was seen between heavy atom count and molecular weight.

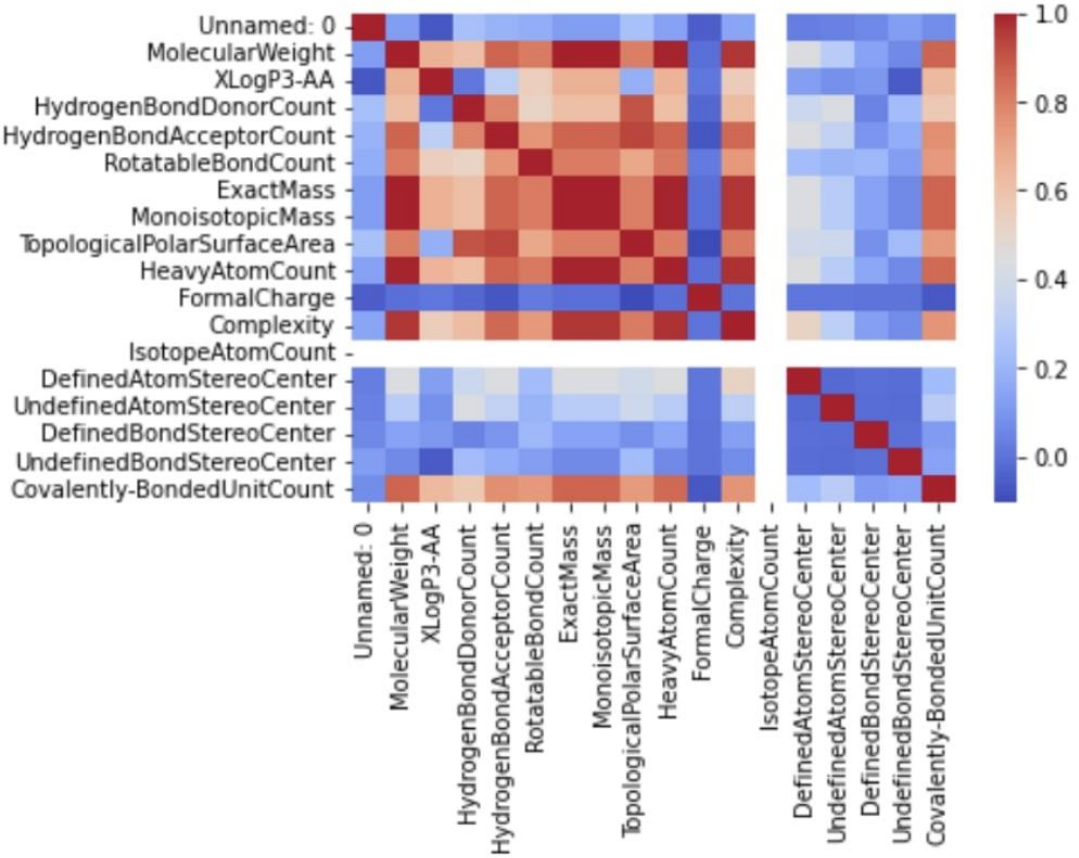

Figure 3. Feature Correlation Plot for Weightage Calculation (Original Image)

The complete data undergoes normalization next. The distribution of the training set was evaluated to extract how to bring the mean to zero and the variance to one (univariance). This methodology to normalize is transposed with new drug compounds that enter the model. Following preprocessing, the data is forwarded to the predictive logBB model. It was trained

across the dataset derived from PubChem and its error was determined using mean squared error compared to the known logBB value. When the predicted logBB value is generated, this path is halted and the neuroinflammation channel initiates. The CRP value is brought to a threshold checkpoint. If the value is less than 0.0025 g/L, then the inflammation level is deemed insignificant and the original logBB value is the final output. If not, then inflammation can be attributed to a non-healthy BBB and thereby altered restrictiveness. To adjust the drug permeation value, the initial logBB value is sent into the neuroinflammation model, a machine learning quadratic polynomial regression model, along with the CRP level to produce a more accurate logBB value for the drug in trial. The final output layer is a continuous output that predicts the logBB value for the user, also geared to minimize overfitting. The two-channel system's output is designed to represent the numerical permeation value through the BBB of a patient with an inflammation level relatively close to the CRP value.

Training and Validation

Small batch sizes were used to train the predictive logBB model and to ensure that the network was learning and not memorizing. Small batches also ensured that the network was more efficient and trained faster. Multiple batch sizes were experimented with to observe loss and error for each epoch, and the best one was highlighted. It was observed under a batch size of 32 that accuracy increased linearly with each epoch. This consequently proved that the network was not overfitting with the given dataset, so applicability was less of a concern. The neuroinflammation model had a more unique process in both training and validation. Since no present research has been made in identifying an equation rather than a correlation, there was no data available to train an *in silico* model. To acquire data for training the regression model, a common CRP level distribution was pulled from prior research (Zhao et. al., 2010). From this, statistical simulations of logBB adjustments were made based on inputs of logBB and CRP levels using a hand-developed Monte Carlo simulation method. 250 compounds were simulated and trimmed down based on available logBB information to 128 compounds that were pooled into a dataset for training. The summary statistics for the simulation are as follows: average CRP level was 0.0055 g/L, the standard deviation was 0.0012, and the number of simulations was 129. Mean squared error was once again used to predict the efficiency of the neuroinflammation model and charts for both models are presented below.

Table 1. Sample Training and Test Data with logBB Outputs and Delta Values for Comparison

Molecule Name	Expected logBB	Predicted logBB	Delta	Training or Test
Cimetidine	-1.42	-1.27	0.15	Training
Zolantide	0.14	0.29	0.15	Training
Carbamazepine	0	-0.198	0.198	Training
Temelastine	-1.88	0.35	0.35	Training
Codeine	0.55	0.38	0.38	Training
2-Methylheptane	0.86	0.72	0.14	Test
2-Methyloctane	0.98	0.976	0.004	Test
2-Methylnonane	1.05	1.27	0.22	Test
3-Methylpentane	1.01	1.149	0.139	Test
Cyclopropane	0.11	0.08	0.03	Test

The average delta, absolute value difference between expected and predicted logBB value, for the sample training molecules was 0.2456 and for the sample test molecules was 0.1066.

Results

The predictive logBB model made 150 passes through the training set and updated the model every 32 sample predictions. Five-fold cross-validation was used to summarize the skill of a model and limit unforeseen bias in the dataset. The model error chart and associated hyperparameters are detailed below.

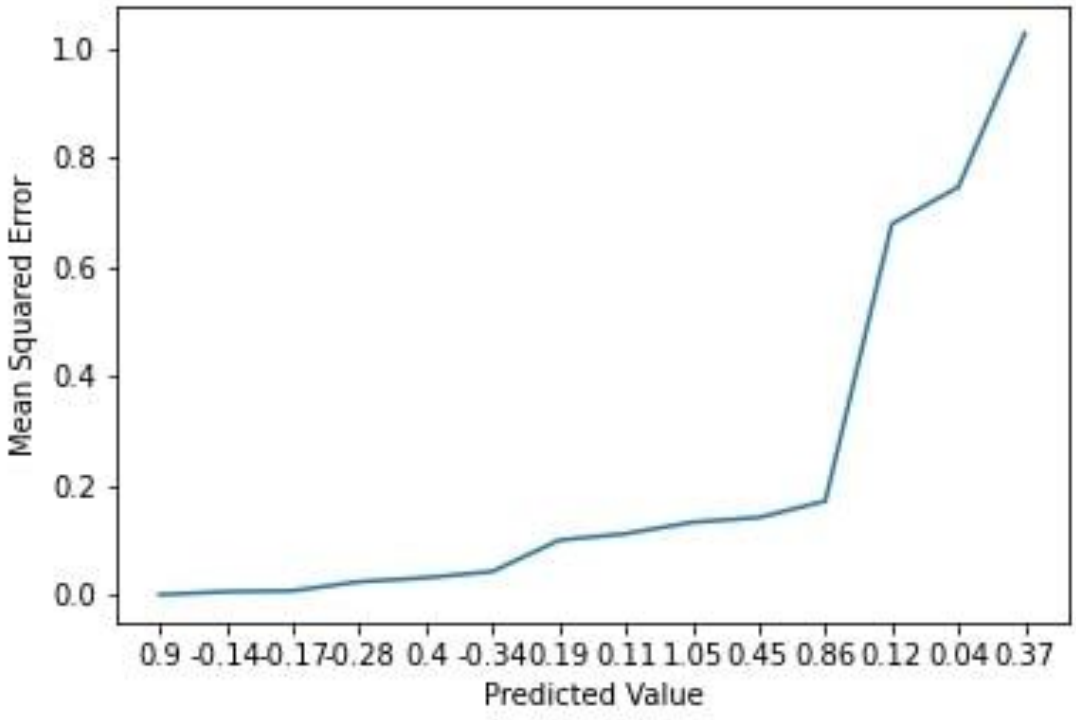

Figure 4. Predictive logBB Model Error. Part 1 of 2-Part Model (Original Image)

Predictive logBB Model Hyperparameters:

1. Number of Epochs: 150
2. Batch Size: 32
3. Fold Validation: Five-fold cross validation

The neuroinflammation model countered the lack of defined equations relating CRP levels to logBB values by using machine learning estimators. 50 estimators, or equations, were used to take the CRP distribution and associated effect on the restrictiveness of the BBB to simulate drug permeation and predict logBB values. Using Sklearn's Pipeline functionality, multiple regression models were added on a loop to pick the one that performed the best. Each of the 50 regression models was an estimator. Since there was no data to compare inflamed logBB values to, the error was based on a comparison to the same dataset employed in the first model. The purpose of this was to determine if the correlation can be quantified.

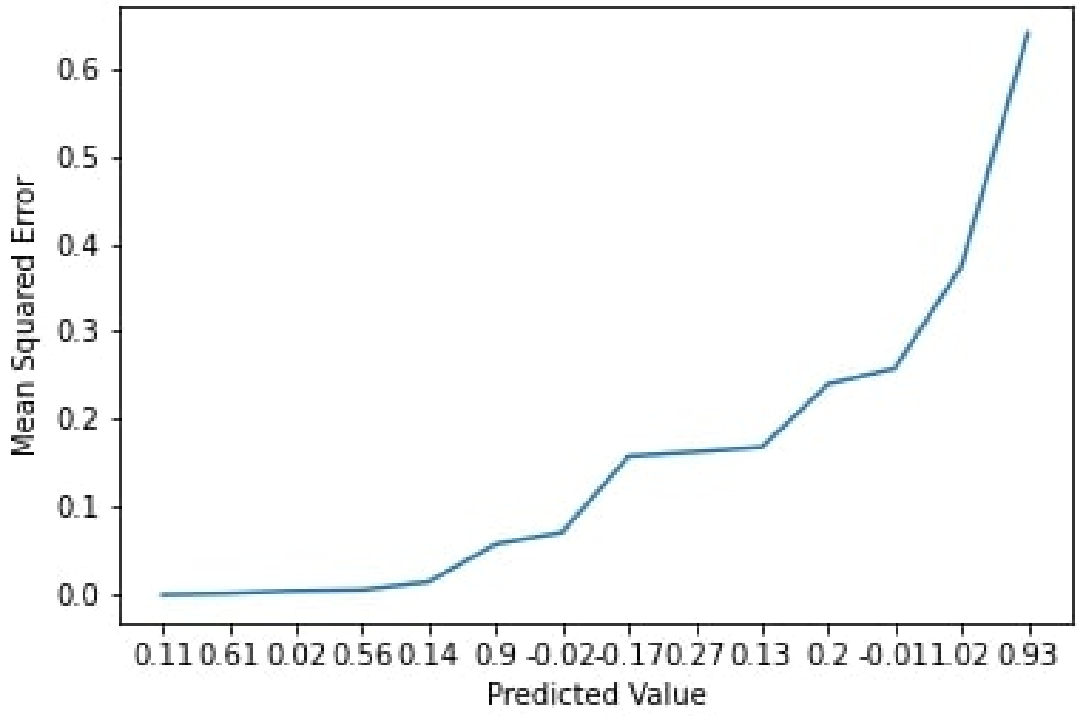

Figure 5. Neuroinflammation Model Error. Part 2 of 2-Part Model (Original Image)

Neuroinflammation Model Hyperparameters:

1. Number of Estimators: 50

Our results show that the predictive logBB model achieved a mean squared error of 0.112 and the neuroinflammation model achieved a mean squared error of 0.3. The logBB model surpassed the vast majority of models in modern research, suggesting that the hand-selected dataset features were appropriate for predicting BBB permeability. The neuroinflammation model, without any prior baseline, has achieved an error comparable to most cited prediction models. This suggests that second and third-order derivative correlation of CRP levels to logBB values is mathematically present and quantifiable given a database for drug permeability through inflamed BBBs.

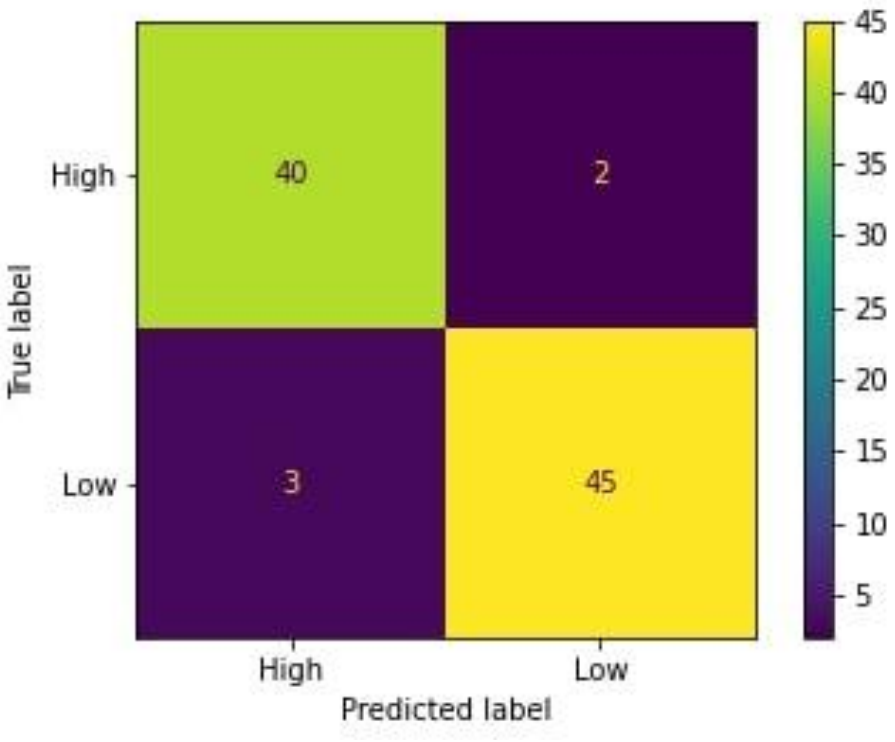

Figure 6. Confusion Matrix for Predictive logBB Model (Original Image)

The current model could not be assessed on accuracy because it was not classification-based. To gain further insight into the model's efficacy, a research-backed logBB threshold of 0.3 units was used to categorize predicted values as permeable (>0.3 units) or impermeable (<=0.3 units). From here, the classification could be compared to the true permeation from the dataset. The confusion matrix demonstrates a strong ability to determine if a drug compound can permeate across the BBB. The overarching accuracy of the test set was 94.4% For impermeable compounds, the precision and recall were 95.7% and 93.8%, respectively. For permeable compounds, the precision and recall were 93.0% and 95.2%, respectively. The minimal false positives and false negatives further demonstrates consistency across the range of potential logBB values.

Discussion

The model developed in this work either performed as well or outperformed nearly every model that was explored across both traditional computation and machine learning. Those that did

perform better had limitations such as being unable to assign permeability failure to one reason. While the model achieved a notable accuracy for the passive diffusion of small molecules, it is important to recognize that this model does not address permeation via active diffusion. Another area to be explored would be to compare this model with one that used drug side effects for prediction to include non-passive diffusion predictions. Beyond this, taking away more constants within the BBB would bring models closer to a realistic human brain BBB physiology. More diverse approaches presently serve as the best approach to improving accuracy and applicability.

The future work on this model will involve organ-on-a-chip biotechnology. Using the simulated blood-brain barrier, the model's applicability can be evaluated in a wet-lab setting. Furthermore, components of this project can be stripped and put into new endeavors in the neuroscience field. One such application is the neuroinflammation readings through CRP levels being used for the diagnosis of neurodegenerative diseases. Lastly, the complete model has the potential to open a path into precision and individualized medicine.

Conclusion

This project offers a unique add-on to modern day research on BBB permeability of drug compounds. Its design enables targeted clinical trials that takes away one of many assumptions made by pharmaceutical scientists in the drug development process. The ability to gauge within certain inflammation ranges opens a new avenue of drugs that could be introduced for diseases that cause inflammation in these ranges. Rather than an overarching medication that is suited for all patients, various small range neurotherapeutics will mitigate the effects of neurological diseases in a much smaller time period. Paralleled with the current work, this will be a notable stride in the neuroinformatics and drug discovery industries.

References

1. Abbott, N. J., Patabendige, A. A.k., Dolman, D. E.m., Yusof, S. R., & Begley, D. J. (2010). Structure and function of the blood–brain barrier. *Neurobiology of Disease, 37*(1), 13-25. https://doi.org/10.1016/j.nbd.2009.07.030
2. Alsenan, S., Al-turaiki, I., & Hafez, A. (2020). A recurrent neural network model to predict blood–brain barrier permeability. *Computational Biology and Chemistry*, *89*, 107377. https://doi.org/10.1016/j.compbiolchem.2020.107377
3. Bechmann, I., Galea, I., & Perry, V. H. (2007). What is the blood–brain barrier (not)? *Trends in Immunology*, *28*(1), 5-11. https://doi.org/10.1016/j.it.2006.11.007
4. Broccatelli, F., Salphati, L., Plise, E., Cheong, J., Gobbi, A., Lee, M.-L., & Aliagas, I. (2016). Predicting passive permeability of drug-like molecules from chemical structure: Where are we? *Molecular Pharmaceutics*, *13*(12), 4199-4208. https://doi.org/10.1021/acs.molpharmaceut.6b00836
5. Casella, G., Tontini, G. E., Bassotti, G., Pastorelli, L., Villanacci, V., Spina, L., Baldini, V., & Vecchi, M. (2014). Neurological disorders and inflammatory bowel diseases. *World Journal of Gastroenterology*, *20*(27), 8764-8782. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4112885/
6. Chen, K., Li, F., Li, J., Cai, H., Strom, S., Bisello, A., Kelley, D. E., Friedman-einat, M., Skibinski, G. A., Mccrory, M. A., Szalai, A. J., & Zhao, A. Z. (2006). Induction of leptin resistance through direct interaction of c-reactive protein with leptin. *Nature Medicine, 12*(4), 425-432. https://doi.org/10.1038/nm1372
7. Christodoulou, E., Ma, J., Collins, G., Steyerberg, E., Verbakel, J., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over

logistic regression for clinical prediction models. *Journal of Clinical Epidemiology*, *110*, 12-22. https://doi.org/10.1016/j.jclinepi.2019.02.004

8. Crivori, P., Cruciani, G., Carrupt, P.-A., & Testa, B. (2000). Predicting blood−brain barrier permeation from three-dimensional molecular structure. *Journal of Medicinal Chemistry*, *43*(11), 2204-2216. https://doi.org/10.1021/jm990968+
9. Daneman, R., & Prat, A. (2015). The blood–brain barrier. *Cold Spring Harbor Perspectives in Biology*, *7*(1). https://doi.org/10.1101/cshperspect.a020412
10. DiMasi, J. (2016). Cost of developing a new drug. *Journal of Health Economics*, *47*, 20-33. https://csdd.tufts.edu/cost-study
11. Dorronsoro, I., Chana, A., Abasolo, M., Castro, A., Gil, C., Stud, M., & Martinez, A. (2004). CODES/Neural network model: A useful tool for in silico prediction of oral absorption and blood-brain barrier permeability of structurally diverse drugs. *QSAR & Combinatorial Science*, *23*(23), 89-98. https://doi.org/10.1002/qsar.200330858
12. Dube, M. G., Torto, R., & Kalra, S. P. (2008). Increased leptin expression selectively in the hypothalamus suppresses inflammatory markers CRP and il-6 in leptin-deficient diabetic obese mice. *Peptides*, *29*(4), 593-598. https://doi.org/10.1016/j.peptides.2008.01.001
13. Geldenhuys, W., Mohammad, A., Adkins, C., & Lockman, P. (2015). Molecular determinants of blood–brain barrier permeation. *Therapeutic Delivery*, *6*(8). https://doi.org/10.4155/tde.15.32
14. Gerebtzoff, G., & Seelig, A. (2006). In silico prediction of blood−brain barrier permeation using the calculated molecular cross-sectional area as main parameter.

Journal of Chemical Information and Modeling, *46*(6), 2638-2650. https://doi.org/10.1021/ci0600814

15. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. (2017). On collaboration of modern neural networks. *Proceedings of Machine Learning Research, 70*, 1321-1330. http://proceedings.mlr.press/v70/guo17a.html
16. Hsuchou, H., He, Y., Kastin, A. J., Tu, H., Markadakis, E. N., Rogers, R. C., Fossier, P. B., & Pan, W. (2008). Obesity induces functional astrocytic leptin receptors in hypothalamus. *Brain*, *132*(4), 889-902. https://doi.org/10.1093/brain/awp029
17. Hsuchou, H., Kastin, A. J., Mishra, P. K., & Pan, W. (2012). C-Reactive protein increases BBB permeability: Implications for obesity and neuroinflammation. *Cellular Physiology and Biochemistry*, *30*(5), 1109-1119. https://doi.org/10.1159/000343302
18. Huang, JC., Ko, KM., Shu, MH. *et al.* Application and comparison of several machine learning algorithms and their integration models in regression problems. *Neural Comput & Applic* 32**,** 5461–5469 (2020). https://doi.org/10.1007/s00521-019-04644-5
19. Johansson, B., Starmark, A., Berglund, P., Rödholm, M., & Rönnbäck, L. (2009). A self-assessment questionnaire for mental fatigue and related symptoms after neurological disorders and injuries. *Brain Injury*, *24*(1), 2-12. https://doi.org/10.3109/02699050903452961
20. Li, H., Yap, C. W., Ung, C. Y., Xue, Y., Cao, Z. W., & Chen, Y. Z. (2005). Effect of selection of molecular descriptors on the prediction of blood−brain barrier penetrating and nonpenetrating agents by statistical learning methods. *Journal of Chemical Information and Modeling*, *45*(5), 1376-1384. https://doi.org/10.1021/ci050135u

21. Liebner, S., Dijkhuizen, R. M., Reiss, Y., Plate, K. H., Agalliu, D., & Constantin, G. (2018). Functional morphology of the blood-brain barrier in health and disease. *Acta neuropathologica*, *135*(3), 311–336. https://doi.org/10.1007/s00401-018-1815-1

22. Liu, X., Tu, M., Kelly, R. S., Chen, C., & Smith, B. J. (2004). Development of a computational approach to predict blood-brain barrier permeability. *Drug Metabolism and Disposition*, *32*(1), 132-139. https://doi.org/10.1124/dmd.32.1.132

23. Lochhead, J., Yang, J., Ronaldson, P., & Davis, T. (2020). Structure, function, and regulation of the blood-brain barrier tight junction in central nervous system disorders. *Frontiers in Physiology*. https://doi.org/10.3389/fphys.2020.00914

24. Miao, R., Xia, L.-Y., Chen, H.-H., Huang, H.-H., & Liang, Y. (2019). Improved classification of blood-brain-barrier drugs using deep learning. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-44773-4

25. Muegge, I., & Mukherjee, P. (2015). An overview of molecular fingerprint similarity search in virtual screening. *Expert Opinion on Drug Discovery*, *11*(2), 137-148. https://doi.org/10.1517/17460441.2016.1117070

26. Muehlbacher, M., Spitzer, G. M., Liedl, K. R., & Kornhuber, J. (2011). Qualitative prediction of blood–brain barrier permeability on a large and refined dataset. *Journal of Computer-Aided Molecular Design*, *25*(12), 1095-1106. https://doi.org/10.1007/s10822-011-9478-1

27. Nicolazzo, J. A., Charman, S. A., & Charman, W. N. (2006). Methods to assess drug permeability across the blood-brain barrier. *Journal of Pharmacy and Pharmacology*, *58*(3), 281-293. https://doi.org/10.1211/jpp.58.3.0001

28. Ooms, F., Weber, P., Carrupt, P.-A., & Testa, B. (2002). A simple model to predict blood–brain barrier permeation from 3D molecular fields. *Biochimica Et Biophysica Acta (BBA) - Molecular Basis of Disease, 1587*(2-3), 118-125. https://doi.org/10.1016/S0925-4439(02)00074-1

29. Pan, W., Hsuchou, H., He, Y., Sakharkar, A., Cain, C., Yu, C., & Kastin, A. J. (2008). Astrocyte leptin receptor (ObR) and leptin transport in adult-onset obese mice. *Endocrinology*, *149*(6), 2798-2806. https://doi.org/10.1210/en.2007-1673

30. Pardridge, W.M. The blood-brain barrier: Bottleneck in brain drug development. *Neurotherapeutics* 2, 3–14 (2005). https://doi.org/10.1602/neurorx.2.1.3

31. Plisson, F., & Piggott, A. (2019). Predicting blood–brain barrier permeability of marine-derived kinase inhibitors using ensemble classifiers reveals potential hits for neurodegenerative disorders. *Marine Drugs, 17*(2), 81. https://doi.org/10.3390/md17020081

32. Puscas, I., Bernard-patrzynski, F., Jutras, M., Lécuyer, M.-A., Bourbonnière, L., Prat, A., Leclair, G., & Roullin, V. G. (2019). IVIVC assessment of two mouse brain endothelial cell models for drug screening. *Pharmaceutics, 11*(11), 587. https://doi.org/10.3390/pharmaceutics11110587

33. Radchenko, E. V., Dyabina, A. S., & Palyulin, V. A. (2020). Towards deep neural network models for the prediction of the blood–brain barrier permeability for diverse organic compounds. *Molecules*, *25*(24), 5901. https://doi.org/10.3390/molecules25245901

34. Réda, C., Kaufmann, E., & Delahaye-duriez, A. (2020). Machine learning applications in drug development. *Computational and Structural Biotechnology Journal*, *18*, 241-252. https://doi.org/10.1016/j.csbj.2019.12.006

35. Reichel, A. (2006). The role of blood-brain barrier studies in the pharmaceutical industry. *Current Drug Metabolism*, *7*(2), 183-203. https://doi.org/10.2174/138920006775541525

36. Wang, S.-C. (2003). Artificial neural network. *Interdisciplinary Computing in Java Programming*, 81-100. https://doi.org/10.1007/978-1-4615-0377-4_5

37. White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, *1*(4), 425-464. https://doi.org/10.1162/neco.1989.1.4.425

38. Zhao, Y., Wang, R., Ma, X., Yan, X., Zhang, Z., He, X., & He, J. (2010). Distribution of c-reactive protein and its association with cardiovascular risk factors in a population-based sample of chinese. *Disease Markers*, *28*(6), 333-342. https://doi.org/10.3233/DMA-2010-0713