IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 13 December 2022, accepted 24 December 2022, date of publication 29 December 2022,
date of current version 1 February 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3233110

RESEARCH ARTICLE

Re-Routing Drugs to Blood Brain Barrier: A Comprehensive Analysis of Machine Learning Approaches With Fingerprint Amalgamation and Data Balancing

MOHAMMED YUSUF ANSARI[1], VAISALI CHANDRASEKAR[2], AJAY VIKRAM SINGH[3], AND SARADA PRASAD DAKUA[2]

[1]Electrical and Computer Engineering, Texas A&M University, College Station, TX 77843, USA
[2]Hamad Medical Corporation, Doha, Qatar
[3]German Federal Institute for Risk Assessment (BfR), 10609 Berlin, Germany

Corresponding author: Sarada Prasad Dakua (SDakua@hamad.qa)

This work was supported in part by the Qatar National Research Fund (a member of Qatar Foundation) under Grant NPRP-11S-1219-170106; in part by Hamad Medical Corporation, Doha, Qatar, under Grant IRGC-05-SI-18-360; and in part by the Medical Research Center, Hamad Medical Corporation.

ABSTRACT Computational drug repurposing is an efficient method to utilize existing knowledge for understanding and predicting their effect on neurological diseases. The ability of a molecule to cross the blood-brain barrier is a primary criteria for effective therapy. Thus, accurate predictions by employing Machine learning models can effectively identify the drug candidates that could be repurposed for neurological conditions. This study comprehensively analyzes the performance of the well-known machine learning models on two different datasets to overcome dataset-related biases. We found that random forest and extratrees (i.e., tree-based ensembled models) have the highest accuracy with mol2vec fingerprint for BBB permeability prediction, attaining AUC_ROC of 0.9453 and 0.9601 on BBB and B3DB dataset, respectively. Additionally, we have analyzed the impact of the data balancing technique (i.e., SMOTE) to improve the specificity of the models. Finally, we have explored the impact of different fingerprint combinations on accuracy. By employing SMOTE and fingerprint combination, SVC attains the highest AUC_ROC of 0.9511 on BBB dataset. Finally, we used the best-performing models of the B3DB dataset to evaluate the BBB permeability for drugs intended to be used for repurposing. Model validation for repurposing predicted the non-passage for most antihypertensive drugs and passage for CYP17A1 cancer drugs.

INDEX TERMS Blood brain barrier, drug permeability, drug repurposing, empirical study, machine learning.

I. INTRODUCTION

An increasing number of neurological diseases and a rapidly ageing population with several neuro-disorders has substantiated the escalation of healthcare/drug development expenditures. This has fueled the research toward expediting drug discovery and/or drug repurposing by utilizing existing knowledge on the structure and function of the central nervous system (CNS), drug-target interactions and pharmacological properties. The lower success rates of CNS drugs can be attributed to the insufficient knowledge of the pathophysiology of complex neuro-diseases, the presence of the blood-brain barrier (BBB) and poor target engagement, which could result in 15-19 years for drug advancement from discovery to regulatory approval [1]. The most significant challenge is the presence of BBB, a highly selective semipermeable barrier that protects the CNS from external insults, thus rendering physiologically effective drugs non-practical

The associate editor coordinating the review of this manuscript and approving it for publication was Shadi Alawneh.

for CNS applications [2], [3]. BBB acts as a physical and metabolic barrier with simultaneous transport and secretory functions. Clinical determination of the BBB permeability of compounds, though accurate is time-consuming, cost-ineffective and impractical for diverse drug candidates [4], [5], [6]. Alternatively, *in vitro* and *in vivo* animal models are employed to identify the permeability properties. However, poor mimicking of *in vitro* models and clinical differences in the drug-target interactions between animal models and human data have hindered drug development progress [7], [8]. Consequently, determining the physicochemical properties of the drug associated with the BBB permeability has not been tackled enough in the *in vitro* or *in vivo* studies.

Over the years, progresses in high-throughput screening (HTS) and omics technologies have resulted in enormous data availability in chemical activity [9], [10], [11]. Processing such large-scale data for faster drug development has been the focus of several research groups in CNS drug discovery. Predicting and forecasting the BBB transport by using computational models including artificial intelligence (AI) and machine learning (ML) can effectively accelerate the drug development process for neurological conditions [1], [12], [13]. Recent years have seen unprecedented applications of AI/ML methods in addressing diverse problems ranging from medical image analysis [14], [15], [16] to drug discovery [17]. Several AI/ML-based models have been proposed to facilitate expeditious CNS drug discovery/repurposing by minimizing the number of laborious and time-consuming BBB permeability studies [4], [18], [19], [20], [21]. Several approaches for the identification and optimal generation of key molecular properties that are involved in BBB permeability have been reported. Some of the preliminary *in silico* models predicting BBB permeability were based on quantitative structure-activity relationships (QSAR) data [20]. QSAR modelling helps in the prediction of specific characteristics from the molecular structure of compounds. Some of the common molecular descriptors used for this application are molecular property-based descriptors like 1D (Molecular formula), 2D (atom connection), 3D descriptors (Molecular shape) and Fragment-based descriptors (fingerprints) [21]. Several ML and deep learning (DL) techniques have been applied to predicting BBB permeability in recent years [22]. While ML/DL models would not necessarily provide the exact understanding of why some drugs cross or do not cross the BBB, its significance lies in the integrative practical applications, thus serving as an initial screening filter in high throughput screening of chemicals. ML techniques like support vector machine (SVM), random forest (RF), k-nearest neighbours (KNN), multidimensional linear regression and linear discriminant analysis (LDA) have been used for prediction [23], [24], [25].

SVM based models were found to be more popular and effective in predicting BBB permeability in several studies [1], [21], [25], [26]. For instance, SVM with four discrimination models was implemented on a dataset of 625 compounds with more than 85% accuracy, specificity and sensitivity in prediction [27]. SVM can work effectively with linear and non-linear datasets by mapping with higher dimensions. Hence, SVM has been commonly used for binary classification problems such as biological barrier permeability prediction. In 2018, Wang and team developed BBB permeability prediction algorithms using SVM, RF, kNN and multilayer perceptron (MLP) neural network for a dataset of 2358 compounds [28]. RF is an ensemble learning model that circumvents the drawback of overfitting seen in the decision tree (DT) model that works by constructing a binary tree of decision nodes for either regression or classification [19]. A recent study employed DL algorithms with 5-fold cross-validation to minimize redundancies involving undetected overfitting and overestimation. This model was found to achieve an accuracy of 0.97, an AUC_ROC of 0.98 and an F1 score of 0.92, which is a benchmark result among CNS drug studies. A couple of the key concerns regarding the applications of AI models are access and usage control over the clinical data [29]. While most of the present models have been studied with publicly available data, adaptations of the same with secured patient data needs to be treated with utmost care to maintain privacy and security. For this purpose several usage control models could probably be utilized to address the diverse security requirements [29].

The first and foremost challenge in applying ML/AI algorithms for BBB permeability prediction is data collection. BBB is a highly complex biological system that varies significantly within individuals and species. Hence, the experimental data on the BBB permeability of drug molecules exhibit poorly balanced datasets with several discrepancies [30]. A direct comparison of results from different studies cannot be done, as the data collection strategy, descriptor/fingerprint used and the model applied, vary significantly with each study. To address this, the current study will focus on analysing the model performances of two datasets (BBB and B3DB). As the model performances reported in the literature are biased by the limitations set by the datasets, a systematic study on the performance of different ML models could answer some of the discrepancies related to some of the best models for predicting BBB permeability. Therefore, the present work will compare and comprehensively analyse the model performance of different types of ML models across multiple datasets. Data imbalance is an another key challenge in BBB datasets. Therefore, we also evaluate the impact of data balancing technique (specifically SMOTE) on the classification accuracy. We believe that our empirical study will facilitate the progress of ML-driven applications for drug repurposing in neurological conditions.

Parallel to this, the predicting capability of the models is also dependent on the representations that best describe a chemical compound. Though several types of molecular descriptors and fingerprints exit, no single descriptor type can represent all attributes of the compounds. As an attempt to capture all aspects of the compounds to improve the predicting accuracy, the ML models will be exhaustively run

for multiple descriptors/fingerprints and their combinations. To the best of our knowledge, this is the first report to analyse the performance of several models over different descriptors, fingerprints and their combinations on two different BBB datasets. Lastly, the validation of the developed ML model will be done by venturing into the drug repurposing concept by predicting the BBB permeability of drugs of different applications.

II. METHODOLOGY

Drugs consumed to treat various diseases are typically 2D/3D molecules that can interact with the proteins (target) of the infected region of interest (ROI) [31]. By binding with the target, the drugs change the chemical properties of the ROI, thereby initiating healing. During the drug discovery, clinical trials determine whether the drug crosses the semi-permeable endothelial cells of the BBB and enters into the extracellular fluids of the CNS [32]. These trials are expensive, convoluted, time-consuming, and need biological and bioinformatics expertise for their success.

The recent success of machine learning (ML) models for classification and regression tasks has encouraged the bioinformatics community to develop ML models for predicting the drug BBB permeability. These models serve as a second opinion to the bioinformaticians, assisting in drug selection for the clinical trials, thereby minimizing the time and expense of clinical trials and improving the efficiency of the drug discovery process.

The first step in training an ML system for drug discovery is to represent the molecules in a standard format that can be used for feature extraction. A conventional molecular-input line-entry system (SMILES) comprehensively represents the drug's atoms, chemical bonds, cycles, and functional groups in an ASCII string representation. Next, the SMILES are processed to extract numerical handcrafted descriptors or different molecular fingerprints. Finally, the trained ML models map the numerical features of drugs to a probability of BBB permeability. Figure 1 provides an overview of the ML pipeline for drug BBB permeability prediction. This section provides information about the different molecular fingerprints, ML models, datasets, evaluation metrics, and implementation details of our empirical study.

A. DATASETS

We evaluate the fingerprints and ML models on two BBB datasets to overcome the biases to present findings consistent across datasets.

1) BBB DATASET

Wang et al. [28] compiled a list of compounds with LogBB concentrations from four different sources [33], [34], [35], [36]. The compounds were processed with MacroModel 11.1 to generate populated neutral tautomers at PH 7.0. Further processing with Open Babel standardized the dative bonds and assisted in generating SMILES representation for the compounds. The LogBB measure was used to generate the ground truth of whether the drug crosses the BBB. A compound was tagged BBB+ if LogBB$> = -1$ and BBB− if LogBB< -1.

We utilized the list of chemicals provided by Wang et al. [28] to generate the SMILES from Pubchem and the SMILES were further used to extract fingerprints and 1D/2D descriptors. RDKit raised errors on 54 SMILES during fingerprint extraction, resulting in a dataset with 2304 molecules. Of these molecules, 1766 are labelled BBB+ and 538 BBB−.

2) B3DB DATASET

Meng et al. [30] compiled compounds and their BBB permeability ground truth from fifty published articles or open-access datasets. Systematic data collection, data cleaning, and data curation protocols were followed during the dataset generation phase to provide a standardized benchmark for the BBB−related classification and regression problems. The resulting dataset contains 7807 compounds with BBB+/BBB− ground truth and 1058 compounds with LogBB concentration values. Of the 7807 molecules, 4956 are labelled BBB+ and 2851 BBB−.

B. MOLECULAR FINGERPRINTS

Molecular fingerprints are binary vectors that aim to capture different physical features of the molecules. Conventionally, fingerprints capture the molecules' path-based (e.g., FP2) and substructure-based (e.g., FP3, FP4, MACCS) features. The path-based fingerprints capture features that quantify the series of atoms in different regions of molecules. In contrast, the substructure-based fingerprints determine whether a mini-molecular fragment (with specific functional groups and bonds) is part of the large molecule. The substructure-based fingerprints have had a better success rate in drug discovery because of their ability to quantify functional group, bond, and atomic positioning information effectively. Some hybrid fingerprints (e.g., Avalon) calculate both the path-based and substructure-based features to comprehensively quantify the physical molecular properties. Additionally, manually handcrafted 1D/2D descriptors (e.g., PaDEL descriptors) of the molecules are also a popular way to capture molecular properties. Recently, unsupervised machine learning techniques for natural language processing have been employed to generate latent space representation of the molecular substructures. We extract the fingerprints from multiple chemistry toolkits (open babel [38], cdk [39], RDKit) and comprehensively evaluate the impact of molecular fingerprints, 1D/2D descriptors, and latent space representation of molecules by training several machine learning models and comparing their performance. Since most of the present studies explore ML and DL models with individual fingerprints are descriptors, there is always scope for better representation of the molecules for high predictive ability. While this could be addressed by developing new encoders, the existing fingerprints need to be systematically studied to identify the best representative option for the molecule. Further, to the best of our knowledge

TABLE 1. Summary of descriptors, fingerprints, and latent space vector representation of molecules along with their dimensions for BBB dataset.

Feature Name	Feature type	Description	Dimension	Dimension after PCA (95% variance retained)
1D_2D_imputed_KNN	1D/2D descriptors	Manually handcrafted 1D/2D descriptors extracted using PaDEL descriptors. Missing Null values have been imputed using the SKlearn KNN imputer	1441 (BBB)	94 (BBB)
FP2	Babel	Linear/Daylight style fingerprint that extract path-based features	1024	363
FP3	Babel	Substructure fingerprint that represents whether certain SMARTS pattern are present in the molecule	1024	20
FP4	Babel	Substructure fingerprint that represents whether certain SMARTS pattern are present in the molecule	1024	68
Avalon	RDK	Hybrid fingerprint that captures path-based and substructure-based features like an atom, bond, and ring patterns and their paths	512	308
Morgan	RDK	Captures chemical features like donor, acceptor, aromatic, acidic, and basic properties of the molecules along with their atomic connectivity	1024	604
RDK-MACCS	RDK	Substructure fingerprint with handcrafted 166 SMARTS pattern for drug molecules	167	74
Rdkit	RDK	Substructure fingerprint with properties of daylight fingerprints to capture atom and bond types	2048	843
Estate	CDK	The Estate fingerprint captures the electro-topological state (E-state) that contains information on functional group, Kier-Hall electro-negativity of atoms, and graph topology. [37]	79	23
Extended	CDK	Similar to Standard fingerprint but emphasizes on the ring structures and atomic properties	1024	386
Graph	CDK	Similar to Standard fingerprint but emphasizes on the connectivity of atoms	1024	134
Hybridization	CDK	Similar to Standard fingerprint but emphasizes on the hybridization state	1024	352
Klekota-roth	CDK	Substructure fingerprint with biological activity of the molecules	4860	326
MACCS	CDK	CDK implementation of MACCS, with some minor changes in SMARTS patterns and molecule aromaticity	166	75
Pubchem	CDK	Substructure fingerprint that represents whether certain SMARTS pattern are present in the molecule	881	142
Standard	CDK	Hashed fingerprints generated by examining paths of different lengths from key functional groups	1024	377
Substructure	CDK	Substructure fingerprint that represents whether certain SMARTS pattern are present in the molecule	1024	70
mol2vec	Unsupervised ML	Latent space vector representation of the molecules inspired by unsupervised natural language processing models.	300	33

this is the first paper to address cross-toolkit fingerprint analysis. While this can provide redundancy in fingerprints, the SMILES processing by each toolkit provides a different output. Additionally, when the outputs of SMILES reading by different toolkit are close to each other, they are still expected to be different. This could be attributed to the ability of different toolkits capacity to process the records and the difference in their aromaticity. The records accessed by one toolkit might not be handled by another toolkit. However, the real analysis should include records that could be handled by both toolkits. A summary of fingerprints and other descriptors is present in Table 1.

C. EMPIRICAL STUDY DESCRIPTION

We design a comprehensive empirical study to understand the impact of different fingerprints and synthetic over-sampling techniques on the classification accuracy of ML models. First, we extract eighteen fingerprints (described in Table 1) to create a dataset. The 1D/2D PaDEL descriptors contained missing values. We employed KNN imputer to fill the missing features of the descriptors. We observed that fingerprints are sparse binary vectors. To condense the fingerprint information, we use principal component analysis (PCA) to reduce individual fingerprint dimensions, retaining 95% of fingerprint variability (Figure 1). PCA is preferred over other dimensionality reduction (e.g., Kernel PCA) and feature selection methods (e.g., feature ranking-based selection) because it provides a straightforward mechanism to control the variability in the transformed data relative to the original data. Fundamentally, PCA projects the data on principle components/ Eigenvectors (i.e., directions that capture the variability in data), thereby suggesting directions in multi-dimensional space with the most information. Thus, only retaining the components with high variance can decrease the dimension of the data without significant loss of information, resulting in improved model efficiency and better generalizability on the test set. Next, we normalize the feature vectors by removing the mean and scaling it to unit variance, ensuring that the individual feature range is limited. We apply PCA to the entire dataset to reduce the dimensionality, the PCA enforced dataset was further subjected to parameter choice with Grid search to systematically choose the right dimensions and parameters.

Parameter choice has a significant impact on the accuracy of ML models. At the same time, a biased dataset split may not accurately evaluate the prediction capability of ML models for BBB permeability. We pair grid search with

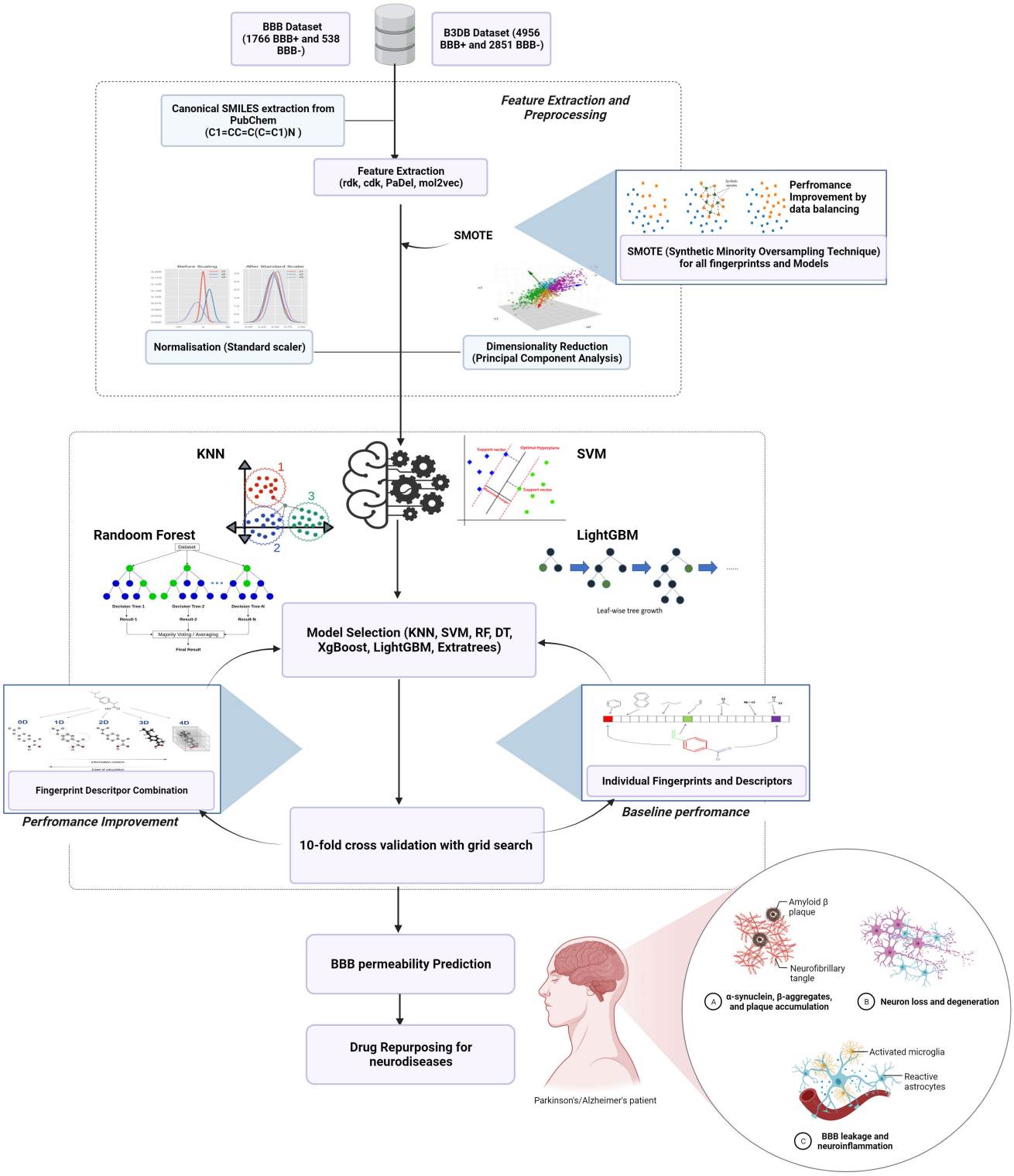

FIGURE 1. Empirical study pipeline summarizing, pre-processing, training, and inference stages.

10-fold cross-validation to select the parameters that provide the best results across the folds (Figure 1). In each step of the cross-validation, 1 fold of the dataset is treated as the test set and the remaining 9 folds are used for model training. We report the performance of the model which is the average test performance of the 10 folds, allowing

TABLE 2. Summary of hyperparameters employed for the gridsearch of ML models.

Machine Learning Model	Gridsearch Parameters
K-nearest neighbours (KNN)	n_neighbors: [5,10,15,20,25,35,50], weights: [uniform, distance], metric: [minkowski, euclidean, manhattan]
Support Vector Classifier (SVC)	C: [0.1,1,10], gamma: [1,0.1,0.01,0.001], kernel: [linear, poly, rbf, sigmoid]
Decision Trees (DT)	max_features: [auto, sqrt, log2], ccp_alpha: [0.1, 0.01, 0.001, 0.0001], max_depth: [5, 10, 15, 20, 35, 50, 100], criterion: [gini, entropy]
Random Forest (RF)	n_estimators: [10, 50, 100, 200, 1000], max_features: [auto, sqrt, log2], max_depth: [4,10,15,20,25], criterion: [gini, entropy]
Adaboost Classifier	n_estimators: [50,100,200,400,800], learning_rate: [10,1,0.5,0.1,0.001]
Extratrees Classifier	n_estimators: [10, 50, 100, 200, 1000], max_features: [sqrt, log2], max_depth: [4,10,15,20,25], criterion: [gini, entropy], class_weight: [balanced, balanced_subsample]
Gradient Boosting Classifier (GBC)	loss: [deviance, exponential], n_estimators: [100, 200, 500], max_features: [sqrt, log2], max_depth: [15,25], criterion: [friedman_mse, squared_error]
LightGBM	learning_rate: [0.01], n_estimators: [10, 20, 50], num_leaves: [6, 8, 12, 16], colsample_bytree: [0.5, 0.75, 0.9], subsample: [0.7, 0.75], reg_alpha: [1, 5], reg_lambda: [1, 5]
XgBoost	subsample: [0.7, 1], colsample_bytree: [0.7, 0.9], min_child_weight: [1,5,10], learning_rate: [0.1, 0.03]

us to evaluate the model robustly. We also employ Grid search to exhaustively search the parameter space for the best parameters for the 10 folds. Table 2 summarizes the hyper-parameter space searched for each ML model. In this initial experiment, we train and evaluate nine ML models with eighteen individual fingerprints without upsampling the minority class (i.e., BBB−), providing a baseline performance for ML models and fingerprints. We also examine the top-4 fingerprints for each model on the both datasets. Additionally, to interpret the performance variations of different fingerprints, we generate similarity matrix using centered kernel alignment (CKA) [40]. CKA has been proposed for comparing the neural network representations generated from different initializations. We employ it because it provides seamless and statistically correct method for comparing fingerprints of different dimensions.

In the next phase of the study, we introduce the resampling techniques to address the data imbalance. Resampling techniques can be employed to either incorporate under-sampling or oversampling the datasets. While under-sampling eliminates data from majority class, oversampling adds data to the minority class. Considering the criticality of each data point provided by diverse drugs, oversampling the minority class will retain such crucial information in the majority class. We also analysed the existing work to weigh the choice of using under-sampling techniques which indicates better performances of oversampling techniques in comparison to under-sampling [41], [42].Threfore, synthetic minority oversampling technique (SMOTE) has been employed to over-sample the minority class (Figure 1). SMOTE is initiated by applying KNN to the existing minority class instance (say, P). A new vector (say, V) is calculated between instance P and one of its neighbors. V is then multiplied by a random number between 0 and 1 to generate a synthetic sample with the same ground truth label as P. We noted during our literature review that a few papers applied SMOTE to the entire dataset before splitting it into train/val/test sets [20], [28], [43]. Applying SMOTE on the entire dataset introduces synthetic samples in the val/test sets, thereby not providing the actual performance of the ML models on the drugs of the val/test sets. In our study, we apply SMOTE only to the training set in each cross-validation fold, allowing us to understand actual performance gain with SMOTE. We also thoroughly compare the best performing fingerprints for each ML model with/without using SMOTE. We quantify the impact of SMOTE by computing the percentage increase and decrease in specificity and sensitivity, respectively. Additionally, we recommend the best fingerprint for each ML model, based on specificity and AUC_ROC, (it may be noted that these three measures are relatively correlated).

We observed in the literature that fingerprint combinations (including 1D/2D descriptors, MACCS, Klekota-Roth, and Pubchem) are heavily utilized for BBB permeability prediction tasks [4], [28], [43]. In the final phase of the study, we combine the best performing fingerprints for each model to validate whether merging multiple fingerprints improve the classification accuracy of the ML models. Additionally, we recommend the best combination of fingerprints for every ML model.

To showcase the use our empirical study, we employ the highly accurate models and their best performing fingerprints to predict the BBB permeability of drugs that are generally not intended for neurological applications as a part of repurposing strategy. The drugs were chosen carefully, considering their physiological implications to indicate the importance of computational models in drug repurposing. Additionally, we discuss our finding in detail and relate the BBB permeability to the drugs molecular characteristics.

D. MACHINE LEARNING MODELS

We incorporate nine different ML models in our study to understand the performance of models for the individual fingerprints and their combinations. A short description of ML models is as follows:

1) KNN: KNN works by calculating K nearest points to the query instance (i.e., test instance) using a distance measure. The query point is then assigned the majority label of the neighbors.
2) Support Vector Classifier (SVC): SVC finds an optimal plane in the high dimensional feature space that maximizes the margin between the instances of two classes.

3) Decision Tree (DT): DT is a tree-based ML model that uses features to generate decision rules at the nodes and classification outcomes at the leaves.
4) RF: RF is an ensemble learning method that outputs the class determined by most decision trees in the forest.
5) Adaboost Classifier: Adaboost employs a base estimator (e.g., DT) to fit the training set. Subsequently, Adaboost fits additional copies of the base estimator with higher emphasis (i.e., greater weight) on the incorrect samples to improve model accuracy.
6) Gradient Boosting Classifier (GBC): GBC builds a boosting ML model using an additive technique. In the case of binary classification, a single regression tree is introduced initially to fit the negative gradient of the loss function.
7) Extratrees (ET): ET is an ensemble-based technique that fits the training data on randomized decision trees using various sub-samples of the dataset. Additionally, it employs averaging across the decision trees to improve the prediction performance and reduce overfitting.
8) LightGBM: LightGBM is an open-source gradient boosting framework by Microsoft. Two crucial components of LightGBM are gradient-based one side sampling (GOSS) and exclusive feature bundling (EFB). GOSS allows LightGBM to track the under-trained instances, giving the model significant information gain. EFB relies on the sparsity of higher dimensional spaces to select a bundle of mutually exclusive features, thereby improving memory complexity and training time for the model.
9) XgBoost: XgBoost is a distributed gradient boosting library which maximizes computational speed and performance by building the trees in parallel. Furthermore, it adopts a level-wise strategy by evaluating the partial gradient sums to understand the quality of the possible dataset splits.

E. METRICS

We evaluate ML models using a suite of metrics. For balanced datasets, accuracy (i.e., *no. of correctly classified instances/ total no. of instances*) highlights the actual performance of the models. However, in the case of unbalanced datasets (i.e., BBB and B3DB), accuracy may not highlight the capability of the model to differentiate between the BBB+ and BBB− classes. Therefore, we evaluate the model specificity (i.e., *true negative/(true negative + false positive)*) and sensitivity (i.e., *true positive/(true positive + false negative)*) to measure the recall of the negative and positive classes. We also compute the area under the receiver operator characteristic curve (AUC_ROC) to quantify the model's ability to differentiate between the two BBB+/BBB− classes.

F. IMPLEMENTATION DETAILS

For extracting the fingerprints, we utilize pyfingerprint (python library). pyfingerprint is a wrapper around CDK and RDK, allowing fingerprint extraction with a single function call. It also provides a pre-trained mol2vec model for latent space vector generation using SMILES. Additionally, we employ padelpy (python wrapper library) for extracting 1D/2D PaDEL descriptors. Altogether, we extract 18 different numerical representations of SMILES for the BBB and B3DB dataset.

We utilize the Scikit-learn library to load the ML models and perform grid search with cross-validation (using gridsearchCV). The LightGBM and XgBoost models are loaded from their independent packages. We also employ imbalearn package for creating a pipeline that applies SMOTE only to the training sets for each cross-validation fold.

The ML models are trained on an HP Z8 workstation with 128 GB of RAM and 64 core Intel® Xeon(R) Silver 4216 CPU with a 2.10 GHz base clock.

III. RESULTS

In our empirical study, we process the individual fingerprints, 1D/2D features, and unsupervised ML representation of SMILES (i.e., mol2vec) using PCA to reduce data dimensionality. A brief description and dimensions of the extracted features are highlighted in Table 1. The empirical study contains three phases. In the first phase, we train the ML models with every feature and establish a baseline. In the second phase, we analyze the impact of oversampling (i.e., SMOTE) on the performance metrics of the ML models. In the final phase, we combine the best individual fingerprints for ML models to investigate whether fingerprint combination improves classification accuracy. The last section of the results addresses the validation of the developed model for drug repurposing application of anti-hypertensives, cancer, and anti-inflammatory drugs.

A. MODEL PERFORMANCE ON IMBALANCED BBB DATA

In the first phase of our study, we train nine ML models with individual features (i.e., fingerprints, descriptors) for predicting BBB permeability without rectifying the data imbalance. The ML models can be categorized into four categories, allowing us to reason about the classification accuracy for the BBB task. The categories are as follows: 1) In-memory-based (i.e., KNN) 2) Vector separation-based (i.e., SVC) 3) Tree-based (i.e., DT, RF, ET) 4) Boosting-based (i.e., Adaboost, GBC, LightGBM, XgBoost). We employ the grid-search with 10-fold cross-validation for training to find the hyper-parameters that allow ML models to perform best across the folds. The results for the ML model trained with individual fingerprints can be found in the supplementary material.

The ML model's performance without SMOTE and best-performing fingerprints for BBB and B3DB datasets are summarized in Tables 3 and 4, respectively. KNN performs lowest in terms of specificity among all the ML models for the BBB dataset. The performance comparison of KNN with that of the literature indicated similar results including the reduced specificity [44], [45]. Considering the

TABLE 3. Performance summary of ML models with and without SMOTE on the BBB dataset. The best performing ML models and their corresponding fingerprints is highlighted in gray gradients.

		Without smote				With Smote			
ML model	**Fingerprint**	**Accuracy**	**Specificity**	**Sensitivity**	**AUC_ROC**	**Accuracy**	**Specificity**	**Sensitivity**	**AUC_ROC**
KNN	1D/2D Descriptors	0.8885	0.5523	0.9909	0.9289	0.8633	0.8422	0.8698	**0.9312**
	mol2vec	0.8958	0.5912	0.9887	0.9316	0.8711	0.8198	0.8868	0.9272
	pubchem	0.8950	0.5951	0.9864	0.9310	0.8615	0.8216	0.8737	0.9239
	rdkit	0.9058	0.7028	0.9677	0.9207	0.7678	*0.9035*	0.7265	0.9157
SVC	Avalon	0.9102	0.6619	0.9859	0.9366	0.9171	0.7622	0.9643	**0.9406**
	MACCS	0.9084	0.6879	0.9757	0.9352	0.9093	0.7900	0.9457	0.9362
	mol2vec	0.9102	0.6861	0.9785	0.9327	0.9037	0.7882	0.9389	0.9348
	Estate	0.8824	0.6546	0.9519	0.9015	0.8555	*0.7919*	0.8749	0.9027
DT	FP3	0.8759	0.6282	0.9513	0.8288	0.8433	*0.7138*	0.8828	**0.8341**
	mol2vec	0.8459	0.6416	0.9083	0.8043	0.8242	0.6879	0.8658	0.8118
	FP4	0.8399	0.6490	0.8981	0.7947	0.7817	0.7044	0.8052	0.7981
	Estate	0.8329	0.5745	0.9117	0.7934	0.7969	0.7049	0.8250	0.8060
RF	mol2vec	0.8997	0.6210	0.9847	0.9375	0.9123	*0.7362*	0.9660	**0.9438**
	1D/2D Descriptors	0.8893	0.5541	0.9915	0.9340	0.9089	0.6675	0.9825	0.9420
	Avalon	0.8872	0.5449	0.9915	0.9308	0.9050	0.6508	0.9825	0.9405
	pubchem	0.8898	0.5692	0.9875	0.9304	0.9006	0.6730	0.9700	0.9385
Adaboost	1D/2D Descriptors	0.8932	0.6340	0.9723	0.9145	0.8820	0.7623	0.9185	0.9165
	Avalon	0.8958	0.6507	0.9705	**0.9213**	0.8654	*0.7677*	0.8952	0.9118
	FP4	0.8941	0.6320	0.9740	0.9141	0.8781	0.7603	0.9139	0.9101
	rdkit	0.9041	0.6620	0.9779	0.9176	0.8902	0.7436	0.9349	0.9100
Extratrees	mol2vec	0.9054	0.6396	0.9864	0.9453	0.9097	0.6841	0.9785	**0.9466**
	1D/2D Descriptors	0.8937	0.5782	0.9898	0.9415	0.9050	0.6451	0.9841	0.9384
	Avalon	0.8972	0.5951	0.9892	0.9410	0.8993	0.6174	0.9853	0.9381
	Estate	0.8750	0.6490	0.9439	0.9040	0.8742	*0.6937*	0.9292	0.9083
GBC	mol2vec	0.9050	0.6452	0.9842	0.9349	0.9106	*0.7195*	0.9689	**0.9387**
	1D/2D Descriptors	0.8993	0.6155	0.9858	0.9306	0.9145	0.6935	0.9819	0.9352
	Avalon	0.8963	0.6100	0.9836	0.9259	0.9058	0.6527	0.9830	0.9331
	pubchem	0.9002	0.6286	0.9830	0.9271	0.9015	0.6582	0.9757	0.9317
LightGBM	mol2vec	0.8932	*0.7882*	0.9253	**0.9368**	0.8958	0.7715	0.9338	0.9342
	1D/2D Descriptors	0.9010	0.7789	0.9383	0.9360	0.8998	0.7585	0.9428	0.9335
	MACCS	0.8919	0.7695	0.9292	0.9310	0.8946	0.7658	0.9338	0.9317
	pubchem	0.9019	0.7566	0.9462	0.9335	0.8967	0.7509	0.9411	0.9311
XgBoost	1D/2D Descriptors	0.9037	0.6545	0.9796	0.9413	0.9119	0.7418	0.9638	**0.9423**
	mol2vec	0.9032	0.6675	0.9751	0.9400	0.9067	0.7436	0.9564	0.9419
	MACCS	0.9019	0.6580	0.9762	0.9336	0.9045	0.7435	0.9536	0.9356
	substructure	0.9071	0.6897	0.9734	0.9302	0.9002	*0.7510*	0.9456	0.9295

similarity-based KNN algorithm used, it is not surprising that the imbalanced data predicted more BBB+ than BBB−. Interestingly for the B3DB dataset, KNN achieves similar and better specificity as other ML models. As a result, the AUC_ROC for KNN is higher for the B3DB dataset. The performance discrepancy of KNN between the dataset is due to the higher imbalance in the BBB dataset and the distribution of training instances in the high-dimensional feature space. SVC attains higher specificity and similar sensitivity on the BBB dataset compared to the KNN model, thereby improving the AUC_ROC. We could not train SVC for the B3DB dataset because of its poor scalability due to its cubic training complexity (i.e., $O(n^3)$, where n is no. of dataset instances). SVC is one of the widely explored model in literature for BBB prediction with different approaches to improve the model performance. In SVC, the model is built to assign unclassified new compounds to BBB+ or BBB−, thus making it a non-probabilistic binary linear classifier. Among all the models, DT has the lowest AUC_ROC for both datasets. The lower performance of DT can be explained due to its simple tree-based structure, which may not be able to utilize all the features of the input data due to tree-depth/leaf constraints. Interestingly, the DT model is preferred in pharmaceutical practice due to its simple structure. DT model has been popularly employed in understanding the mechanistic data on pharmacodynamic and pharmacokinetic properties [46]. However, their applicability in BBB permeability prediction was not effective due to their low specificity score [24]. Among the tree-based models, ET and RF achieve the highest AUC_ROC for the BBB and B3DB datasets, respectively. The higher performance of these tree-based models is due to their ensembled nature, allowing them to train individual DT on subsets of instances/features and combining their results with sophisticated strategies to

TABLE 4. Performance summary of ML models with and without SMOTE on the B3DB dataset. The best performing ML models and their corresponding fingerprints is highlighted in gray gradients.

		Without smote				With Smote			
ML model	**Fingerprint**	**Accuracy**	**Specificity**	**Sensitivity**	**AUC_ROC**	**Accuracy**	**Specificity**	**Sensitivity**	**AUC_ROC**
KNN	MACCS	0.8750	0.7710	0.9348	**0.9435**	0.8813	0.8629	0.8918	0.9421
	mol2vec	0.8651	0.7061	0.9566	0.9420	0.8763	0.8141	0.9120	0.9421
	B3DB features	0.8651	0.7106	0.9540	0.9418	0.8795	0.8359	0.9046	0.9421
	rdkit	0.8701	0.7629	0.9318	0.9397	0.8610	*0.9004*	0.8384	0.9367
DT	mol2vec	0.8435	0.7478	0.8985	**0.8888**	0.8293	0.7969	0.8479	0.8869
	MACCS	0.8293	0.7510	0.8743	0.8740	0.8236	0.8046	0.8345	0.8783
	B3DB features	0.8385	0.7682	0.8789	0.8741	0.8241	*0.8060*	0.8345	0.8753
	substructure	0.8302	0.7668	0.8666	0.8719	0.8245	0.8004	0.8384	0.8755
RF	mol2vec	0.8851	0.7801	0.9455	**0.9601**	0.8887	*0.8232*	0.9264	0.9594
	B3DB features	0.8746	0.7412	0.9514	0.9581	0.8813	0.8043	0.9255	0.9582
	MACCS	0.8754	0.7513	0.9467	0.9569	0.8874	0.8222	0.9249	0.9568
	Avalon	0.8604	0.6899	0.9584	0.9529	0.8743	0.7576	0.9415	0.9545
Adaboost	rdkit	0.8656	0.7797	0.9151	0.9341	0.8700	*0.8190*	0.8993	0.9361
	morgan	0.8650	0.7675	0.9211	**0.9369**	0.8603	0.7945	0.8981	0.9333
	klekota-roth	0.8647	0.7678	0.9205	0.9328	0.8649	0.8081	0.8975	0.9312
	Avalon	0.8641	0.7846	0.9098	0.9301	0.8615	0.8116	0.8902	0.9280
Extratrees	mol2vec	0.8824	0.7629	0.9512	**0.9554**	0.8846	0.7840	0.9425	0.9547
	B3DB features	0.8758	0.7510	0.9475	0.9540	0.8801	0.7804	0.9374	0.9542
	rdk-MACCS	0.8791	0.7731	0.9401	0.9539	0.8848	*0.8022*	0.9324	0.9536
	Avalon	0.8632	0.7022	0.9558	0.9535	0.8697	0.7324	0.9487	0.9535
GBC	mol2vec	0.8841	0.7875	0.9397	**0.9545**	0.8896	*0.8327*	0.9223	0.9538
	B3DB features	0.8804	0.7794	0.9385	0.9534	0.8854	0.8201	0.9229	0.9534
	MACCS	0.8825	0.7794	0.9419	0.9511	0.8881	0.8306	0.9211	0.9500
	Avalon	0.8688	0.7327	0.9471	0.9514	0.8741	0.7643	0.9372	0.9504
LightGBM	mol2vec	0.8676	0.8317	0.8882	0.9429	0.8676	0.7871	0.9138	**0.9434**
	Avalon	0.8645	*0.8380*	0.8797	0.9411	0.8665	0.8127	0.8975	0.9407
	FP4	0.8651	0.8359	0.8820	0.9403	0.8626	0.7955	0.9011	0.9401
	rdk-MACCS	0.8665	*0.8380*	0.8830	0.9386	0.8659	0.8120	0.8969	0.9387
XgBoost	mol2vec	0.8873	0.8008	0.9370	**0.9588**	0.8856	0.8236	0.9213	0.9582
	B3DB features	0.8793	0.7780	0.9377	0.9569	0.8820	0.8239	0.9155	0.9574
	Avalon	0.8788	0.7808	0.9352	0.9550	0.8813	0.8110	0.9217	0.9558
	rdk-MACCS	0.8810	0.7847	0.9364	0.9556	0.8902	*0.8401*	0.9191	0.9556

generate the final prediction. We also trained the well-known ML models (i.e., Adaboost, GBC, LightGBM, and XgBoost) that boost weak learners to maximize classification accuracy. For both the datasets, XgBoost follows the performance ensemble-based models and achieves the highest AUC_ROC among the models that employ boosting strategy. Furthermore, XgBoost acquires the highest specificity among all the models for the B3DB dataset. LightGBM attains the best specificity among all the ML models (i.e., 0.7882 for BBB and 0.838 for B3DB). The higher specificity of LightGBM is because it internally handles unbalanced datasets and provides regularization parameters for the estimators to prevent the model from favoring the majority class. The performance of LightGBM is in line with the results of Shaker et al. [47], which were calculated on a dataset similar to B3DB. Overall, by comparing the ML models across two datasets, we find that the performances of the top three models (i.e., RF, ET, XgBoost) are almost similar in terms of AUC_ROC irrespective of the dataset used.

There are several representations to describe drug molecules. However, no one descriptor can comprehensively capture all aspects of the molecule. Molecular descriptors are formulated for holistic representations, including molecular size, weight, and shape. Molecular fingerprints encode topological geometrical, thermodynamic, electronic, and constitutional information. To capture the impact of descriptors and fingerprints, each ML model is exhaustively trained with individual molecular representations to identify the best feature set for BBB predictability. To some extent, BBB permeability depends on the lipophilicity of the molecules for effective transport passive diffusion [7], [48], which in turn depends on the functional groups. Fingerprints and descriptors that are the best representative of this property are expected to show high performance. Information provided by the fingerprints and descriptors determines the performance of the models since the ML models cannot engineer new features out of the input. We analyze the fingerprints consistently in the top-4 ranked by AUC_ROC for the ML models in both datasets. We observed that the representation

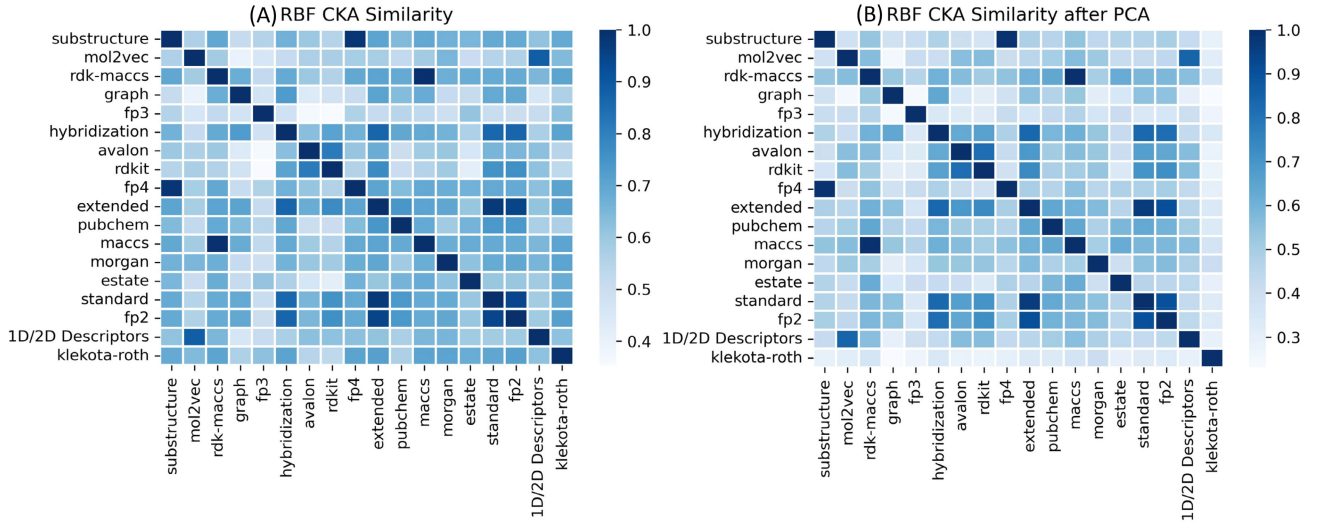

FIGURE 2. Radial basis function (RBF) CKA similarity matrix for descriptors and fingerprints before (A) and after PCA (B).

of molecular SMILES generated by the unsupervised ML (i.e., mol2vec) performed the best for 8 of 9 models for BBB and 7 of 8 models for the B3DB, suggesting that neural network-generated representations are more effective for training ML models as compared to handcrafted descriptors and other molecular fingerprints. 1D/2D PaDEL descriptors and B3DB features follow the performance of mol2vec for most models across both datasets. Experts in physical and organic chemistry have carefully crafted these descriptors to highlight the different molecular properties, allowing the ML models to make BBB predictions based on the properties. MACCS (CDK or RDK implementation) is also in the top-4 fingerprints for 7 of 8 models on the B3DB dataset. Another popular fingerprint is Avalon, which is present in 5 of 9 models for BBB and 6 of 8 models for B3DB. The presence of MACCS and Avalon in the top-4 fingerprint suggests that capturing essential substructures of the molecules or combining path-based and substructure-based features can help boost the performance of ML models. Interestingly, we observed that FP3 has a low variance. Thus, PCA significantly decreases the dimensions of FP3 (from 1024 to 20) for the BBB dataset. Due to a limited number of features, FP3 has the lowest performance for all models except DT. The performance of FP3 for DT is acceptable because DT can effectively use a small set of features.

Figure 2 shows a heat-map of CKA similarities between different fingerprints and features before and after applying PCA. It is evident from the heat-map that PCA conserves the information present in the features, and similarities between the fingerprints are retained. But, fingerprints with lower similarities have a lower CKA metric because of 5% decrease in variance. The following fingerprints have high similarity according to the CKA matrix: 1) substructure and FP4 have a higher magnitude of CKA because they search for similar SMART patterns in drug molecules. 2) Similarly, standard, fp2, and extended have high similarity because they compute similar substructures. 3) Relative to other fingerprints, mol2vec shows high similarity to 1D/2D descriptors, thus explaining the similar performance of the two features for both datasets. 4) rdk-MACCS and MACCS also have high similarities. This phenomenon is expected because the two fingerprints nearly search the same molecular substructure, slightly differing in their implementation (i.e., CDK and RDK). 5) Avalon and rdkit display high CKA suggesting that rdkit implementation is inspired by the Avalon toolkit. Hybridization shows proximity to both Avalon and rdkit, indicating that hybridization states are part of the Avalon and rdkit fingerprints.

Interestingly, we observed after PCA that klekota-roth shows lower similarity to other extracted fingerprints. FP3 shows dissimilarity to all fingerprints, specifically with Avalon and rdkit, because the former is a substructure-based fingerprint, and the latter are hybrid fingerprints. Furthermore, the substructures searched by FP3 may be very different from those computed by Avalon and rdkit. mol2vec shows high dissimilarity to the graph, indicating the ring structures and atomic connectivity may be lacking from mol2vec.

Key findings:

1) Tree-based ensemble models (i.e., ET and RF) achieve the highest AUC_ROC for BBB permeability classification on BBB and B3DB datasets.
2) LightGBM attains the highest specificity for both datasets among all ML models.
3) mol2vec is the best performing fingerprint for most ML models, followed by 1D/2D descriptors / B3DB features. The CKA similarity of the two features supports the finding.
4) Substructure-based fingerprints like MACCS perform well for most models on the both dataset. Additionally, standard, extended, and FP4 fingerprints performs

poorly across both datasets. These fingerprint show high similarity as evidenced by CKA value due to potential overlaps in the SMART patterns.

B. IMPACT OF SMOTE

One primary challenge for the ML models for BBB permeability is the class imbalance between BBB+ and BBB− instances, resulting in the inferior predictive capability of true negatives or low specificity. Albeit the considerably higher prediction accuracy than those reported in the literature, the imbalanced dataset results in a low recall ratio for the minority BBB− class (i.e., low specificity). In the case of BBB permeability, the minority class is crucial as it represents the molecular features restricted explicitly by the BBB phenotypes. While the application of grid search parameters and 10-fold cross-validation addressed this issue to a certain extent, the prediction capability of the minority class remains considerably low. In order to address this issue, SMOTE resampling technique is applied as suggested in the empirical study description. SMOTE ensures that number of BBB+ and BBB− instances are the same in the training set of each cross validation fold. SMOTE is an oversampling technique proposed by Chawla et al. [49] by constructing synthetic minority samples through the interpolation between minority data and its k-nearest neighbors. SMOTE was initially employed on the BBB dataset by Wang et al. (2018) [28] to improve the specificity score up to 88.6%. In our empirical study, the application of SMOTE enhanced the specificity for all models except LightGBM across both datasets. The performance gain with SMOTE is particularly evident in KNN, with more than a 52% increase in specificity in the BBB dataset. The imbalance ratio in BBB dataset (23.3% for BBB−) is higher than B3DB (36.5% for BBB−); as a result, the improvement in specificity is more evident in the BBB dataset. Among the tree-based models, RF observes the highest gain in specificity. In contrast, the ET experience the lowest increment in specificity.

Even though it is expected that AUC_ROC will improve after using SMOTE, we observe that it slightly decreases for all ML models, except LightGBM for the B3DB dataset. This phenomenon can be explained by the gain in specificity and a slight drop in sensitivity experienced by all models except LightGBM. On the other hand, every ML model experiences a slight improvement in AUC_ROC except LightGBM on the BBB dataset. By analyzing these differences across the dataset, we can deduce that performance gains in AUC_ROC after SMOTE are associated with the degree of imbalance in data, thus assisting the ML models on the BBB dataset more than the B3DB dataset. Interestingly, LightGBM suffers a loss of specificity after SMOTE application for both datasets, suggesting that oversampling techniques may not assist the models that internally handle class imbalance. Altogether, ET achieves the highest AUC_ROC of 94.66% after SMOTE application on the BBB dataset, whereas RF attains the highest AUC_ROC of 96.01% without SMOTE application on the B3DB dataset.

In addition to SMOTE, studies have incorporated other resampling methods like adaptive synthetic sampling (ADASYN), random under sampler (RUS) [50]. These studies found that SMOTE is more effective in dealing with class imbalance than other techniques employed in the literature. However, the application of oversampling needs to be done cautiously. Employing SMOTE to the entire dataset as suggested in the literature [43] leads to synthetic samples in the test set, which could result in a model overfitting and inaccurate evaluation of the model's predictive capability. We observed that incorrect application of SMOTE may provide high specificity and AUC_ROC (as noted by [43]) that may hide the model's performance on the actual drug molecules while highlighting it on synthetic samples. Furthermore, resampling prior to feature selection or dimensionality reduction could result in the selection or elimination of crucial dimensions, as seen in [50]. Our study has carefully analyzed the appropriate application of resampling techniques (i.e., SMOTE) to avoid incorrect model evaluation and overfitting while providing its impact on all the well-known ML models.

It is also crucial to evaluate whether the best fingerprints for ML models alter after applying SMOTE. The results in Supplementary Tables give the overall evaluation of models with each fingerprint after SMOTE application. While the best performing fingerprints after SMOTE have not changed for most models, their order of effectiveness changed to a certain extent. Tables 3 and 4 show the ranked fingerprints for ML models after the application of SMOTE. While SMOTE has been effectively used in literature and proved to be advantageous in overcoming the limitations set by imbalanced data, the choice to employ it needs to be assessed carefully. Specifically, drug design and development approaches must carefully analyze the intended drug's application. For instance, a drug designed for treating neurological conditions is required to cross the BBB to reach the site of pathological origin. In such cases, information on the chemical aspects of a molecule that restricts transport across BBB can be avoided during drug design for targeted delivery. As this information is mainly from the minority class data, applying SMOTE to the data or using models that address the data imbalance like LightGBM can provide better drug design outcomes for neurological conditions.

Key findings:

1) Application of SMOTE on the data improves specificity of models at a slight cost of sensitivity except LightGBM.
2) Extent of class imbalance impacts the performance gains observed with the application of SMOTE.
3) Oversampling techniques (i.e., SMOTE) must be applied only on the training set (when using train-test split or K-fold cross validation) to avoid the presence of synthetic samples in the test set.
4) The best performing fingerprint remain the same for most ML models before/after application of SMOTE with slight changes in their ranking.

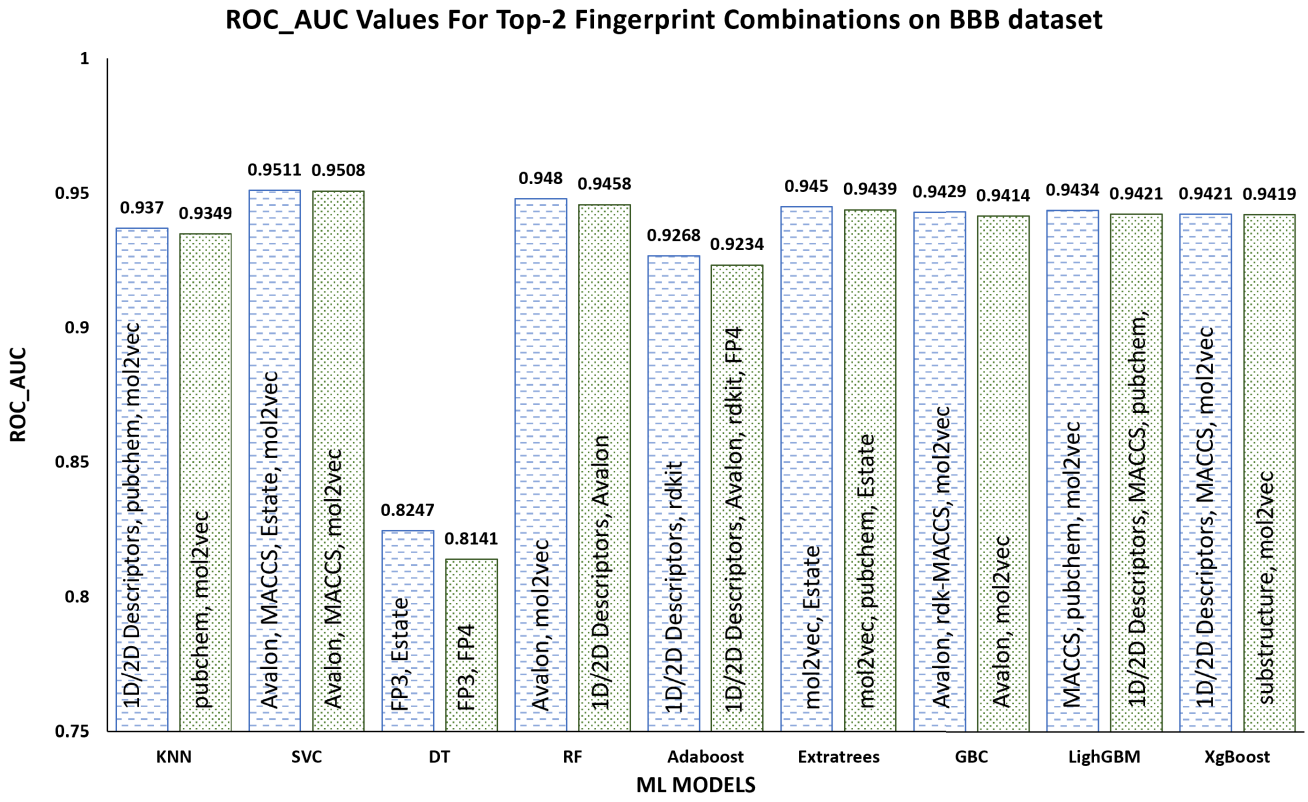

FIGURE 3. Performance summary of ML models with top fingerprint combinations using SMOTE on the BBB dataset.

C. IMPACT OF FINGERPRINT COMBINATION AND SMOTE

The existing datasets on BBB permeability of drugs contain molecules that cross BBB by several transport properties. The predictability of the methods mostly relies on the passive diffusion across BBB driven by the molecular properties. However, crucial molecules like glucose and insulin occur via highly selective transporters/receptors interaction like ATP-binding cassette, efflux transporters etc. Such mechanisms are poorly defined by the physicochemical properties of the compounds. Thus the use of physicochemical descriptors in the form of 1D/2D handcrafted vectors and unsupervised generation of molecular vectors by mol2vec showed better predictability in all the models employed in the current study. The similarly between these two features have already been established using CKA similarity search (Figure 1).

To overcome this limitation we propose the combined use of property-based features influencing passive diffusion and molecular fingerprints influencing receptor interactions like uptake, efflux and binding. Figures 3 and 4 summarize the best performing fingerprint combinations for each model on both datasets. Overall the highest performance was obtained using SVC model for BBB dataset with 95.1% predictability. A combination of Avalon, mol2vec and MACCS fingerprints further improved the specificity in SVC model indicating the integrated learning of the model to predict true negatives using more fingerprint information. While this result cannot be compared with B3DB dataset due to poor scalability, it is consistent with Yuan et al. (2018) [21]. The authors identified the use of 1D/2D descriptors or fingerprints individually for the SVM classification model resulted in 91.7% and 96.8% predictability, respectively. However, the combination of these two improved the predictability to 97.5%. While essentially, physicochemical features of the molecule have more contribution to the BBB permeability, fragment-based features add more value to the molecular properties responsible for permeability. As mentioned previously, MACCS is a substructure fingerprint that include predefined atom symbols, bonds, atom properties and environment. Specifically, MACCS provides information like the presence of nitrogen heterocycle which is a major contributing factor to BBB permeability [19]. The presence of this nitrogen heterocycle can further influence physicochemical properties like lipophilicity, polarity and hydrogen bonding capacity. This could be attributed to the higher performance of the combination of MACCS and mol2vec in SVC model. For both datasets, most models experienced a mild increase in predictability (AUC_ROC) upon using the combination of top fingerprints but a significant reduction in the specificity. This can be due to the suppression of information essential for

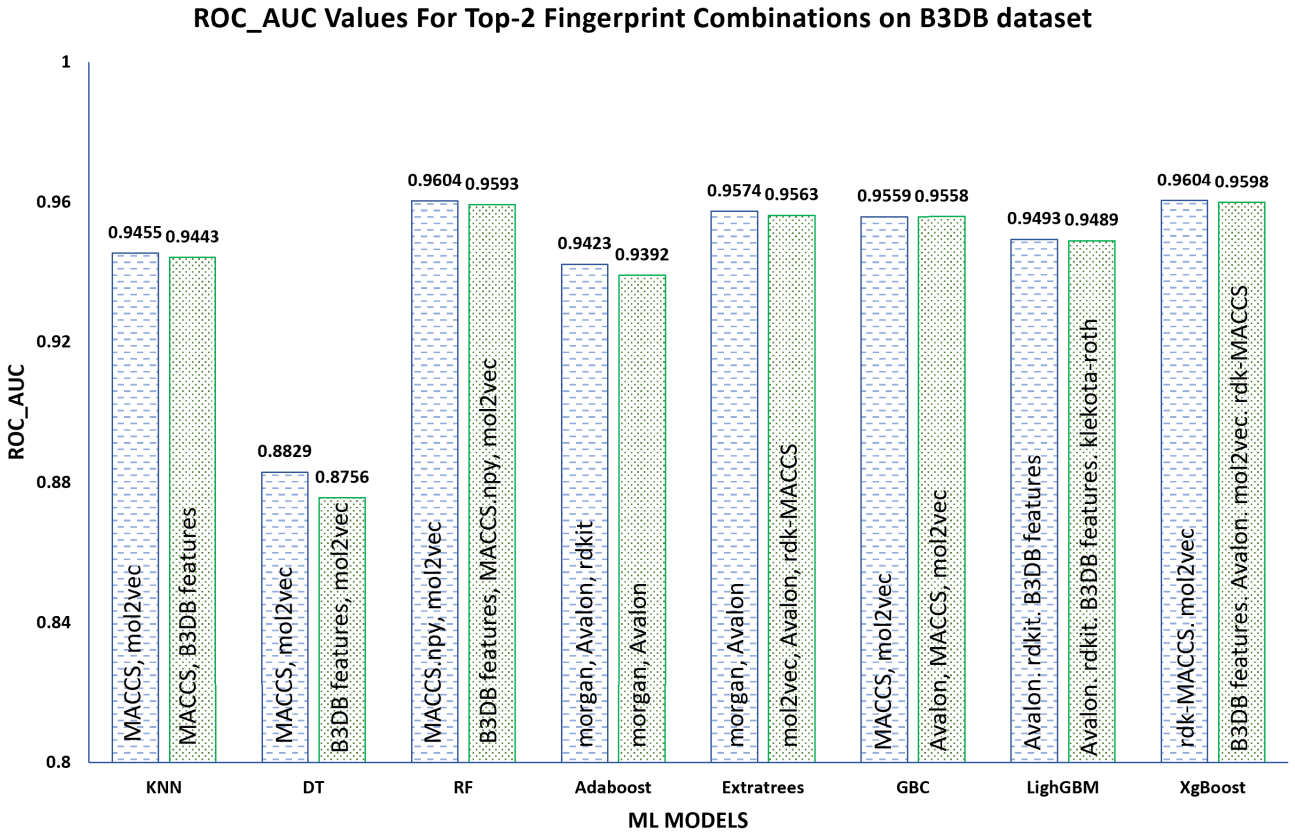

FIGURE 4. Performance summary of ML models with top fingerprint combinations using SMOTE on the B3DB dataset.

BBB− predictability when PCA is applied on a combination of fingerprints (e.g., mol2vec and MACCS). Interestingly, AUC_ROC of ET models decreased slightly with fingerprint combination relative to individual fingerprints. The performance of XgBoost remained the same after fingerprint combination for BBB and B3DB datasets, respectively. These observation suggest that every model may not benefit from training with combination of fingerprints. Our findings imply that SVC and RF are effective in predicting the passive diffusibility of drugs attributed to the physicochemical features rather than the fragment-based information contributing to the other transport properties.

Key findings:

1) For B3DB dataset, MACCS and mol2vec combination works well for most model, whereas for the BBB data, mol2vec or 1D/2D features is present in the best performing fingerprint combination.
2) For both datasets, the specificity of the models decrease with fingerprint combination. Nevertheless, KNN achieves the highest specificity among the trained models.
3) For BBB and B3DB dataset, some models performed similarly with and without fingerprint combinations. ET experienced a slight dip in accuracy with fingerprint combinations.

D. VALIDATION OF MODELS FOR DRUG RE-PURPOSING

To validate the performance of the top ML models as an application for drug repurposing, a set of 30 drugs is selected from the literature (Table 5). Some of the drugs are chosen based on the results of the network medicine approach using ML for neurological disease drug repositioning reported by Dias et al. [51]. The study maps the genes responsible for neurological diseases and the gene targets of drug candidates from the last 50 years. While genetic targets of the drug are mapped using the ML model in the mentioned study, predicting the BBB crossing ability of these drugs could further predict the actual effectiveness of these drugs in neurological conditions. This study is chosen as a representative of the huge applications and opportunities provided by machine learning tools in the drug development process. We eliminate the redundant drugs from this list that are present in the B3DB dataset. Further, hypertension-associated neurological effects have been long identified to be linked with dementia, potentiating Alzheimer's pathology, pre-eclampsia etc. [52], [53]. Hypertensive drugs like beta-blockers, calcium channel blockers, and renin-angiotensin system (RAS) drugs have been studied to lower dementia risk [54]. Though RAS is primarily involved in maintaining homeostasis, the presence of RAS in the central nervous system indicates its importance in cognition and neuronal functions. Thus, RAS drugs that

TABLE 5. BBB permeability prediction with models trained using B3DB data for recently proposed drugs targeting neurological diseases and hypertension. Here, 1 and 0 represent the BBB+ and BBB− classes, respectively.

Pubchem CID	Drug	Mechanism of Action/ Intended Application	KNN	RF	ET
4485	Nifedipine	Calcium Channel Blocker/ hypertention	1	1	0
119607	Valdecoxib	NSAID/ COX-2 Inhibitor/ osteoarthirtis	0	0	1
151166	Lumiracoxib	NSAID/ COX-2 inhibitor/ osteoarthiritis	0	0	1
6857724	Platensimycin	Antibiotic/ 3-oxoacyl-[acyl-carrier-protein] synthase 2 targets	1	0	1
9568614	Esomeprazole	Proton Pump Inhibitor/ treatment of gastric reflux	0	0	1
10345214	Ronacaleret	Calcium sensing receptor antagonist	0	0	1
11338033	AT7519	CDK9 inhibitor – supresses MCL1 expression/ anti neoplastic activity	0	0	1
12004316	PSI-694	P-Selectin (cell adhesion molecule) inhibitor	1	0	1
3760	Isoconazole	Antifungal	1	1	1
39147	Nadolol	Beta blockers/ hypertention	1	1	1
179344	Eslicarbazepine acetate	Inhibition of voltage-gated sodium channels/ anti convusing drug	1	1	1
448043	H-1152	Rho inhibitor and cAMP dependent protein kinase inhibitor/glaucoma	1	1	1
3064778	Hydroxyfasudil	Rho inhibitor and cAMP dependent protein kinase inhibitor/glaucoma	1	1	1
5311505	Onapristone	Inhibits Fibronectin production/ cancer treatment	1	1	1
6419718	Omigapil	Targets GAPDH-SIAH1 mechanism/muscular dystropy	1	1	1
9796590	Orteronel	CYP17A1 inhibitor/ cancer	1	1	1
71301276	OR-12741	Adrenoreceptor ADRAC2 antagonist/ Alzheimers	1	1	1
78357816	Seviteronel	CYP17A1 inhibitor/ cancer	1	1	1
4474	Nicardipine	Calcium Channel Blocker/ hypertention	0	0	0
56330	Cilazapril	RAS inhibitor/ ACE inhibitor/ hypertention	0	0	0
60846	Valsartan	RAS inhibitor/ Angiotensin II receptor blockers/ hypertension	0	0	0
92400	Zofenopril	RAS inhibitor/ Angiotensin II receptor blockers/ hypertension	0	0	0
5311447	Spirapril	RAS inhibitor/ ACE inhibitor/ hypertension	0	0	0
5362123	Benazepril hydrochloride	RAS inhibitor/ ACE inhibitor/ hypertension	0	0	0
5464343	Imidapril	RAS inhibitor/ ACE inhibitor/ hypertension	0	0	0
5493444	Aliskiren	RAS inhibitor/ renin inhibitor/ hypertension	0	0	0
9821849	Abiraterone acetate	CYP17A1 inhibitor/ cancer	0	0	0
11984597	Olmesartan medoxomil	RAS inhibitor/ Angiotensin II receptor blockers/ hypertension	0	0	0
25210270	Edarbyclor	RAS inhibitor/ Angiotensin II receptor blockers/ hypertension	0	0	0
70675710	Prestalia	RAS inhibitor/ ACE inhibitor/ hypertension	0	0	0
135000000	Azilsartan	RAS inhibitor/ Angiotensin II receptor blockers/ hypertension	0	0	0

could cross BBB can be expected to improve Alzheimer's, Parkinson's, and Huntington's diseases that are characterised by a cognitive decline [51], [55]. Hence, most of the hypertensive drugs with more focus on RAS-acting drugs (Table 5) are also studied for ML model validation as an effort toward their repurposing for neurological diseases.

To accurately determine the BBB permeability of the above-mentioned drugs, we train the best performing ensemble tree-based models (i.e, RF and ET) on the B3DB dataset (with SMOTE enabled) using the mol2vec fingerprint. We use mol2vec because it is one of the best-performing fingerprints for these models (Table 4). Additionally, We also infer using the KNN model due to its higher specificity and lower false positive rate. By considering the predictions of three different ML models, we utilized majority voting to further minimize the chance of incorrectly predicting BBB permeability.

Table 5 presents the list of drugs for which the selected models had 100% agreement in predicting BBB+ and BBB−. It can be seen that CYP17A1 gene inhibitors like seviteronel and orteronel commonly intended for cancer treatment indicated a majority value for BBB permeability. A recent study on the Chinese Han population indicated the involvement of the CYP17A1 rs743572 allele in the late onset of Alzheimer's disease [56]. The BBB permeability of these drugs can indicate a possible new application of these drugs for preventing Alzheimer's disease. As mentioned previously, the presence of heterocycle is one of the major contributing factors to BBB permeability. Thus, the aromatic heteropolycyclic compounds, seviteronel and orteronel are predicted to pass BBB by all the models. In contrast, a sterol-based CYP17A1 inhibitor (Abiraterone) is predicted as BBB− indicating the high specificity of the developed model. Similarly, anti-hypertensive drugs are gaining popularity as repurposing targets for neurological diseases [57], [58], [59]. However, the majority of the anti-hypertensive drugs are predicted to possess poor BBB permeability irrespective of their mechanism of action. As brain RAS is involved in several cognitive functions, BBB is mostly equipped to selectively transport the molecular components of this system to maintain the functions. Since they are crucial compounds of neuronal functions, their transport is mediated by highly selective transporters like p-glycoprotein and efflux transporters [60]. However, beta-blocker anti-hypertensive

like nadolol is predicted as BBB+ by all the models based on their physicochemical properties (Table 5).

Table 5 also highlights the list of drugs for which the models disagreed in terms of BBB permeability. We can observe that RF and KNN have more agreement in this class of drugs relative to ET. This is because ET has the highest sensitivity and lowest specificity among the models used for inference, resulting in more BBB+ inference than BBB−. This can be understood from the case of non-steroidal anti-inflammatory drugs, where the ET model predicted BBB+ permeability while KNN and RF models predicted BBB−. As KNN and RF models have relatively higher specificity in comparison with ET (Tables 3&4), the prediction of false positives is lower, implying the non-crossing property of these drugs. Several studies have explored the potential application of NSAIDs in the prevention and treatment of Alzheimer's and Parkinson's and found them to be ineffective [61]. This can be attributed to their poor BBB permeability as identified in this study. Another example indicating the high specificity of KNN and RF models is the prediction of the transport of anti-hypertensive drug Nifidepine (calcium channel blocker). This is in contrast to similar calcium channel blocker nicardipine, that is predicted as BBB−. This can be attributed to the relatively smaller size of nifidepine in comparison with nicardipine. Some of the primary criteria for BBB passage is small molecules and lipophilic nature [7].

IV. CONCLUSION

In this paper, we conduct a comprehensive empirical study to evaluate the performance of ML models with different molecular fingerprints and handcrafted descriptors, to establish the best performing ML model and corresponding fingerprints for BBB permeability prediction. We ensure to eliminate dataset bias by conducting the study on two different datasets and analyzing the performance with several fingerprint types. A correlation among best-performing fingerprints has been made using the CKA similarity measure. We further evaluate the performance impact of the data balancing technique (i.e., SMOTE) on all models for both datasets. The developed model indicates the correct usage of such resampling techniques depending on the type of model and the intended application of the model. Additionally, we examine whether the combination of different fingerprints improves the performance of ML models. Finally, we utilize the models trained in our study to infer BBB permeability on drugs proposed for repurposing/repositioning.

AI/ML-assisted drug development for neurological applications has seen unprecedented success in recent years. It is to be noted that the use of these models in drug discovery/repurposing needs to be carefully applied considering the limitations and bias in the datasets. Thus, a proper amalgamation of the research knowledge, BBB physiology, and existing permeability models can help in developing robust models with better accuracy, thereby making drug development a rapid process. Hence, we envision the exploitation of computational AI by using neural networks and deep learning models as an application-driven screening framework for neuro-oncology, CNS infections, and neurodegenerative disorders such as Alzheimer, Parkinsonâ and multiple sclerosis. With increasing scope for nanotechnology-based targeted delivery across BBB, we further visualize the huge prospects of AI in designing engineered nanomaterials (ENMs) for CNS theranostics development.

ACKNOWLEDGMENT

The findings herein reflect the work are solely the responsibility of the authors. The Open Access funding is provided by the Qatar National Library. *(Mohammed Yusuf Ansari and Vaisali Chandrasekar contributed equally to this work.)*

REFERENCES

[1] S. Vatansever, A. Schlessinger, D. Wacker, H. Ü. Kaniskan, J. Jin, M. Zhou, and B. Zhang, "Artificial intelligence and machine learning-aided drug discovery in central nervous system diseases: State-of-the-arts and future directions," *Medicinal Res. Rev.*, vol. 41, no. 3, pp. 1427–1473, May 2021.

[2] S. Doniger, T. Hofmann, and J. Yeh, "Predicting CNS permeability of drug molecules: Comparison of neural network and support vector machine algorithms," *J. Comput. Biol.*, vol. 9, no. 6, pp. 849–864, Dec. 2002.

[3] G. C. Terstappen, A. H. Meyer, R. D. Bell, and W. Zhang, "Strategies for delivering therapeutics across the blood–brain barrier," *Nature Rev. Drug Discovery*, vol. 20, no. 5, pp. 362–383, May 2021.

[4] R. Kumar, A. Sharma, A. Alexiou, A. L. Bilgrami, M. A. Kamal, and G. M. Ashraf, "DeePred-BBB: A blood brain barrier permeability prediction model with improved accuracy," *Frontiers Neurosci.*, vol. 16, May 2022, Art. no. 858126.

[5] S. C. Massey, J. C. Urcuyo, B. M. Marin, J. N. Sarkaria, and K. R. Swanson, "Quantifying glioblastoma drug response dynamics incorporating treatment sensitivity and blood brain barrier penetrance from experimental data," *Frontiers Physiol.*, vol. 11, p. 830, Aug. 2020.

[6] Y. Mi, Y. Mao, H. Cheng, G. Ke, M. Liu, C. Fang, and Q. Wang, "Studies of blood–brain barrier permeability of gastrodigenin in vitro and in vivo," *Fitoterapia*, vol. 140, Jan. 2020, Art. no. 104447. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0367326X19319987

[7] A. V. Singh, V. Chandrasekar, P. Janapareddy, D. E. Mathews, P. Laux, A. Luch, Y. Yang, B. Garcia-Canibano, S. Balakrishnan, J. Abinahed, A. Al Ansari, and S. P. Dakua, "Emerging application of nanorobotics and artificial intelligence to cross the BBB: Advances in design, controlled maneuvering, and targeting of the barriers," *ACS Chem. Neurosci.*, vol. 12, no. 11, pp. 1835–1853, Jun. 2021.

[8] A. V. Singh, P. Laux, A. Luch, S. Balkrishnan, and S. P. Dakua, "Bottom-UP assembly of nanorobots: Extending synthetic biology to complex material design," *Frontiers Nanosci. Nanotechnol.*, vol. 5, no. 1, pp. 1–2, 2019.

[9] A. V. Singh, M. H. D. Ansari, D. Rosenkranz, R. S. Maharjan, F. L. Kriegel, K. Gandhi, A. Kanase, R. Singh, P. Laux, and A. Luch, "Artificial intelligence and machine learning in computational nanotoxicology: Unlocking and empowering nanomedicine," *Adv. Healthcare Mater.*, vol. 9, no. 17, Sep. 2020, Art. no. 1901862.

[10] A. V. Singh, D. Rosenkranz, M. H. D. Ansari, R. Singh, A. Kanase, S. P. Singh, B. Johnston, J. Tentschert, P. Laux, and A. Luch, "Artificial intelligence and machine learning empower advanced biomedical material design to toxicity prediction," *Adv. Intell. Syst.*, vol. 2, no. 12, Dec. 2020, Art. no. 2000084.

[11] S. Albaradei, M. Thafar, A. Alsaedi, C. Van Neste, T. Gojobori, M. Essack, and X. Gao, "Machine learning and deep learning methods that use omics data for metastasis prediction," *Comput. Structural Biotechnol. J.*, vol. 19, pp. 5008–5018, Jan. 2021.

[12] V. Chandrasekar, A. V. Singh, R. S. Maharjan, S. P. Dakua, S. Balakrishnan, S. Dash, P. Laux, A. Luch, S. Singh, and M. Pradhan, "Perspectives on the technological aspects and biomedical applications of virus-like particles/nanoparticles in reproductive biology: Insights on the medicinal and toxicological outlook," *Adv. NanoBiomed Res.*, vol. 2, no. 8, Aug. 2022, Art. no. 2200010.

[13] S. Kamboj, A. Rajput, A. Rastogi, A. Thakur, and M. Kumar, "Targeting non-structural proteins of hepatitis C virus for predicting repurposed drugs using QSAR and machine learning approaches," *Comput. Structural Biotechnol. J.*, vol. 20, pp. 3422–3438, Jan. 2022.

[14] M. Y. Ansari, A. Abdalla, M. Y. Ansari, M. I. Ansari, B. Malluhi, S. Mohanty, S. Mishra, S. S. Singh, J. Abinahed, A. Al-Ansari, S. Balakrishnan, and S. P. Dakua, "Practical utility of liver segmentation methods in clinical surgeries and interventions," *BMC Med. Imag.*, vol. 22, no. 1, pp. 1–17, May 2022.

[15] Y. Akhtar, S. P. Dakua, A. Abdalla, O. M. Aboumarzouk, M. Y. Ansari, J. Abinahed, M. S. M. Elakkad, and A. Al-Ansari, "Risk assessment of computer-aided diagnostic software for hepatic resection," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 6, pp. 667–677, Jul. 2022.

[16] M. Y. Ansari, Y. Yang, S. Balakrishnan, J. Abinahed, A. Al-Ansari, M. Warfa, O. Almokdad, A. Barah, A. Omer, A. V. Singh, P. K. Meher, J. Bhadra, O. Halabi, M. F. Azampour, N. Navab, T. Wendler, and S. P. Dakua, "A lightweight neural network with multiscale feature enhancement for liver CT segmentation," *Sci. Rep.*, vol. 12, no. 1, p. 14153, Aug. 2022, doi: 10.1038/s41598-022-16828-6.

[17] A. V. Singh, R.-S. Maharjan, A. Kanase, K. Siewert, D. Rosenkranz, R. Singh, P. Laux, and A. Luch, "Machine-learning-based approach to decode the influence of nanomaterial properties on their interaction with cells," *ACS Appl. Mater. Inter.*, vol. 13, no. 1, pp. 1943–1955, Jan. 2021.

[18] S. Cherian Parakkal, R. Datta, and D. Das, "DeepBBBP: High accuracy blood-brain-barrier permeability prediction with a mixed deep learning model," *Mol. Informat.*, vol. 41, no. 10, Oct. 2022, Art. no. 2100315.

[19] T.-H. Yu, B.-H. Su, L. C. Battalora, S. Liu, and Y. J. Tseng, "Ensemble modeling with machine learning and deep learning to provide interpretable generalized rules for classifying CNS drugs with high prediction power," *Briefings Bioinf.*, vol. 23, no. 1, Jan. 2022, bbab377.

[20] S. Alsenan, I. Al-Turaiki, and A. Hafez, "A recurrent neural network model to predict blood–brain barrier permeability," *Comput. Biol. Chem.*, vol. 89, Dec. 2020, Art. no. 107377.

[21] Y. Yuan, F. Zheng, and C.-G. Zhan, "Improved prediction of blood–brain barrier permeability through machine learning with combined use of molecular property-based descriptors and fingerprints," *AAPS J.*, vol. 20, no. 3, pp. 1–10, May 2018.

[22] P. Carracedo-Reboredo, J. Liñares-Blanco, N. Rodríguez-Fernández, F. Cedrón, F. J. Novoa, A. Carballal, V. Maojo, A. Pazos, and C. Fernandez-Lozano, "A review on machine learning approaches and trends in drug discovery," *Comput. Structural Biotechnol. J.*, vol. 19, pp. 4538–4558, 2021.

[23] A. I. Khan, Q. Lu, D. Du, Y. Lin, and P. Dutta, "Quantification of kinetic rate constants for transcytosis of polymeric nanoparticle through blood-brain barrier," *Biochimica Biophysica Acta (BBA) Gen. Subjects*, vol. 1862, no. 12, pp. 2779–2787, Dec. 2018.

[24] C. Suenderhauf, F. Hammann, and J. Huwyler, "Computational prediction of blood-brain barrier permeability using decision tree induction," *Molecules*, vol. 17, no. 9, pp. 10429–10445, Aug. 2012.

[25] L. Liu, L. Zhang, H. Feng, S. Li, M. Liu, J. Zhao, and H. Liu, "Prediction of the blood–brain barrier (bbb) permeability of chemicals based on machine-learning and ensemble methods," *Chem. Res. Toxicology*, vol. 34, no. 6, pp. 1456–1467, 2021.

[26] D. Saxena, A. Sharma, M. H. Siddiqui, and R. Kumar, "Blood brain barrier permeability prediction using machine learning techniques: An update," *Current Pharmaceutical Biotechnol.*, vol. 20, no. 14, pp. 1163–1171, Nov. 2019.

[27] L. Jiang, J. Chen, Y. He, Y. Zhang, and G. Li, "A method to predict different mechanisms for blood–brain barrier permeability of CNS activity compounds in Chinese herbs using support vector machine," *J. Bioinf. Comput. Biol.*, vol. 14, no. 1, Feb. 2016, Art. no. 1650005.

[28] W. Zhuang, H. Yang, Z. Wu, T. Wang, W. Li, Y. Tang, and G. Liu, "In silico prediction of blood–brain barrier permeability of compounds by machine learning and resampling methods," *ChemMedChem*, vol. 13, no. 20, pp. 2189–2201, 2018.

[29] P. Rajkumar, S. K. Ghosh, and P. Dasgupta, "Concurrent usage control implementation verification using the spin model checker," in *Proc. Int. Conf. Netw. Secur. Appl.* Cham, Switzerland: Springer, 2010, pp. 214–223.

[30] F. Meng, Y. Xi, J. Huang, and P. W. Ayers, "A curated diverse molecular database of blood–brain barrier permeability with chemical descriptors," *Sci. Data*, vol. 8, no. 1, pp. 1–11, Oct. 2021. [Online]. Available: https://www.nature.com/articles/s41597-021-01069-5

[31] A. Tiwari Pandey, I. Pandey, A. Kanase, A. Verma, B. Garcia-Canibano, S. Dakua, S. Balakrishnan, and M. Singh, "Validating anti-infective activity of pleurotus opuntiae via standardization of its bioactive mycoconstituents through multimodal biochemical approach," *Coatings*, vol. 11, no. 4, p. 484, Apr. 2021. [Online]. Available: https://www.mdpi.com/2079-6412/11/4/484

[32] E. Pimentel, K. Sivalingam, M. Doke, and T. Samikkannu, "Effects of drugs of abuse on the blood-brain barrier: A brief overview," *Frontiers Neurosci.*, vol. 14, p. 513, May 2020.

[33] I. F. Martins, A. L. Teixeira, L. Pinheiro, and A. O. Falcao, "A Bayesian approach to in silico blood-brain barrier penetration modeling," *J. Chem. Inf. Model.*, vol. 52, no. 6, pp. 1686–1697, Jun. 2012.

[34] J. Shen, F. Cheng, Y. Xu, W. Li, and Y. Tang, "Estimation of ADME properties with substructure pattern recognition," *J. Chem. Inf. Model.*, vol. 50, no. 6, pp. 1034–1041, Jun. 2010.

[35] M. Muehlbacher, G. M. Spitzer, K. R. Liedl, and J. Kornhuber, "Qualitative prediction of blood–brain barrier permeability on a large and refined dataset," *J. Comput.-Aided Mol. Design*, vol. 25, no. 12, pp. 1095–1106, Dec. 2011.

[36] W. Wang, M. T. Kim, A. Sedykh, and H. Zhu, "Developing enhanced blood–brain barrier permeability models: Integrating external bio-assay data in QSAR modeling," *Pharmaceutical Res.*, vol. 32, no. 9, pp. 3055–3065, Sep. 2015.

[37] D. C. Elton, Z. Boukouvalas, M. S. Butrico, M. D. Fuge, and P. W. Chung, "Applying machine learning techniques to predict the properties of energetic materials," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, Jun. 2018.

[38] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open babel: An open chemical toolbox," *J. Cheminformatics*, vol. 3, no. 1, pp. 1–14, Dec. 2011.

[39] E. L. Willighagen, J. W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliazkova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C. T. Evelo, R. Guha, and C. Steinbeck, "Erratum to: The chemistry development kit (CDK) v2.0: Atom typing, depiction, molecular formulas, and substructure searching," *J. Cheminformatics*, vol. 9, no. 1, pp. 1–19, Dec. 2017.

[40] S. Kornblith, M. Norouzi, H. Lee, and G. Hinton, "Similarity of neural network representations revisited," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 3519–3529.

[41] R. Mohammed, J. Rawashdeh, and M. Abdullah, "Machine learning with oversampling and undersampling techniques: Overview study and experimental results," in *Proc. 11th Int. Conf. Inf. Commun. Syst. (ICICS)*, Apr. 2020, pp. 243–248.

[42] J. Luengo, A. Fernández, S. García, and F. Herrera, "Addressing data complexity for imbalanced data sets: Analysis of SMOTE-based oversampling and evolutionary undersampling," *Soft Comput.*, vol. 15, no. 10, pp. 1909–1936, 2011.

[43] S. Alsenan, I. Al-Turaiki, and A. Hafez, "A deep learning approach to predict blood-brain barrier permeability," *PeerJ Comput. Sci.*, vol. 7, p. e515, Jun. 2021.

[44] L. Zhang, H. Zhu, T. I. Oprea, A. Golbraikh, and A. Tropsha, "QSAR modeling of the blood–brain barrier permeability for diverse organic compounds," *Pharmaceutical Res.*, vol. 25, no. 8, pp. 1902–1914, Aug. 2008.

[45] D. Roy, V. K. Hinge, and A. Kovalenko, "To pass or not to pass: Predicting the blood–brain barrier permeability with the 3D-RISM-KH molecular solvation theory," *ACS Omega*, vol. 4, no. 16, pp. 16774–16780, 2019.

[46] E. P. Chen, R. W. Bondi, and P. J. Michalski, "Model-based target pharmacology assessment (mTPA): An approach using PBPK/PD modeling and machine learning to design medicinal chemistry and DMPK strategies in early drug discovery," *J. Medicinal Chem.*, vol. 64, no. 6, pp. 3185–3196, Mar. 2021.

[47] B. Shaker, M.-S. Yu, J. S. Song, S. Ahn, J. Y. Ryu, K.-S. Oh, and D. Na, "LightBBB: Computational prediction model of blood–brain-barrier penetration based on LightGBM," *Bioinformatics*, vol. 37, no. 8, pp. 1135–1139, May 2021.

[48] R. Yang, T. Wei, H. Goldberg, W. Wang, K. Cullion, and D. S. Kohane, "Getting drugs across biological barriers," *Adv. Mater.*, vol. 29, no. 37, Oct. 2017, Art. no. 1606596.

[49] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jul. 2018.

[50] Z. Shi, Y. Chu, Y. Zhang, Y. Wang, and D.-Q. Wei, "Prediction of blood-brain barrier permeability of compounds by fusing resampling strategies and eXtreme gradient boosting," *IEEE Access*, vol. 9, pp. 9557–9566, 2021.
[51] T. Lüscher Dias, V. Schuch, P. C. B. Beltrão-Braga, D. Martins-de-Souza, H. P. Brentani, G. R. Franco, and H. I. Nakaya, "Drug repositioning for psychiatric and neurological disorders through a network medicine approach," *Transl. Psychiatry*, vol. 10, no. 1, pp. 1–10, May 2020.
[52] D. M. Kelly and P. M. Rothwell, "Blood pressure and the brain: The neurology of hypertension," *Practical Neurol.*, vol. 20, no. 2, pp. 100–108, Apr. 2020.
[53] A. Fournier, R. Oprisiu-Fournier, J.-M. Serot, O. Godefroy, J.-M. Achard, S. Faure, H. Mazouz, M. Temmar, A. Albu, R. Bordet, O. Hanon, F. Gueyffier, J. Wang, S. Black, and N. Sato, "Prevention of dementia by antihypertensive drugs: How AT1-receptor-blockers and dihydropyridines better prevent dementia in hypertensive patients than thiazides and ACE-inhibitors," *Exp. Rev. Neurotherapeutics*, vol. 9, no. 9, pp. 1413–1431, Sep. 2009.
[54] J. K. Ho, F. Moriarty, J. J. Manly, E. B. Larson, D. A. Evans, K. B. Rajan, E. M. Hudak, L. Hassan, E. Liu, N. Sato, N. Hasebe, D. Laurin, P.-H. Carmichael, and D. A. Nation, "Blood-brain barrier crossing renin-angiotensin drugs and cognition in the elderly: A meta-analysis," *Hypertension*, vol. 78, no. 3, pp. 629–643, Sep. 2021.
[55] J. K. Ho and D. A. Nation, "Memory is preserved in older adults taking AT1 receptor blockers," *Alzheimer's Res. Therapy*, vol. 9, no. 1, pp. 1–14, Dec. 2017.
[56] L. Xie, H. Yan, L. Shi, Y. Kong, M. Huang, J. Li, J. Li, J. Zheng, Y. Zhao, and S. Zhao, "Association between *CYP17A1* rs3824755 and rs743572 gene polymorphisms and Alzheimer's disease in the Chinese Han population," *Neurosci. Lett.*, vol. 618, pp. 77–82, Apr. 2016.
[57] F. Gouveia, A. Camins, M. Ettcheto, J. Bicker, A. Falcão, M. T. Cruz, and A. Fortuna, "Targeting brain renin-angiotensin system for the prevention and treatment of Alzheimer's disease: Past, present and future," *Ageing Res. Rev.*, vol. 77, May 2022, Art. no. 101612.
[58] S. Hussain, A. Singh, S. O. Rahman, A. Habib, and A. K. Najmi, "Calcium channel blocker use reduces incident dementia risk in elderly hypertensive patients: A meta-analysis of prospective studies," *Neurosci. Lett.*, vol. 671, pp. 120–127, Apr. 2018.
[59] P. J. Tully, O. Hanon, S. Cosh, and C. Tzourio, "Diuretic antihypertensive drugs and incident dementia risk: A systematic review, meta-analysis and meta-regression of prospective studies," *J. Hypertension*, vol. 34, no. 6, pp. 1027–1035, 2016.
[60] M. Mogi and M. Horiuchi, "Remote control of brain angiotensin II levels by angiotensin receptor blockers," *Hypertension Res.*, vol. 33, no. 2, pp. 116–117, Feb. 2010.
[61] M. Terzi, G. Altun, S. Şen, A. Kocaman, A. A. Kaplan, K. K. Yurt, and S. Kaplan, "The use of non-steroidal anti-inflammatory drugs in neurological diseases," *J. Chem. Neuroanatomy*, vol. 87, pp. 12–24, Jan. 2018.

MOHAMMED YUSUF ANSARI received the B.Sc. degree in computer science from Carnegie Mellon University and the M.Sc. degree in data science from Hamad Bin Khalifa University. He is currently pursuing the Ph.D. degree in computer engineering with Texas A&M University. He also works as a Research Associate at Hamad Medical Corporation.

VAISALI CHANDRASEKAR received the Ph.D. degree in industrial biotechnology from the National Institute of Technology Karnataka, India, in 2018. She is currently a second-year Post-doctoral Fellow at Hamad Medical Corporation, Qatar. Her research interests include in vitro disease modeling of neurological diseases for nano-based drug and food delivery systems, enzymology, and novel bioactive synthesis.

AJAY VIKRAM SINGH received the M.Sc. degree in biotechnology from Pune University, India, in 2005, and the Ph.D. degree in medical nanotechnology from the European School of Molecular Medicine (SEMM), University of Milan, in 2012. Since 2018, he has been a Senior Scientist at the Department of Chemical and Product Safety, German Federal Institute for Risk Assessment (BfR), Berlin. His research interests include chemical and nanotoxicology, micronanorobotics, neurobiology, and antibacterial surfaces and understanding the biophysicochemical interactions at nano biointerface.

SARADA PRASAD DAKUA received the M.B.A. degree from the University of Leicester, U.K., and the Ph.D. degree in medical image processing from the Indian Institute of Technology Guwahati, India. He is currently working as a Senior Research Scientist with the Department of Surgery, Hamad Medical Corporation, Qatar. He has more than 15 years of research experience in computer vision. He is a Certified PMP.