IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 20 March 2023, accepted 25 April 2023, date of publication 4 May 2023, date of current version 5 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3272987

RESEARCH ARTICLE

Investigating the Use of Machine Learning Models to Understand the Drugs Permeability Across Placenta

VAISALI CHANDRASEKAR[1], MOHAMMED YUSUF ANSARI[2], AJAY VIKRAM SINGH[3], SHAHAB UDDIN[4], KIRTHI S. PRABHU[4], SAGNIKA DASH[5], SOUHAILA AL KHODOR[6], ANNALISA TERRANEGRA[6], MATTEO AVELLA[6], AND SARADA PRASAD DAKUA[1]

[1]Department of Surgery, Hamad Medical Corporation, Doha, Qatar
[2]Electrcial and Computer Engineering, Texas A&M University, College Station, TX 77843, USA
[3]German Federal Institute for Risk Assessment (BfR), 10589 Berlin, Germany
[4]Translational Research Institute, Academic Health System, Hamad Medical Corporation, Doha, Qatar
[5]Department of Obstetrics and Gynecology, Apollo Clinic, Doha, Qatar
[6]Maternal and Child Health Department, Research Branch, Sidra Medicine, Ar-Rayyan, Doha, Qatar

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

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

ABSTRACT Owing to limited drug testing possibilities in pregnant population, the development of computational algorithms is crucial to predict the fate of drugs in the placental barrier; it could serve as an alternative to animal testing. The ability of a molecule to effectively cross the placental barrier and reach the fetus determines the drug's toxicological effects on the fetus. In this regard, our study aims to predict the permeability of molecules across the placental barrier. Based on publicly available datasets, several machine learning models are comprehensively analysed across different fingerprints and toolkits to find the best suitable models. Several dataset analysis models are utilised to study the data diversity. Further, this study demonstrates the application of neural network-based models to effectively predict the permeability. K-nearest neighbour (KNN), standard vector classifier (SVC) and Multi-layer perceptron (MLP) are found to be the best-performing models with a prediction percentage of 82%, 86.4% and 90.8%, respectively. Different models are compared to predict the chosen set of drugs, drugs like Aliskiren, some insulin secretagogues and glucocorticoids are found to be negative while predicting the permeability.

INDEX TERMS Placenta barrier, machine learning, drug permeability, developmental toxicity.

I. INTRODUCTION

With the growing number of pregnant women, who have pre-existing medical conditions such as type II diabetes, thyroid disorders, psychiatric disorders, asthma, and hypertension, the scenario of either limited or no medication during pregnancy is now changed [1]. Over the years, there has been an increase of over 60% in the use of prescription drugs during the first trimester [2]. In addition, the risk of pregnancy complications and the exacerbation of the existing ones demand further therapy. When drugs are administered, the active pharmaceutical ingredient is distributed throughout the body via the bloodstream. During this process, the ingredient encounters various biological barriers such as the blood-brain barrier (BBB), blood-nerve barrier (BNB), blood-retinal barrier, pulmonary barrier, and placental barrier (in the case of pregnancy) [3], [4], [5]. While the intended drug interaction is easily achieved, there are always reported side effects because of the ability of drugs to cross these biological barriers and interact with untargeted regions [6]. For instance, thalidomide' a nonaddictive, nonbarbiturate sedative was prescribed to treat morning sickness in pregnant women. However, it was later identified that the drug developed the symptoms of peripheral neuropathy and severe birth defects [7]. This

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

indicates that the drug could cross the placental barrier and reach the fetus, along with the ability to cross the BBB and BNB. Following this disastrous tragedy of the 20th century, several studies have started emerging on identifying the ability of a drug to penetrate biological barriers.

The placenta barrier plays a crucial role in the maintenance of fetal health and development by ensuring the transport of nutrients and growth factors while eliminating xenobiotics. The placental permeability of chemicals/drugs is a crucial aspect of developmental toxicology [8]. To ensure safety during pregnancy, the application of in vitro, in vivo and ex vivo models of the placenta are adapted to study the ability of drugs or xenobiotics to pass through the placental barrier and reach the fetus [9]. However, their outcomes are not completely reliable due to the difference in the models, their simplicity and high cost. In this regard, computational algorithms could throw some light to predict the fate of xenobiotics in the placental atmosphere and the toxicity effects. These algorithms could also be used to design new drug candidates [9], [10].

In contrast to traditional classification models [11], [12], [13], [14], [15], [16], [17], [18], [19], machine learning algorithms [20] have recently been applied to predict BBB permeability. One of the applications of the machine learning (ML) models is to eliminate the tedious early stages of drug discovery cycle by providing quick inference on the permeability across the barrier to the bioinformaticians and drug developers [21]. The data-inductive nature of the ML models allows them to learn patterns in the data associated with drugs that permeate through the barrier, thereby allowing them to generalize on other drugs and new chemical compounds. We believe that a similar strategy could also be utilized to train the models for placenta barrier permeability. Unlike BBB, the quantitative structure-activity relationship (QSAR) models for the placenta barrier are extremely limited owing to poor data availability. Some QSAR models are based on ex vivo placental perfusion [22], [23], while others are based on the maternal-fetal blood concentration ratio [24]. To the best of our knowledge, limited studies have discussed the application of machine learning models for predicting the placental transport of chemicals [2].

By employing data from the basic science of drug structure and physicochemical properties, clinical decisions on the drug prescription during pregnancy-related complications can be considered using such permeability prediction algorithms. Since translational medicine facilitates medical advancement from scientists to clinician, such clinical decision-making algorithms become imperative. After the Thalidomide incident, a typical two-way bedside to bench-side communication, safer medications for morning sickness has been recognised. However, the development of predictive algorithms could fasten this decision making process from bench-side to effective bed-side communication.

Risk categorization for drugs during pregnancy was initially introduced in 1978 by Sweden followed by US Food and Drug Administration (FDA), which initiated the risk stratification representation using letters A to D along with an X category in 1979 [25]. The representations are as follow: Category A: No risk in human studies; catergory B: No risk in animal studies; category C: Risk cannot be ruled out due to non-satisfactory studies; Category D: Evidence of risk and finally, Category X: Contraindicated with risks of drugs outweighing potential benefits. FDA categorization needs robust data like controlled studies during pregnancy to categorize a drug as safe. Such studies are unlikely to be taken due to ethical and safety concerns. As a result, most of the drugs have been assigned as category C i.e., ''risk cannot be ruled out''. With the increase in the use of over-the-counter (OTC) medications, there is an urgent need to understand the fate of drugs and their permeability across the placenta. For instance, a recent study has implicated paracetamol, one of the most common OTC used during pregnancy affecting certain neurodevelopmental, reproductive and urogenital disorders [26].

The primary rationale behind this study is to have a better understanding of the ability of certain commonly used drugs to pass through the placental barrier. Assimilation of such information can be extended to decode the drug's effects on fetal development and its toxicity. For this purpose, our study has been designed to predict the passage of a few common drugs anti-diabetic, anti-allergic, and anti-hypertensive drugs. While many of these drugs are not generally prescribed during pregnancy, it is still crucial to understand their effects due to an increasing trend in self-medication [27]. To achieve this, we develop ML models with existing datasets to predict the transport of molecules across the placenta. The performance of the models is further analysed to predict the transport of the above-mentioned class of drugs.

In this study, we aim to develop and explore the application of machine learning and deep learning models for predicting the placental permeability of different drug classes. Since machine learning applications for such placental permeability have not been explored extensively in literature [2], this study has been performed to explore different models and molecular representations to understand the ability computer algorithms and correlate patterns in the drug structural aspects and its passage across the barrier. While Di Filippo et al. explore the suitability of genetic algorithm for critical feature selection, our study aims to comprehensively analyse the suitability of different models including neural networks with cross toolkit testing for this prediction application. We further employ the best models for permeability prediction of drugs that can be explored for pregnancy complications like hypertension, diabetes, and allergies. Another important rationale behind the current study is to understand the fate of some of the drugs that are generally prescribed during non-pregnancy. We envision correlating this permeability prediction with our developed machine learning models as a step towards a new approach for developmental toxicity prediction.

II. METHODOLOGY

At the core of our methodology, we conduct a thorough experimental study with several ML and deep learning (DL)

models to highlight the potential of utilizing the supervised learning algorithms for the placenta permeability. Specifically, we design a study to maximize the performance of the ML models by employing techniques that can overcome the challenges in the dataset. We further analyze the impact of these techniques on classification performance. Finally, we show an application of the trained models on the drugs commonly prescribed during pregnancy.

A. DATASET

Filippo et al. [2] compile a list of 248 drugs for placenta barrier permeability. The dataset also contains the fetal-maternal blood concentration ratio (F/M ratio) and clearance index (CI) of drug molecules. A drug is labelled as crossing (PB+) the placenta barrier if the F/M ratio is higher than 0.3 or CI is higher than 0.8. On the other hand, a drug is labelled as not crossing (PB-) the placenta barrier if the F/M ratio is lower than 0.3 or CI is less than 0.8. The drugs that had an F/M ratio between 0.15 and 0.3 were omitted from the dataset due to ambiguity in the permeability label. Altogether, the dataset contains 213 (86%) molecules with the label PB+ and 48 (14%) molecules with PB- labels. We utilize Pubchem [28], to represent the drugs in a simplified-molecular-input-line-entry system (SMILES). The SMILES representation allows for fingerprint and 1D/2D descriptors extraction through various chemical feature extraction frameworks. RDKit framework employed for feature extraction has classified 11 molecular SMILES as invalid, resulting in a final dataset size of 237.

B. MOLECULAR FINGERPRINTS

One of the aims of our experimental study is to evaluate the suitability of different molecular fingerprints for the placenta barrier permeability. Our study includes fingerprints that are path-based (e.g., FP2), substructure-based (e.g., FP3), hashing-based (e.g., Standard), Hybrid (e.g., Avalon), and model-generated (e.g., mol2vec). By selecting the well-known molecular fingerprints in each category, we aim to analyze the impact of the fingerprints on the classification accuracy, hoping the future research to focus on ML/DL model innovation without needing to test across all molecular fingerprint representations. We use several computational chemistry toolkits including RDKit, CDK, and Open Babel to extract the diverse molecular representations for this study.

C. EXPERIMENTAL STUDY OVERVIEW

We conduct a comprehensive empirical study of the placenta barrier permeability to identify the best parameters for different stages of the ML pipeline. First, we extract representative fingerprints from well-known categories to analyze the fingerprint that provides adequate information to the ML/DL models for permeability prediction. Specifically, we extract five different fingerprints, as explained earlier. After analyzing the fingerprints, we observe that most fingerprints are sparse binary vectors (e.g., FP3, Standard) except the ones generated by ML models as a learned representation of the molecules (e.g., mol2vec). Based on the sparse nature of fingerprints, we perform feature transformation using principal component analysis (PCA)(Table 1).

PCA allows us to convert high-dimensional data vectors into low-dimensional space without losing variance information. We chose to do PCA over other dimensionalilty reduction methods like kernel PCA and other feature selection methods as it provides a straightforward mechanism to control variability in the transformed data. PCA projects the data on principle components/Eigenvectors (i.e., directions that capture the variability in data) thereby suggesting directions in the multi-dimensional space while retaining most of the information.

This transformation makes the learning procedure easier for the ML models because the low-dimensional data needs fewer parameters for effective generalization while needing smaller training datasets. Another crucial advantage of PCA is that it allows us to control information loss in the dataset during transformation by specifying the variance of the transformed dataset. Additionally, we apply standard scalar normalization before PCA to effectively transform the dataset instances to avoid significant differences in feature ranges.

Once the necessary steps are taken to prepare the dataset, we explore the different ML models for training and effective generalization in real-world applications. To elaborate, we select a representative model from popular categories, such as in-memory-based (e.g., K-nearest neighbour (KNN)) classifiers, linear and kernel-based classifiers (e.g., support vector classifier (SVC)), tree and ensemble-based classifiers (e.g., random forest (RF)), boosting classifiers (e.g., Light gradient boosting machine (LightGBM)), and deep neural networks (e.g., multi-layer perceptron (MLP)).

KNN works by calculating K nearest instances to the query points (i.e. test) using the measure of distance. Alternatively SVC finds an optimal dimension in the high dimensional feature space to maximise the margin between two instances. Here, we chose to work with non-linear SVC from the SVM module to be able to implement the kernel hack and for better classification of drugs that fall between the positive and negative classes. To have a combination of decision trees, we chose random forest, an ensemble learning method that outputs the class determined by mist decision trees.LightGBM is an open-source gradient boosting framework; gradient-based one side sampling (GOSS) and exclusive feature bundling (EFB) are two essential parts of LightGBM. GOSS enables LightGBM to track the instances that are not properly trained, greatly enhancing the model's knowledge gain. EFB uses the sparsity of higher dimensional spaces to choose a group of features that are mutually exclusive, reducing the model's training time and memory complexity. MLP is a multilayer feedforward artificial neural network that generates output classes from input sets. The above ML models were chosen based on their successful application in previous studies [21]. For instance, KNN is one of the basic methods of ML modes and most frequently used with easy application. Similarly, MLP is another well

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

Feature Name	Feature type	Description	Dimension	Dimension after PCA (95% variance retained)
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
Rdkit	RDK	Substructure fingerprint with properties of daylight fingerprints to capture atom and bond types	2048	843
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
mol2vec	Unsupervised ML	Latent space vector representation of the molecules inspired by unsupervised natural language processing models.	300	33

explored models for a variety of applications and is continuously evolving.

One of the critical aspects of training the ML/DL models is to appropriately hyperparameterize the tuning for the dataset of interest. We conduct a comprehensive grid search for different hyperparameters of each model with built-in cross-validation (i.e., GridSearchCV) to ensure effective generalization across different folds of a limited dataset. Parameter choice can impact the accuracy of the chosen ML models with limitations contributed by biased dataset. To ensure appropriate parameter choice, we pair grid search with 10-fold cross-validation to select parameters that provide the best results across the folds. 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, thus allowing us to evaluate the models robustly. Grid search is strategically employed to exhaustively search for the best parameters for the 10 folds. In the initial set of experiments, we train and evaluate different ML modess with different fingerprints without upsampling the minority class to provide a baseline performance for the models.

One of the significant challenges of the placenta permeability dataset is the imbalance between PB+ and PB- classes (86:14). Therefore, necessary measures are needed to prevent model bias towards the PB+ class. To this end, we employ the synthetic minority oversampling technique (SMOTE) to generate instances of PB- class in the train set of each cross-validation during grid search. SMOTE generates the sample of PB- class by employing KNN and on one of the instances of PB- class. The new synthetic sample is generated by computing a vector between the selected instance of PB- and one of its neighbours. We employ SMOTE to analyze whether the data balancing technique can improve model performance by reducing the false positive rate (FPR). To elaborate, we examine the model's effectiveness for correctly predicting the minority class (i.e., PB-) without labelling them as PB+. Lower FPR provides higher confidence for drugs when inferred using the ML/DL models, encouraging further lab testing and release of newer drugs in the real world to treat different chronic conditions that women encounter during pregnancy.

In the final stage of our study, we construct a dataset comprising drugs prescribed for treating allergies, hypertension, and diabetes to predict their ability to cross the barrier using our models. By doing so, we expect to add value to the existing knowledge on toxicity predicted for drugs and in turn, provide a faster preliminary screening for drug safety during pregnancy. We employ the best-performing models and corresponding fingerprints for placenta barrier permeability. We infer the permeability using multiple models (ensemble approach) to minimize the likelihood of misprediction since ensemble models have lower error rates than individual ML/DL models. Next, we thoroughly analyze the inference of ML/DL models and discuss our findings.

D. METRICS

We exhaustively evaluate different classification aspects of the study's ML/DL models by evaluating their ability to correctly predict the samples from PB+ and PB- classes. We employ overall accuracy, sensitivity (i.e., *true positive/(true positive + false negative)*), and specificity (i.e., *true negative/(true negative + false positive)*) to quantify the prediction capabilities for PB+ and PB- classes.

Additionally, we utilize the area under the receiver operator characteristic curve (AUC_ROC) for combining sensitivity and specificity into a single measure. We also use the f_beta score (specifically, f_0.5) for taking the harmonic mean of precision and recall. The f_0.5 metric places higher weightage on precision, thereby ensuring a higher f_0.5 correspond

to a lower FPR. During the initial establishment of baseline performance of different models, we analyzed the f_0.5 score of the models in addition to AUC_ROC. However, we chose not to discuss the results of f beta score in the manuscript for better comparison with the existing articles on the barrier permeability. The source code, raw data and dataset details can be found in the GitHub library using the link "https://gitlab.com/chip23618/PlacentaBarrier"

E. IMPLEMENTATION DETAILS

We utilize the Pyfingerprint library for extracting fingerprints from different chemical fingerprint extraction libraries (e.g., RDKit, CDK, and Open Babel). We employ the SMOTE implementation made available by the imbalearn package in Python to handle the dataset imbalance. The imbalearn package supports Scikit-learn (sklearn) pipelines, which assist in grid search and model training with unbalanced dataset. Additionally, we utilize the standard sklearn implementation of ML models and MLP for our empirical study. The Xgboost and LightGBM models have been trained from their independent Python packages (compatible with sklearn). The grid search with cross-validation is also implemented using the built-in GridSearchCV method. The empirical study is conducted on an HP Z8 workstation in a Linux operating system, with 64-core Intel® Xeon(R) Silver 4216 CPU @2.10 GHz (boost to 2.6GHz) and 128 GB of RAM.

III. RESULTS

A. DATASET ANALYSIS

The placenta dataset utilized in the current study is the direct adaptation of the dataset used in the study conducted by Di Filippo et al. [2]. The authors create the dataset based on clearence index (CI), fetal/maternal drug concentration ratio (F/M) and literature evidence on whether a molecule passes the barrier or not. As the current study focuses on creating prediction models, the dataset from Di Filippo et al. was directly adapted without any modifications to it, thus ensuring a better comparison of the results with their study. In case of F/M ratio ≤ 0.15, the molecule is considered negative for placental barrier (PB-) permeability and molecules with F/M ≥ 0.3 are termed as positive for placental barrier permeability (PB+). Similarly, a CI >0.8 is termed PB+. Di Filippo et al. choose this threshold based both CI and F/M data. The dataset presents a total of 248 molecules with 213 PB+ and 35 PB- compounds. Unlike other common barrier permeability prediction dataset, the placenta dataset is extremely limited owing to poor data availability due to poor experimental models [29]. For building a reliable model, a thorough understanding of the available dataset is needed. In line with this, a scatter plot is developed to understand the chemical space occupied by the positive and negative samples of the placenta dataset. Such a plot would graphically represent the diversity of the molecules in the dataset. As can be observed in Figure 1, the chemical diversity of the compounds in the dataset is reasonably wide and the distribution of positive and negative molecules also indicates a diverse chemical space.

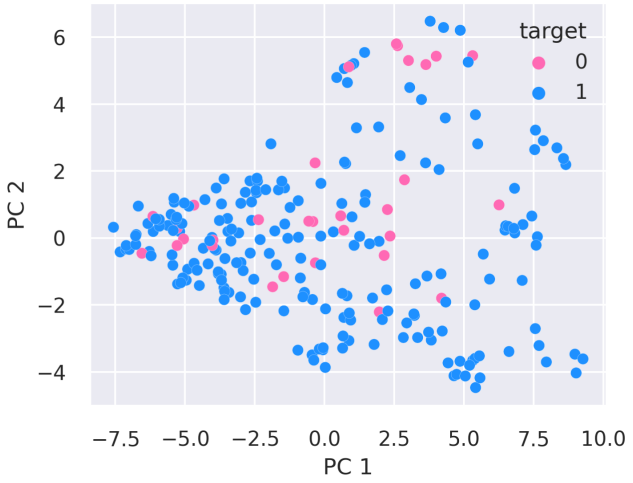

FIGURE 1. Chemical diversity analysis of the placenta dataset after PCA.

Further, understanding the similarity between two molecules is a routine and key task in cheminformatic analysis. In the current scenario, by studying similarity index of the molecules in the dataset, an understanding on the diversity of the molecules contributing to the dataset could be understood. A typical metric for calculating the separation or similarity of molecules is the Tanimoto coefficient. It could provide an understanding on the extent to which the information captured by these descriptors overlap. To analyse the similarity index of the molecules in the dataset and the selected set of drugs, the Tanimoto similarity index is calculated for each pair of molecules by using the Avalon fingerprint. The choice to study the similarity index using the Avalon fingerprint is made based on the preliminary results of several ML models application for predicting placental permeability across different fingerprints. As Avalon shows maximum prediction percentage for the majority of models employed, it is chosen as an obvious choice for studying the similarity index. A distribution plot is made to compare the similarity index of placenta dataset with anti-hypertensive drugs, anti-allergic drugs, and anti-diabetic drugs individually. Finally, a distribution plot of similarity between the three classes of drugs is also generated to understand the drug data (Figure 2).

The Tanimoto similarity index is calculated for each pair using Avalon fingerprint of the placenta dataset and the average similarity index is found to be 0.26, whereas the allergy drug list shows an average similarity index of 0.29 (Figure 2). Similarly, the similarity indices of diabetes drugs and allergy drugs are found to be 0.31 and 0.25, respectively. While the low similarity index of the placenta dataset indicates the structurally diverse compounds in the data thereby facilitating the development of a reliable model, the drugs chosen for analysing the ML models performances are also equally diverse. The similarity of the chosen drugs to that of placenta dataset (Figure 2D) indicates a similar diverse set. However, it can safely be assumed from these plots that the prediction model developed from placenta datset could effectively be

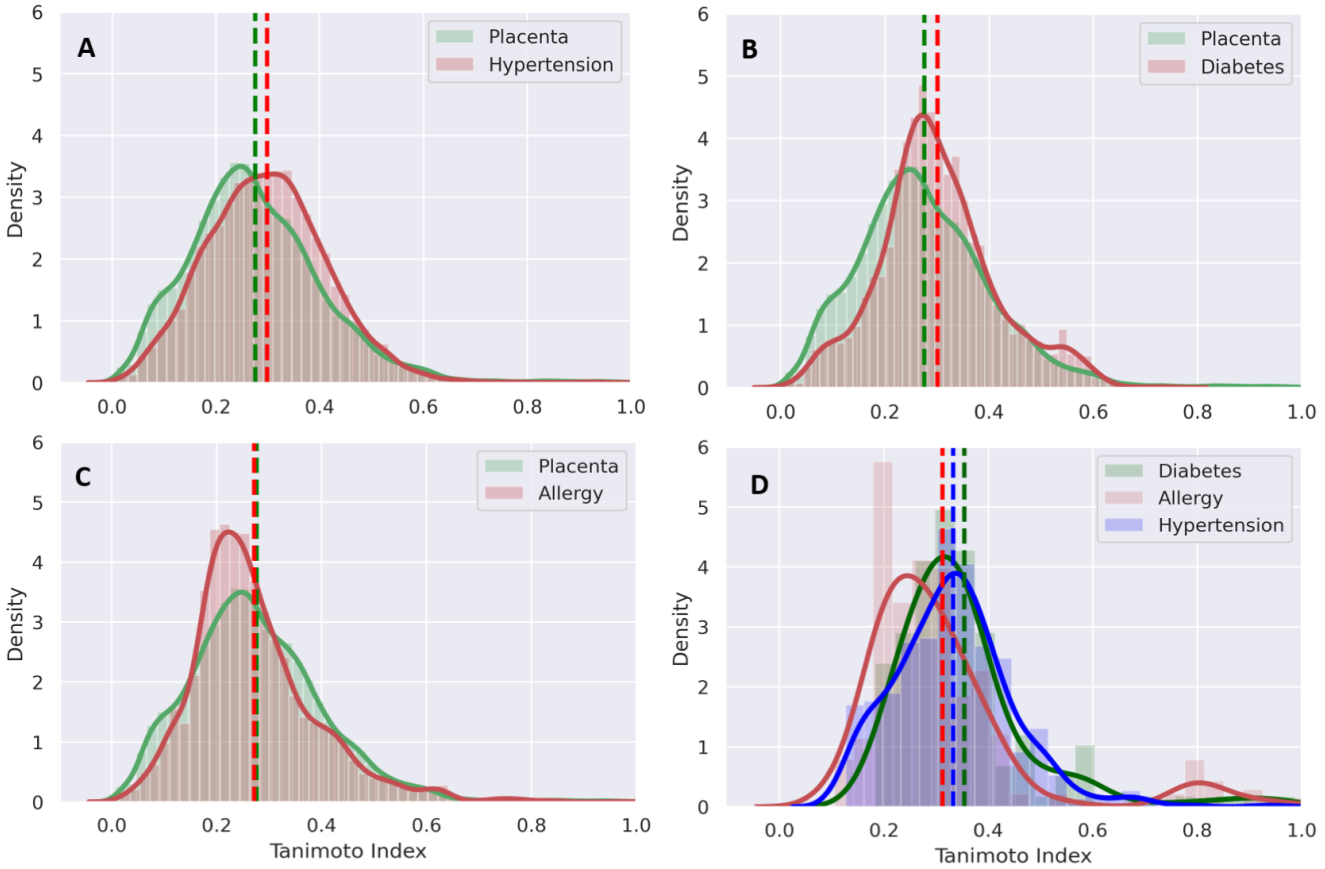

FIGURE 2. Tanimoto similarity index of (A) Placenta barrier chemicals and Hypertension drugs (B) Placenta function barrier chemicals and Diabetes drugs (C) barrier chemicals and and Allergy drugs (D) Hypertension, diabetes and allergy drugs. The x-axis represents tanimoto similarity index and y-axis represents the 10 times of the probability of the density of each tanimoto similarity index.

applied to predict the fate of the chosen drugs in placental atmosphere.

B. MODEL PERFORMANCES

The baseline performances of 10 different ML models for placenta permeability are studied (as provided in the supplementary file). While binary classifiers are generally analysed based on metrics like sensitivity, specificity, and receiver operating characteristics, the precision-recall plots can help provide with an accurate prediction of potential classification. This is attributed to their ability to evaluate the fraction of true positives among positive predictors [22]. In the current study, we have analysed the prediction performance using such a precision-recall curve Figure 3. While AUC_ROC curves are more popular, the interpretation of these metrices should cautiously be conducted when using imbalanced dataset.

With the aim of identifying ML models capable of classifying whether a drug component can cross the placental barrier, we have developed several ML models commonly used for classification problems and used different molecular representations toolkits. We have initially tested different ML models with several fingerprints and descriptor types to establish a baseline performance. Of the 10 models initially analysed with 17 fingerprint types, 5 best performing ML models are chosen further for improving performances. As can be seen from Figure 3& 4 and Table 2, the prediction performance of all the models are approximated around 70% to 86%. Although using different toolkits to generate similar fingerprints tend to offer certain level of redundancy, we choose to analyse all the fingerprints and descriptors from different toolkits, as an effort to perform cross-toolkit analysis. Consequently, the SMILES processing by different toolkits provides variable output, owing to the ability of their processing capacity and the difference in the aromaticity. However, due to the highly imbalanced nature of the datasets, the specificity of the models is identified to be very low. Interestingly, the KNN model show high sensitivity and relatively low specificity Table 2. Such low specificity has been reported previously in several KNN models with imbalanced datasets [22], [30]. As the KNN model is primarily a similarity-based classification model, such low specificities are expected with highly imbalanced datasets [21]. Similar results are obtained for all the models with very low specificity and high sensitivity. Interestingly, LightGBM model has performed well with some fingerprints owing to the in-built data balancing feature and provided regularization parameters for the estimators in the model to prevent them from favoring the majority class.

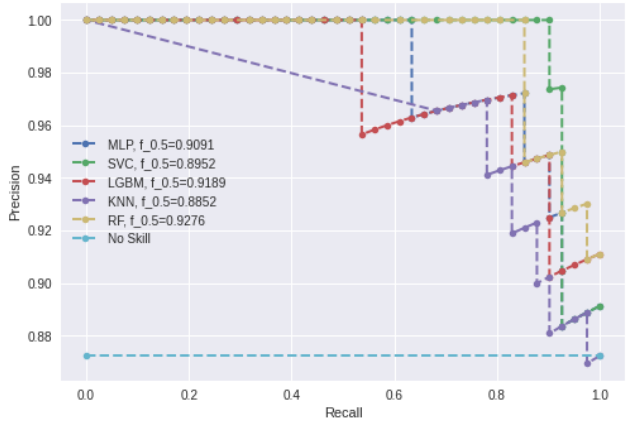

FIGURE 3. Precision recall curve for fold 4 of the models in the cross-validation study.

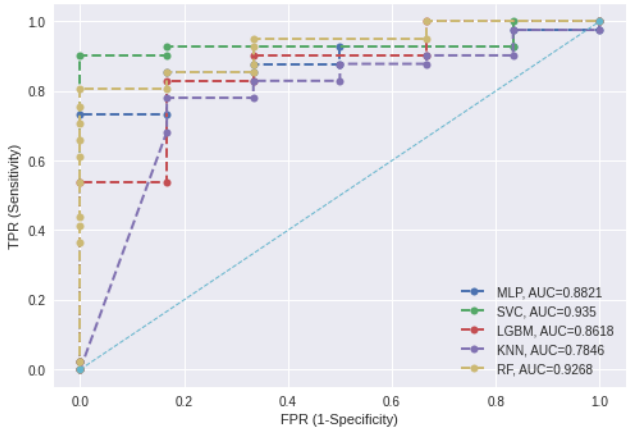

FIGURE 4. Receiver operating characteristic curve fold 4 of the models in the cross-validation study.

In this study, we have analysed the performance of a neural network model' multi-layer perceptron (MLP). MLP is a form of multilayer feed-forward neural network (FFNN) consisting of one or more hidden layers and the input of each hidden or output layer is considered as an inner product of previous output layer and weights [31]. As a result, MLP shows the highest prediction capability with better specificity and sensitivity when compared to other models without any data balancing stage 2. MLP is chosen among several deep learning models as their application has been previously proven to have outperformed other models used for similar classifier problems [21], [32], [33].

Based on our previous work with blood-brain barrier [21], it is noted that data balancing techniques in such a classifier model plays a crucial role in determining the model outcome and prediction capability. Albeit the high prediction accuracy than the present literature, the limited data on PB- class have resulted in a low recall ratio of the minority class (low specificity). In order to address this issue, oversampling techniques are commonly used. Though resampling techniques can involve either oversampling or undersampling, we choose to utilize oversampling methods as it adds more data points to the minority class without eliminating any critical data points offered by the majority class. A review of the existing work on re-sampling techniques further indicates better performances of oversampling techniques in comparison to undersampling [21], [33]. Hence, SMOTE is introduced as a data balancing technique to ensure that the number of PB+ and PB- instances are equal in the training and validation folds. This is achieved by interpolation between minority data and its k-nearest neighbours [34]. The application of SMOTE significantly has improved the specificity in most of the models (as provided in supplementary files). The improvement in the performance of KNN is very evident in case of standard fingerprint (Table 2). The extremely low specificity and high sensitivity seen without data balancing in all the models get improved significantly after the implementation of SMOTE. Although the prediction is expected to improve with data balancing techniques, some models like random forest fail to significantly improve the prediction accuracy. It is clear that accuracy would not be a suitable score for the classification job due to the imbalance of the dataset classes. It has been demonstrated that computing precision and recall (Figure 3), a standard metric in classification jobs, provides superior insight on the classificator's performance for imbalanced sets than the Receiver Operating Characteristic curve (Figure 4) [35].

The performance gain after SMOTE application is highly evident across all the models. It can be deduced that the higher the data imbalance, the greater is the perfromance gain after SMOTE application. While models like LighGBM have in-built data balancing modules, their effectiveness is not obvious without SMOTE 2. The imbalance ratio in the current dataset with positive class (86%) and negative class (14%) is extremely high, thus resulting in significant performance improvement by all models. Many studies have incorporated other data balancing techniques like adaptive synthetic sampling (ADASYN), random under sampler (RUS) [3] and have concluded SMOTE to be the most advantageous option. However, care should be taken to employ SMOTE appropriately, as studies have incorporated SMOTE to the entire dataset resulting in synthetic samples in the test set leading to model over-fitting [21].

C. PERMEABILITY PREDICTION OF DIFFERENT DRUG CLASSES

Use of prescriptions medications during pregnancy has become a common practice with prevalence ranging from 27 to 99% [36]. Though pregnant women are not included in pre-clinical studies, the current knowledge on the risk-benefit profile of drugs in pregnancy is mostly analysed through post-authorization studies. Anti-hypertensives, anti-diabetics, antibiotics, anti-allergens are some of the common drug groups used during pregnancy to treat and manage different conditions. As 0.6 to 2% of women are known to have chronic hypertension during pregnancy, medications to treat it are commonly prescribed [37] and 14% of maternal

TABLE 2. Performance of machine learning models with and without data balancing techniques (SMOTE).

Without Data Balancing						With Data Balancing using SMOTE			
Model Name	Features	Accuracy	Specificity	Sensitivity	Auc_Roc	Accuracy	Specificity	Sensitivity	Auc_Roc
KNN	Standard	0.87766	0.061905	1	0.825155	0.801684	0.680952	0.820674	0.861595
	FP3	0.890426	0.357143	0.971196	0.831959	0.78945	0.680952	0.80662	0.813811
	Avalon	0.873493	0.033333	1	0.818912	0.371454	0.9	0.291521	0.791715
	FP2	0.87766	0.061905	1	0.835221	0.696011	0.833333	0.674332	0.852333
	mol2vec	0.877748	0.066667	1	0.854955	0.725621	0.780952	0.718351	0.83463
SVC	Standard	0.894504	0.290476	0.985482	0.822687	0.88617	0.32381	0.970848	0.835492
	FP3	0.864982	0	0.995122	0.832404	0.239096	0.966667	0.128571	0.854346
	Avalon	0.869238	0	1	0.85542	0.844149	0.709524	0.864344	0.846748
	FP2	0.890248	0.257143	0.985366	0.812892	0.751064	0.709524	0.757143	0.808053
	mol2vec	0.911525	0.42381	0.985598	0.845219	0.882004	0.228571	0.98072	0.827449
RF	Standard	0.869238	0.061905	0.99036	0.804336	0.890337	0.357143	0.970848	0.777458
	FP3	0.88617	0.357143	0.966434	0.811285	0.857004	0.52381	0.908595	0.821254
	Avalon	0.88617	0.128571	1	0.787195	0.864894	0.285714	0.951336	0.634108
	FP2	0.860727	0.161905	0.966086	0.68555	0.87766	0.22381	0.975726	0.736353
	mol2vec	0.881915	0.161905	0.990244	0.835153	0.869415	0.395238	0.941928	0.814208
Light GBM	Standard	0.852394	0.357143	0.92741	0.794841	0.848227	0.609524	0.883856	0.82651
	FP3	0.852926	0.614286	0.889315	0.829249	0.835904	0.585714	0.874448	0.810763
	Avalon	0.886259	0.290476	0.975958	0.83614	0.852305	0.380952	0.922184	0.830314
	FP2	0.827216	0.480952	0.878978	0.781543	0.840071	0.32381	0.91777	0.753842
	mol2vec	0.839982	0.561905	0.883508	0.828339	0.839894	0.561905	0.883624	0.80873
MLP	Standard	0.886082	0.290476	0.975726	0.784079	0.87766	0.385714	0.951568	0.754162
	FP3	0.894947	0.585714	0.942393	0.804975	0.869592	0.52381	0.92288	0.773926
	Avalon	0.881738	0.480952	0.942044	0.737476	0.869415	0.614286	0.908014	0.860647
	FP2	0.864716	0.385714	0.936818	0.71686	0.873493	0.280952	0.961324	0.758672
	mol2vec	0.873493	0.485714	0.932288	0.77619	0.864894	0.485714	0.9223	0.746438

deaths have been attributed to hypertension [38]. However, their safety assessment needs to be addressed prior to prescription. While treating pregnancy-related hypertension with anti-hypertensives has been reported to increase the risk of pre-term birth, low birth weight, growth restrictions, untreated pregnancy-related hypertension have their own risk making therapy indispensable. Likewise, gestational diabetes is another common complication of pregnancy affecting over 15% of pregnancies worldwide. Consequently, the need for anti-diabetic medication during pregnancy has become mandatory [39]. Until recently, insulin was the only drug recommended for use in pregnancy. Metformin has been identified to have equal effectiveness as insulin without any risk to the fetus [40]. Though insulin is highly effective, it requires multiple daily injections thereby reducing patient adherence. This calls for more studies on alternate oral anti-diabetic agents for gestational diabetes and its effects on fetus. While the current study predicts the permeability of many drugs in these categories, we have chosen to discuss on few interesting drugs which were predicted positive or negative by majority of the models.

1) ANTI-HYPERTENSIVE DRUGS

Acetazolamide use in pregnancy is generally not recommended owing to its suspected teratogenic risks. Many congenital malformations like exencephaly, cleft lip, retarded incisor teeth development, and microphthalmia have been reported based on animal studies [39], [40]. However, poor evidence on the actual undesirable effects of acetazolamine use during pregnancy is still unknown. All our models predict the permeability of acetazolamine across the placental barrier, indicating its ability to reach the fetus (Table 3). This is consensus with the case report by Al-Saleem and Al-Jobair [41], where several congenital diseases are reported in 12-year-old boy who was exposed to maternal acetazolamide before and during the first trimester of the pregnancy. As a result, FDA has classified this drug as class C indicating adverse effects based on animal studies but no well-controlled human studies [41]. Similarly, angiotensin receptor blocker is a class of anti-hypertensive therapeutics commonly prescribed for hypertension. However, their use during pregnancy is not allowed owing to the possible congenital effects on the fetus [42]. However, pregnancy-related hypertensive disorders like pre-eclampsia need proper medical care as it is known to affect the fetus resulting in premature delivery, growth retardation, and death. While the regulations for high blood pressure therapy are constantly updated, the definition and treatment recommendations for pregnancy-related hypertension have not evolved considerably [42]. According to the US National High Blood Pressure Education Program (NHBPEP), methyldopa, labetalol, beta-blockers (other than atenolol), nifedipine with slow release profile and a diuretic for pre-existing hypertension are considered safe treatment during pregnancy [43]. In agreement with this, none of our

TABLE 3. Antihypertensive drug permeability prediction across placenta for different models trained with best fingerprints. Here, 1 and 0 represent the +ve permeability and −ve permeability classes respectively.

CID	Compound Name	Intended use	KNN (Standard)	MLP (Avalon)	SVC (Avalon)
1798	Acetazolamid	Idiopathic intracranial hypertension	1	1	1
3749	Irbesartan	RAS inhibitor/Angiotensin II receptor blockers/hypertension	1	1	1
4499	Nisoldipine	Calcium Channel Blocker/hypertention	1	1	1
4828	Pindolol	Beta blockers /hypertention	1	1	1
5833	Spironolactone	aldosteron antoagonist/potassium sparring diuretic/fluid drainage	1	1	1
54675783	Minocycline	Idiopathic intracranial hypertension	1	1	1
5493444	Aliskiren	RAS inhibitor/renin inhibitor/Hypertension	0	0	0
3869	Labetalol	Beta blockers /hypertention	0	0	0
2162	Amlodipine	Calcium Channel Blocker/hypertention	0	1	1
2405	Bisoprolol	Beta blockers /hypertention	1	1	0
2541	Candesartan	RAS inhibitor/Angiotensin II receptor blockers/hypertension	0	1	1
2720	Chlorothiazide	Diuretic/fluid draingage/hypertension	0	1	1
3333	Felodipine	Calcium Channel Blocker/hypertention	0	1	1
3702	Indapamide	thiazide-like diuretic/hypertension	0	1	1
3784	Isradipine	Calcium Channel Blocker/hypertention	0	1	1
4170	Metolazone	Diuretics/Hypertension	0	1	1
4474	Nicardipine	Calcium Channel Blocker/hypertention	0	1	1
4946	Propranolol	Beta blockers /hypertention	1	1	0
65999	Telmisartan	RAS inhibitor/Angiotensin II receptor blockers/hypertension	0	1	1
107807	Perindopril	RAS inhibitor/ ACE inhibitor/ Hypertension	0	1	1
5281037	Eprosartan	RAS inhibitor/Angiotensin II receptor blockers/hypertension	0	1	1
39147	Nadolol	Beta blockers /hypertention	1	1	0
2520	Verapamil	Calcium Channel Blocker/hypertention	0	0	1
41781	Torsemide	blocking the chloride-binding site of the Na+/K+/2Cl- cotransport	0	0	1
2471	Bumetanide	Diuretic/fluid draingage/hypertension	0	1	0

models predict the permeability of labetalol (Table 3), a beta-blocker thereby indicating their safety for pregnancy-related hypertensive disorders.

However, Aliskiren is a direct renin-angiotensin inhibitor; it has been classified as a pregnancy category C agent for the first trimester and category D drug for the second plus third trimesters [44]. Though aliskiren has not been evaluated in pregnant women and the categorization has been based on other renin-angiotensin inhibitors, our models predict the non-passage of these drugs across the placenta barrier. This could be attributed to either lack of actual data related to the drug permeability or lack of case reports on human effects from this drug. Though NHBPEP has reported beta-blockers relatively safe, their use during pregnancy has attained more controversies than ever, owing to the increase in the proportion of pregnant women with hypertension and the risk of congenital malformations in offspring.

Alternatively, calcium channel blockers are commonly used to treat pregnancy related hypertension as they are generally thought to have a favourable safety profile in pregnancy. However, a large study has observed the calcium channel blockers exposure during the third trimester and it is associated with a high risk of neonatal seizures. This is attributed to the plausibility of the permeability of calcium channel blockers across the placenta resulting in a decrease in intracellular calcium [45]. Similarly, amlodipine, which was predicted positive for permeability across the placenta by the majority of our models (Table 3) has been identified to have no association with fetal malformations as compared to other anti-hypertensive medications [46]. However, the study was based on the small sample size. thus, more studies on the actual passage of this drug and its effects during pregnancy need to be conducted to be able to tag them safe. Similarly, nifedipine has been reported to be favourable, though most guidelines do not recommend it for pregnancy use. There is also the possibility of their infiltration into the breast milk, however, no adverse fetal effects have been reported in this regard [47].

2) ANTI-DIABETIC DRUGS

Owing to the unparalleled efficacy, safety and lack of well-studied alternatives drugs, insulin continues to be the gold standard treatment for gestational diabetes. As for any diabetes type, medical nutrition therapy continues to be the starting point for diabetic therapy. However, the need for rapid control of glycemic levels demands alternate drug therapies [48]. The use of oral anti-diabetic drugs is not recommended by FDA, whereas the UK National Institute of Health and Care Excellence (NICE) considers metformin and glyburide as safe anti-diabetic medications. Metformin is a biguanide that decreases intestinal glucose absorption and increases insulin sensitivity [49]. Finally, it is metabolized by the CYP450 pathway and excreted in urine with a half-life of 6.2 hours. As recommended by NICE, insulin secretagogues

TABLE 4. Anti-diabetic drug permeability prediction across placenta for different models trained with best fingerprints. Here, 1 and 0 represent the +ve permeability and −ve permeability classes respectively.

CID	Compound name	Intended Use	KNN (Standard)	MLP (Avalon)	SVC (Avalon)
3476	Glimepiride	insulin secretagogue/anti-diabetic	0	0	0
3478	Glipizide	insulin secretagogue/anti-diabetic	0	0	0
5505	Tolbutamide	insulin secretagogue/anti-diabetic	0	0	0
4829	Pioglitazone	Agonist to Peroxisome proliferator-activated receptor gamma/anti-diabetic	1	1	1
10096344	Linagliptin	insulin secretagogue/anti-diabetic	1	1	1
11243969	Saxagliptin	DPP4-inhibitors	1	1	1
11450633	Alogliptin	DPP4-inhibitors	1	1	1
11949646	Empagliflozin	inhibits the sodium-glucose contransporter 2(SGLT2)	1	1	1
45588096	Exenatide	glucacon-like peptide-1(GLP-1) receptor agonist	0	1	1
145994868	Albiglutide	glucacon-like peptide-1(GLP-1) receptor agonist	0	1	1
16134956	Liraglutide	GLP-1 receptor agonist	0	1	1
65981	Repaglinide	inhibitor of ATP-sensitive potassium channels in a glucose-dependent manner/anti-diabetic	0	1	1
77999	Rosiglitazone	agonist at peroxisome proliferator activated receptors (PPAR)	0	1	1
4369359	Sitagliptin	DPP4-inhibitors	0	1	1
5311309	Nateglinide	non-sulfonylurea insulin secretagogue	0	1	1
9887712	Dapagliflozin	inhibits the sodium-glucose contransporter 2(SGLT2)	0	1	1
41774	Precose	alpha-glucosidase inhibitors/anti-diabetic	0	1	0
24812758	Canagliflozin	inhibits the sodium-glucose contransporter 2(SGLT2)	0	1	0

like glimepiride, glipizide and tolbutamide are predicted negative for placental permeability by all our models (Table 4). Based on the review by Kalra et al. [50], the choice of such glyburide use during pregnancy must be approached pragmatically after analyzing the target abnormality (insulin resistance or glucose intolerance), the pathophysiology of the person's diabetic condition, fetal and maternal safety. Similar results on the fetal safety of insulin secretagogues is reported by Moore et al. [51] based on a randomized control trial comparing the efficacy of metformin and glyburides for gestational diabetes.

Sulphonylurea based secretagogues have been proven to show effectiveness in treating hyperglycemia. However, their use during pregnancy result in increased incidence of neonatal hypoglycemia [52]. However, these conclusions on the resultant hypoglycemia are considered flawed owing to the lack of data on fetal metabolic profile. Another study by Towner et al. [53] rectify this lack of information by including metabolic profile and identify glycemic control as the sole independent risk factor for poor fetal outcomes. While risk assessment of more such drugs in the second and third trimester has started emerging, their assessment in the first trimester is to be addressed yet. Similarly, thiazolidinediones that activate the nuclear peroxisome proliferator activator receptor gamma (PPARγ) have been reported to cross the placenta and are consistent with our model prediction of pioglitazone and rosiglitazone (Table 4). However, their impact is studied only in vitro and has been reported to have no teratogenic effects. Although, owing to the role of PPARγ in placental maturation, there is a suspected effect of growth retardation [54]. With the advent of their potential application in stimulating ovulation in cases of polycystic ovary syndrome, more research on their application during pregnancy can make these drug candidates as an alternate option for managing gestational diabetes. Parallelly, glucosidase inhibitors have been predicted to permeate across placenta by only one model, while the other two models have predicted negative permeation (Table 4). In line with this, glucosidase inhibitors are reported to not have any adverse fetal outcomes in several studies [52]. However, most of the GLP-1 agonists have been predicted to cross the placental barrier by the majority of our models indicating their adverse effects on the fetus (Table 4), which is further consistent with the findings by Young and Anwar [52].

3) ANTI-ALLERGIC DRUGS

Allergic diseases are one of the most prevalent conditions in the urban world affecting 18 to 30% of women of childbearing age [55]. In pregnancy, allergic asthma has been related to adverse pregnancy outcomes. During pregnancy, the mother is in a state of immunotolerance to prevent the rejection of paternal antigens in the fetus [56]. Similarly, placentation and fetal growth require a specific immune environment. Hence, there needs to be a constant balance in this immune system to retain the foreign DNA and prevent pathogens or allergens from affecting it. Ideally, there should be no pharmacologic therapy during pregnancy, especially during the first trimester. However, to avoid adverse effects due to untreated conditions like allergic asthma, management of allergic conditions like asthma, allergic rhinitis, and allergic bronchitis is crucial.

In the current chosen pool of allergic medications, the majority of them were predicted to cross the placental barrier by most of our models (Table 5). Among the different types of glucocorticoids, ciclesonide and budesonide were predicted negative for permeability, whereas other glucocorticoids were predicted positive (Table 5). Generally, inhaled corticosteroids are considered relatively safe during pregnancy when used in lower doses. However, many continue to show apprehension about their use during pregnancy. This could be due to their use is considered safe as it prevents adverse outcomes of untreated conditions [57]. In the next class of anti-allergic medications, antihistamines are very popular and can be acquired as over-the-counter medications.

TABLE 5. Anti-allergic drug permeability prediction across placenta for different models trained with best fingerprints. Here, 1 and 0 represent the +ve permeability and −ve permeability classes respectively.

CID	Compound Name	Intended Use	KNN (Standard)	SVC (Avalon)	MLP (Avalon)
2200	Antazoline	anticholinergic activity/ allergic conjunctivitis	1	1	1
2267	Azelastine	histamine H1-receptor antagonist	1	1	1
21700	Beclomethasone dipropionate	synthetic corticosteroid	1	1	1
9865442	Loteprednol	glucocorticoid	1	1	1
5284514	Acrivastine	triprolidine analog antihistamine	1	1	1
5281071	Olopatadine	histamine H1 antagonist and mast cell stabilizer	1	1	1
444036	Fluticasone propionate	synthetic glucocorticoid	1	1	1
82153	Flunisolide	corticosteroid with anti-inflammatory	1	1	1
27503	Cromolyn sodium	degranulation of mast cells,	1	1	1
31307	Triamcinolone	Conrticosteroid	1	1	1
6918155	Ciclesonide	glucocorticoid	0	0	0
5281004	Budesonide	glucocorticoid that is a mix of the 22R and 22S epime	0	0	0
3241	Epinastine	histamine H1- and H2-receptors antaonist	1	1	1
3348	Fexofenadine	selective peripheral H1-antagonist	1	1	0
2913	Cyproheptadine	antagonist of both serotonin and histamine receptors	0	1	1
3219	Emedastine	allergic conjunctivitis	0	1	1
3957	Loratadine	selective peripheral H1-antagonist	0	1	1
4636	Oxymetazoline	alpha (α)-adrenergic agonist	0	1	1
26987	Clemastine	histamine H1 antagonist	0	1	1
124087	Desloratadine	selective peripheral H1-antagonist	0	1	1
1549000	Levocetirizine	histamine H1 antagonist	0	1	1
5282408	Ketotifen fumarate	histamine H1 antagonist and mast cell stabilizer	0	1	1
2678	Cetirizine	histamine H1 antagonist	0	1	1

This demands a more serious review of the safety of its use during pregnancy. In line with this, our models have predicted all the anti-histamines to be positive to permeate through the placental barrier. Anti-histamines like histamine-type 1 (H1) antagonists are some of the first generation of antihistamines used for allergic conditions. While none of these drugs has been reported to increase fetal risk if used during pregnancy [58], there are some recent contradicting reports which claim the risk of pre-eclampsia [59]. Owing to such irregularities in the literature, a more thorough amalgamation of computational models, in vitro models and data from other upcoming models is needed for a better understanding of their safety. Alternative to the first-generation antihistamines, the second generation is preferred owing to their reduced effect on the central nervous system (drowsiness). These second-generation antihistamines like cetirizine, fexofenadine, and loratadine are considered relatively safe based on a meta-analysis of different groups prescribed with these drugs posing no apparent fetal risks [58]. Though there is strong evidence of the safety of the use of anti-allergic medications during pregnancy and they are considered safe to use at moderate doses. However, given the current state of the research, it is still difficult to determine whether there are actual effects of these medication on pregnancy outcomes and if they are due to the underlying condition or due to the medication. Hence, more studies addressing these are needed to estimate the appropriate drug safety profile during pregnancy.

While the current study has been demonstrated to effectively predict the different drugs classes permeability across the placental barrier, there are certain limitations that need to be addressed to advance further in this field of study. For instance, the dataset utilised in the current study is a mixture of several in vitro perfused placental drug transfer data and blood analysis during pregnancy to determine drug concentration. While these two models are the most reliable data sources, it results only in a limited set of data points (drugs), as very limited drugs are administered to pregnant women. Alternatively, in vitro and in vivo data can be adapted individually and studies can be performed to identify the key differences between the data points. Thus, an ensemble of in vitro and in vivo study specifically for data generation of placental transfer of drugs can be considered to overcome the limitations pertaining to dataset.

IV. CONCLUSION

Inherent to pregnancy traits, it is extremely difficult to gather clinical data on fetal exposure to medications. Additionally, it is also difficult to conduct research and predict the level of fetal exposure. As a result, drug transfer across the placenta barrier and the overall quantitative information is highly sparse. In this regard, machine learning and deep learning models can be exploited to predict the fate of drugs across the placenta barrier. In this study, we have constructed the best-performing machine learning models and neural networks. While traditional ML models like KNN, and SVC have performed well in this classifier problem, tree-based and boosting algorithms could not perform well. We believe that this study is one of the first to use neural networks to analyze placental permeability. We further identified the application of

data balancing techniques like SMOTE improved the performance of all the models. Interestingly, data balancing worked only for few fingerprint types indicating the robustness of the models developed. In conclusion Avalon fingerprint worked based for most of the models and hence they were used for further studies on understanding the drug permeability of different drug classes.

The extremely little dataset and its severe imbalance could be attributed for the subpar results of tree-based and boosting algorithm models. The limits in datasets caused by ethical concerns could be addressed through efforts to enhance the dataset by fusing this work with in vitro model-based data. Future studies can concentrate on identifying the compounds and their representative qualities especially necessary for transport through receptor/transporter proteins because the current dataset have minimal information on the type of transport properties.

ML-assisted drug development for pregnancy complications and toxicological screening for developmental toxicology has seen enormous growth in the recent years. Owing to the highly dynamic nature of the pregnancy period, the pharmacodynamic and pharmacokinetic behaviour of a drug can vary during applications during pregnancy. Hence, use of such models could work towards predicting the fate of the drug in pregnant women. However, the dataset limitations can be addressed prior to the progress in this field. Future studies can include a combination of in vitro data generated using perfused placenta or other placenta barrier models like placenta-on-chip that are specifically developed to study a particular pregnancy condition and the drugs that could be used to treat them. Data from such studies could provide more details on the genomic and proteomic implications of a drug crossing the placenta and the models can aid in predicting the passage as well as the toxicological effects.

ACKNOWLEDGMENT

The findings herein reflect the work, and are solely the responsibility of the authors.

(Vaisali Chandrasekar and Mohammed Yusuf Ansari contributed equally to this work.)

REFERENCES

[1] B. Narayan and C. Nelson-Piercy, "Medical problems in pregnancy," *Clin. Med.*, vol. 17, no. 3, p. 251, 2017.

[2] J. I. D. Filippo, M. Bollini, and C. N. Cavasotto, "A machine learning model to predict drug transfer across the human placenta barrier," *Frontiers Chem.*, vol. 9, Jul. 2021, Art. no. 714678.

[3] Z. Shi, Y. Chu, Y. Zhang, Y. Wang, and D. 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.

[4] 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.

[5] A. V. Singh, V. Chandrasekar, P. Laux, A. Luch, S. P. Dakua, P. Zamboni, A. Shelar, Y. Yang, V. Pandit, V. Tisato, and D. Gemmati, "Micropatterned neurovascular interface to mimic the blood–brain Barrier's neurophysiology and micromechanical function: A BBB-on-CHIP model," *Cells*, vol. 11, no. 18, p. 2801, Sep. 2022.

[6] E. C. Ailes, J. Zimmerman, J. N. Lind, F. Fan, K. Shi, J. Reefhuis, C. S. Broussard, M. T. Frey, J. D. Cragan, E. E. Petersen, K. D. Polen, M. A. Honein, and S. M. Gilboa, "Using supervised learning methods to develop a list of prescription medications of greatest concern during pregnancy," *Maternal Child Health J.*, vol. 24, no. 7, pp. 901–910, Jul. 2020.

[7] N. Vargesson, "Thalidomide-induced teratogenesis: History and mechanisms," *Birth Defects Res. C, Embryo Today, Rev.*, vol. 105, no. 2, pp. 140–156, Jun. 2015.

[8] P. Myllynen, M. Pasanen, and O. Pelkonen, "Human placenta: A human organ for developmental toxicology research and biomonitoring," *Placenta*, vol. 26, no. 5, pp. 361–371, May 2005.

[9] 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.

[10] R. F. Barghash, I. M. Fawzy, V. Chandrasekar, A. V. Singh, U. Katha, and A. A. Mandour, "In silico modeling as a perspective in developing potential vaccine candidates and therapeutics for COVID-19," *Coatings*, vol. 11, no. 11, p. 1273, Oct. 2021.

[11] S. P. Dakua, "Use of chaos concept in medical image segmentation," *Comput. Methods Biomech. Biomed. Eng., Imag. Visualizat.*, vol. 1, no. 1, pp. 28–36, Mar. 2013, doi: 10.1080/21681163.2013.765709.

[12] S. P. Dakua, "LV segmentation using stochastic resonance and evolutionary cellular automata," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 29, no. 3, May 2015, Art. no. 1557002, doi: 10.1142/S0218001415570025.

[13] S. P. Dakua, "Performance divergence with data discrepancy: A review," *Artif. Intell. Rev.*, vol. 40, no. 4, pp. 429–455, Dec. 2013.

[14] S. P. Dakua, "AnnularCut: A graph-cut design for left ventricle segmentation from magnetic resonance images," *IET Image Process.*, vol. 8, no. 1, pp. 1–11, Jan. 2014. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2013.0088

[15] S. P. Dakua, J. Abinahed, and A. Al-Ansari, "A PCA-based approach for brain aneurysm segmentation," *Multidimensional Syst. Signal Process.*, vol. 29, no. 1, pp. 257–277, Jan. 2018.

[16] S. P. Dakua and J. Abi-Nahed, "Patient oriented graph-based image segmentation," *Biomed. Signal Process. Control*, vol. 8, no. 3, pp. 325–332, May 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809412001292

[17] S. P. Dakua, "Towards left ventricle segmentation from magnetic resonance images," *IEEE Sensors J.*, vol. 17, no. 18, pp. 5971–5981, Sep. 2017.

[18] S. P. Dakua and J. S. Sahambi, "Modified active contour model and random walk approach for left ventricular cardiac MR image segmentation," *Int. J. Numer. Methods Biomed. Eng.*, vol. 27, pp. 1350–1361, Sep. 2011, doi: 10.1002/cnm.1430.

[19] S. P. Dakua, J. Abinahed, and A. A. Al-Ansari, "Pathological liver segmentation using stochastic resonance and cellular automata," *J. Vis. Commun. Image Represent.*, vol. 34, pp. 89–102, Jan. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047320315002138

[20] 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.

[21] M. Y. Ansari, V. Chandrasekar, A. V. Singh, and S. P. Dakua, "Re-routing drugs to blood brain barrier: A comprehensive analysis of machine learning approaches with fingerprint amalgamation and data balancing," *IEEE Access*, vol. 11, pp. 9890–9906, 2023.

[22] 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. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18553217/

[23] 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. 2070125.

[24] T. Takaku, H. Nagahori, Y. Sogame, and T. Takagi, "Quantitative structure–activity relationship model for the fetal–maternal blood concentration ratio of chemicals in humans," *Biol. Pharmaceutical Bull.*, vol. 38, no. 6, pp. 930–934, 2015.

[25] R. Sannerstedt, P. Lundborg, B. R. Danielsson, I. Kihlström, G. Alván, B. Prame, and E. Ridley, "Drugs during pregnancy an issue of risk classification and information to prescribers," *Drug Saf.*, vol. 14, no. 2, pp. 69–77, Feb. 1996.

[26] A. Z. Bauer, D. Kriebel, M. R. Herbert, C.-G. Bornehag, and S. H. Swan, "Prenatal paracetamol exposure and child neurodevelopment: A review," *Hormones Behav.*, vol. 101, pp. 125–147, May 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29341895/
[27] N. Shenai, J. Shulman, P. Gopalan, E. Cheng, and J. M. Cerimele, "Fetal outcomes in intentional over-the-counter medication overdoses in pregnancy," *Psychosomatics*, vol. 59, no. 4, pp. 400–404, Jul. 2018.
[28] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, "PubChem in 2021: New data content and improved web interfaces," *Nucleic Acids Res.*, vol. 49, no. D1, pp. D1388–D1395, Jan. 2021.
[29] Y. Pu, J. Gingrich, and A. Veiga-Lopez, "A 3-dimensional microfluidic platform for modeling human extravillous trophoblast invasion and toxicological screening," *Lab Chip*, vol. 21, no. 3, pp. 546–557, 2021.
[30] 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.
[31] R. Zabihi, M. Schaffie, and M. Ranjbar, "The prediction of the permeability ratio using neural networks," *Energy Sources, A, Recovery, Utilization, Environ. Effects*, vol. 36, no. 6, pp. 650–660, Mar. 2014.
[32] Z. Wang, 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, Oct. 2018.
[33] 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.
[34] 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, Jun. 2002.
[35] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0118432.
[36] F. E. Pisa, A. Casetta, E. Clagnan, E. Michelesio, L. Vecchi Brumatti, and F. Barbone, "Medication use during pregnancy, gestational age and date of delivery: Agreement between maternal self-reports and health database information in a cohort," *BMC Pregnancy Childbirth*, vol. 15, no. 1, pp. 1–14, Dec. 2015.
[37] L. Cea Soriano, B. T. Bateman, L. A. G. Rodríguez, and S. Hernández-Díaz, "Prescription of antihypertensive medications during pregnancy in the U.K.," *Pharmacoepidemiol. Drug Saf.*, vol. 23, no. 10, pp. 1051–1058, Oct. 2014.
[38] C. A. Fitton, M. F. C. Steiner, L. Aucott, J. P. Pell, D. F. Mackay, M. Fleming, and J. S. McLay, "In-utero exposure to antihypertensive medication and neonatal and child health outcomes: A systematic review," *J. Hypertension*, vol. 35, no. 11, pp. 2123–2137, 2017.
[39] N. Kojima, M. Naya, and T. Makita, "Effects of maternal acetazolamide treatment on body weights and incisor development of the fetal rat," *J. Veterinary Med. Sci.*, vol. 61, no. 2, pp. 143–147, 1999.
[40] F. G. Biddle, "Genetic differences in the frequency of acetazolamide-induced ectrodactyly in the mouse exhibit directional dominance of relative embryonic resistance," *Teratology*, vol. 37, no. 4, pp. 375–388, Apr. 1988.
[41] A. I. Al-Saleem and A. M. Al-Jobair, "Possible association between acetazolamide administration during pregnancy and multiple congenital malformations," *Drug Design, Develop. Therapy*, vol. 10, pp. 1471–1476, Apr. 2016.
[42] C. M. Brown and V. D. Garovic, "Drug treatment of hypertension in pregnancy," *Drugs*, vol. 74, no. 3, pp. 283–296, Mar. 2014.
[43] National High Blood Pressure Education Program Working Group on High Blood Pressure in Pregnancy, "Report of the national high blood pressure education program working group on high blood pressure in pregnancy," *Amer. J. Obstetrics Gynecol.*, vol. 183, no. 1, pp. s1–s22, 2000.
[44] J. Cromer and S. Peker, "Aliskiren (Tekturna), a novel antihypertensive approach to inhibition of the renin–angiotensin–aldosterone system," *Pharmacy Therapeutics*, vol. 33, no. 2, p. 92, 2008.
[45] R. L. Davis, D. Eastman, H. McPhillips, M. A. Raebel, S. E. Andrade, D. Smith, M. U. Yood, S. Dublin, and R. Platt, "Risks of congenital malformations and perinatal events among infants exposed to calcium channel and beta-blockers during pregnancy," *Pharmacoepidemiol. Drug Saf.*, vol. 20, no. 2, pp. 138–145, Feb. 2011.
[46] A. Mito, A. Murashima, Y. Wada, M. Miyasato-Isoda, C. A. Kamiya, M. Waguri, J. Yoshimatsu, N. Yakuwa, O. Watanabe, T. Suzuki, N. Arata, M. Mikami, and S. Ito, "Safety of amlodipine in early pregnancy," *J. Amer. Heart Assoc.*, vol. 8, no. 15, Aug. 2019, Art. no. e012093.
[47] L. Malha and P. August, "Safety of antihypertensive medications in pregnancy: Living with uncertainty," *J. Amer. Heart Assoc.*, vol. 8, no. 15, Aug. 2019, Art. no. e013495.
[48] C. E. Cesta, J. M. Cohen, L. Pazzagli, B. T. Bateman, G. Bröms, K. Einarsdóttir, K. Furu, A. Havard, A. Heino, S. Hernandez-Diaz, K. F. Huybrechts, Ø. Karlstad, H. Kieler, J. Li, M. K. Leinonen, H. L. Gulseth, D. Tran, Y. Yu, H. Zoega, and I. Odsbu, "Antidiabetic medication use during pregnancy: An international utilization study," *BMJ Open Diabetes Res. Care*, vol. 7, no. 1, Nov. 2019, Art. no. e000759.
[49] J. S. Refuerzo, "Oral hypoglycemic agents in pregnancy," *Obstetrics Gynecol. Clinics North Amer.*, vol. 38, no. 2, pp. 227–234, Jun. 2011.
[50] S. Kalra, Y. Gupta, B. Kalra, and R. Singla, "Use of oral anti-diabetic agents in pregnancy: A pragmatic approach," *North Amer. J. Med. Sci.*, vol. 7, no. 1, p. 6, 2015.
[51] L. E. Moore, D. Clokey, V. J. Rappaport, and L. B. Curet, "Metformin compared with glyburide in gestational diabetes: A randomized controlled trial," *Obstetrics Gynecol.*, vol. 115, no. 1, pp. 55–59, 2010.
[52] J. Young and A. Anwar, "Diabetic medications in pregnancy," *Current Diabetes Rev.*, vol. 5, no. 4, pp. 252–258, Nov. 2009.
[53] D. Towner, S. L. Kjos, B. Leung, M. M. Montoro, A. Xiang, J. H. Mestman, and T. A. Buchanan, "Congenital malformations in pregnancies complicated by NIDDM: Increased risk from poor maternal metabolic control but not from exposure to sulfonylurea drugs," *Diabetes Care*, vol. 18, no. 11, pp. 1446–1451, Nov. 1995.
[54] L. Y.-S. Chan, J. H.-K. Yeung, and T. K. Lau, "Placental transfer of rosiglitazone in the first trimester of human pregnancy," *Fertility Sterility*, vol. 83, no. 4, pp. 955–958, Apr. 2005.
[55] I. Pali-Schöll, J. Namazy, and E. Jensen-Jarolim, "Allergic diseases and asthma in pregnancy, a secondary publication," *World Allergy Org. J.*, vol. 10, p. 10, Jan. 2017.
[56] F. Colucci, "The immunological code of pregnancy," *Science*, vol. 365, no. 6456, pp. 862–863, Aug. 2019.
[57] L. Smy, A. C. Chan, P. Bozzo, and G. Koren, "Is it safe to use inhaled corticosteroids in pregnancy?" *Can. Family Physician*, vol. 60, no. 9, pp. 809–812, 2014.
[58] M. So, P. Bozzo, M. Inoue, and A. Einarson, "Safety of antihistamines during pregnancy and lactation," *Can. Family Physician*, vol. 56, no. 5, pp. 427–429, 2010.
[59] A. K. Sande, E. A. Torkildsen, R. K. Sande, I. Dalen, K. C. Danielsson, and N.-H. Morken, "Use of antihistamines before or during pregnancy and risk of early-onset pre-eclampsia in allergic women: A population-based cohort study," *BMJ Open*, vol. 12, no. 10, Oct. 2022, Art. no. e061837.

VAISALI CHANDRASEKAR received the Ph.D. degree in industrial biotechnology from the National Institute of Technology Karnataka, India, in 2018. She was a Postdoctoral Researcher in modeling smart warehouse system projects. She is currently a third-year Postdoctoral Fellow with Hamad Medical Corporation, Qatar. Her research interests include *in vitro* disease modeling of neurological diseases for nano-based drug and food delivery systems, structure-activity relationship analysis for natural products, enzymology, and novel bioactive synthesis.

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 is a Research Associate with Hamad Medical Corporation.

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 with the Department of Chemical and Product Safety, German Federal Institute for Risk Assessment (BfR), Berlin. His research interests include chemical and nanotoxicology, micro nanorobotics, neurobiology, and antibacterial surfaces and understanding the bio physicochemical interactions at the nano-bio interface.

SHAHAB UDDIN received the Ph.D. degree in biochemistry from Aligarh Muslim University, India, in 1988. He did the first postdoctoral training with Ohio State University, Columbus, OH, USA, and the second postdoctoral training with Loyola University Chicago, Chicago, IL, USA. He has been a part of different healthcare institutions with diverse experiences in cancer genomics. He is currently a Principal Research Scientist and the Head of the Molecular Pathophysiology Section, HMC Translational Research Institute, Hamad Medical Corporation, Doha, Qatar. He has published over 260 peer-reviewed articles in high-impact factor journals. His research interests include identifying dysregulated signaling molecular that are involved in the survival and proliferation of malignant cells and utilized for prognosis biomarker identification in crucial malignancies. He received many awards/honors for his research, including the Hamdan bin Al-Makhtoom, Dubai, for his original research paper, in 2006, the Adeela bint Abdullah, Riyadh, for humanitarian for children cancer research, in 2011, and the Research Excellence from Hamad Medical Corporation, in 2016, 2017, and 2018. He has been listed in the world's top 2% of scientists in a global list released by Stanford University, USA.

KIRTHI S. PRABHU received the master's degree in pharmacy from Manipal University, India, in 2006, and the Ph.D. degree from Manipal University, in 2010, under the National Doctoral Fellowship Program sponsored and funded by All India Council of Technical Education, India. During this tenure, she also worked on various research projects funded by the pharmaceutical industry. She later joined the cadre of Assistant Professor with the Gahlot College of Pharmacy, India, wherein she taught undergraduate students various subjects related to pharmaceutical sciences. She moved to Qatar in 2014 and is currently a Postdoctoral Research Scientist in Translational Research Institute, Academic Health Systems, Hamad Medical Corporation. As a Postdoctoral Research Scientist, she works in collaboration with Qatar University and Weill Cornell College, Doha, Qatar, on various projects funded by the Medical Research Centre. In these projects, she has played an active role in investigating the potential of various known and newly isolated natural and synthetic compound/s and elucidating their role in various molecular pathways like PI3K/AKT/XIAP, SKP2-P21 axis, JAK-STAT pathway, and many others using cell lines of different lineages.

SAGNIKA DASH received the M.S. degree in obstetrics and gynecology. She has received her training in obstetrics and gynecology from Kolkata, India. She became a member of the Royal College of Obstetrics and Gynecology (MRCOG). She is currently a Gynaecologist with a vision to improve maternal health. While working in the private health sector in the state of Qatar, her passion for improving public health has led her to deliver health talks in schools and various cultural groups. Her research interests include ethics, research, medical education, and perinatal mental health.

SOUHAILA AL KHODOR received the bachelor's degree in medical laboratory technology from the Faculty of Public Health, Lebanese University, in 2001, the first master's degree in microbiology and immunology from the American University of Beirut, in 2004, and the second master's and Ph.D. degrees in microbiology and immunology from the University of Louisville, Louisville, KY, USA, in 2008. She has been the Director of the Maternal and Child Health Department, Research Branch, Sidra Medicine, Qatar, since July 2019, and an Investigator-Associate Level, since January 2015. She is in charge of the Microbiome and Biomarkers Discovery Laboratory. Before joining Sidra Medicine, she was a Postdoctoral Fellow with the Signaling Systems Unit, Laboratory of Systems Biology, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH), Bethesda, MD, USA. She is currently an Adjunct Assistant Professor with the College of Health and Life Science, Hamad Bin Khalifa University, and an Adjunct Assistant Professor with the Department of Biomedical Sciences, College of Health Sciences, Qatar University. She has over 48 peer-reviewed publications. She has been serving as an Associate Editor for the *Journal of Translational Medicine*, since 2017.

ANNALISA TERRANEGRA received the M.Sc. degree (cum laude) in biological sciences from the University of Siena, Italy, in 2000, and the Ph.D. degree in molecular medicine and the Postgraduate Diploma degree (cum laude) in nutritional sciences from the University of Milan, Italy, in 2007 and 2015, respectively. She achieved strong experience in molecular biology and genetics during her career with the University of Milan, as a Postgraduate Fellow (2001–2003), a Ph.D. Fellow (2003–2006), and a Postdoctoral Fellow (2007–2010). She was a Consultant in genetics with the San Raffaele Hospital, Milan (2010–2012), and a Research Fellow in nutrition with the San Paolo Hospital, Milan (2013–2014). She joined Sidra Medicine, in 2014. Since 2015, she has been an Adjunct Assistant Professor with the College of Health and Life Sciences, Hamad Bin Khalifa University, Qatar. Since 2018, she has been an Adjunct Assistant Professor with the College of Health Sciences, Qatar University, Qatar.

MATTEO AVELLA received the B.A. degree in biology from the University of Milan, Bicocca, Italy, and the Ph.D. degree from the Polytechnic University of Marche, Ancona, Italy, studying the intestinal microbiota's effects on fish's early development and reproduction. He conducted postdoctoral research with the National Institutes of Health, Bethesda, MD, USA, where he switched his studies to the mammalian system and focused his research efforts on understanding the molecular mechanisms regulating fertilization in mice and humans. He is currently a Principal Investigator with the Division of Maternal and Child Health, Sidra Medicine, and the Head of the Laboratory of Reproductive Biology. Before joining Sidra Medicine, he was an Assistant Professor with the Department of Biological Science, University of Tulsa, Tulsa, OK, USA, and the School of Health Professions, Eastern Virginia Medical School, Norfolk, VA, USA.

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. He has 15 years of research experience in computer vision. He is currently a Senior Research Scientist with the Department of Surgery, Hamad Medical Corporation, Qatar. He is a Certified PMP.