Drug Permeability Across Placenta using Machine Learning

ST. TERESA'S COLLEGE (AUTONOMOUS)

AFFILIATED TO MAHATMA GANDHI UNIVERSITY

PROJECT REPORT

In partial fulfilment of the requirements for the award of the degree of

BCA (CLOUD TECHNOLOGY & INFORMATION SECURITY MANAGEMENT)

By

Athira Mani -SB21BCA006

&

Riya Fathima -SB21BCA030

III DC BCA (CLOUD TECHNOLOGY AND INFORMATION SECURITY MANAGEMENT)

Under the guidance of

Maria Neethu Titus

DEPARTMENT OF BCA (CLOUD TECHNOLOGY & INFORMATION SECURITY MANAGEMENT)

MARCH 2024

Drug Permeability Across Placenta using Machine Learning

ST. TERESA'S COLLEGE (AUTONOMOUS)

AFFILIATED TO MAHATMA GANDHI UNIVERSITY

PROJECT REPORT

In partial fulfilment of the requirements for the award of the degree of

BCA (CLOUD TECHNOLOGY & INFORMATION SECURITY MANAGEMENT)

By

Athira Mani -SB21BCA006

&

Riya Fathima -SB21BCA030

III DC BCA (CLOUD TECHNOLOGY AND INFORMATION SECURITY MANAGEMENT)

Under the guidance of

Maria Neethu Titus

DEPARTMENT OF BCA (CLOUD TECHNOLOGY & INFORMATION SECURITY MANAGEMENT)

MARCH 2024

DECLARATION

We, undersigned, hereby declare that the project report, " Drug Permeability Across Placenta using Machine Learning", submitted for partial fulfilment of the requirements for the award of degree of BCA (Cloud Technology and Information Security Management) at St. Teresa's College (Autonomous), Ernakulam (Affiliated to Mahatma Gandhi University), Kerala, is a Bonafede work done by us under the supervision of Miss. Maria Neethu Titus. This submission represents our ideas in our own words and where ideas or words of others have not been included. We have adequately and accurately cited and referenced the original sources. We also declare that we have adhered to the ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Ernakulam
March 2024

Athira Mani- SB21BCA006
Riya Fathima -SB21BCA030

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

BCA (CLOUD TECHNOLOGY AND INFORMATION SECURITY MANAGEMENT)

DEPARTMENT OF BCA

CERTIFICATE

This is to certify that the report entitled "Drug Permeability Across Placenta using Machine Learning", submitted by Athira Mani and Riya Fathima to the Mahatma Gandhi University in partial fulfilment of the requirements for the award of the Degree of BCA (Cloud Technology and Information Security Management) is a Bonafede record of the project work carried out by them under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Internal Supervisor

Ms. Archana Menon P

Head of the Department

25/04/24

External Supervisor

ST. TERESA'S COLLEGE
DEPT. OF COMPUTERS
ER...

ACKNOWLEDGMENT

First and foremost, we thank God Almighty for his blessings. We take this opportunity to express our gratitude to all those whole helped us in completing this project successfully. I wish to express our sincere gratitude to the Manager Rev. Dr. Sr. Vinitha CSST and the Principal Dr. Alphonsa Vijaya Joseph for providing all the facilities.

We express our sincere gratitude towards the Head of the Department Ms. Archana Menon P. We deeply express sincere thanks to our project guide Ms. Maria Neethu Titus for her proper guidance and support throughout the project work.

We are indebted to our beloved teachers whose cooperation and suggestion throughout the project which helped us a lot. We thank all our friends and classmates for their support.

We convey our hearty thanks to our parents for the moral support, suggestion and encouragement

ABSTRACT

Predicting the permeability of molecules across the placental barrier is indeed a crucial aspect of ensuring the safety of drugs for both the mother and the fetus. The use of computational algorithms and machine learning models to predict drug permeability across the placental barrier offers several advantages, including reducing the need for animal testing and providing insights into the potential toxicological effects on the fetus. Our approach of analysing publicly available datasets and employing various machine learning models across is commendable. It allows for a comprehensive evaluation of the predictive performance and helps in identifying the most suitable models for the task. The reported performance metrics of the machine learning models, including K-nearest neighbor (KNN), standard vector classifier (SVC), and Multi-layer perceptron (MLP), are impressive. Achieving prediction percentages of 82%, 86.4%, and 90.8%, respectively, indicates the effectiveness of these models in predicting drug permeability across the placental barrier. Moreover, the identification of specific drugs like Aliskiren, insulin secretagogues, and glucocorticoids as having negative predictions for permeability underscores the practical implications of your study. Such findings can guide healthcare professionals and pharmaceutical researchers in making informed decisions regarding drug use during pregnancy. Overall, our study contributes significantly to the field of pharmacology and toxicology by offering a valuable alternative approach to evaluating drug safety during pregnancy. It also highlights the potential of computational methods and machine learning in advancing our understanding of drug interactions with the placental barrier.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 GENERAL BACKGROUND

The general background of our study appears to center around the use of computational models to predict drug permeability across the placental barrier. This is an important area of research because ethical and practical issues limit the amount of direct drug testing that can be conducted on pregnant women. Computational models, including machine learning algorithms, provide a non-invasive alternative that can help predict how drugs interact with the placenta, and by extension, their potential effects on a fetus. These predictive models are vital for assessing the safety of drug administration during pregnancy, aiming to prevent teratogenic effects or other harmful consequences to the developing child. Moreover, by analyzing and comparing the performance of various machine learning models, our study seeks the most effective approach to accurately predict the placental transfer of pharmaceutical compounds. Our exploration leverages publicly available datasets to train different machine learning models, such as K-nearest neighbor, support vector classifier, and multi-layer perceptron, to predict the likelihood of various molecules crossing the placental barrier. Such predictions are based on comparing different chemical fingerprints unique representations of the molecules' structures assisting in the identification of safe medications for pregnant women and reducing potential drug-related risks. This multidisciplinary field intersects pharmacology, toxicology, computer science, and bioinformatics. Scientists and researchers developing these models and algorithms require a deep understanding of both the biological aspects of drug transport across the placenta and the technical considerations of machine learning.

1.2 Drug Development and Safety

Drug development is a meticulous and multi-faceted process that spans several stages, starting from the identification of potential drug candidates to their eventual regulatory approval and availability in the market. The preclinical phase initiates this journey, where potential drug compounds undergo thorough laboratory testing to evaluate their safety, efficacy, and pharmacokinetic properties. This phase involves extensive in vitro studies, animal testing, and toxicity assessments to gather crucial data. Subsequently, promising candidates progress to clinical trials, which entail testing in human subjects to assess

safety, dosage, effectiveness, and potential side effects. Clinical trials are conducted in sequential phases, starting with small-scale Phase I trials and advancing to larger Phase II and Phase III trials to gather comprehensive evidence. Following successful completion of clinical trials, drug developers submit applications to regulatory agencies, such as the U.S. Food and Drug Administration (FDA) or the European Medicines Agency (EMA), for approval to market the drug to healthcare providers and patients. Ensuring drug safety is paramount throughout the entire drug development process and beyond. Safety assessments involve evaluating potential risks and benefits associated with pharmaceutical products, considering factors such as adverse reactions, toxicity, drug interactions, and long-term effects on patient health. Post-market surveillance becomes crucial once a drug is available in the market, involving continuous monitoring to detect and assess any adverse effects that may arise in real-world clinical settings. Regulatory agencies play a pivotal role in safeguarding drug safety by establishing guidelines and standards for drug development, approval, and post-market surveillance. They review safety data, conduct inspections, and may take regulatory action, such as issuing warnings or recalls, to protect public health. Additionally, pharmaceutical companies and healthcare providers implement risk management strategies to minimize potential risks associated with drug use. These strategies may include patient education, prescribing guidelines, monitoring protocols, and risk mitigation measures, collectively contributing to the overarching goal of ensuring drug safety throughout the lifecycle of pharmaceutical products.

1.3 Placental Barrier

The placental barrier stands as a pivotal determinant of drug permeability during pregnancy, delineating the extent to which medications traverse from the maternal bloodstream to the fetal circulation. Comprising layers of maternal and fetal tissues, including the syncytiotrophoblast, the placental barrier orchestrates a complex exchange process vital for fetal nourishment and protection. Drug molecules navigate this barrier through diverse mechanisms, including passive and active transport, influenced by their molecular properties such as size, lipid solubility, and charge. Consequently, drugs with greater lipid solubility and smaller molecular size typically exhibit heightened permeability, potentially impacting fetal exposure and susceptibility to therapeutic effects or adverse reactions. The assessment of drug permeability across the placental barrier underscores its profound implications for maternal therapy and fetal health, guiding

healthcare decisions and risk management strategies during pregnancy. Understanding the interplay between drug properties and placental transport mechanisms is paramount in optimizing maternal treatment regimens while safeguarding fetal development.

1.4 Traditional Testing Limitations

Traditional testing methodologies encounter several limitations when evaluating drug permeability across the placental barrier, prompting the need for alternative approaches to assess drug safety during pregnancy. Firstly, reliance on animal models may not fully capture the intricacies of human placental physiology and drug transport mechanisms, potentially leading to inaccurate predictions of drug effects in pregnant women. Furthermore, conducting experiments directly on pregnant women poses ethical concerns and practical challenges, limiting the availability of human data and complicating the interpretation of results. The complexity and variability of placental function further exacerbate these challenges, necessitating innovative strategies that leverage computational algorithms and predictive models to predict drug permeability accurately. By addressing these limitations, researchers can enhance our understanding of drug interactions with the placental barrier and improve the safety assessment of medications for pregnant populations.

1.5 Computational Approaches

Computational approaches offer a promising avenue for overcoming the limitations of traditional drug testing methods, particularly in assessing drug permeability across the placental barrier. These approaches leverage computational algorithms, mathematical models, and advanced data analysis techniques to simulate drug interactions, predict pharmacokinetic properties, and assess potential toxicological effects. Machine learning algorithms, including neural networks, support vector machines, and decision trees, analyze large datasets to identify patterns and predict drug permeability across the placental barrier. These algorithms can integrate diverse molecular and clinical data to develop predictive models with high accuracy and specificity.

1.6 Machine Learning

Machine learning (ML) is a discipline of artificial intelligence (AI) that provides machines with the ability to automatically learn from data and past experiences while identifying patterns to make predictions with minimal human intervention. Machine learning methods enable computers to operate autonomously without explicit programming. ML applications are fed with new data, and they can independently learn, grow, develop, and adapt. Machine learning derives insightful information from large volumes of data by leveraging algorithms to identify patterns and learn in an iterative process. ML algorithms use computation methods to learn directly from data instead of relying on any predetermined equation that may serve as a model. The performance of ML algorithms adaptively improves with an increase in the number of available samples during the 'learning' processes.

1.6.1 Types of Machine Learning

There are four basic types of machine learning:

1. Supervised learning
2. Unsupervised learning
3. Semi supervised learning
4. Reinforcement learning.

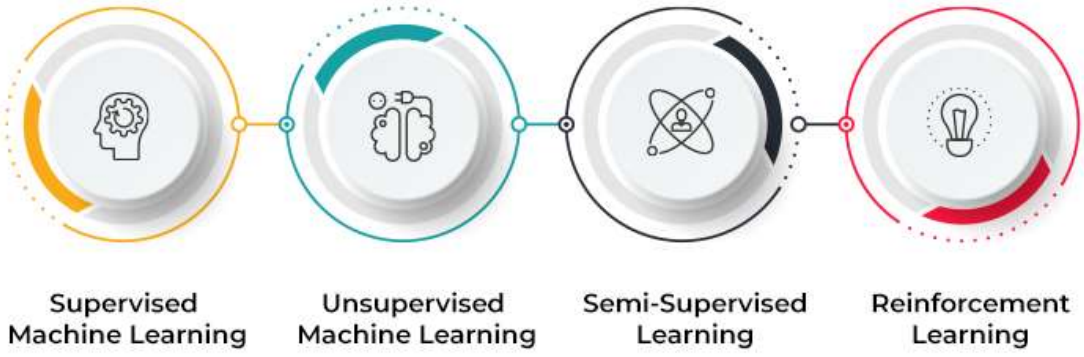


Fig 1.1 Types of Machine Learning

In our project, supervised machine learning methodologies have been employed extensively.

1.6.2 Supervised Learning

Supervised learning is a category of machine learning that uses labelled datasets to train algorithms to predict outcomes and recognize patterns. Unlike unsupervised learning, supervised learning algorithms are given labelled training to learn the relationship between the input and the outputs. Supervised machine learning algorithms make it easier for organizations to create complex models that can make accurate predictions. As a result, they are widely used across various industries and fields, including healthcare, marketing, financial services, and more. Supervised learning in machine learning is generally divided into two categories: classification and regression.

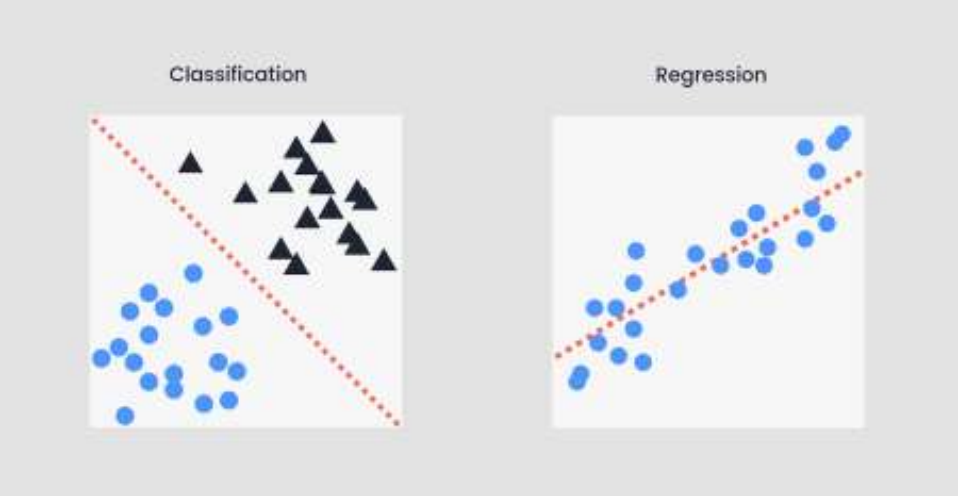


Fig 1.2 Types of Supervised Learning

Regression

Regression algorithms are used to predict a real or continuous value, where the algorithm detects a relationship between two or more variables. A common example of a regression task might be predicting a salary based on work experience. For instance, a supervised learning algorithm would be fed inputs related to work experience (e.g., length of time, the industry or field, location, etc.) and the corresponding assigned salary amount. After the model is trained, it could be used to predict the average salary based on work experience.

Classification

Classification algorithms are used to group data by predicting a categorical label or output variable based on the input data. Classification is used when output variables are categorical, meaning there are two or more classes. One of the most common examples of classification algorithms in use is the spam filter in your email inbox. Here, a supervised learning model is trained to predict whether an email is spam or not with a dataset that contains labelled examples of both spam and legitimate emails. The algorithm extracts information about each email, including the sender, the subject line, body copy, and more. It then uses these features and corresponding output labels to learn patterns and assign a score that indicates whether an email is real or spam.

CHAPTER 2

LITERATURE SURVERY

Drug permeability across physiological barriers, such as the placenta and the blood-brain barrier (BBB), presents a significant challenge in pharmaceutical research and development. Understanding and predicting the ability of drugs to cross these barriers is crucial for ensuring the safety and efficacy of therapeutic interventions, particularly during pregnancy and in the treatment of neurological disorders. In recent years, interdisciplinary approaches integrating computational modeling, machine learning (ML) algorithms, advanced biotechnological techniques, and nanotechnology have emerged as promising strategies to address this challenge.

The placenta serves as a critical interface between the maternal and fetal circulatory systems, regulating the exchange of nutrients, waste products, and drugs between the mother and the developing fetus. However, the placental barrier also presents a formidable obstacle to the passage of exogenous substances, including therapeutic compounds. Predicting the placental transfer of drugs is essential for assessing fetal exposure and potential adverse effects during pregnancy. Computational models leveraging physicochemical properties, structural features, and experimental data can provide valuable insights into drug permeability across the placental membrane.

One approach to predicting placental drug transfer involves the development of ML models trained on comprehensive databases of compounds with experimental information on placental transfer. For example, researchers have compiled databases of compounds characterized by a diverse set of descriptors, including physicochemical properties and structural features. By evaluating different ML classifiers and implementing feature selection algorithms, such as genetic algorithms, researchers can identify robust models capable of accurately classifying compounds based on their ability to cross the placental barrier.

In a study focused on placental drug transfer prediction, a low-dimensional ML model was developed and trained on a database of 248 compounds with experimental placental transfer data. The model utilized a set of approximately 5.4 thousand descriptors, including structural features and physicochemical properties, to classify compounds

according to their ability to cross the placental barrier. By optimizing feature selection and model parameters, researchers identified a Linear Discriminant Analysis (LDA) model trained with only four structural features as a robust predictor of placental drug transfer. This model exhibited high accuracy and minimized false positives, making it a valuable tool for predicting drug permeability during pregnancy.

Similarly, predicting the ability of drugs to traverse the BBB is essential for developing treatments for neurological diseases. The BBB is a highly selective barrier that regulates the passage of molecules into the brain, protecting it from potentially harmful substances. However, this barrier also limits the delivery of therapeutic agents to the central nervous system, posing a challenge for drug development. Computational approaches leveraging molecular structure features and ML algorithms offer a promising strategy for predicting BBB permeability and identifying candidate drugs for neurological disorders.

In a study focused on BBB permeability prediction, researchers extracted molecular structure features from drug molecule SMILE strings using computational methods. By integrating feature selection algorithms and employing ML techniques, such as XGBoost, researchers identified 33 chemical structure features with significantly discriminant performance. These features were used to construct multiple discriminant models, with the XGBoost model selected as the final prediction model due to its high accuracy on training and validation datasets. Through data preprocessing and parameter optimization, the XGBoost model achieved robust performance in predicting BBB permeability, demonstrating its potential for facilitating drug development for neurological conditions.

Advancements in nanotechnology have also revolutionized drug delivery strategies, particularly for crossing physiological barriers like the BBB. Nanoparticles-based biotechnological engineering offers precise control, targeting, and delivery of theranostic payloads across the BBB, thereby enhancing drug efficacy and minimizing off-target effects. By designing nanoparticles capable of traversing the BBB, researchers aim to improve the delivery of therapeutic agents for brain cancer and neurological disorders.

In a review focused on nanoparticles-based biotechnological engineering for BBB traversal, researchers highlighted the potential of nanorobots to deliver theranostic payloads across the BBB. By integrating computational modeling, ML tools, and artificial intelligence (AI) with robotics, researchers aim to predict and design the next generation of nanorobots capable of crossing the BBB without causing harm. These interdisciplinary approaches offer new opportunities for targeted drug delivery in neurological diseases, with the potential to revolutionize pharmaceutical design and BBB research.

In parallel, the development of microfluidic organ-on-a-chip platforms has enabled the creation of in vitro models that mimic the complexities of physiological barriers like the placenta and the BBB. These platforms offer controlled environments for studying cellular interactions, barrier functions, and drug permeability, facilitating toxicological screening and drug development processes. By recapitulating key elements of the placental microenvironment or BBB in a controlled setting, researchers can gain insights into the regulatory pathways governing placental development or BBB function and assess potential toxicological effects of drug candidates.

For example, a placenta-on-a-chip platform was designed to evaluate trophoblast invasion within a 3D microfluidic chip. This platform incorporates endothelial and trophoblast cells layered with an extracellular matrix, allowing for real-time monitoring, imaging, and evaluation of trophoblast cell invasion. Coupled with fluorescent cell tagging and flow cytometry, this platform enables the collection of invasive cells, providing insights into pathways regulating trophoblast invasion and potential toxicological effects on placental development.

Furthermore, computational drug repurposing methodologies offer an efficient approach to identify potential therapeutic candidates for neurological disorders. By leveraging ML models and high-throughput datasets, researchers can predict the efficacy of existing drugs for repurposing in treating neurological conditions. These studies have demonstrated the effectiveness of tree-based ensembled models, such as random forest and extratrees, in accurately predicting BBB permeability, thereby facilitating the selection of candidate drugs for repurposing efforts.

In summary, the integration of computational modeling, ML algorithms, advanced biotechnological techniques, and nanotechnology offers new opportunities for understanding and predicting drug permeability across physiological barriers like the placenta and the BBB. These interdisciplinary approaches hold promise for revolutionizing drug discovery and development processes, personalized medicine, and targeted therapeutic interventions for complex diseases. By combining computational predictions with experimental validation, researchers can accelerate the identification of safe and effective therapeutic options for pregnancy-related complications and neurological disorders.

CHAPTER 3
EXISTING SYSTEM

The existing system in this project involves the development and application of computational algorithms to predict the permeability of molecules across the placental barrier. This system is designed to address the limitations of traditional drug testing methods, particularly in pregnant populations, where ethical and practical constraints limit the ability to conduct extensive experimental studies. By leveraging publicly available datasets and employing machine learning techniques, researchers aim to create predictive models that can estimate the likelihood of a molecule crossing the placental barrier and reaching the fetus.

The project begins by collecting and curating relevant datasets containing information on drug properties and their permeability across the placental barrier. These datasets serve as the foundation for training and evaluating machine learning models. Various machine learning algorithms are then applied to the datasets, including K-nearest neighbor (KNN), support vector classifier (SVC), and multi-layer perceptron (MLP), among others. These algorithms are chosen based on their suitability for classification tasks and their ability to handle complex datasets.

In addition to exploring different machine learning algorithms, the project also investigates various fingerprinting methods and toolkits to represent the chemical structures of the molecules. Fingerprinting techniques capture important molecular features that may influence their ability to cross the placental barrier. By comparing the performance of different fingerprinting methods, researchers aim to identify the most effective approach for predicting permeability.

To ensure the robustness and reliability of the predictive models, the project includes comprehensive analyses of the datasets to understand their diversity and potential biases. Data preprocessing techniques may be applied to address any imbalances or inconsistencies in the datasets, ensuring that the models are trained on high-quality data.

Once the predictive models are developed and validated, they are used to predict the permeability of a chosen set of drugs across the placental barrier. These predictions

provide valuable insights into the potential toxicological effects of drugs on the fetus and help inform clinical decision-making regarding drug use during pregnancy. Drugs that are predicted to have low permeability may be considered safer choices for pregnant individuals, while those with high permeability may warrant closer monitoring or alternative treatment options.

Overall, the existing system in this project combines advanced computational techniques with comprehensive dataset analysis to predict the permeability of molecules across the placental barrier. By leveraging machine learning algorithms and chemical fingerprinting methods, researchers aim to develop accurate and reliable predictive models that can contribute to safer pharmacotherapy during pregnancy.

3.1 Drawbacks of the existing system

- **Data Availability and Quality**

  The success of machine learning models heavily relies on the availability and quality of data. In this case, publicly available datasets may be limited in size and scope, leading to potential biases or gaps in the training data. Additionally, the quality and reliability of the data may vary, affecting the performance of the predictive models.
- **Complexity of Biological Systems**

  The placental barrier is a complex biological system influenced by numerous factors, including maternal and fetal physiology, placental structure, and drug properties. Computational models may oversimplify or overlook certain aspects of this complexity, leading to inaccuracies in predictions. For example, the interaction between drugs and placental transporters or metabolic enzymes may not be fully captured by the models.
- **Ethical Considerations**

  While computational models offer an alternative to animal testing, ethical considerations regarding the use of human data and the potential risks associated with drug exposure during pregnancy must be carefully considered. The reliance on predictive models may raise concerns about the safety and efficacy of drugs, particularly in vulnerable populations such as pregnant individuals and fetuses.

CHAPTER 4

PROPOSED SYSTEM

Safeguarding the health of a developing fetus during pregnancy presents a unique challenge when it comes to testing medications. Traditional drug testing methods often raise ethical concerns due to the involvement of pregnant women. The placenta, a vital organ acting as a protective barrier, filters substances entering the fetus from the mother's bloodstream. While this filtration system shields the fetus from harmful elements, certain drugs can permeate this barrier, potentially impacting fetal development. To address this challenge, our project proposes a groundbreaking machine learning system designed to predict a drug's permeability across the placenta.

This system tackles the limitations of conventional drug testing in pregnant women by leveraging the power of machine learning. We will utilize publicly available datasets that contain information on various drugs and their documented placental permeability (whether they can cross the barrier or not). The first step involves meticulously preparing this data to ensure its quality and consistency. This might involve addressing missing data points, standardizing data formats, and potentially even applying techniques known as feature engineering to extract even more relevant information from the data.

With the data prepared, we delve into the exciting realm of machine learning models. Our exploration will be comprehensive, encompassing models that have shown promise in previous research (K-nearest neighbor, Support Vector Classifier, Multi-layer Perceptron) alongside a range of additional models boasting impressive accuracy (XGBoost, Decision Trees, Random Forest, Gaussian Naive Bayes, Logistic Regression). To further enhance the model's learning process, we might also experiment with various methods of mathematically representing the drug data, known as fingerprints and toolkits.

Once we have a suite of trained models, we can put their abilities to the test. Each model will be evaluated using a separate set of drugs to assess its effectiveness in accurately classifying whether a drug can permeate the placenta. The champion of this evaluation will be the model demonstrating the highest accuracy in predicting permeability. This victor will then be integrated into our user-friendly system.

The final system will be designed for ease of use. Users will be able to input information about new drugs, and the chosen machine learning model will then predict the likelihood of these drugs crossing the placental barrier. This offers a significant advantage for researchers, allowing them to assess the potential impact of new drug candidates on fetal development before extensive clinical trials. Healthcare professionals can also benefit from this system by having access to a valuable tool that can aid in making informed decisions regarding medication use during pregnancy.

It's important to remember that ethical considerations are paramount throughout this project's development. The system will be built upon publicly available datasets that have been ethically sourced. Additionally, we will ensure clear communication regarding the system's limitations, emphasizing the crucial role of real-world testing alongside the predictions generated by the model.

In conclusion, this project presents a groundbreaking approach to predicting drug permeability across the placenta using machine learning. This system has the potential to be a game-changer in our understanding of drug safety during pregnancy. By providing researchers and healthcare professionals with a powerful tool, we can pave the way for the development of safer and more effective medications for pregnant women, ultimately contributing to improved maternal and fetal health outcomes.

4.1 Merits of the Proposed system

- **Enhanced Prediction Accuracy**
  Utilizes machine learning algorithms and data-driven techniques to accurately predict drug permeability across the placenta, potentially outperforming traditional experimental methods.

- **Efficiency and Cost-effectiveness**
  Offers a more efficient and cost-effective means of assessing drug safety during pregnancy compared to lengthy and expensive clinical trials, saving time and resources.

- **User-friendly Interface**
  Designed to be user-friendly, allowing researchers and healthcare professionals to easily input drug information and obtain predictions on placental permeability, facilitating informed decision-making.

CHAPTER 5

SYSTEM DESIGN ARCHITECTURE

5.1 ARCHITECTURE DIAGRAM

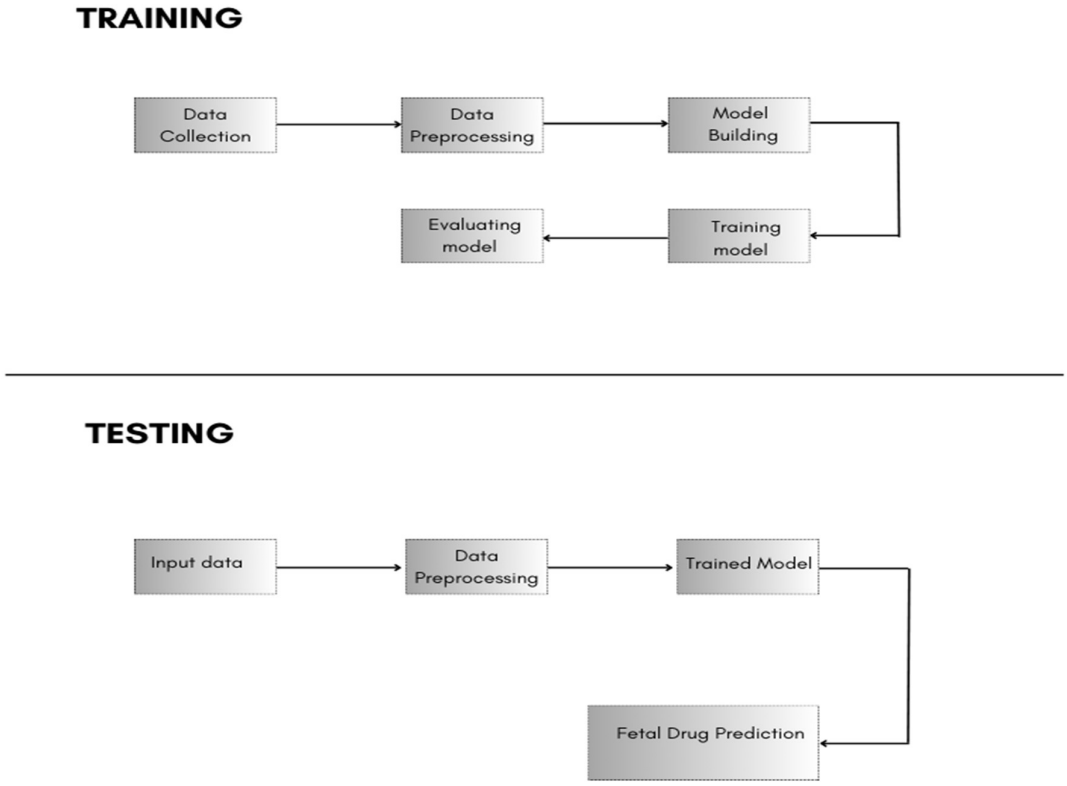
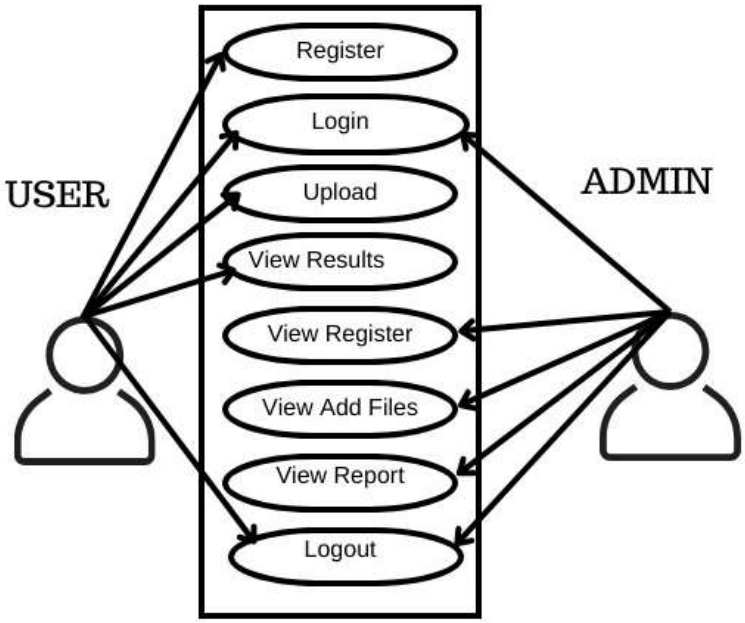
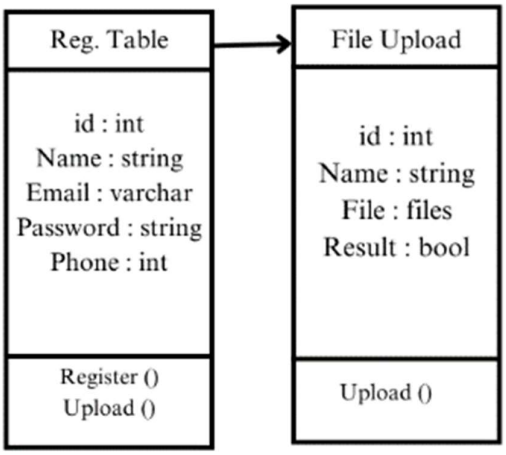


Fig 5.1 Proposed System

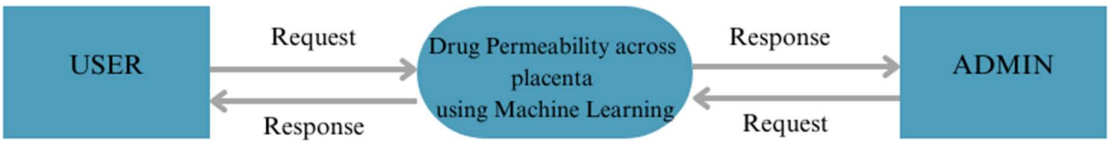
5.2 USE CASE DIAGRAM



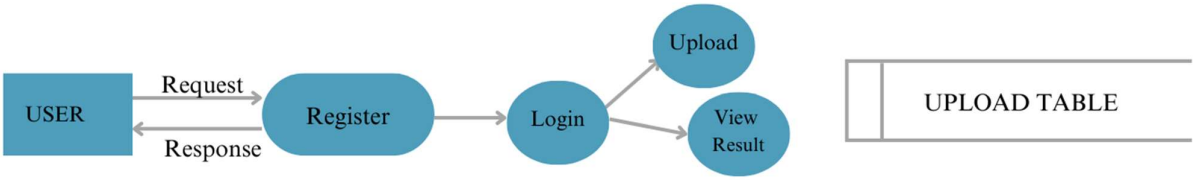
5.3 Class Diagram



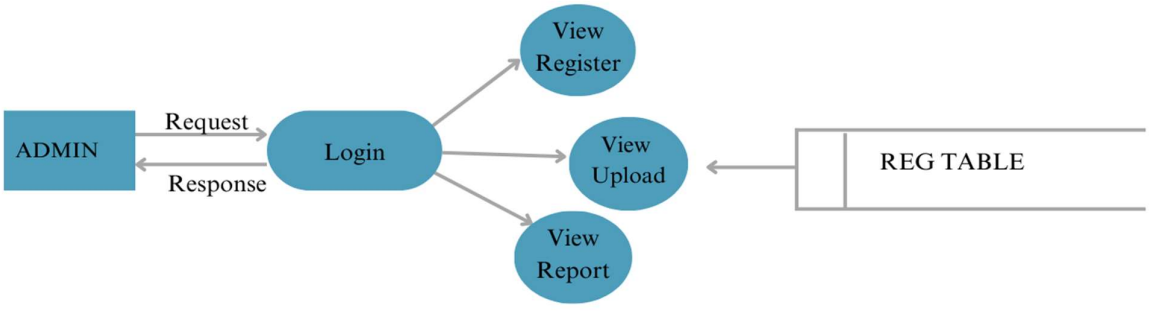
5.4 DFD LEVEL 0



5.5 DFD LEVEL 1- USER



5.6 DFD LEVEL 1- ADMIN



CHAPTER 6

SYSTEM REQUIREMENTS

6.1 SOFTWARE REQUIREMENTS

(i) Operating System

Windows10 or above

(ii) IDE

Notepad ++

(iii) Front End

- HTML
- CSS
- Js

(iv) Back End

- Python
- MYSQL
- Django

(v) Tool kit

XAMPP

6.2 Hardware Requirements

(i) System

- Processor: Intel Core i3 – 3220 (3.3 Ghz) or above
- Ram: 4 GB or above
- Storage: 512 GB or above
- Other Keyboard and Mouse

CHAPTER 7

MODULE DESCRIPTION

(i) Admin Module

- Login/Logout
- View Report
- View Register
- View Add files

(ii) User Module

- Register
- Login/Logout
- Upload
- View Result

(iii) Training Module

The algorithms used are listed below:

- XGBOOST
- GaussianNB
- Multilayer Perceptron (MLPC)
- K-Nearest Neighbour
- Support Vector Classifier (SVC)
- Decision Trees
- Random Forest
- Logistic Regression

The Algorithms and their accuracy score:

XGBOOST	99%
GaussianNB	96%
Multilayer Perceptron (MLPC)	83%
K-Nearest Neighbour (KNN)	98%
Support Vector Classifier (SVC)	92%
Decision Trees	99%
Random Forest	99%
Logistic Regression	92%

(iv) Testing Module

- Model Evaluation

CHAPTER 8
IMPLEMENTATION

The implementation of the project "Drug Permeability Across Placenta Using Machine Learning" involves leveraging various machine learning models to predict the permeability of drugs across the placental barrier. These models include XGBoost, Gaussian Naive Bayes (GaussianNB), Multilayer Perceptron (MLP), k-Nearest Neighbors (KNN), Support Vector Classifier (SVC), Decision Trees, Random Forest, and Logistic Regression.

To facilitate the utilization of these models and provide a user-friendly interface, a website named "The Placenta Permeability Hub" has been developed using Django, a high-level Python web framework. The website serves two primary user roles: administrators and regular users.

User Functionality:

- ✓ Registration: Users can register themselves on the website by providing necessary details such as username, email, and password.
- ✓ Login/Logout: Registered users can securely log in and out of the system using their credentials.
- ✓ File Upload: Users can upload CSV files containing relevant data for drug permeability analysis. These files likely include information such as drug properties and experimental conditions.
- ✓ View Results: After uploading files, users can view the results generated by the machine learning models. These results may include predictions of drug permeability and associated confidence scores.

Admin Functionality:

- ✓ Login/Logout: Administrators have secure access to the system through their unique credentials.
- ✓ View Reports: Admins can view comprehensive reports summarizing the performance of the machine learning models. These reports may include metrics such as accuracy, precision, recall, and F1-score for each model.
- ✓ User Management: Admins can manage user accounts, including registration and deletion of users if necessary.
- ✓ File Management: Admins have the ability to add files to the system, which may include updating datasets or adding new data sources for model training.

Implementation Details:

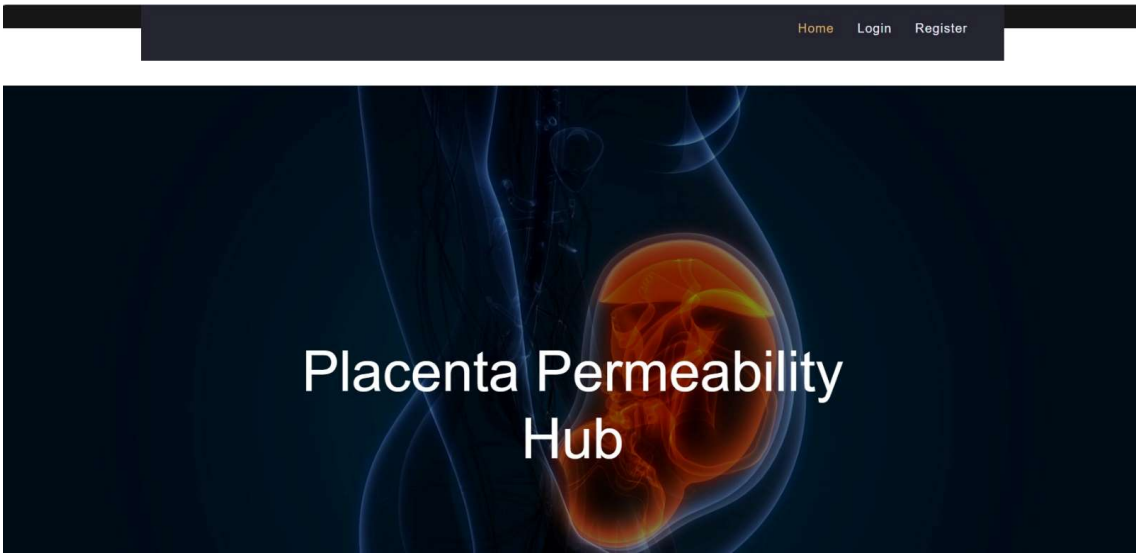
- ✓ Django Framework: The website is built using Django, which provides a robust and scalable foundation for web development in Python.
- ✓ Machine Learning Integration: Machine learning models trained on drug permeability data are integrated into the Django backend. These models are invoked upon user file uploads to generate predictions.
- ✓ Security Measures: The system ensures secure user authentication and data transmission to protect sensitive information.
- ✓ Responsive Design: The website features a responsive design, ensuring optimal user experience across various devices and screen sizes.
- ✓ Error Handling: Comprehensive error handling mechanisms are implemented to provide users with meaningful feedback in case of errors during file upload or other interactions.

By providing a centralized platform like "The Placenta Permeability Hub," researchers and clinicians can efficiently analyse drug permeability across the placental barrier, aiding in the development of safer and more effective medications for pregnant individuals.

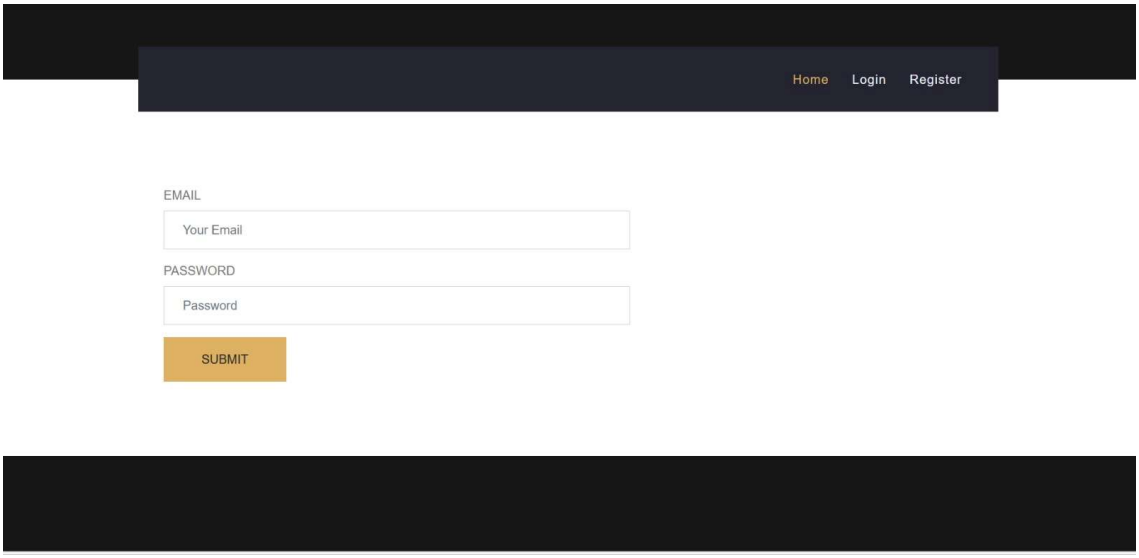
Below are the steps to utilize the features of "The Placenta Permeability Hub" website effectively:

(i) The User Functionality

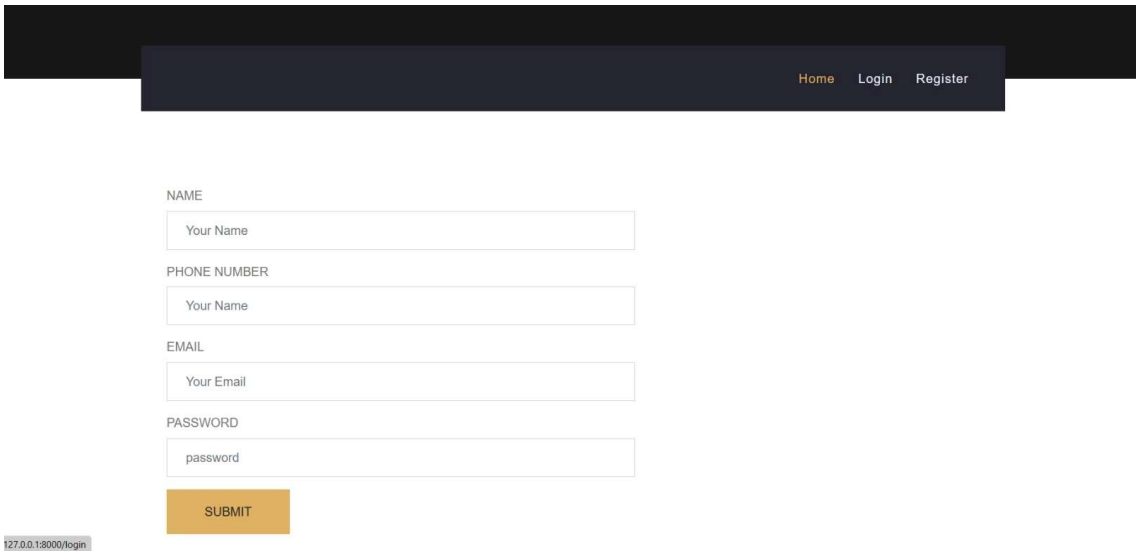
1. Click on the URL: http://127.0.0.1:8000/
2. Upon clicking the URL, users will be directed to the website's landing page, where they will encounter navigation options such as Home, Login, and Register.



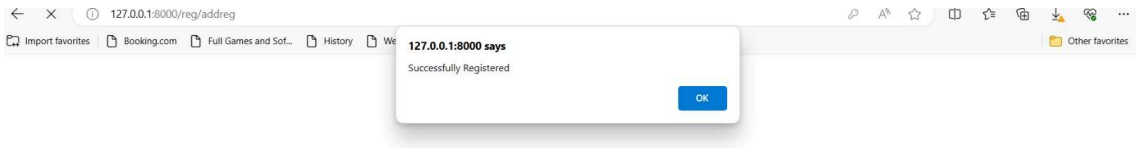
3. The login page displays the option to login by entering the email id and password.



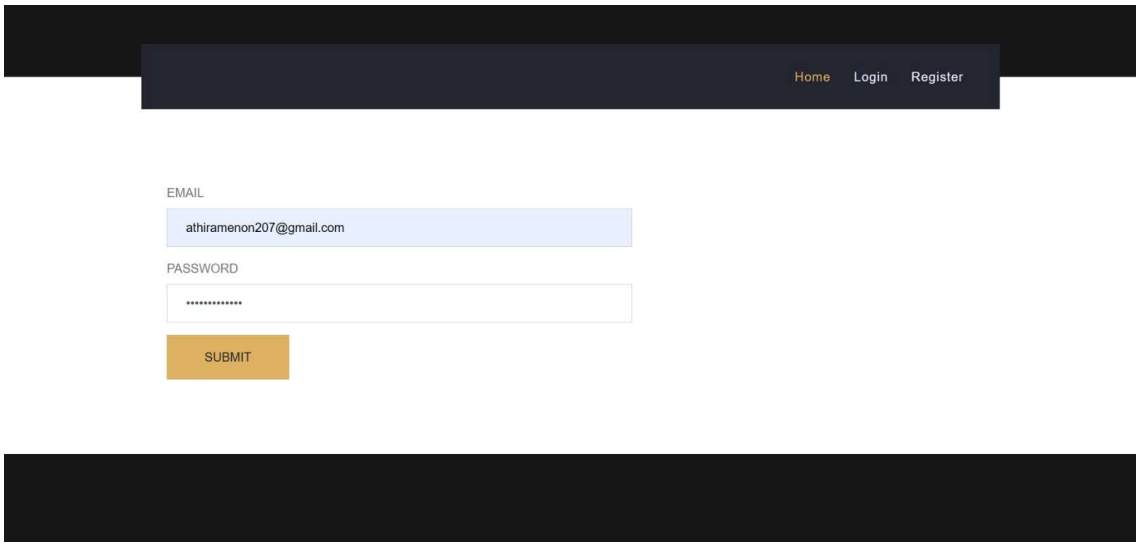
4. For new users, the “Register” option can be selected where they are required to register themselves using their name, phone number, email address, and password.



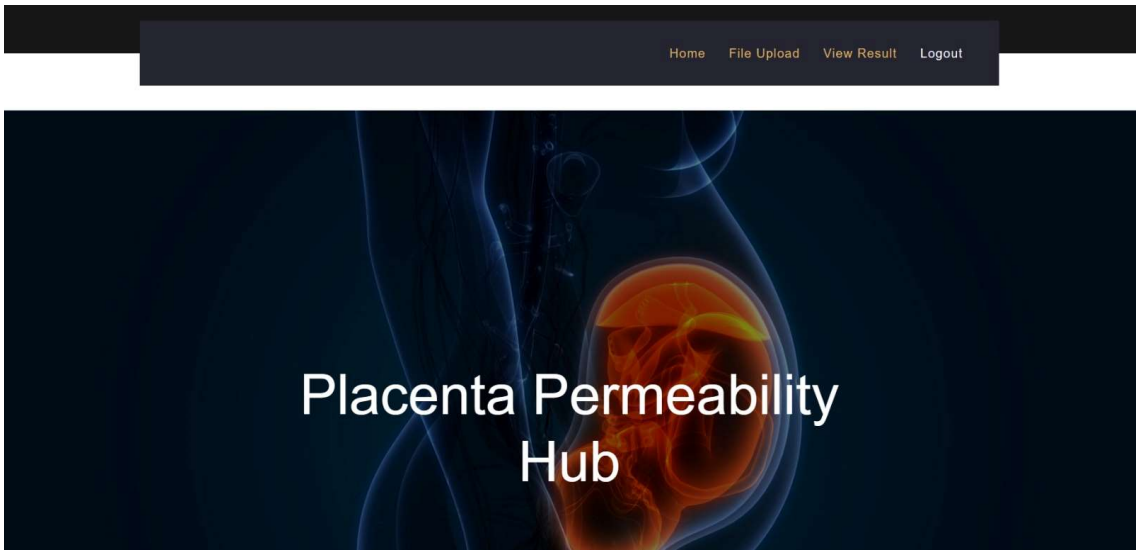
5. After entering the details, a “successfully registered” alert is displayed to the users.



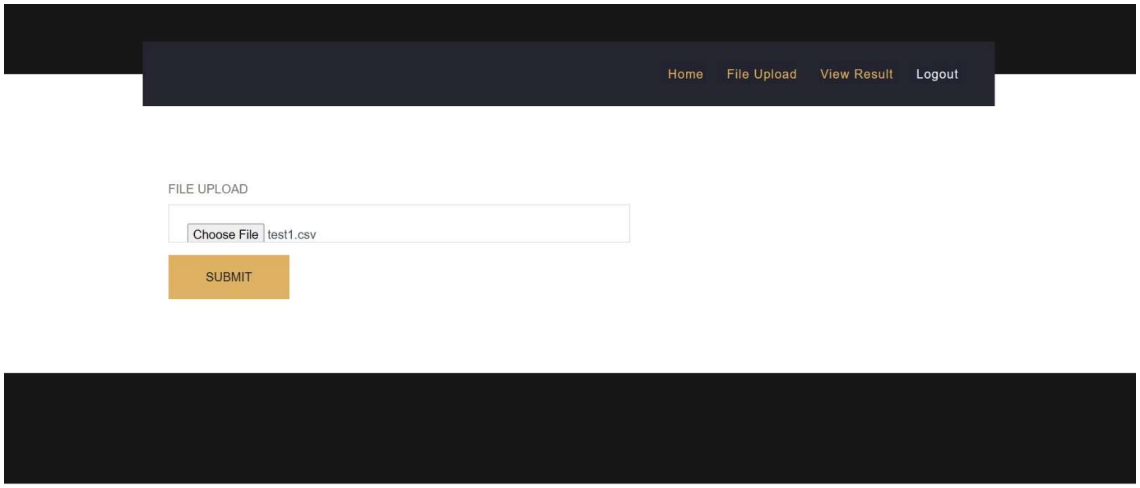
6. Now, the registered users can enter their email id and password to login.



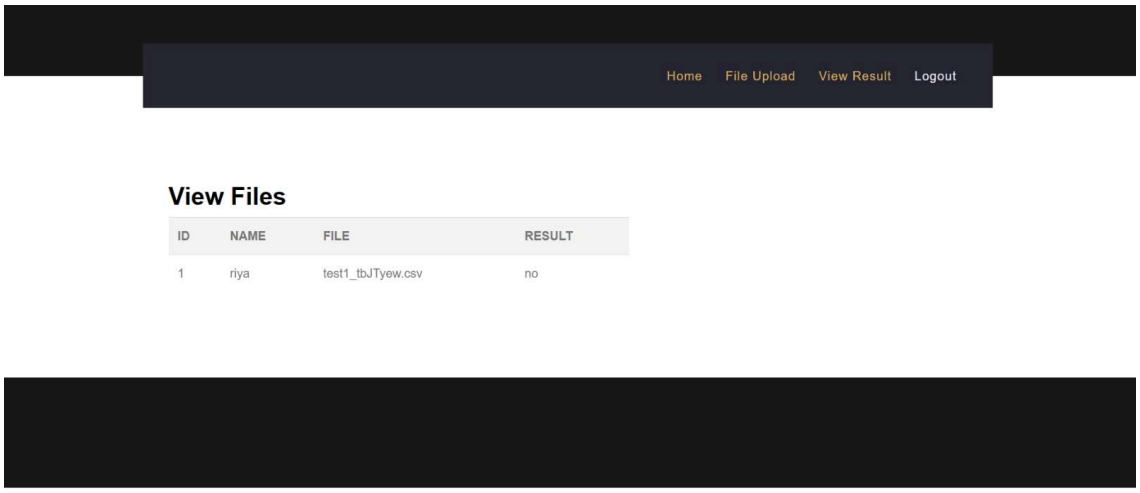
7. After logging into the system, the login page displays several options such as Home, File Upload, View Results, and Logout.



8. Next, select the "File Upload" option to upload your CSV file, then click on the submit button.

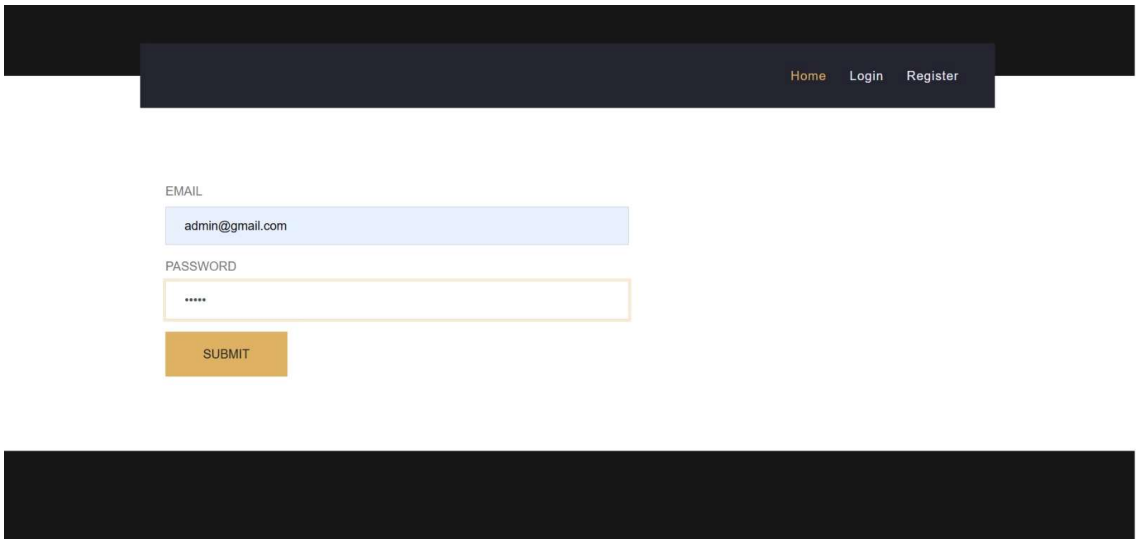


9. After submitting the CSV file, select the "View Result" option to check the outcome of the file you've uploaded.

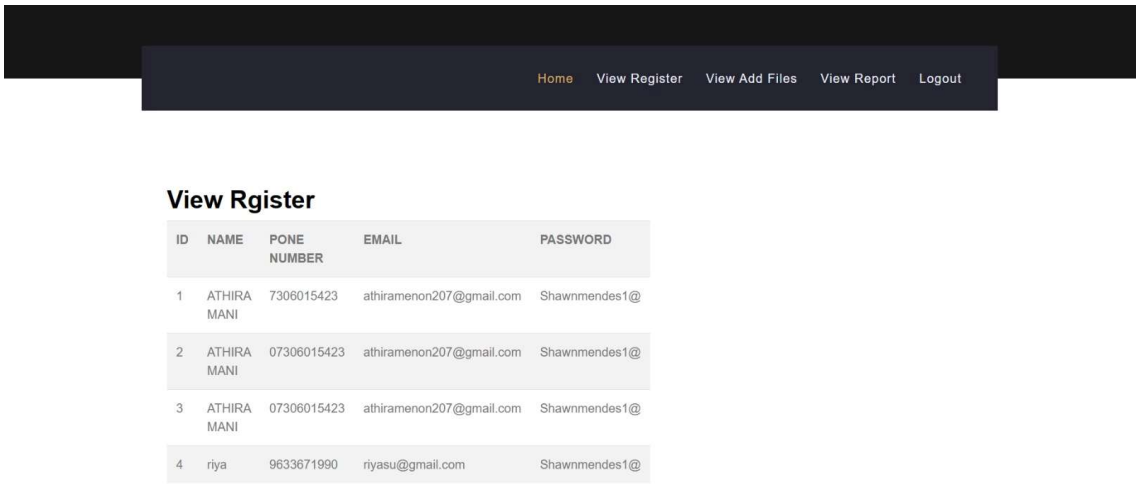


(ii) The Admin Functionality

1. The admins can log in to the system using their email address and password.



2. After the Admin logs into the system, the login page presents options like Home, View Register, View Add Files, View Reports and Logout. Click on 'View Register' to see the list of registered users.



3. Next, click on the “View Reports” to see the reports.

CHAPTER 9
CONCLUSION

In our study on drug permeability across the placenta, we addressed the critical need for alternative approaches to traditional animal testing due to limited testing possibilities in pregnant populations. Our focus was on developing computational algorithms to predict the fate of drugs in the placental barrier, thereby determining their potential toxicological effects on the fetus.

Using publicly available datasets, we employed various machine learning models to comprehensively analyze the data diversity and identify the most suitable models for prediction. Our study demonstrated the effectiveness of neural network-based models in predicting drug permeability across the placental barrier.

Among the models tested, XGBOOST, Decision Trees, and Random Forest emerged as the best-performing models, achieving a prediction accuracy of 99.2%. Additionally, other models such as KNN, GaussianNB, and SVC exhibited high accuracy scores ranging from 92.25% to 98.45%. However, the MLPC model showed the lowest accuracy at 83.72%.

Furthermore, our analysis of the chosen set of drugs revealed interesting findings. Drugs like Aliskiren, certain insulin secretagogues, and glucocorticoids were predicted to have negative permeability across the placenta. This suggests that these drugs may have limited passage through the placental barrier, potentially reducing their impact on fetal development.

In conclusion, our study highlights the importance of computational algorithms in predicting drug permeability across the placental barrier. By leveraging machine learning models, we provide valuable insights into the potential toxicological effects of drugs on the fetus, ultimately contributing to safer pharmaceutical development for pregnant individuals.

APPENDICES

(i) Training Code

```
# coding: utf-8

# In[3]:



import pandas as pd
df=pd.read_csv('drug.csv')
df.head()



# In[4]:



df.isna().sum()



# In[5]:



df.shape



# In[6]:



df.nunique()
```

```
# In[7]:


print(df.columns)


# In[8]:


df.info()


# In[9]:


maps={}
def find_category_mappings(df,variable):
    return{k:i for i,k in enumerate(df[variable].unique())}
def integer_encode(df,variable,ordinal_mapping):
    df[variable]=df[variable].map(ordinal_mapping)
for variable in ["Drug (Commercial)","Cross","Cla","F/M"]:
    mappings=find_category_mappings(df,variable)
    integer_encode(df,variable,mappings)
    maps[variable]=mappings



# In[10]:


with open('map.txt', 'w', encoding='utf-8') as f:
```

```
    f.write(str(maps))


# In[11]:


df.isna().sum()


# In[12]:


y=df.iloc[:,-1:]
y.head()


# In[13]:


x=df.iloc[:,:-1]
x.tail()


# In[14]:


y=df['Cross']
from collections import Counter#checking the dataset balance
Counter(y)
```

In[15]:

```
from imblearn.over_sampling import RandomOverSampler
ros = RandomOverSampler()
xres,yres = ros.fit_resample(x, y)
```

In[16]:

```
Counter(yres)
```

In[17]:

```
from sklearn.model_selection import train_test_split
xtrain,xtest,ytrain,ytest=train_test_split(xres,yres,train_size=0.6)
```

In[18]:

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier
from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.svm import SVC
from sklearn.neural_network import MLPClassifier
```

```
from sklearn.naive_bayes import GaussianNB
import xgboost as xgb


# In[19]:


clf_xgb = xgb.XGBClassifier()
clf_gnb = GaussianNB()
clf_mlpc = MLPClassifier()
clf_svc=SVC()
clf_knn = KNeighborsClassifier()
clf_dt=DecisionTreeClassifier()
clf_rf=RandomForestClassifier()
clf_lr=LogisticRegression()


# In[20]:


clf_xgb.fit(xtrain,ytrain)
clf_gnb.fit(xtrain,ytrain)
clf_mlpc.fit(xtrain,ytrain)
clf_svc.fit(xtrain,ytrain)
clf_knn.fit(xtrain,ytrain)
clf_dt.fit(xtrain,ytrain)
clf_rf.fit(xtrain,ytrain)
clf_lr.fit(xtrain,ytrain)


# In[21]:
```

```
from sklearn.metrics import accuracy_score,classification_report,confusion_matrix


# In[22]:



ypred_xgb=clf_xgb.predict(xtest)
ypred_gnb=clf_gnb.predict(xtest)
ypred_mlpc=clf_mlpc.predict(xtest)
ypred_knn=clf_knn.predict(xtest)
ypred_svc=clf_svc.predict(xtest)
ypred_dt=clf_dt.predict(xtest)
ypred_rf=clf_rf.predict(xtest)
ypred_lr=clf_lr.predict(xtest)


# In[23]:



acc_xgb=accuracy_score(ytest,ypred_xgb)
acc_gnb=accuracy_score(ytest,ypred_gnb)
acc_mlpc=accuracy_score(ytest,ypred_mlpc)
acc_knn=accuracy_score(ytest,ypred_knn)
acc_svc=accuracy_score(ytest,ypred_svc)
acc_dt=accuracy_score(ytest,ypred_dt)
acc_rf=accuracy_score(ytest,ypred_rf)
acc_lr=accuracy_score(ytest,ypred_lr)
```

In[24]:

```
cr_xgb=classification_report(ytest,ypred_xgb)
cr_gnb=classification_report(ytest,ypred_gnb)
cr_mlpc=classification_report(ytest,ypred_mlpc)
cr_knn=classification_report(ytest,ypred_knn)
cr_svc=classification_report(ytest,ypred_svc)
cr_dt=classification_report(ytest,ypred_dt)
cr_rf=classification_report(ytest,ypred_rf)
cr_lr=classification_report(ytest,ypred_lr)
```

In[25]:

```
cm_xgb=confusion_matrix(ytest,ypred_xgb)
cm_gnb=confusion_matrix(ytest,ypred_gnb)
cm_mlpc=confusion_matrix(ytest,ypred_mlpc)
cm_knn=confusion_matrix(ytest,ypred_knn)
cm_svc=confusion_matrix(ytest,ypred_svc)
cm_dt=confusion_matrix(ytest,ypred_dt)
cm_rf=confusion_matrix(ytest,ypred_rf)
cmlr=confusion_matrix(ytest,ypred_lr)
```

In[26]:

```
print("**********ACCURACY SCORE**********\n")
print("\nXGBOOST Accuracy Score:",acc_xgb)
print("\nGaussianNB Accuracy Score:",acc_gnb)
print("\nMLPC Accuracy Score:",acc_mlpc)
print("\nKNN Accuracy Score:",acc_knn)
print("\nSVC Accuracy Score:",acc_svc)
print("\nDECISION TREE Accuracy Score:",acc_dt)
print("\nRANDOM FOREST Accuracy Score:",acc_rf)
print("\nLogisticRegression Accuracy Score:",acc_lr)



# In[27]:



print("**********CLASSIFICATION
REPORT**********\n")

print("\nXGBOOST Classification Report:\n",cr_xgb)
print("\nGaussianNB Classification Report:\n",cr_gnb)
print("\nMLPC Classification Report:\n",cr_mlpc)
print("\nKNN Classification Report:\n",cr_knn)
print("\nSVC Classification Report:\n",cr_svc)
print("\nDECISION TREE Classification Report:\n",cr_dt)
print("\nRANDOM FOREST Classification Report:\n",cr_rf)
print("\nLogisticRegression Classification Report:\n",cr_lr)



# In[28]:



print("**********CONFUSION MATRIX**********\n")
```

```
print("\nXGBOOST Confusion Matrix:\n",cm_xgb)
print("\nGaussianNB Confusion Matrix:\n",cm_gnb)
print("\nMLPC Confusion Matrix:\n",cm_mlpc)
print("\nKNN Confusion Matrix:\n",cm_knn)
print("\nSVC Confusion Matrix:\n",cm_svc)
print("\nDECISION TREE Confusion Matrix:\n",cm_dt)
print("\nRANDOM FOREST Confusion Matrix:\n",cm_rf)
print("\nLogisticRegression Confusion Matrix:\n",cm_rf)


# In[29]:


import joblib
joblib.dump(clf_xgb,"xgb.pkl")
joblib.dump(clf_gnb,"gnb.pkl")
joblib.dump(clf_mlpc,"mlpc.pkl")
joblib.dump(clf_knn,"knn.pkl")
joblib.dump(clf_svc,"SVC.pkl")
joblib.dump(clf_dt,"Decision.pkl")
joblib.dump(clf_rf,"Random.pkl")
joblib.dump(clf_rf,"LogisticRegression.pkl")


# In[30]:


df.iloc[2:3].to_csv("test1.csv")
df.iloc[230:231].to_csv("test2.csv")
```

(ii) Testing Code

```
# coding: utf-8

# In[32]:



import pandas as pd
import numpy as np
import joblib
dd=pd.read_csv('test2.csv')
dd



# In[37]:



df=dd.iloc[:,1:-1]
df



# In[38]:



df.isna().sum()



# In[39]:



df.info()
```

```
# In[40]:


import numpy as np

with open('map.txt', 'r', encoding='utf-8') as f:
   content = f.read()

content = content.replace("nan", "np.nan")
maps = eval(content)


# In[41]:


model=joblib.load("Decision.pkl")


# In[42]:


pred=model.predict(df)
pred[0]


# In[43]:
```

```
if pred[0] == 1:
   print("yes")
else:
   print("no")
```

REFERENCES

[1] 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.

[2] 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

[3] 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 manoeuvring, and targeting of the barriers,’’ ACS Chem. Neurosci., vol. 12, no. 11, pp. 1835–1853, Jun. 2021.

[4] 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

[5] 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

[6] 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

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

[8] Y. Pu, J. Gingrich, and A. Veiga-Lopez, ‘‘A 3-dimensional microfluidic platform for modelling human extra villous trophoblast invasion and toxicological screening,’’ Lab Chip, vol. 21, no. 3, pp. 546–557, 2021.

[9] 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.

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

[11] 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.

[12] 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