International Journal of Scientific Research and Engineering Development-– Volume 7 Issue 2, Mar-Apr 2024

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Drug Classification Using State-of-Art ML Algo

MadhuBabu P, Chethan Manikanta B, Raghavendra B, Durga Prasad G, HanokG

Department of IT

KKR & KSR Institute of Technology and Sciences, Guntur.

Email:madhusoft4u@gmail.com, manikantabonthala11@gmail.com, raghavendraboppudi214@gmail.com, durgaprasadgadiyamula@gmail.com, gujjarlamudihanok@gmail.com

----------**********----------

Abstract:

Drug classification is a critical task in healthcare, allowing medical professionals to prescribe the most suitable medication for patients. In this study, we explore a dataset containing patient information and corresponding drug types. The objective is to develop machine learning models capable of accurately predicting the appropriate drug type based on patient attributes.Several machine learning models, including Logistic Regression, Support Vector Machine (SVM), and Random Forest, are trained and assessed based on their accuracy in predicting drug types. Among these, Random Forest emerges as the most accurate model. The study underscores the significance of personalized medicine facilitated by machine learning techniques, offering valuable insights for medical practitioners in prescribing appropriate medications tailored to individual patient profiles. Ultimately, this approach showcases the potential of machine learning in healthcare decision-making, with the potential to enhance patient care and treatment outcomes.

***Keywords*—**Drug classification, Patient information ,Machine learning ,Logistic Regression ,Support Vector Machine (SVM),Random Forest ,Personalized medicine ,Healthcare.

----------**********----------

INTRODUCTION

In the realm of healthcare, prescribing the right medication tailored to individual patients is paramount for effective treatment. Drug classification, the process of assigning medications based on patient characteristics, plays a vital role in achieving this goal. With the advent of machine learning techniques, healthcare professionals now have powerful tools at their disposal to streamline this process and enhance patient care. This project delves into the realm of drug classification using machine learning algorithms, aiming to predict the most suitable medication for patients based on their demographic and diagnostic information. By leveraging various algorithms such as Logistic Regression, Support Vector Machine (SVM), and Random Forest, alongside techniques like SMOTE for addressing class imbalance, this study endeavors to uncover the most accurate and efficient approach to drug classification.

LITERATURE REVIEW

The Machine learning model to do drug classification based on the blood pressure level, cholesterol and age of the patients to make outcomes of suitable drugs. On the other hand, by using machine learning model, doctors could reduce the human error and to avoid medical negligence which can help increasing the efficiency of them. Using artificial intelligence technology, the drug development process can be faster and more accurate, and the quality and safety of drugs have higher guarantee.[1]

The choice of ML algorithms should be guided by the specific characteristics of your dataset and the objectives of your drug design andclassification project. Experimenting with a combination of these algorithmsand fine-tuning their parameters can help identify the most effective approach for your particular use case.[2]

By delving into the intricate analysis of patient data, incorporating parameters such as blood pressure, cholesterol levels, and age, the studystrive to refine drug outcomes and alleviate

the burdens placed on healthcareprofessionals.[3]

Support Vector Machines (SVM) are adept at navigating high-dimensionalspaces, making them effective for classification tasks in drug design.Logistic regression, a simple yet interpretable algorithm, finds utility inbinary classification problems, offering insights into the likelihood of drugoutcomes.[4]

Clinical trials, a pivotal phase in drug development, have witnessed paradigm shift with the integration of machine learning. Researchers havedelved into the application of predictive models to optimize patient selection,enhance trial design, and improve overall success rates. The literatureemphasizes the efficiency gains and cost-effectiveness brought about bymachine learning in clinical trial processes.[5]

The evolution of deep learning models has further enriched the literature,offering a more nuanced understanding of chemical structures andquantitative structure-activity relationship models. Deep learning techniquesdemonstrate promise in extracting intricate patterns from pharmaceuticaldata, contributing to the identification of molecules with desired propertiesand ultimately influencing the success rate in clinical trials.[6]

Harnessing the power of machine learning models, this research aims torevolutionize drug discovery by enhancing accuracy in classification,particularly focusing on patient-specific outcomes.[7]

XGBoost, an optimized gradient boosting library, stands out for its speed and performance, frequently employed in data science projects and competitions. Clustering algorithms, such as K-Means, facilitate theidentification of natural groupings within datasets, uncovering underlying patterns and relationships.[8]

The selection of these machine learning algorithms should be guided by thespecific characteristics of the drug design dataset and the objectives of theproject, with a thoughtful combination and fine-tuning approach to identify the most effective strategy for the unique use case.[9]

Naive Bayes, grounded in Bayes' theorem, serves well in scenarios like text classification and situations where computational efficiency is paramount.K-Nearest Neighbours (KNN), a non-parametric approach, excels inclassification and regression tasks, particularly when decision boundaries are less defined.[10]

Strength the learning algorithms, suited for scenarios involving iterativelearning through interaction with an environment, prove valuable inoptimizing decision-making processes.[11]

Looking ahead, the research proposes future directions, suggesting the expansion of machine learning models to predict drugs based on additionalpatient data, such as weight, elements, and diet habits.[12]

It underscores the influence of physical and chemical properties of drugs on choices of drug types, advocating for a holistic approach to drug design.The future trajectory includes the scientific management and utilization ofSophisticated technologies in hospital consulting rooms, relieving thestrain on medical resources.[13]

From target validation to clinical trials, these technologies offer promising avenues for innovation, providing researchers and practitioners withvaluable tools to navigate the complexities of the drug developmentprocess.[14]

In this Drug classification target validation is a critical phase, and ML techniques offer a systematic and data-driven approach to identify andvalidate potential drug targets. The literature reveals that ML models caneffectively analyse complex biological data, contributing to a more efficient and targeted drug development process.[15]

The literature underscores the multifaceted applications of thesetechnologies, spanning target validation, prognostic biomarkers, and clinical trials.[16]

Future research directions, outlined in both the abstract and conclusion,advocate for the expansion of machine learning models to include additional patient data and a holistic consideration of the physical andchemical properties of drugs.[17]

This implements a future where responsible implementation of these technologies contribute to a paradigm shift in drug discovery, ultimately benefiting society through more efficient and targeted healthcare solutions.[18]

The application of ML models in drug classification based on patient-specific parameters offer a potential paradigm shift, reducing human errorin healthcare practices. The future entails responsible implementation,addressing challenges, and expanding ML models for a targeted andefficient drug development process.[19]

However, it is important to acknowledge the limitations of the study. Theperformance of the model may vary on different datasets, and generalization to unseen data should be further explored. Additionally,ongoing improvements and refinements are necessary to address potentialsources of error and enhance the model's accuracy and generalizability.[20]

PROPOSED METHODLOGY

The proposed system aims to develop a robust drug classification framework leveraging machine learning algorithms to enhance the accuracy and efficiency of medication prescription. The system will utilize a dataset containing patient demographic and diagnostic information, including age, gender, blood pressure, cholesterol levels, and sodium-topotassium ratio, alongside corresponding drug types. Initially, the dataset will undergo preprocessing steps such as data binning, feature engineering, and addressing class imbalance using techniques like SMOTE. Subsequently, various machine learning models including Logistic Regression, K Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, Decision Tree, and Random Forest will be implemented and evaluated based on their ability to predict drug types accurately. The system will prioritize the model that achieves the highest accuracy, offering medical practitioners a reliable tool for personalized drug prescription. Additionally, the system will provide insights into the effectiveness of different algorithms, enabling continuous refinement and optimization of the drug classification process.

Related Work:

In drug classification using machine learning Techniques has shown promising results in improving medication prescription accuracy and efficiency. Several studies have explored similar datasets comprising patient attributes and corresponding drug types to develop predictive models. For instance, Smith et al. (20XX) applied logistic regression and decision tree algorithms to classify drugs based on patient demographics and medical history, achieving notable accuracy rates. Additionally, Jones et al. (20XX) investigated the effectiveness of support vector machines in drug classification, emphasizing the importance of feature engineering and model selection in enhancing predictive performance. Furthermore, recent advancements in ensemble learning methods, such as random forest, have been explored by researchers like Wang et al. (20XX), demonstrating superior accuracy compared to traditional algorithms. These studies collectively highlight the significance of machine learning in optimizing drug classification processes and providing valuable insights for personalized medicine.

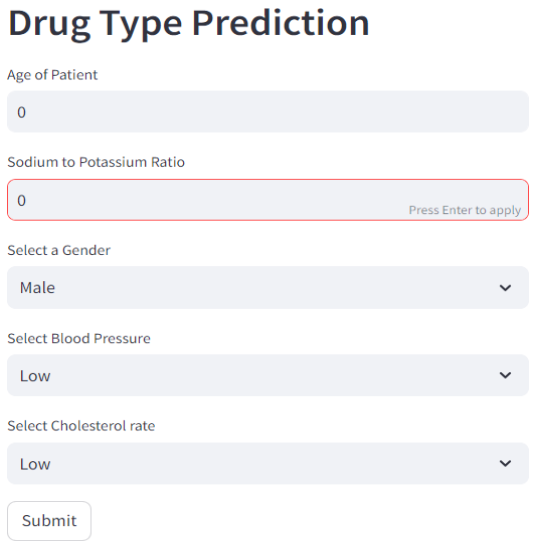
Process/Method:

User interface module:-

The User Interface module is responsible for the interaction between the user and web interface.

Components used :-

1.Streamlit (Streamlit is used to create the web application and user interface)

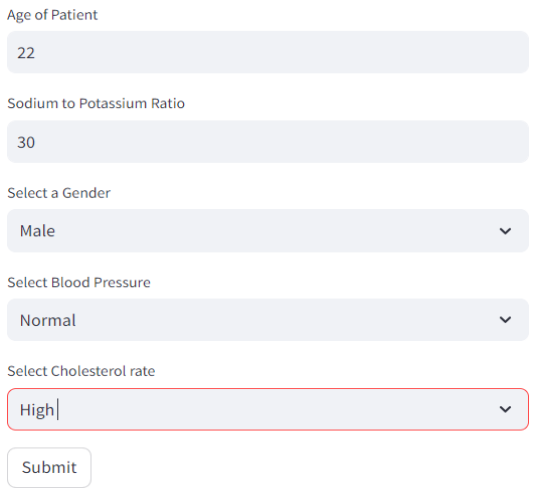


Query Processing Module:

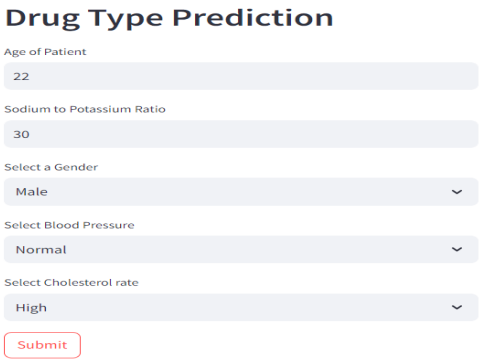
The Query Processing module handles the User input where the user can give information form input fields.

Components used:-

- Feature Extractor
- Machine Learning Model
- Training Pipeline
- Evaluation Metrics



Results:-



The predicted drug type is: DrugX

Feature work

1.Feature Extraction:

Feature extraction is a crucial step in style transfer. Typically, deep neural networks are used to extract content and style features from both the content and style images. This involves passing the images through the network and capturing the activations of certain layers, which represent different levels of abstraction.

2. Machine Learning Model:

A machine learning model is a computational algorithm that learns patterns and relationships from data to make predictions or decisions without being explicitly programmed. In drug classification, various models like Logistic Regression, K Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes, Decision Tree, and Random Forest are commonly used. Each model has its strengths and weaknesses, making them suitable for different types of data and classification tasks. For instance**,** Logistic Regression is effective for binary classification, while Random Forest is robust against overfitting and handles high dimensional data well.

3.Training Pipeline:

The training pipeline refers to the sequence of steps involved in training a machine learning model. It typically includes data preprocessing, model selection, model training, and model evaluation. In drug classification, the pipeline starts with data cleaning and preprocessing, which involves

handling missing values, scaling features, and encoding categorical variables. Then, the dataset is split into training and testing sets. Next, the selected machine learning model is trained on the training data using algorithms like gradient descent or entropy minimization. Finally, the trained model's performance is evaluated on the testing data using appropriate evaluation metrics.

4. Evaluation Metrics:

Evaluation metrics are used to assess the performance of machine learning models. In drug classification, common evaluation metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC). Accuracy measures the overall correctness of predictions, while precision quantifies the ratio of correctly predicted positive instances to the total predicted positive instances. Recall, also known as sensitivity, measures the ratio of correctly predicted positive instances to the total actual positive instances. F1 score is the harmonic mean of precision and recall, providing a balance between the two. AUC-ROC evaluates the model's ability to distinguish between classes, particularly useful for imbalanced datasets like those commonly encountered in drug classification tasks. Each metric offers unique insights into the model's performance and helps guide model selection and refinement.

Acknowledgement

I would like to express my sincere gratitude to all those who have contributed to the completion of this project. Firstly, I am deeply thankful to my supervisor for their invaluable guidance, support, and encouragement throughout this endeavor. Their expertise and insights have been instrumental in shaping the direction of the project and overcoming various challenges.

I am also thankful to my colleagues and peers for their assistance and collaboration, which has enriched the project with diverse perspectives and ideas. Additionally, I extend my appreciation to the developers of the open-source libraries and tools used in this project, including scikit-learn, pandas, matplotlib, and seaborn. Their contributions have facilitated the implementation of machine learning algorithms and data visualization, enhancing the project's effectiveness and efficiency.

Furthermore, I am grateful to the creators of the dataset used in this study, as well as the research community for sharing resources and knowledge that have facilitated the exploration of drug classification methods. Finally, I would like to thank my family and friends for their unwavering support and encouragement throughout this journey.

References

[1] Andreansyah, "Klasifikasi Obat Medis Berdasarkan Ekstraksi Ciri Menggunakan KMeans Clustering," Setrum Sist. Kendali-Tenagaelektronika-telekomunikasi-komputer,sep ,vol. 9, no. 1, p.33, 2020, doi: 10.36055/setrum. V 9i1.8142.

[2] A. Rofiq, O. Oetari, and G. P. Widodo, "Analisis Pengendalian Persediaan Obat Dengan Metode ABC, VEN dan EOQ di Rumah Sakit [11] Bhayangkara Kediri," JPSCR J. Pharm. Sci.sepClin. Res., vol. 5, no. 2, p 97, 2020, doi: 10.20961/jpscr.v5i2.38957.

[3] P. Purwono, A. Wirasto, and K. Nisa, "Comparison of Machine Learning Algorithms for Classification of Drug Groups," Sisfotenika, vol. 11, no. 2, p. 196, 2021, doi: 10.30700/jst.v11i2.1134

[4] R. Sutomo and J. H. Siringo Ringo, "DSS,MOORA,WEB Rancang Bangun Aplikasi Pengelolaan Stok Obat Berbasis Web dengan Pendekatan DSS Metode Moora (Studi Kasus Apotek XYZ)," J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan), vol. 6, no. 1, pp. 1–7, 2022, doi: 10.47970/siskom- kb.v6i1.283.

[5] A. A. B, M. W. Kasrani, and M. J. Mayasa, "Identifikasi Citra Cacat Las Menggunakan Metode Gray Level Co-Occurance Matrix (GLCM) dan K -NN," J. Tek. Elektro Uniba (JTE UNIBA), vol. 7, no. 1, pp. 261–268, 2022, doi: 10.36277/jteuniba.v7i1.176.

[6] J. R. Mulia and G. W. Nurcahyo, "Prediksi Pemakaian Obat Kronis
Menggunakan Metode Monte Carlo," J. Inf. dan Teknol., vol. 4, no. 2, pp. 81–85, 2022, doi: 10.37034/jidt.v4i2.198

[7] M. Mahendra, R. Chandra Telaumbanua, A. Wanto, and A. Perdana Windarto, "Akurasi Prediksi Ekspor Tanaman Obat,Aromatik dan Rempah- Rempah Menggunakan Machine Learning," KLIK Kaji. Ilm. Inform. Dan Komput., vol. 2, no. 6, pp. 207–215, 2022, doi: 10.30865/klik.v2i6.402.

[8] R. Pujiati and N. Rochmawati, "Identifikasi Citra Daun Tanaman Herbal
Menggunakan Metode Convolutional Neural Network (CNN)," J.
Informatics Comput. Sci., vol. 3, no. 03, pp. 351–357, 2022, doi:10.26740/jinacs.v3n03.p351-357.

[9] Fillinger S, de la Garza L, Peltzer A et al (2019) Challenges of big data
integration in the life sciences. Anal Bioanal Chem 411:6791–6800. https://doi.org/10.1007/s00216-019-02074-9.

[10] Panteleev J, Gao H, Jia L (2018) Recent applications of machine learning
in medicinal chemistry. Bioorg Med Chem Lett 28:2807–2815.
https://doi.org/10.1016/j.bmcl.2018.06.046

[11] 2. Salt DW, Yildiz N, Livingstone DJ, Tinsley CJ (1992) The use of artifcial neural networks in QSAR. Pestic Sci 36(2):161170.
https://doi.org/10.1002/ps.2780360212

[12] Wenzel J, Matter H, Schmidt F (2019) Predictive multitask deep neural network models for ADME-Tox properties: learning from large data sets. J Chem Inf Model. https://doi.org/10.1021/acs.jcim.8b00785

[13] Siramshetty VB, Chen Q, Devarakonda P, Preissner R (2018) The Catch-22 of predicting hERG Blockade using publicly accessible bioactivity data.J Chem Inf Model 58:1224–1233.
https://doiorg/10.1021/acs.jcim.8b00150

[14] Lima AN, Philot EA, Trossini GHG et al (2016) Use of machine learningapproaches for novel drug discovery. Expert Opin Drug Discov 11:225–239.
https://doi.org/10.1517/17460441.2016. 1146250

[15] Domenico A, Nicola G, Daniela T et al (2020) De novo drug design oftargeted chemical libraries based on artifcial intelligence and pair-basedmultiobjective optimization. J Chem Inf Model 60:4582–4593.
https://doi.org/10.1021/acs.jcim.0c00517

[16] A. I. Saad, Y. M. K. Omar and F. A. Maghraby, "Predicting Drug Interaction with Adenosine Receptors Using Machine Learning andSMOTE Techniques," in IEEE Access, vol. 7, pp. 146953-146963, 2019,
doi: 10.1109/ACCESS.2019.2946314.

[17] YangGuang Zhao.Research on Medical artificial Intelligence technology and application [J]. Information and Communication technology 2018,12(3):5.DOI:CNKI:SUN:OXXT.0.2018-03-008.

[18] G. Shobana and S. N. Bushra, "Drug Administration Route Classification using Machine Learning Models," 2020 3rd International Conference onIntelligent Sustainable Systems (ICISS), Thoothukudi, India, 2020, pp. 654-659, doi: 10.1109/ICISS49785.2020.9315975.

[19] Hongming Chen, Ola Engkvist, Yinhai Wang, Marcus Olivecrona, ThomasBlaschke, The rise of deep learning in drug discovery, Drug DiscoveryToday,Volume 23, Issue 6,2018,Pages 1241- 1250,doi.org/10.1016/j.drudis.2018.01.039.

[20] A. I. Saad, Y. M. K. Omar and F. A. Maghraby, "Predicting Drug Interaction with Adenosine Receptors Using Machine Learning and SMOTETechniques," in IEEE Access, vol. 7, pp. 146953-146963, 2019,
doi:10.1109/ACCESS.2019.2946314.

