IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13372

DRUG CONSUMPTION DETECTION USING MACHINE LEARNING

Ajay Shetty[1], Chirag V[2], Darshan U Shetty[3], Disha P[4], Dr. Babu Rao K[5]

Students, Dept of Computer Science and Engineering, Mangalore Institute of Technology & Engineering Karnataka, India[1,2,3,4]

Professor, Dept of Computer Science and Engineering, Mangalore Institute of Technology & Engineering Karnataka, India[5]

Abstract: Diagnosing and monitoring drug use creates important challenges in the medical and social fields. Traditional methods have relied heavily on self-reports, which can be unreliable due to various factors such as social bias and memory bias. In recent years, there has been a growing interest in using machine learning techniques to augment or replace traditional approaches to drug detection. This paper provides a comprehensive overview of the current state of the art in machine learning-based drug use diagnosis.

Describes common preprocessing steps for cleaning and preparing data for analysis, including feature extraction and dimensionality reduction techniques. We then take a closer look at various machine learning algorithms and models used for drug detection, including random forests, deep learning architectures, and ensemble techniques. We discuss the strengths and weaknesses of each approach and highlight recent advances and challenges.

Additionally, we discuss ethical considerations for using machine learning in this context, including privacy concerns, algorithmic bias, and the impact of false positives and negatives.

Finally, we identify potential avenues for future research, including developing more robust and interpretable models, integrating multiple data methods to improve accuracy, and exploring real-time monitoring systems.
Overall, this review highlights the potential of machine learning to revolutionize drug use diagnosis and highlights the importance of interdisciplinary collaboration to address the complex challenges inherent in this field.

I. INTRODUCTION

The rapid development and spread of new drugs, including illicit substances, poses a significant challenge to public health and safety. Accurate and efficient drug detection methods are crucial in a variety of settings such as forensics, clinical toxicology, and drug abuse monitoring. Traditional drug detection techniques, such as immunoassays, often rely on targeting specific classes of drugs and can suffer from limitations such as false positives and the inability to identify novel drugs. Mass spectrometry (MS) has emerged as a powerful tool for drug analysis due to its high sensitivity, specificity, and ability to identify unknown compounds. MS analyzes the mass-to-charge (m/z) ratio of molecules and provides a unique fingerprint for each compound. This fingerprint enables identification and quantification of the drugs present in the sample. However, analyzing raw MS data for drug detection can be complex and requires specialized expertise. Machine learning (ML) offers a promising approach to automate drug identification from RS data. ML algorithms can be trained on a dataset of known illegal drugs and their corresponding RS properties. The trained model can then predict the presence or absence of illegal drugs in an unknown sample based on its RS profile. This project explores the application of machine learning (ML), specifically the Random Forest classifier, in conjunction with RS to automate drug identification from mzML files. MzML is a standard MS data storage format that ensures compatibility with various software tools. Machine learning offers an efficient approach to solving problems such as spectral complexity, data volume, specificity, etc. Supervised learning algorithms can be trained on a dataset of corresponding mzML files. to various drugs. These training datasets include the M (mass) and M+proton values associated with each drug, allowing the model to learn the unique spectral signatures of different compounds. Once trained, the model can be used to analyze mzML files from unknown samples and predict the presence and identity of the drug(s).

DOI: 10.17148/IJARCCE.2024.13372

II. LITERATURE SURVEY

Chien-Hua Chiang *et al..*,[1] This study applied thermal desorption electrospray ionization/mass spectrometry (TD–ESI/MS) for the rapid analysis of regulated substances. By redistributing the mass spectrum sources, TD–ESI/MS enables direct analysis that offers both qualitative screening and quantitative confirmation without pretreatment. Analysis of the four-component drug standard via 60-MRM demonstrated selectivity with comparable results to 8-MRM analysis. Sequential analyzes of the tablets showed no cross-contamination and achieved a sampling rate of two samples per minute. Even minute drug concentrations (<2 mg g-1) were detected, indicating sensitivity. TD–ESI/MS with its switchable ion source design has potential as a pretreatment-free qualitative screening tool to complement LC–MS/MS for illicit drug analysis.

Claude Mallet *et al..*,[2] This article presents a new approach to speed up the identification of illicit drugs in forensic toxicology cases. By combining multidimensional chromatography with microextraction techniques, sample preparation time is significantly reduced while maintaining the integrity of the separation. Various illicit drugs enriched in human urine were efficiently extracted using different protocols, including 2D optimized and 2D sequential extraction modes. Using a comprehensive protocol for the development of automated methods, the extraction was completed within 20 minutes. Analysis of the extracted samples revealed a limit of detection of 100 pg/ml for all drugs, with some analytes showing robust signals even at 10 pg/ml. This method holds promise for fast and reliable forensic drug analysis.

Jeff C. Eichhorst *et al..,*[3] This article discusses advances in drug screening methods and highlights the evolution from immunoassay systems to liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Immunoassays, although widely used, lack specificity and require confirmation by GC/MS. LC-MS/MS offers higher specificity, allowing accurate identification of the drugs present. The method outlined in the paper analyzes 40 drugs using labeled internal standards, allowing rapid analysis of urine samples with short processing times. This approach holds promise for more reliable, high-throughput drug screening, potentially replacing traditional immunoassay systems in forensic laboratories

David S Wishart *et al..,*[4] The article highlights the significant progress of DrugBank, a comprehensive database of molecular drug information. DrugBank 5.0, the latest update, represents the most significant update in more than a decade and shows a substantial increase in data content, including investigational drugs, drug interactions, and drug effects associated with SNPs. In addition, new data on pharmacoetabolomics, pharmacotranscriptomics and pharmacoproteomics have been added, along with information on clinical trials and drug transformation. Improvements in data quantity, quality, and consistency, as well as improvements to the website's interface and performance, promise to increase the usefulness of DrugBank in pharmacological research, pharmaceutical science, and drug education.

Armen G Beck *et al..,*[5] This paper explores the intersection of mass spectrometry (MS) and machine learning (ML), showcasing recent advancements in ML methods, particularly artificial neural networks (ANN) and deep learning architectures. It highlights the growing significance of ML in various MS applications, including mass spectrometry imaging and proteomics. The paper offers an introductory overview of ML methodology relevant to MS, providing insights into practical aspects and recent developments. By presenting a comprehensive review and discussing future directions, it aims to elucidate the evolving landscape of ML integration with MS-based techniques, offering valuable perspectives for researchers in the field.

Hannes L. Rost *et al..,***[**6**]** This paper introduces pyOpenMS, an open-source Python interface to the C++ OpenMS library, facilitating MS-based proteomic analysis. It provides easy access to various functions including file access, signal processing and comprehensive data analysis. With Python bindings that allow raw access to OpenMS algorithms, pyOpenMS enables rapid prototyping and development of workflows, especially for non-C++ researchers. The framework supports interactive Python interpretation and comes with an autowrap tool for Cython code generation. Both pyOpenMS and autowrap are freely available under the 3-clause BSD license, which promotes accessibility and makes it easy for other projects to create similar bindings.

Fabian Pedregosa *et al..,*[7] The paper introduces scikit-learn, a Python module that integrates a variety of state-of-the-art machine learning algorithms for medium-scale supervised and unsupervised problems. It emphasizes accessibility to non-specialists through a user-friendly, high-level language while prioritizing ease of use, performance, documentation, and API consistency. With minimal dependencies and distributed under the simplified BSD license, scikit-learn encourages adoption in academic and commercial settings. Users can access source code, binaries, and documentation from http://scikit-learn.sourceforge.net, making it a versatile and widely accessible tool for machine learning applications.

DOI: 10.17148/IJARCCE.2024.13372

Lennart Martens *et al..,*[8] The article discusses the development of mass spectrometry data formats, which culminated in the development of mzML by the HUPO PSI consortium. Aiming to standardize mass spectrometry data for better sharing and analysis, mzML incorporates desirable attributes from previous formats while introducing improvements such as a controlled vocabulary and validation tools for consistent use. In particular, mzML offers improved support for selected reaction monitoring data and facilitates rapid adoption through readily available implementations. As a well-tested open source format, mzML serves as a key tool for the life sciences community and adapts effectively to advances in mass spectrometry technology.

Ravi Manne *et al..,*[9] This paper explores the key role of machine learning and deep learning techniques in revolutionizing drug discovery at various stages, including target validation, prognostic biomarkers, and clinical trials. These technologies, using large data sets from large databases, offer significant opportunities for advancement in the pharmaceutical industry. By leveraging machine learning algorithms, researchers can extract valuable insights and patterns from complex data to facilitate more efficient and effective drug development processes. In particular, deep learning techniques hold promise for improving predictive modeling and accelerating the identification of potential drug candidates. Overall, this analysis underscores the transformative impact of artificial intelligence in driving innovation in the pharmaceutical sector.

Pradeep K Sinha *et al..,*[10] This paper presents advances in Random Forest classifiers that aim to improve accuracy and reduce learning and classification time. By proposing five new approaches, including disjoint partitioning, pruning, optimal subset selection, and parallel algorithms, the research seeks to address key issues such as the strength of individual decision trees and the correlation between underlying trees. Through empirical analysis and experimentation, the study demonstrates better performance in learning and classification tasks using the Random Forest algorithm. These improvements place Random Forest as a competitive option among ensemble techniques such as bagging and boosting, and demonstrate its effectiveness in data mining applications.

III. SUMMARY AND OBSERVATION

The literature review provided highlights advances in drug detection methods using mass spectrometry (MS) and machine learning (ML). Traditional methods such as immunoassays have limitations in specificity and require confirmation using techniques such as GC/MS. Articles by Eichhorst et al. [5] and Mallet et al. [4] address these limitations by investigating liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) for drug screening. This approach offers higher specificity and enables rapid analysis with shorter processing times.

Beck et al. [7] delve into the growing synergy between MS and ML, particularly the use of artificial neural networks (ANNs) and deep learning architectures. This integration holds significant promise for various MS applications, including drug identification. The work of Rost et al. [8] presents pyOpenMS, an open-source Python interface for MS-based proteomic analysis. This framework facilitates data analysis and workflow development, especially for researchers without C++ expertise.

Finally, the work of Manne et al. [12] examines the transformative role of machine learning and deep learning in drug discovery. These techniques using large data sets present opportunities for identifying potential drug candidates and improving drug development processes. Overall, the literature survey highlights the evolving landscape of drug detection methods, with RS and ML playing an increasingly important role in accurate, efficient, and high-throughput analysis.

IV. CONCLUSION

In conclusion, this literature survey highlights the evolving landscape of drug detection methods and the growing importance of RS and machine learning in this field. Traditional techniques such as immunoassays are being superseded by more specific and high-throughput methods such as LC-MS/MS, which offer faster and more reliable drug screening. Advances in sample preparation techniques, such as multidimensional chromatography with microextraction, further increase the efficiency of drug identification in forensic toxicology. The integration of machine learning with MS holds immense promise for the future of drug analytics. Techniques such as artificial neural networks and deep learning offer powerful tools for extracting valuable insights from complex MS data, which can lead to improved drug identification accuracy and efficiency. Open source software frameworks such as pyOpenMS further contribute to this progress by making MS analysis tools available to a wider range of researchers. By taking advantage of these advances, researchers can develop more sophisticated and efficient methods for drug detection and analysis.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13372

REFERENCES

[1]. Fei Wang, Daniel Pasin, Michael A. Skinnider, Jaanus Liigand, Jan-Niklas Kleis, David Brown, Eponine Oler, Tanvir Sajed, Vasuk Gautam, Stephen Harrison, Russell Greiner, Leonard J. Foster, Petur Weihe Dalsgaard, and David S. Wishart, "Deep Learning-Enabled MS/MS Spectrum Prediction Facilitates Automated Identification Of Novel Psychoactive Substances", Analytical Chemistry 2023 95 (50), 18326-18334 DOI: 10.1021/acs.analchem.3c02413

[2]. Chien-Hua Chiang, Hei-Hwa Lee, Bai-Hsiun Chen, Yi-Ching Lin, Yu-Ying Chao, Yeou-Lih Huang. "Using ambient mass spectrometry and LC–MS/MS for the rapid detection and identification of multiple illicit street drugs." Journal of Food and Drug Analysis, Volume 27, Issue 2, 2019, Pages 439-450, ISSN 1021-9498, DOI: 10.1016/j.jfda.2018.11.003.

[3]. Claude Mallet and Sabra Botch-Jones, "Illicit Drug Analysis Using Two-Dimension Liquid Chromatography/Tandem Mass Spectrometry", Journal of Analytical Toxicology, Volume 40, Issue 8, October 2016, Pages 617–627, DOI:10.1093/jat/bkw082

[4]. Jeff C. Eichhorst, Michele L. Etter, Nadine Rousseaux, Denis C. Lehotay, "Drugs of abuse testing by tandem mass spectrometry: A rapid, simple method to replace immunoassays," Clinical Biochemistry, Volume 42, Issue 15, 2009, Pages 1531-1542, ISSN 0009-9120, DOI:10.1016/j.clinbiochem.2009.07.019.

[5]. David S Wishart, Yannick D Feunang, An C Guo, Elvis J Lo, Ana Marcu, Jason R Grant, Tanvir Sajed, Daniel Johnson, Carin Li, Zinat Sayeeda, Nazanin Assempour, Ithayavani Iynkkaran, Yifeng Liu, Adam Maciejewski, Nicola Gale, Alex Wilson, Lucy Chin, Ryan Cummings, Diana Le, Allison Pon, Craig Knox, Michael Wilson, "DrugBank 5.0: a major update to the DrugBank database for 2018", Nucleic Acids Research, Volume 46, Issue D1, 4 January 2018, Pages D1074–D1082, DOI:10.1093/nar/gkx1037

[6]. Röst, H.L.; Schmitt, U.; Aebersold, R.; Malmström, L. pyOpenMS: A Python-based interface to the OpenMS mass-spectrometry algorithm library. Proteomics 2014, 14, 74–77. DOI: 10.1002/pmic.201300246

[7]. Armen G. Beck, Matthew Muhoberac, Caitlin E. Randolph, Connor H. Beveridge, Prageeth R. Wijewardhane, Hilkka I. Kenttämaa, and Gaurav Chopra, "Recent Developments in Machine Learning for Mass Spectrometry", ACS Measurement Science Au 0, 0, pp, DOI: 10.1021/acsmeasuresciau.3c00060

[8]. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay, "Scikit-learn: Machine Learning in Python" J. Mach. Learn. Res. 12, null (2/1/2011), 2825–2830.

[9]. Martens, Lennart & Chambers, Matthew & Sturm, Marc & Kessner, Darren & Levander, Fredrik & Shofstahl, Jim & Tang, Wilfred & Römpp, Andreas & Neumann, Steffen & Pizarro, Angel & Montecchi-Palazzi, Luisa & Tasman, Natalie & Coleman, Mike & Reisinger, Florian & Souda, Puneet & Hermjakob, Henning & Binz, Pierre-Alain & Deutsch, Eric. (2011). "mzML—a Community Standard for Mass Spectrometry Data". Molecular & cellular proteomics : MCP. 10. R110.000133. 10.1074/mcp.R110.000133.

[10]. Manne, Ravi. (2021). Machine Learning Techniques in Drug Discovery and Development. 7. 21-28. 10.22271/allresearch.2021.v7.i4a.8455.DOI: 10.22271/allresearch.2021.v7.i4a.8455