Assignment 1 – Literature Review [2,174 words]

The World Health Organization defines Adverse Drug Reaction (ADR) as “a response to a drug that is noxious and unintended, and which occurs at doses normally used” (World Health Organization, 2020). Many research projects are undertaken to ascertain whether newly discovered drugs can have negative side-effects. Several factors in addition to taking a pill can result in adverse health conditions and it is impossible to keep all the non-drug factors constant. These studies can benefit from the application of machine learning and deep learning, which is especially useful in situations where multiple factors can affect an outcome (Tukey, 1977; Bruce, 2020: 30-42). A literature review was conducted on papers dealing with machine learning (ML) and deep learning (DL) approaches to ADR and is presented here.

Searching Google Scholar for the words: Adverse Drug Reaction Machine Learning produced 249,000 results. Grouping words together: “Adverse Drug Reaction” AND “Machine Learning” produced 6,270. Specifying “Review Articles” and Since 2021 narrowed it down to 668 results. As one of our textbooks warns, “a literature search… usually ‘snowballs’ as it continues” (Dawson, 2015: 68). Therefore, I have selected some promising titles from the 668 results, meaning that this review cannot be exhaustive.

I excluded papers I could not access, only using ones I could download (see references below) and for each one I made comments about the author(s) in small print, enclosed by square brackets, to be deleted in the final version. I limited the papers to those written by authors at named universities or colleges. Some papers had two entries in Google Scholar, one for a pre-print which was not peer-reviewed and the other for a later version in a journal. I discarded the pre-publication versions.

The ten papers I was left with can be divided roughly into review studies and research projects, five of each, with some overlap. The reviews, which all have similar titles involving ML and ADR because of my search, included Basnet (2023), Hu et al. (2024), Kim et al. (2022), and Lee & Chen (2019). One paper, Nguyen, D.A. et al. (2021), used a survey. Because Hu et al. (2024) performed a statistical analysis on data drawn from multiple papers, it is also of the meta-analysis type. The five research projects have more specialised titles. Two of them, Dey et al. (2018) and Dong (2021), are experimental projects, whereas the other three, Anjali & Ravi Kumar (2022), Islam et al. (2018), and McMaster et al. (2019) are empirical by nature but examine the results of other projects.

The papers covered different aspects of ADR analysis, each having an emphasis on one or more of: data acquisition from electronic health records (EHRs) and from social media; statistical methods and the use of machine learning and deep learning; and practical applications of these analyses.

Data acquisition from EHRs was an important theme in Hu et al. (2024), which is itself a review of ten papers on research into ML applied to EHRs for ADR analysis. The authors have a medical and academic background, being at West China Hospital of Sichuan University, so this can be considered a serious source. The use

of EHRs is promoted, as they are in the form of structured data, which can (relatively) easily be engineered into features which can be ingested into ML systems for training, testing and eventually predicting adverse drug reactions. By its meta-analysis of statistically aggregating results, this paper ensures that it is a comprehensive study, and I agree with its call for standardisation. On the other hand, it has the potential to be biased in favour of studies that produced positive results, and it does not consider other possible sources of data.

Data acquisition from social media was the subject of Anjali & Ravi Kumar (2022), itself a review of studies, published as a highly reputable IEEE Conference paper. The authors, both of whom are in the Computer Science Department at Adichunchanagiri University, India, highlight the advantage of working with up-to-date data, enabling ADRs in new drugs to be detected quickly. The problem of converting human verbal data to machine learning features is to be solved by the use of natural language processing agents. Reviewing multiple ML approaches gave the paper depth, while the lack of warnings about information quality in social media, or comparison with EHR-sourced data, were weaknesses, in my view.

Statistical methods were a theme of Kim et al. (2022), a review published in Medicine and written by researchers from the Biomedicine depts of two South Korean universities. They found 72 papers about ADR analysis: 51 papers about statistical methods and 21 papers about machine learning methods. They mined data from an online database (DrugBank, 2025), which had been extracted from EHRs, and found it to be “the most used database for machine learning” (Kim et al., 2022). They found that the vast majority of statistical studies were in the form of regression analysis, with the k-nearest neighbour method, or KNN (Bruce et al., 2020: 238-241) also being popular. In the ML analyses, researchers felt free to work with a variety of variables. This paper is unusual in combining statistical approaches with ML approaches, and reviews the papers systematically. However, there is no synthesis of data and meta-analysis, unlike Hu et al. (2024) above, and because the selection of papers is not very transparent, there is potential for bias. There was also no experimental validation.

Another paper lacking experimental validation was Lee & Chen (2019), which goes into some depth in its analysis of ML methods in ADR detection. Written by a researcher currently pursuing her Master’s degree at La Trobe University, Melbourne, with 3 publications to her name, this paper focuses on pharmacovigilance, and how ML can monitor drug safety. Different ML models are compared, and factors affecting ADR prediction, such as the quality of data available, feature selection decisions and interpretability of results are discussed. There is no standardisation in the selection of studies and no outline of how these ML techniques could be applied to the real world. The paper is a little dated now.

A more up to date review is Basnet (2023), which looks at the same subject matter and comparing ML models, giving a good overview of the state of the art. The author works in the Department of Pharmacy Practice, Aditya College of Pharmacy. He has attempted to identify trends in the application of ML to ADR prediction and also gaps in the research so far. The weaknesses in this paper are the lack of consistent

criteria in the selection of studies, the reliance on secondary sources for analysis, the author's potential lack of experience, as he is described as an intern, and also the recent arrival on the scene of the Journal of Pharma Insights and Research, which has a limited track record as an authoritative peer-reviewed journal.

Three researchers from Kyoto University examine the application of ML to "three tasks: drug-ADR benchmark data creation, drug–ADR prediction and ADR mechanism analysis" (Nguyen et al., 2021). The ML methods include KNN (Bruce et al., 2020: 238-241), kernel methods like support vector machines (SVMs) "for two-group classification problems" (Cortes & Vapnik, 1995). The researchers compare these methods: "Different from k-NN, the kernel methods learn weights from a training process, which depends on both drugs and ADRs, while weights in k-NN are calculated only from drug similarities" (Nguyen et al., 2021). This paper provides the formulae for each type of ML considered, putting it clearly in the context of ADR. It identifies gaps in current ADR prediction and points to ADR mechanisms as the area to focus on in future studies. Weaknesses are the lack of a meta-data study which might have validated these ML methods on a larger scale, the lack of a discussion of deep learning methods, and the possibility of bias because of unclear selection criteria.

A research paper from a medical institution, Zhejiang Pharmaceutical College, NingBo, China, is the 4-page conference paper by Dong (2021), which attempts to "classify the [drug] characteristics according to the differences of [their] molecular structure" (Dong, 2021). Unfortunately, the paper is more specific about the feature engineering than the actual ML models used, and it tends to assume the reader is already familiar with the study. It does not properly evaluate any models, is limited in scope and lacks novelty.

There is more novelty in another paper from a medical institution: Dey et al. (2018), which goes into deep learning methods applied to ADR prediction. The lead author is at the Center for Computational Health, IBM T.J. Watson Research Center, NY. The methodology of this research paper is to gather compatible data from biomedical databases, clean it and perform feature engineering, i.e. handling missing values and normalizing values, then using binary encoding and training a multiple layer neural network, tuning some of the parameters. The models are scored using classification metrics, including accuracy, precision, F1-score and recall (Bruce et al., 2020: 219-223). This study benefits from its concentration on the interpretability of the results in the real world, and is pioneering in its application of deep learning. On the other hand, it does not make comparisons with more traditional ML methods. The heavy computational requirements of deep learning are not addressed.

Another short conference paper is Islam et al. (2018), by four computer scientists at BRAC University, Dhaka, Bangladesh. This is novel in its focus on the prediction of the actual severity of ADRs instead of merely their likelihood of occurrence. The study is transparent in its data gathering from the Food and Drug Administration, feature engineering, using "python's built-in panda library called 'Feature Selection'" (Islam et al., 2018), encoding methods, using "One-Hot Encoding, Vectorization and Label Encoding" (Islam et al., 2018), analysis methods and evaluation techniques.

Fig 1 shows the type of data extracted and Fig 2 shows the methods of data mining and machine learning used:

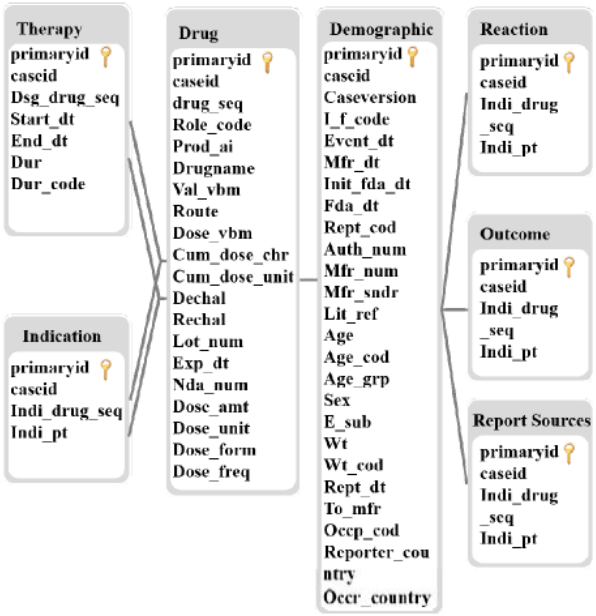

Fig. 1. Dataset Tables

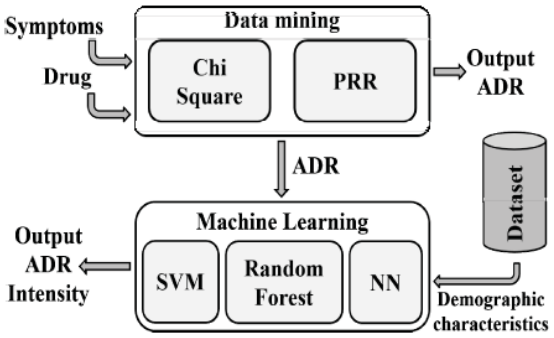

Fig. 2. Hybrid Model

A failing in Islam et al. (2018) is that unlike Nguyen et al. (2021) or Dey et al. (2018), it does not provide an extensive comparison of different ML models. It also fails to indicate the scale of the dataset used.

The author of the final paper in this literature review, McMaster, C. et al. (2019), is at the Department of Medicine, University of Melbourne. He begins with concerns that many ADRs are not being recognised, and conducts a study by assigning an ICD-10 code (World Health Organization, 2016) to each admitted inpatient. This is an easy way to codify medical conditions using a worldwide standard. Values are assigned to "ADR probability ('doubtful', 'possible', 'probable', 'definite') and severity ('mild', moderate', 'severe', 'fatal')" (McMaster, C. et al. (2019). The probabilities are encoded to 0 for 'doubtful' and 1 for anything else, thereby losing some detail. The models to be trained are of the logistic regression (Bruce et al., 2020: 233-236) and support vector machine (Cortes & Vapnik, 1995) types with random forest classifiers (Bruce et al., 2020: 259-265). There is a recognition of the lack of regulatory integration on the subject patient data ethics. On the weak side, the paper does not draw comparisons with other ML methods.

I believe I have selected a strong body of literature because of the range of ML models investigated (Dey et al., 2018; Dong, 2021; Islam et al., 2018; Lee & Chen, 2019) and the diversity of the data sources used (Anjali & Ravi Kumar, 2022; Hu et al., 2024; Lee & Chen, 2019; McMaster et al., 2019), as well as the recognition of remaining key challenges (Basnet, 2023; Dey et al., 2018; Hu et al., 2024; Kim et al., 2022) and emerging trends (Islam et al., 2018; Kim et al., 2022; Lee & Chen, 2019).

Discrepancies encountered included inconsistencies in the areas of methodologies, findings, data sources, and evaluation metrics. There is no consensus on whether deep learning or machine learning handle the data more efficiently and accurately,

although deep learning methods were thought to be less interpretable (Dey et al., 2018; Lee & Chen, 2019). Data sources are viewed differently. Some papers favour social media because it gives up to date information (Anjali & Ravi Kumar, 2022), whereas others criticise social media for its lack of clinical validation and high incidence of false positives (Hu et al., 2024).

The main findings of the ten papers as a group were that the accuracy of ADR detection improves with the use of ML, and even more so with DL. Interpretability of the results is a problem with ML and the need for large datasets and computing resources is a problem with DL. Standardization is needed in electronic health records and social media data requires advances processing to convert into standard data. Comparisons between machine learning projects would be made easier if there were a universal set of metrics by which they could be evaluated, as suggested in a number of papers, including one from a team of university researchers at Turku, Finland (Rainio, Teuho & Klén, 2024). These metrics would be easier to develop within one field or ontology, such as that of "ADR."

References

Anjali, B.V. & Ravi Kumar, G.K. (2022) 'A Broad Review on Adverse Drug Reaction Detection using Social Media Data', *6th International Conference on Intelligent Computing and Control Systems (ICICCS)*. Madurai, India, 25-27 May 2022. Piscataway, NJ, USA: IEEE. DOI: https://doi.org/10.1109/ICICCS53718.2022.9788381

Basnet, S. (2023) A Systemic Review of Machine Learning Approaches for Adverse Drug Reaction Detection: Novel Perspective and Challenges. *Journal of Pharma Insights and Research*, *1*(2): 179–185. https://jopir.in/index.php/journals/article/view/21

Bruce, P., Bruce, A., & Gedeck, P. (2020) *Practical Statistics for Data Scientists*. 2nd ed. Sebastopol: O' Reilly.

Cortes, C. & Vapnik, V. (1995) Support-Vector Networks. *Machine Learning* 20: 273-297. DOI: https://doi.org/10.1023/A%3A1022627411411

Dawson, C.W. (2015) *Projects in Computing and Information Systems: A Student's Guide*. 3rd ed. Harlow: Pearson Education Ltd.

Dey, S., et al. (2018) Predicting adverse drug reactions through interpretable deep learning framework. BMC Bioinformatics (19)476. DOI: https://doi.org/10.1186/s12859-018-2544-0

Dong, L. (2021) Study on Machine Learning and Prediction Model of Adverse Drug Reactions. *13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, Beihai, China. 518-521. DOI: 10.1109/ICMTMA52658.2021.00119.

DrugBank (2025) DrugBank. Available from: https://go.drugbank.com/ [Accessed 11 March 2025].

Hu, Q. et al. (2024) Machine learning to predict adverse drug events based on electronic health records: a systematic review and meta-analysis. *Journal of International Medical Research* 52(12): 1-19. DOI: http://dx.doi.org/10.1177/03000605241302304

Islam, T., Hussain, N., Islam, S. & Chakrabarty, A. (2018) Detecting adverse drug reaction with data mining and predicting its severity with machine learning. *2018 IEEE region 10 humanitarian technology conference.* Sri Lanka Institute of Information Technology. DOI: https://doi.org/10.1109/R10-HTC.2018.8629806

Kim, H.R. et al. (2022) Analyzing adverse drug reaction using statistical and machine learning methods: A systematic review. Medicine 101(25): e29387. DOI: 10.1097/MD.0000000000029387

Lee, C.Y. & Chen, Y-P.P. (2019) Machine learning on adverse drug reactions for pharmacovigilance. *Drug Discovery Today* 24(7): 1332-1343. DOI: https://doi.org/10.1016/j.drudis.2019.03.003

McMaster, C. et al. (2019) A Machine-Learning Algorithm to Optimise Automated Adverse Drug Reaction Detection from Clinical Coding. *Drug saf* 42: 721–725. DOI: https://doi.org/10.1007/s40264-018-00794-y

Nguyen, D.A., Nguyen, C.H. & Mamitsuka, H. (2021) A survey on adverse drug reaction studies: data, tasks and machine learning methods. *Briefings in Bioinformatics* 22(1): 164–177. DOI: 10.1093/bib/bbz140

Rainio, O., Teuho, J. & Klén, R. (2024) Evaluation metrics and statistical tests for machine learning. *Scientific Reports 14*(1): 6086. Available from: https://www.nature.com/articles/s41598-024-56706-x [Accessed 23 February 2025].

Tukey, J.W. (1977) *Exploratory data analysis*. Reading: Addison-Wesley.

World Health Organization. (2016) *International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10).* Geneva: WHO.

World Health Organization. (2020) Adverse Drug Reaction Definition by World Health Organization. Available from: https://www.publichealth.com.ng/adverse-drug-reaction-definition-by-world-health-organization/ [Accessed 20 February 2025].