**Presentation’s transcript**

Table of Contents

[First slide’s transcript 2](#_Toc134976057)

[Second slide’s transcript 2](#_Toc134976058)

[Third slide’s transcript 2](#_Toc134976059)

[Fourth slide’s transcript 3](#_Toc134976060)

[Fifth slide’s transcript 3](#_Toc134976061)

[Sixth slide’s transcript 4](#_Toc134976062)

[Seventh slide’s transcript 8](#_Toc134976063)

[Eighth slide’s transcript 9](#_Toc134976064)

[Ninth slide’s transcript 9](#_Toc134976065)

[Tenth slide’s transcript 10](#_Toc134976066)

[References 11](#_Toc134976067)

# First slide’s transcript

Hello, thank you for listening to my presentation today on how Machine Learning-driven technologies can be leveraged to enhance the prediction of the safety of medications for depression regardless of its degree of severity, thus building on the previous literature review that was instead focused on major depressive disorder or severe depression alone.

# Second slide’s transcript

In this presentation, I will describe the contribution of the research discussed in this proposal, as well as articulate a clear question underpinning this research, which motivates it and provides a hypothesis to assess via quantitative data based on which to focus the research endeavours. Thereafter, I will outline the aims and objectives of the research question, as well as summarise the key methodological steps, and the ethics and risk assessment to be considered and evaluated prior to undertaking such research. I will also illustrate the end-goal that is to create a software artefact and a timeline of the three main phases of the research required to create it.

# Third slide’s transcript

Algorithms in the literature have either resulted in very high predictive accuracy and reliability for only the most severe stage of depression, known as major depressive disorder, via Deep Learning-based models that are not interpretable or have sacrificed predictive performance to guarantee the required explainability, whilst always limiting their applicability to aid in the treatment of one stage of severity (Benrimoh et al., 2018; Fu et al., 2022). Nevertheless, in both cases, such academic research-oriented algorithmic techniques have not led to translational applications in a clinical setting.

Considering the importance of achieving an acceptable trade-off between predictive performance and interpretability for all stages of depression (Hatherley et al., 2022; Chen et al., 2023), especially the earliest ones that are easier to manage if treated promptly, the key contribution of this research would be the development of an explainable high-performing Machine Learning-driven algorithm that can complement traditional statistical methods to enhance the prediction of safety of medications used to treat depression across its various stages of severity.

# Fourth slide’s transcript

The key research question is: Can a Machine Learning-driven algorithm be as explainable as reliable to enhance the prediction of safety of medications across the various stages of depression to complement traditional statistical tools currently used to perform it?

# Fifth slide’s transcript

The key aim is to develop a Machine Learning-driven algorithm that is as explainable as reliable to enhance the prediction of safety of medications for depression.

The key objectives are the following ones:

* To ensure the Machine Learning-driven algorithm is explainable:
  + Both technical and end-user explainability will be guaranteed respectively by leveraging an intrinsically interpretable algorithm (Hatherley et al., 2022; Chen et al., 2023), e.g., a decision tree-based one, and by reverse engineering its predictions to the key parameters leading to them, which clinicians could fine tune based on their knowledge of a patient’s case and history, as well as expertise and experience in Psychiatry (Dobias et al., 2022), with focus on treating depression.
* To ensure the Machine Learning-driven algorithm is reliable:
  + Its predictive performance would be evaluated on publicly available benchmark datasets that would guarantee its replicability, reproducibility, and reliability by leveraging gold-standard predictive performance metrics.

# Sixth slide’s transcript

The key literature related to the project builds upon the previous literature review and expands it for filling specific research gaps to enhance the explainability and reliability of such clinically relevant predictions, with focus on identifying a methodology for predicting medication safety that can generalise across the various stages of severity of depression.

In fact, the reliability of predictive algorithms has been limited to their application for predicting the safety of medications of one stage of severity of depression alone (Benrimoh et al., 2018; Dobias et al., 2022). Furthermore, predictive algorithms in the literature for aiding the prediction of medication safety for depression have not been explainable (Hatherley et al., 2022; Chen et al., 2023), thus making it harder to trust, use, and adopt them in a clinical setting. As a result, such research findings have not led to clinical translation, thus being confined within the remits of academic research only so far.

Now, let us analyse the relevant literature and identify a suitable approach for predicting the safety of medications for depression.

Previous research works that developed algorithms for predicting medication safety for patients with depression have mainly focused on only one stage of severity of depression, such as severe depression (Benrimoh et al., 2018; Fu et al., 2022). In fact, limited studies have explored the prediction of safety of medications for patients with mild or moderate depression (Parthipan et al., 2019; Gillet et al., 2020). These two stages do not seem to warrant predictive algorithms for medication safety until either adverse events occur or the severity stage has progressed onto the severe stage or major depressive disorder, where medication safety is of paramount importance, as serious adverse events may occur. Thus, the narrow focus of existing studies limits the reliability and generalisation of these predictive models when considering multiple stages of depression. For instance, the study by Fu et al. (2022) developed a predictive model named ‘Hint’, which is a deep neural hierarchical interaction network, that accurately predicted adverse events in patients with severe depression. This methodology leveraged that machine learning-based model, along with patient-level data collected during randomised controlled clinical trials, such as demographics, family and medical histories, and medication profiles. Nevertheless, further research is required to validate the algorithm's performance across different stages of depression severity, especially mild or moderate depression, especially in observational, non-randomised controlled studies, which are more representative of real-world scenarios and epidemiology.

Moreover, studies assessing the reliability of predictive models for medication safety for treatment-resistant depression (TRD) are also limited (Lucchese et al., 2021; Nwanosike et al., 2022). In particular, TRD involves complex and challenging clinical workflows for managing it, and the reliability of predictive algorithms in this context is even more crucial.

An additional concern from the literature surveyed is the limited assessment of the explainability of models aiding the prediction of medication safety for depression from a clinician’s or end-user’s standpoint (Ward et al., 2021; Chen et al., 2023). In fact, whilst the above-mentioned predictive models have had a promising predictive performance when considering a single stage of depression severity at a time (Benrimoh et al., 2018; Fu et al., 2022), their limited transparency hindered clinical adoption. Nevertheless, end-user explainability of such predictive algorithms is key to build trust and aid clinical decision-making processes in pharmacovigilance (Hatherley et al., 2022) and to ensure that the correct drug and dosage are administered to patients along their medical journey and stages of depression severity.

End-user explainability involves ensuring that users, such as clinicians, can interpret and understand the recommendations derived from a predictive model and the key drivers determining them, as well as being able to adjust their weights to ensure their clinical meaningfulness, adherence, and contextual appropriateness based on the individual patient and medication profiles. Providing an explainability layer in addition to a high-performing predictive model can enhance the clinicians’ trust in using and adopting it in their workflows (Chen et al., 2023) to aid treatment planning and execution strategies. In short, reliability of a predictive algorithm for clinical applicability encompasses both predictive and explainable elements.

To aid explainability of the predictions of medication safety in depression for clinicians, previous research works have leveraged interpretable machine learning-driven models, such as ensembles of decision trees or logistic regression (Chen et al., 2023), from which human-interpretable rules that are learnt from data can be extracted, visualised, and understood by clinicians. Nevertheless, whilst it is more straightforward to explain a logistic regression model based on an equation and weights applied to each determining factor or key driver, ensembles of decision trees have not been explained fairly and comprehensively so far (Ward et al., 2021). In fact, authors often illustrated the average weights of each feature towards a certain prediction across all decision trees (Chen et al., 2023) without showing any learnt rules that clinicians could understand and adjust if needed. Furthermore, previous research studies have focused on explaining a learnt rule from only one branch of one of many (tens or hundreds depending on the required depth of the model for a particular dataset and predictive task) decision trees (Nwanosike et al., 2022). Instead, such weights should be averaged across the same branch of all various decision trees the ensemble (or grouped) model is composed of, such that a single visualisation with key drivers and their weights affecting a specific prediction can be produced to be understandable and usable by clinicians.

# Seventh slide’s transcript

The methodology involves the development and validation of a Machine Learning-driven, decision tree-based algorithm for predictive analytics, along with the evaluation of its predictive performance against publicly available benchmark datasets and clinically relevant and interpretable predictive performance metrics.

The required data will be gathered, pre-processed to remove outliers and impute or fill in missing values appropriately. Thereafter, numerical and categorical features will be respectively rescaled from 0 to +1 via min-max normalisation and transformed via ordinal or integer encoding to prepare them for ingestion into a Machine Learning-driven model.

Once data are pre-processed and encoded, the most relevant features will be selected based on both extrinsic and intrinsic methods, such as the algorithm ‘ReliefF’ and decision tree-based feature importance analysis, aiming to improve downstream predictive performance.

Such selected features will be leveraged to perform feature engineering to create new features that are stronger predictors of safety of medications for major depressive disorder, such as the number and frequency of psychotic episodes, family history of psychiatric conditions, environmental and behavioural factors, such as residential location via a postcode, the highest educational qualification attained, eating habits, sleeping patterns, etc. The engineered features leading to enhanced predictive performance will be retained in the feature set.

The data with the optimal features derived via feature selection and feature engineering will be split into three sets, for training, validation, and testing. The validation set will be used to ensure that the model is trained whilst avoiding overfitting or overtraining, thus ensuring generalisation on unseen data, which are represented by the test set. Therefore, the test set will be leveraged to evaluate the true predictive performance of the model, which will be optimised to improve it.

Once the trained model is optimised and its predictive performance on the unseen/test set is increased, it will be deployed by wrapping it into an API endpoint via the ‘FastAPI’ framework, thus being able to send a json request with subject-specific and drug-related information and obtaining a json response with its predicted safety level. A simple frontend will be developed using the library ‘tkinter’ with the ability to enter such input information and receive the predictions in a user-understandable manner, thus not only returning a safety score, but also an explanation regarding its prediction leveraging the input features and their rank or impact towards the predicted safety level.

# **Eighth slide’s transcript**

As secondary, publicly available datasets will be used, and their licence will be checked to ensure its usage for these research purposes, there is no risk or ethical approval required.

# Ninth slide’s transcript

As this is a research proposal presentation, no software artefact would be created, but the project itself, if pursued beyond this proposal stage, would involve the development and validation of a Machine Learning-driven algorithm. Using Python programming, this software would provide an explainable safety risk score in percentage and the key factors increasing or decreasing it for a medication accounting for the stage of depression severity considered, patient and medication profiles. Clinicians could then use this model and a similar visualisation to assist them in their clinical decision-making processes for aiding treatment planning and execution strategies and provide feedback for continuously improving the solution and driving further adoption and positive impact.

# Tenth slide’s transcript

In conclusion, the development and validation of such a predictive algorithm will be achieved in an Agile release cycle (Jurney, 2017), which includes three fortnightly sprints, i.e., within six weeks.

In the first sprint, data will be pre-processed and encoded to prepare the first version of the cleaned feature set. Thereafter, the most relevant features will be selected, and further ones will be engineered based on such selected features to enhance predictive performance.

In the second sprint, the model will be trained, validated, and evaluated on an unseen data partition or test set, and optimised iteratively until the highest predictive performance on the test set is achieved.

In the third and last sprint, the optimal trained model will be deployed as an API and a simple frontend will be developed to return the required predictions, along with accompanying explanations to ensure user adoption of this tool.

# References

Benrimoh, D., Fratila, R., Israel, S., Perlman, K., Mirchi, N., Desai, S., ... & Aifred Health Team, T. (2018) Aifred health, a deep learning powered clinical decision support system for mental health. In *The NIPS'17 Competition: Building Intelligent Systems* (pp. 251-287). Springer International Publishing.

Chen, J., Ooi, L. Q. R., Tan, T. W. K., Zhang, S., Li, J., Asplund, C. L., ... & Yeo, B. T. (2023) Relationship Between Prediction Accuracy and Feature Importance Reliability: an Empirical and Theoretical Study. *NeuroImage* 120115.

Dobias, M. L., Sugarman, M. B., Mullarkey, M. C., & Schleider, J. L. (2022) Predicting mental health treatment access among adolescents with elevated depressive symptoms: machine learning approaches. *Administration and Policy in Mental Health and Mental Health Services Research* *49*(1): 88-103.

Fu, T., Huang, K., Xiao, C., Glass, L. M., & Sun, J. (2022) Hint: Hierarchical interaction network for clinical-trial-outcome predictions. *Patterns* *3*(4): 100445.

Gillett, G., Tomlinson, A., Efthimiou, O., & Cipriani, A. (2020) Predicting treatment effects in unipolar depression: A meta-review. *Pharmacology & therapeutics* *212*: 107557.

Hatherley, J., Sparrow, R., & Howard, M. (2022) The virtues of interpretable medical artificial intelligence. *Cambridge Quarterly of Healthcare Ethics* 1-10.

Jurney, R. (2017) *Agile data science 2.0: Building full-stack data analytics applications with Spark*. " O'Reilly Media, Inc.".

Lucchese, A. C., Sarin, L. M., Magalhães, E. J. M., Del Sant, L. C., B Puertas, C., Tuena, M. A., ... & B Andreoli, S. (2021) Repeated subcutaneous esketamine for treatment-resistant depression: impact of the degree of treatment resistance and anxiety comorbidity. *Journal of Psychopharmacology* *35*(2): 142-149.

Nwanosike, E. M., Conway, B. R., Merchant, H. A., & Hasan, S. S. (2022) Potential applications and performance of machine learning techniques and algorithms in clinical practice: a systematic review. *International Journal of Medical Informatics* *159*: 104679.

Parthipan, A., Banerjee, I., Humphreys, K., Asch, S. M., Curtin, C., Carroll, I., & Hernandez-Boussard, T. (2019) Predicting inadequate postoperative pain management in depressed patients: a machine learning approach. *PLoS One* *14*(2): e0210575.

Ward, I. R., Wang, L., Lu, J., Bennamoun, M., Dwivedi, G., & Sanfilippo, F. M. (2021) Explainable artificial intelligence for pharmacovigilance: What features are important when predicting adverse outcomes? *Computer Methods and Programs in Biomedicine* *212*: 106415.