Modelling Major Depressive Disorder Antidepressant Treatment Response: A miRNA-based Machine Learning Study

Lok Hua Lee, Cyrus S. H. Ho, Gabrielle W. N. Tay, Cheng-Kai Lu, *Senior Member, IEEE*, and Tong Boon Tang, *Senior Member, IEEE*

***Abstract*— Major depressive disorder (MDD) is a psychiatric disorder but currently defined by symptoms rather than biological mechanism. This in turn sets a huge barrier to effective diagnosis and treatment planning. Investigations were done through neuropathogenesis and neuroimaging analysis as an effort to identify discriminative biomarkers for MDD while understanding the biological dependencies. The literature suggested that microRNA or miRNA transcripts are more likely to deliver substantial predictive power in diagnosis and antidepressant treatment response (ATR) prediction. Yet, there presents discrepancy in unique markers, and such discrepancy might be due to the small sample size over some of the reported studies. This study utilized miRNA as a predictor to model MDD ATR using k-nearest neighbour (kNN). The shortlisted miRNA through feature selection techniques scored 71.20%, 68.13%, 72.13%, and 84.07% for three response levels in accuracy, sensitivity, specificity, and precision, respectively. Synthetic Minority Oversampling TEchnique (SMOTE) was then applied to the shortlisted miRNA and three response levels reported at least 98% in each of the mentioned performance metric.**

***Clinical Relevance*— The discovery of miRNA candidates in this study could potentially narrow down the collections of blood miRNA samples for the treatment response prediction.**

I. Introduction

Major depressive disorder (MDD) as a type of psychiatric disorder, inflicts significant changes in mood and experience in psychophysiological turnovers (e.g., disturbance in sleep, appetite, suicidal thoughts) for at least 2 weeks to an individual [1]. Severe episodes of MDD could induce unproductiveness, guiltiness and hopelessness that eventually guides to mortality. Common practice in psychiatric diagnosis of MDD refers to Diagnostic and Statistical Manual of Mental Disorders (DSM) before treatment trials [2]. However, a study involving trained physicians had revealed an astoundingly high misdiagnosis rate of 65.9% among 229 MDD patients using Mini International Neuropsychiatric Interview (MINI) [3]. This reflects that assessment-based clinical diagnostic method is rather subjective and unreliable under certain circumstances. The implications after MDD clinical submission lies upon the complexity of patient's treatment and its response, which includes but not limited to psychotherapy and pharmacotherapy [4]. Pharmacotherapy is presently the most effective treatment however its responsiveness is not guaranteed and rather sceptical towards treatment resistant depression (TRD) [1], [2], [5] – [7]. ATR across MDDs with various drugs and doses takes up to 10 weeks of trial before the responsiveness gets to be determined, as advocated in [7].

MDD treatment might sound complex, but present studies involving genes had revealed its huge prospect to stipulate the foundation of MDD ATR [8]. Multiple evidence on pre- and post-treatment alterations on gene-expression regulator (hereinafter referred as miRNA) supported the narrative of genes is correlated to MDD ATR [9] – [11]. While miRNA could be a potential biomarker and therapeutically beneficial when it is being targeted for regulation, the replicability of the studies remained a challenge due to heterogeneity in MDD demographics [9]. The issue is further highlighted whereby comorbidities and confounding elements (e.g., depression severity) could contribute to changes in miRNA levels for MDD ATR [12].

The sparse collective information of miRNA based MDD ATR however does not affect the artificial intelligence (AI) ability in improving ATR predictability. Meta-analysis from [13], [14] suggested machine learning (ML) could be a potent approach to discriminate MDD treatment responder. Initiates on modelling ATR with depressive symptom-based predictive model showed how AI-guided treatment could steer the outcome of treatment [15]. Early investigation involving genes as predictors in ML models reported similar results [16]. Qi *et al.* [17] further exploited miRNA in ATR profiling through various ML approaches and analyses to seek for possibility in establishing miRNAs as a reliable predictor. While the reproducibility of unique miRNAs across cohorts of independent datasets for ATR profiling showed problematic persistency, multiple studies still encourage its utilization in AI/ML [18] – [20].

For this study, feature selection algorithms Chi2, Kruskal Wallis, and ANOVA are leveraged to shortlist the best performing miRNA with kNN model. Then SMOTE is incorporated to study the changes in performance if the pool of data becomes larger. The flow of implementation is depicted in Fig. 1 and is executed in MATLAB R2023A.

II. Methodology

A. Dataset

A total of 726 blood miRNA was collected from 70 clinically diagnosed MDD patients, by psychiatrist in compliance with fifth edition DSM (DSM-V). All procedures comply with the Declaration of Helsinki's ethical standards

*Research is supported by Yayasan Universiti Teknologi PETRONAS (YUTP-FRG 015LC0-395) and National University Health System Singapore Seed Fund 2018 (grant number: NR19MRF015).

Lok Hua Lee is with the Centre for Intelligent Signal and Imaging Research, University Teknologi PETRONAS, Perak, Malaysia (e-mail: lee_22003902@utp.edu.my).

Cyrus S. H. Ho, is with Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore (e-mail: pcmhsh@nus.edu.sg).

Gabrielle W. N. Tay, is with Department of Psychological Medicine, Yong Loo Lin School of Medicine, National University of Singapore (e-mail: gtwn@nus.edu.sg).

Cheng-Kai Lu is with the Electrical Engineering Department, National Taiwan Normal University, Taipei City, Taiwan (e-mail: cklu@ntnu.edu.tw).

Tong Boon Tang is with the Centre for Intelligent Signal and Imaging Research, University Teknologi PETRONAS, Perak, Malaysia (e-mail: tongboon.tang@utp.edu.my).

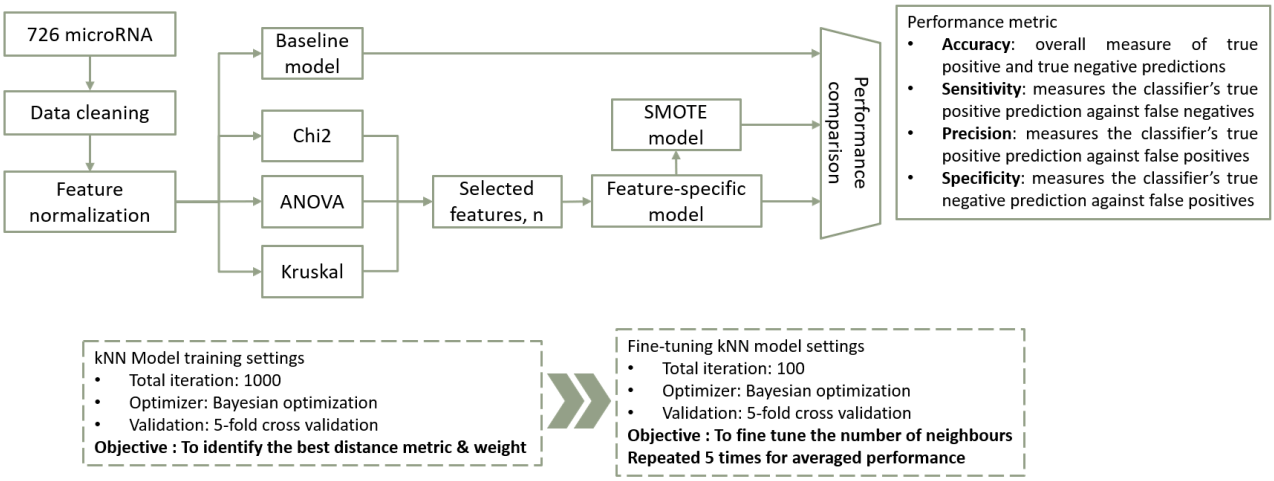

Fig. 1. Grand scheme of proposed methodology

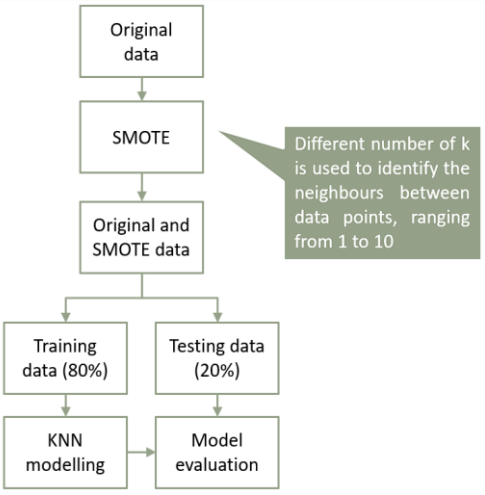

Fig. 2. Implementation of SMOTE algorithm

and the approval was granted by the Domain Specific Review Board of the National Healthcare Group, Singapore (protocol number 2019/00141). The recruited individuals age between 21 to 49 years old, and the gender composition of both entities are 16 males and 54 females. Only 52 out of the 70 are eligible for further analysis after data cleaning. They are further broken down into non-responder (NR), partial-responder (PR) and responder (R) groups for three group analysis. The responder annotation is performed by professional clinicians in comparative to changes in Hamilton Depression Rating Scale (HAM-D) scores before and after a medication trial. The definition of response group refers to percentage of reduction in HAM-D score where below 25% is NR, in between 25% and 50% is PR, and above 50% is R. That sums up to 24 NR, 15 PR and 13 R. Each raw miRNA data comes with different scales of values; hence, some preprocessing is required before proceeding.

B. *Data preprocessing*

Outliers of each column of sampled miRNA are identified and excluded from being recognized as min-max value for data normalization process. Element that is more than three scaled median absolute deviation (MAD) from the median is considered as outlier in this context. The rest of the data is normalized into a range of 0 to 1 by columns while the outliers are being resubstituted and normalized using previously identified min-max values. Normalized outliers where values below 0 and beyond 1 are topped at value 0 or 1. The outliers are not excluded from the analysis to conserve the particularly small sample size.

C. *Feature selection*

Raw data contains 726 miRNA which poses a potential redundancy for a prediction of 52 subjects and may collapse into "curse of dimensionality". To avoid that, the data is subjected to feature selection to identify the best possible miRNAs as features in classifying the groups. Algorithms include parametric evaluation (ANOVA), and non-parametric evaluation (Chi2 and Kruskal Wallis) are considered. The selection of first n number of miRNA is configured, where n = 5 to 25 at a step size of 5 (selection size is predetermined considering the importance scores does not vary beyond the number). The identification of most predictive miRNA is executed concurrently with the kNN model development and is selected based on performance evaluation as below. kNN was particularly selected for this study to benefit the unknown data distribution of current dataset (e.g., linear, non-linear) from the algorithm's non-boundary properties and its simplicity in hyperparameter tuning.

D. *Model optimization strategy*

All the initial kNN models are trained with 1000 iterations with Bayesian optimization and 5-fold cross validation for single run to identify the best distance metric and weight. Then, the distance metric and its weight are inherited for the fine-tuning process running at 100 iterations to obtain the best number of neighbours k. The fine-tuning process is repeated five times to obtain an averaged performance.

E. *Performance evaluation*

All kNN models are 5-fold cross validated and evaluated with performance metrics of accuracy, sensitivity, precision, and specificity that generally provides a good picture of how well the model performs across the groups. The min-max of each metric is also included with the averaged metrics to better assess the model's generalizability. It is also worth noting that the presented results are validation-based metrics as the sample size is simply not capable of rendering any meaningful train-test splitting.

F. *Data generation*

To simulate the performance of the shortlisted miRNA in a larger sample size manner, a classic tabular data generation method is adopted [21]. SMOTE (with different number of k, ranging from 1 to 10 in step size of 1), is applied to the

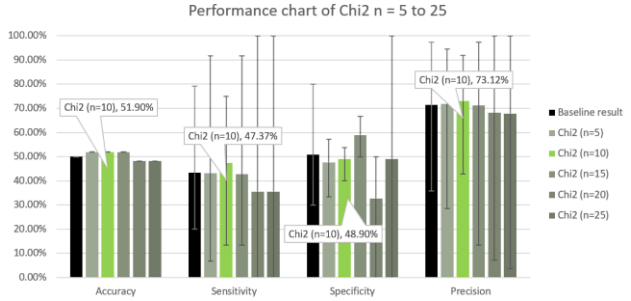

Fig. 3. Performance of miRNAs selected by Chi2 for n = 5 to 25. Averaged values of three groups are plotted on the bar with respective min-max values. The best performing selection is highlighted in green.

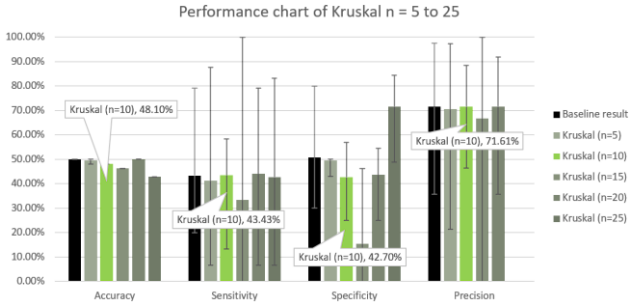

Fig. 4. Performance of miRNAs selected by Kruskal Wallis for n = 5 to 25. Averaged values of three groups are plotted on the bar with respective min-max values. The best performing selection is highlighted in green.

shortlisted miRNAs and generate synthetic data to scale each responding group into at least 100 samples; as closer as the algorithm possibly could. The data is later split into train-test set according to 80/20 rule, where 80% are training data and 20% are testing. The data is generated before splitting to produce relevant interpolations from this size of data. The execution is as illustrated in Fig. 2.

III. Result and Discussion

A. Feature selection result for three group analysis

Fig. 3, 4 and 5 shows the model performance of Chi2, Kruskal Wallis, and ANOVA respectively for three group analysis. MiRNAs selected using Chi2 and Kruskal Wallis does not have any sign of significant improvement over the baseline model (using all 726 miRNAs), and some are even worse. This indicates that for three group analysis, it is not possible to perform group separation with kNN based on the decisions from binning of occurrence frequency of each miRNA (working mechanism of Chi2) and median of each miRNA (working mechanism of Kruskal Wallis). In other words, they might be important for the responder groups considering the p-values of respective hypotheses and discriminable up to certain extent, but the data points simply do not have enough separability in feature space when those measures (i.e., occurrence frequency, median) are considered. ANOVA on the other hand, a mean-based testing algorithm, shows respectable improvement in prediction when first 5 miRNAs are selected. In fact, the overall accuracy is improved by 19.20%, from 50.00% of baseline to 69.20%. Similarly, other metrics are improved by at least 10.78%. This indicates that the shortlisted miRNAs are not only important on a p-value perspective, but also much denser in feature space, allowing the kNN model to discriminate the groups relatively well. To further narrow down the number of miRNAs, another round of selection is performed at n = 1 to 4 for ANOVA and the results are as in Fig. 5. Slight improvement in all performance metrics when n = 3 compared with n = 5 is

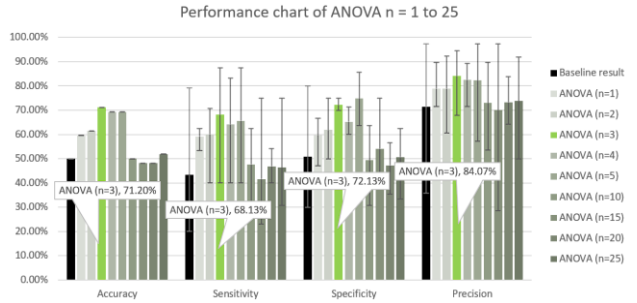

Fig. 5. Performance of miRNAs selected by ANOVA for n = 1 to 25. Averaged values of three groups are plotted on the bar with respective min-max values. The best performing selection is highlighted in green.

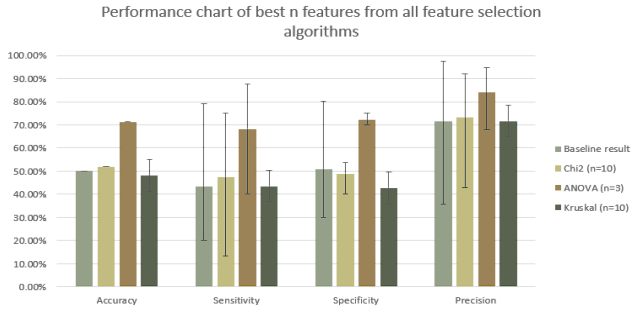

Fig. 6. Performance comparison of all feature selection algorithms. Best performing n miRNAs are selected based on the best averaged values at considerably low variation among groups.

observed, except for specificity where it is marginally dropped by 2.64% with a major decrease in the variations between groups (the min-max values). The comparison of performance for all feature selection algorithms are illustrated in Fig. 6. Those of selected miRNAs by ANOVA are hsa-miR-550b-2-5p, hsa-miR-125a-5p, and hsa-miR-374b-3p in descending order of their importance scores.

B. SMOTE number of k analysis

A series of investigation is carried out on the selection of number k of this classic interpolation method, SMOTE, to create synthetic data based on the shortlisted miRNA above. SMOTE performs linear interpolation based on its random selection of nearest neighbour data points and the total number of nearest neighbours provided for selection is controlled by the parameter k. The validation and testing results of different numbers of k being applied are depicted in Fig. 7. Both the averaged training and testing performance indexes are observed to be declining as the number of k increases. That also comes with increasing variation between the groups (the min-max values) that is not favoured for a generalized model. One could arguably claim that further increasing the number of interpolated data at higher number of k could lead to a similar performance as of current result for k = 1. However, results from a larger pool of synthetic data could just be an ignorance to the fact that higher number of k creates a more sparsely and densely distributed feature space, which might not resemble any close real-world distribution. The comparison of model performance from baseline to SMOTE is summarized in Fig. 8.

C. State-of-the-art comparison

To the best of our knowledge, there is no study that attempted to model miRNA for MDD ATR. The closest possibly we found is utilizing genetics information to perform the binary predictions (NR vs. R) [16], [22]. Based on Table I, as the types of predictors increases, some of the performance

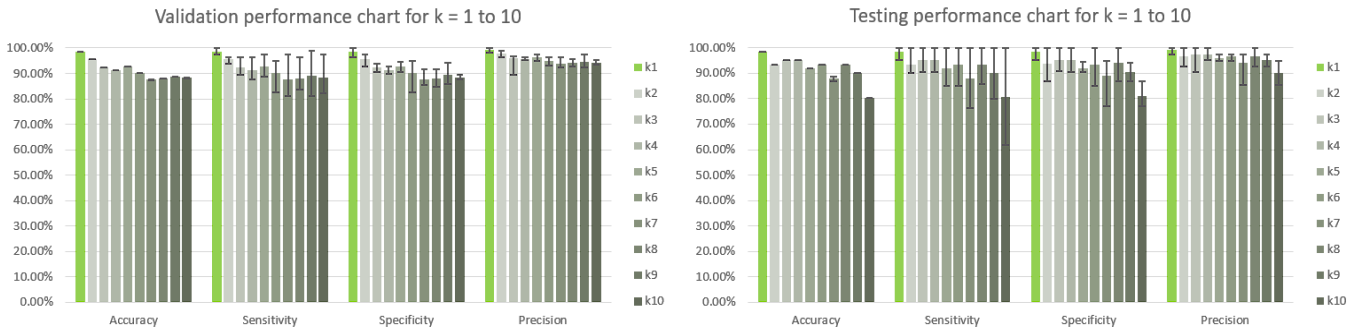

Fig. 7. Validation/training and testing performance using SMOTE with different number of k. The best performing number of neighbours is k = 1 (highlighted in green).

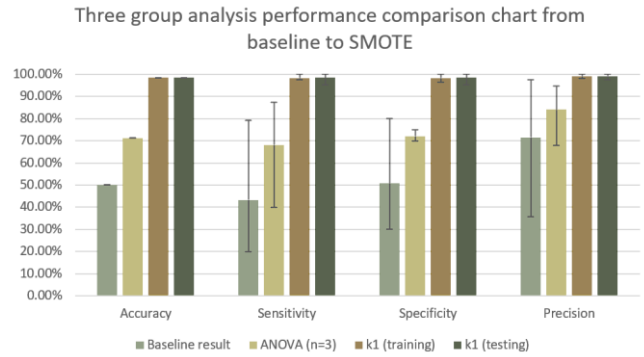

Fig. 8. Performance comparison from baseline to SMOTE application

TABLE I. STATE-OF-THE-ART COMPARISON

Author	Method	Feature	Performance metric			
			Acc. (%)	*Sen. (%)*	*Spec. (%)*	*Prec. (%)*
Joyce *et al.* [16]	PR w/ FS	M/G/CV	77.50	71.00	88.00	-
Lin *et al.* [22]	BE w/ FS	SNP/CV	-	76.51	71.14	-
Kautzky *et al.* [23]	RF w/ FS	CV/PC	69.00	-	-	-
Proposed	kNN w/ FS	miRNA	71.20	68.13	72.13	84.07
Proposed (SMOTE, k = 1)	kNN w/ FS	miRNA	98.40	98.40	98.40	99.17

PR = penalized regression; FS = feature selection; M = metabolomic; G = genes; CV = clinical variable; BE = boosting ensemble; SNP = single nucleotide polymorphism; RF = random forest; PC = physical comorbidity

metrics starts to get improved. Although the proposed model (k = 22) without SMOTE does not surpass any of the state-of-the-art models, it still performs relatively well. The results obtained by Joyce *et al.* [16] could be factored by the large dataset (348 subjects) being utilized in their multivariate study. Based on that, it is expected that the current study shall see improvements as more predictors being introduced in future studies.

IV. CONCLUSION

To summarize the work, a new framework has been proposed to predict three groups of MDD ATR. With the miRNAs (hsa-miR-550b-2-5p, hsa-miR-125a-5p, and hsa-miR-374b-3p) selected by ANOVA, the kNN model achieved scores of 71.20%, 68.13%, 72.13%, and 84.07% in accuracy, sensitivity, specificity, and precision, respectively. The performance is boosted to 98.40% in accuracy, sensitivity, and specificity, and 99.17% in precision when SMOTE is used to create larger pool of data for the prediction to happen. While the SMOTE result appears promising, it is still subjected to further investigation to obtain a more precise explanation on the application's behaviour. The shortlisted miRNAs shall have a more in-depth study on their neurogenesis pathways before it could be clinically adopted for antidepressant treatment response prediction.

REFERENCES

[1] R. H. Belmaker and G. Agam, "Major Depressive Disorder," *New England Journal of Medicine*, vol. 358, no. 1, pp. 55–68, Jan. 2008, doi: 10.1056/nejmra073096.

[2] C. Otte *et al.*, "Major depressive disorder," *Nature Reviews Disease Primers*, vol. 2, no. 1, Sep. 2016, doi: 10.1038/nrdp.2016.65.

[3] M. Vermani, M. Marcus, and M. A. Katzman, "Rates of Detection of Mood and Anxiety Disorders in Primary Care," *The Primary Care Companion for CNS Disorders*, vol. 13, no. 2, Apr. 2011, doi: 10.4088/pcc.10m01013.

[4] M. Fava and K. S. Kendler, "Major Depressive Disorder," *Neuron*, vol. 28, no. 2, pp. 335–341, Nov. 2000, doi: 10.1016/s0896-6273(00)00112-4.

[5] A. M. Kanner, "Is Major Depression a Neurologic Disorder with Psychiatric symptoms?," *Epilepsy & Behavior*, vol. 5, no. 5, pp. 636–644, Oct. 2004, doi: 10.1016/j.yebeh.2004.07.008.

[6] F. W. Lohoff, "Overview of the Genetics of Major Depressive Disorder," *Current Psychiatry Reports*, vol. 12, no. 6, pp. 539–546, Sep. 2010, doi: 10.1007/s11920-010-0150-6.

[7] R. D. Boyce, J. T. Hanlon, J. F. Karp, J. Kloke, A. Saleh, and S. M. Handler, "A Review of the Effectiveness of Antidepressant Medications for Depressed Nursing Home Residents," *Journal of the American Medical Directors Association*, vol. 13, no. 4, pp. 326–331, May 2012, doi: 10.1016/j.jamda.2011.08.009.

[8] Q. S. Li, C. Tian, and D. Hinds, "Genome-wide Association Studies of Antidepressant Class Response and treatment-resistant Depression," *Translational Psychiatry*, vol. 10, no. 1, Oct. 2020, doi: 10.1038/s41398-020-01035-6.

[9] L. Zhou, Y. Zhu, W. Chen, and Y. Tang, "Emerging role of microRNAs in major depressive disorder and its implication on diagnosis and therapeutic response," *Journal of Affective Disorders*, vol. 286, pp. 80–86, May 2021, doi: 10.1016/j.jad.2021.02.063.

[10] G. R. Fries, W. Zhang, D. Benevenuto, and J. Quevedo, "MicroRNAs in Major Depressive Disorder," *Advances in Experimental Medicine and Biology*, vol. 1118, pp. 175–190, Jan. 2019, doi: 10.1007/978-3-030-05542-4_9.

[11] S. He *et al.*, "Alterations of microRNA-124 expression in peripheral blood mononuclear cells in pre- and post-treatment patients with major depressive disorder," *Journal of Psychiatric Research*, vol. 78, pp. 65–71, Jul. 2016, doi: 10.1016/j.jpsychires.2016.03.015.

[12] Y. K. Dwivedi, "Pathogenetic and therapeutic applications of microRNAs in major depressive disorder," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 64, pp. 341–348, Jan. 2016, doi: 10.1016/j.pnpbp.2015.02.003.

[13] A. Pigoni, G. Delvecchio, D. Madonna, C. Bressi, J. Soares, and P. Brambilla, "Can Machine Learning Help Us in Dealing with Treatment Resistant depression? a Review," *Journal of Affective Disorders*, vol. 259, pp. 21–26, Dec. 2019, doi: 10.1016/j.jad.2019.08.009.

[14] A. M. Chekroud *et al.*, "The Promise of Machine Learning in Predicting Treatment Outcomes in Psychiatry," *World Psychiatry*, vol. 20, no. 2, pp. 154–170, May 2021, doi: 10.1002/wps.20882.

[15] M. Browning *et al.*, "The Clinical Effectiveness of Using a Predictive Algorithm to Guide Antidepressant Treatment in Primary Care (PReDicT): an open-label, Randomised Controlled Trial," *Neuropsychopharmacology*, vol. 46, no. 7, Feb. 2021, doi: 10.1038/s41386-021-00981-z.

[16] J. B. Joyce *et al.*, "Multi-omics Driven Predictions of Response to Acute Phase Combination Antidepressant therapy: a Machine Learning Approach with cross-trial Replication," *Translational Psychiatry*, vol. 11, no. 1, Oct. 2021, doi: 10.1038/s41398-021-01632-z.

[17] B. Qi, L. M. Fiori, G. Turecki, and Y. J. Trakadis, "Machine Learning Analysis of Blood microRNA Data in Major Depression: a Case-Control Study for Biomarker Discovery," *International Journal of Neuropsychopharmacology*, vol. 23, no. 8, May 2020, doi: 10.1093/ijnp/pyaa029.

[18] E. Lin and S.-J. Tsai, "Genome-wide Microarray Analysis of Gene Expression Profiling in Major Depression and Antidepressant Therapy," *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 64, pp. 334–340, Jan. 2016, doi: j.pnpbp.2015.02.008.

[19] A. Menke, "Precision pharmacotherapy: Psychiatry's Future Direction in preventing, diagnosing, and Treating Mental Disorders," *Pharmacogenomics and Personalized Medicine*, vol. Volume 11, pp. 211–222, Nov. 2018, doi: 10.2147/pgpm.s146110.

[20] E. Lin and S.-J. Tsai, "Epigenetics and Depression: an Update," *Psychiatry Investigation*, vol. 16, no. 9, Aug. 2019, doi: pi.2019.07.17.2.

[21] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

[22] E. Lin, P.-H. Kuo, Y.-L. Liu, Younger W.-Y. Yu, A. C. Yang, and S.-J. Tsai, "Prediction of Antidepressant Treatment Response and Remission Using an Ensemble Machine Learning Framework," *Pharmaceuticals*, vol. 13, no. 10, pp. 305–305, Oct. 2020, doi: 10.3390/ph13100305.

[23] A. Kautzky *et al.*, "Combining Machine Learning Algorithms for Prediction of Antidepressant Treatment Response," *Acta Psychiatrica Scandinavica*, vol. 143, no. 1, pp. 36–49, Jan. 2021, doi: 10.1111/acps.13250.