Comparison of feature selection approaches in youth depression determination based on handwriting kinematics

Vladimir Džepina
University of Belgrade,
School of Electrical Engineering
Belgrade, Serbia
dzepina.vladimir@gmail.com
[0009-0005-8783-4853]

Blažo Nikolić
Clinic of Neurology and Psychiatry for Children and Youth,
University of Belgrade,
Faculty of Medicine
Belgrade, Serbia
blazo.nikolic@nepsido.rs
[0000-0002-1353-5704]

Milica M. Janković
University of Belgrade,
School of Electrical Engineering
Belgrade, Serbia
piperski@etf.rs
[0000-0002-7506-4995]

Nikola Ivančević
Clinic of Neurology and Psychiatry for Children and Youth,
University of Belgrade,
Faculty of Medicine
Belgrade, Serbia
ivancevicsd@gmail.com
[0000-0003-1497-7245]

Dejan Stevanović
Clinic of Neurology and Psychiatry for Children and Youth,
University of Belgrade,
Faculty of Medicine
Belgrade, Serbia
stevanovic.dejan79@gmail.com
[0000-0001-8236-5246]

Sunčica Rosić
Department of Network and Data Science,
Central European University
Vienna, Austria
rosic_suncica@student.ceu.edu
[0009-0002-4802-5904]

Jasna Jančić
Clinic of Neurology and Psychiatry for Children and Youth,
University of Belgrade,
Faculty of Medicine
Belgrade, Serbia
jasna.jancic.npk@gmail.com
[0000-0002-2935-0798]

***Abstract*—Depressive disorder (DD) in youth is a significant, yet underrecognized mental health issue, often accompanied by psychomotor retardation. Handwriting analysis provides a non-invasive and measurable method for detecting such symptoms. This study explores feature selection approaches to improve machine learning-based classification of DD using kinematic features from a task—repetitively writing the lowercase cursive letter "l". From 177 extracted features, 40 were retained through statistical filtering and further refined using five selection approaches. Logistic regression models were trained and evaluated using subject-wise leave-one-out cross-validation. In this paper, a comparison of different feature selection approaches (Recursive Feature Elimination, Sequential Forward Selection, SHapley Additive exPlanations, Minimum Redundancy Maximum Relevance, and Feature Importance) is presented, considering the occurrence of optimal feature sets as well as the binary classification accuracy of subjects into the DD and control groups.**

***Keywords*— depression, handwriting, graphic tablet; kinematic analysis; machine learning; feature selection**

I. Introduction

Depression is among the most prevalent affective disorders worldwide [1]. Its global impact has been increasingly severe, with the number of cases rising steadily in parallel with population growth [2]. Clinical definition of DD pertains to a persistent low mood and/or anhedonia lasting at least two weeks, accompanied by symptoms such as fatigue, sleep disruption, cognitive impairment, and changes in appetite [3, 4]. Importantly, behavioral markers such as psychomotor disturbances provide a potentially objective basis for diagnosis [5, 6].

Assessment methods of psychomotor retardation include clinical scales, objective movement analyses, and reaction time tests [5, 7–9]. Notably, drawing and handwriting tasks offer a low-cost, quantifiable, and sensitive way to detect motor abnormalities in depression [7, 8, 10]. Handwriting, long studied in graphology [11], has demonstrated utility in identifying traits related to personality and cognitive disorders, including neuroticism, learning difficulties, and dyslexia [12, 13]. Given the close relationship between motor behavior and affective state, handwriting analysis holds promise for detecting mood disorders such as depression. Studies report slower and less forceful handwriting in individuals with DD—both adults [14] and youth [15].

Although handwriting data have been used in depression research, most studies rely on traditional statistical methods to identify significant features [10, 15–18]. Few studies [8, 19–21], to our knowledge, have combined statistical filtering with machine learning to explore the predictive potential of handwriting features in determining depressive disorder. Moreover, research addressing machine learning-supported depression detection specifically in youth populations remains scarce [22].

Feature selection plays a critical role in machine learning, particularly in biomedical applications, by reducing dimensionality, mitigating overfitting, and enhancing model interpretability [23]. This work investigates several complementary feature selection strategies (Recursive Feature Elimination (RFE) [24], Sequential Forward Selection (SFS) [25], SHapley Additive exPlanations (SHAP) [26], Minimum Redundancy Maximum Relevance (mRMR) [27] and Feature Importance (FI) [28] in the context of machine learning-based recognition of psychomotor retardation in youth depression.

II. Methods and Materials

A. Subjects

The study included 40 youths divided into two groups: 20 diagnosed with DD and 20 healthy controls. The dataset used in this study comes from our earlier study [15, 22]. In the DD group, 80% were female, with a mean age of 14.65 years (SD = 1.63), and 90% were right-handed. All were diagnosed according to DSM-5 criteria [3], and 55% had received antidepressant treatment (SSRIs or TCAs) for an average of 15.8 weeks prior to the study. The control group included 50% females, with a mean age of 15.9 years (SD = 0.45), and 85% were right-handed, with no history of neurological or psychiatric disorders. The DD group was tested at the Clinic for Neurology and Psychiatry for Children and Youth in Belgrade, while the control group was tested at "Gimnazija Smederevo" high school in Smederevo, Serbia. All participants and guardians provided written informed consent, and the study was approved by the clinic's ethics committee (Approval No. 21-79/02, 16 July 2018).

B. Experiment description

The handwriting experimental protocol included 3 tasks [29]: drawing single semicircles, drawing composite figures and repetitive handwriting of the lowercase cursive letter "l" within two rectangles of different sizes (40 x 160 mm and 9 x 160 mm). This study focused on the analysis of the third task realized within the smaller rectangle. Tasks were performed on A4 templates affixed to a digital tablet. Handwriting data were collected using a Wacom® Intuos 4 XL tablet and cordless pen (sampling at 200 Hz, ±0.25 mm accuracy, 2048 pressure levels). Signals recorded included pen position (X(t), Y(t)) and pen pressure (p(t)), using custom LabHand 0.7 software [30].

C. Feature dataset

In this study, we used features extracted from recorded pen position and pressure signals (X, Y, p) by the pipeline presented in detail in our previous research [22]. This pipeline includes: 1) segmentation of individual instances of the letter "l," 2) extraction of 177 kinematic handwriting features from segmented instances (165 statistical features and 12 "letter" features that describe the shape and dynamics of each individual letter segment) and 3) reduction of the number of features using statistical approach that combines Bartlett's test [31] for retaining features with similar variances across groups and Mann–Whitney U [32] test with Bonferroni correction [33]. The list of the reduced feature dataset is presented in Table 1.

D. Feature selection and classification

The feature selection and machine learning analysis were performed using a Jupyter notebook written in Python 3.11.4 programming language. The input dataset consists of a reduced feature dataset (40 features, Table 1). To evaluate the model, subject-wise leave-one-out cross-validation (LOOCV) was used, splitting the data into training and test sets accordingly. The optimal feature subset was identified using one of the following feature selection approaches applied to the training set: RFE, SFS, SHAP, mRMR and FI. Logistic regression (LR) was used as the estimator for RFE, SFS, SHAP and FI. To prevent overfitting in the LR model, the regularization technique ElasticNet, which combines L1 and L2 penalty terms, was used. All subset sizes N ($N \in [1, 40]$) were evaluated, resulting in 40 different optimal feature subset lengths.

TABLE I. FEATURE DATASET

Variable	Derivative variable	Feature
X position, X(t)	Velocity per x-axis, Vx(t)	cv, max, pct75, pct90
	Acceleration per x-axis, Ax(t)	med, pct90
	Jerk per x-axis, Jx(t)	med
Y position, Y(t)	Velocity per y-axis, Vy(t)	std, max, min, pct10, pct75, pct90
	Acceleration per y-axis, Ay(t)	min, pct10
	Jerk per y-axis, Jy(t)	std, max, pct25, pct75, pct90
Total	Total velocity, V(t)	med, min, std, max, pct25, pct75, pct90
	Total acceleration, A(t)	std, pct25, pct75, pct90
	Total jerk, J(t)	med, min, pct10, pct25, pct75, pct90
Pressure, p(t)	$\frac{d^2p}{dt^2}$	min, pct10
Letter feature	n.a.	LS
Total features = 40		

cv—coefficient of variation, max—maximum value, pct75—75th percentile, pct90—90th percentile, std—standard deviation, min—minimum value, pct10—10th percentile, med—median, pct25—25th percentile, LS—letter of drawing speed

LR classifier was performed on test sets for the binary classification of valid "letters" into DD and control groups using the obtained optimal feature subsets for each N.

Individual "letter" classification accuracy was calculated across all test sets obtained via subject-wise leave-one-out cross-validation (LOOCV), hereafter referred to as *Letter Accuracy***.** These results were aggregated to compute *Subject Accuracy*, defined as the percentage of subjects correctly classified into either the depressive disorder (DD) or the control group. A subject was deemed correctly classified if at least 50% of their letters were predicted correctly by the model.

III. Results

Fig. 1 presents the comparison of letter classification accuracy obtained by the LR for all five feature selection approaches across the range N ∈ [1, 40]. Black vertical lines highlight three optimal feature subset sizes where the maximal accuracy or "stable" accuracy was obtained. The total elapsed time required for feature selection and letter classification across the range N ∈ [1, 40] was measured for each approach. Total elapsed times were as follows: RFE - 1063 s, SFS - 55018 s, SHAP - 163 s, mRMR - 2237 s, and FI - 115 s.

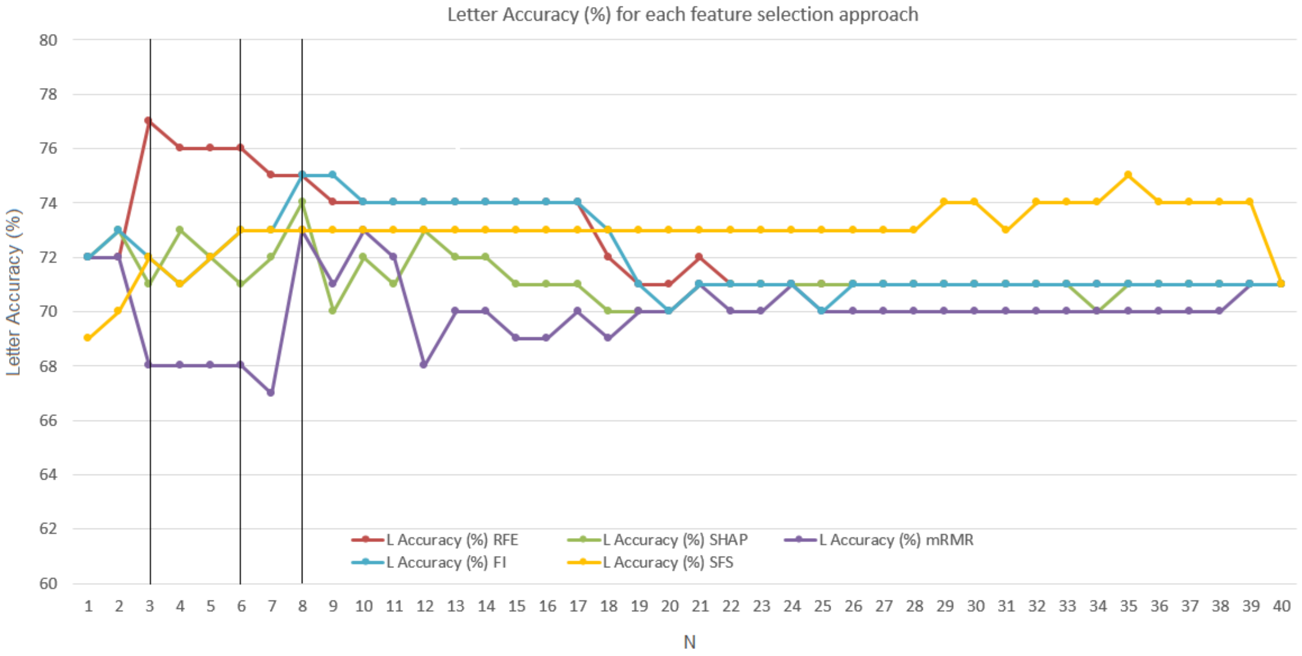

Fig. 1. *Letter accuracy for each feature selection approach*

TABLE II. Accuracies for optimal feature subset sizes

N	Approach	The optimal feature set	Occurrence [%]	Letter Accuracy [%]	Subject Accuracy [%]
3	RFE	**Vy_std, Vy_pct75, V_pct25**	95	76.7	82.5
	SFS	**Vy_std, Vy_pct75, V_pct25**	47.5	72	77.5
	SHAP	Vy_pct90, Vx_max, **Vy_pct75**	60	71	72.5
	mRMR	**Vy_pct75**, Jp_min, Vx_cv	100	68	75
	FI	Vy_pct90, Vx_max, **Vy_pct75**	72.5	72.5	75
6	RFE	**Vx_max, Vy_std, Vy_pct75, V_std, V_pct25, FS**	47.5	75.8	77.5
	SFS	**Vy_std, Vy_pct75**, V_mean, V_max, **V_pct25**, Ax_pct90	7.5	72.3	80
	SHAP	Vy_max, **Vy_std**, Vx_pct90, Vy_pct90, **Vx_max**, **Vy_pct75**	25	71.2	70
	mRMR	**Vy_pct75,** Jp_min, Vx_cv, Ax_med, Vy_pct90, Jx_med	50	72	80
	FI	Vx_pct90, **V_std, Vy_std**, Vy_pct90, **Vx_max, Vy_pct75**	30	73.4	80
8	RFE	**Vx_max, Vy_std, Vy_max, Vy_pct75, V_std, V_max, V_pct25, FS**	25	74.9	77.5
	SFS	**Vy_std, Vy_pct75**, V_med, V_mean, **V_max, V_pct25**, Ax_med, Ax_pct90	5	73.6	80
	SHAP	**V_pct25**, V_pct90, **Vy_max**, **Vy_std**, Vx_pct90, Vy_pct90, **Vx_max, Vy_pct75**	20	74.3	77.5
	mRMR	Vy_pct75, Jp_min, Vx_cv, Ax_med, Vy_pct90, Jx_med, A_pct25, J_pct75	50	70.7	77.5
	FI	**V_pct25, Vy_max**, Vx_pct90, **V_std, Vy_std**, Vy_pct90, **Vx_max, Vy_pct75**	30	74.7	77.5

Table 2 summarizes the classification results obtained using five feature selection methods across three optimal feature set sizes (N ∈ {3, 6, 8}). For each method, the most frequently selected feature subset and its occurrence in percentage across all 40 folds is reported (*Occurrence*), together with the *Letter Accuracy* and *Subject Accuracy* resulting from this approach.

RFE approach for **N**=3, 6, 8 consistently gave the highest *Letter Accuracy* among all approaches - 76.7%, 75.8%, 74.9%, respectively (with *Subject Accuracies* 82.5%, 77.5%, 77.5%, respectively). Therefore, for each N, the feature subset most frequently selected by RFE is bolded in Table 2. Features from this subset that appear in subsets of other feature selection approaches are bolded.

The *Occurrence* of the predominantly selected optimal subset decreases as **N** increases, with the mRMR approach consistently achieving the highest selection *Occurrence* among all approaches for each **N** reported in Table 2.

IV. Discussion

This study evaluated the effectiveness of multiple feature selection strategies for distinguishing youths with depressive disorder (DD) from healthy controls based on kinematic features extracted from a handwriting task.

Fig. 1 illustrates how classification performance, expressed as *Letter Accuracy*, varies with the number of features selected (N) across five feature selection approaches: RFE, SFS, SHAP, mRMR, and Feature Importance (FI). The RFE approach achieves a peak *Letter Accuracy* of 76.7% at N = 3, outperforming others for low N values, up to N = 8. SHAP, mRMR and FI approaches resulted in the highest *Letter accuracies* at N=8. For N=6, the LR began maintaining “stable” accuracy when using the SFS approach. “Stable” accuracy was defined as classification performance that remained consistent over at least three consecutive values of N.

The consistent superiority of RFE in lower dimensions supports its ability to identify highly discriminative features with minimal redundancy. This is in line with findings from previous studies that emphasized the utility of recursive elimination in small-sample, high-dimensional biomedical data [22, 34]. The plots reveal that RFE and FI maintain relatively stable *Letter accuracies* for N ≤ 10, followed by a slight decline. In contrast, SFS exhibits a modest but steady increase in *Letter accuracy* for larger N, reaching a secondary local peak at N = 35. However, this improvement comes at the cost of increased feature set complexity and reduced interpretability.

When examining the total elapsed time required for feature selection and letter classification, SHAP and FI demonstrated a clear computational advantage. This suggests that, particularly in the context of large datasets, these methods may offer significantly faster processing times with marginal compromises in classification accuracy in comparison to RFE. In contrast, SFS was found to be highly time-inefficient, rendering it impractical for real-world applications where scalability and efficiency are critical.

Among all tested methods, RFE consistently demonstrated superior performance across all feature set sizes reported in Table 2. For the subset size (N = 3), RFE achieved the highest *Subject Accuracy* of 82.5% and *Letter Accuracy* of 76.7%, correctly classifying 33 out of 40 individuals. Notably, the same three-feature subset (*Vy_std, Vy_pct75, V_pct25*) was also most frequently selected by SFS, pointing to the robustness and discriminative value of these specific features. These features primarily capture variability and quantiles in vertical velocity and total speed, which aligns with the known psychomotor slowing in depression [6, 14, 19, 22, 35].

The performance of the models showed a slight decline or plateau as the number of features increased, with larger subsets not necessarily resulting in improved classification. This highlights a well-known phenomenon in high-dimensional spaces—the curse of dimensionality—where irrelevant or redundant features may dilute predictive power and lead to overfitting. Additionally, as the subset size increased, the frequency of consistent subset selection decreased for most approaches, indicating greater variability and model instability at higher dimensions.

While RFE led in classification performance, the mRMR approach yielded the most consistent feature subsets across folds, achieving a 100% selection *Occurrence* for N = 3 and maintaining the highest *Occurrences* at N = 6 and N = 8, indicating strong internal agreement. Nevertheless, its predictive performance remained consistently lower than that of RFE and SFS. This suggests that while mRMR optimally balances relevance and redundancy at a statistical level, it may be less effective in capturing task-specific, nonlinear interactions critical to psychomotor assessment.

The SHAP and FI methods, both rooted in model interpretability, performed moderately well but showed greater variability in subset composition and lower selection frequencies. While useful for post hoc model explanation, these techniques may be less stable for small-sample feature selection when used in isolation.

The primary limitation of this study pertains to the representativeness of the sample used for analysis. The dataset comprised 40 subjects, which, while comparable to sample sizes in other handwriting-based clinical studies [8, 18], has the data density advantage by using “letter” segments. Expanding the sample in future research would enhance the generalizability, robustness and external validity of the results.

V. Conclusion

Beyond reducing model complexity, effective feature selection enhances performance, interpretability, and generalizability, especially in small-sample studies where overfitting is a primary concern. The finding that a small, well-chosen subset of features can yield high classification accuracy emphasizes the potential of digital biomarkers such as vertical stroke dynamics for supporting objective, non-invasive screening tools in youth depression. Among the evaluated methods, wrapper-based approaches like RFE proved especially effective in identifying concise, high-performing feature sets, making them particularly wellsuited for the constrained sample sizes typical in clinical research. SHAP and IF achieved slightly lower classification accuracy but offered substantially faster execution times.

Future studies could incorporate larger and more demographically diverse samples to improve the generalizability and clinical applicability of the models. Exploring deep learning models, especially those capable of learning directly from raw handwriting trajectories (e.g., autoencoders, LSTM-based encoders), may uncover nonlinear relationships and latent features that are difficult to capture through handcrafted feature engineering. The comparison could clarify trade-offs between interpretability and performance.

Acknowledgment

This work was partly financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, under contract number 451-03-137/2025-03/200103. The research was partially conducted in the premises of the Palace of Science, Miodrag Kostić Endowment.

References

[1] WHO, “Depressive disorder (depression).” Accessed: Jun. 11, 2024. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/depression

[2] Q. Liu, H. He, J. Yang, X. Feng, F. Zhao, and J. Lyu, “Changes in the global burden of depression from 1990 to 2017: Findings from the Global Burden of Disease study,” *Journal of Psychiatric Research*, vol. 126, pp. 134–140, Jul. 2020, doi: 10.1016/j.jpsychires.2019.08.002.

[3] American Psychiatric Association and American Psychiatric Association, Eds., “Diagnostic and statistical manual of mental disorders: DSM-5,” in *Diagnostic and statistical manual of mental disorders: DSM-5*, 5th ed., Washington, D.C: American Psychiatric Association, 2013, pp. 591–643.

[4] WHO, “ICD-11 for Mortality and Morbidity Statistics.” Accessed: Oct. 06, 2024. [Online]. Available: https://icd.who.int/browse/2024-01/mms/en#821852937

[5] A. Esposito, G. Raimo, M. Maldonato, C. Vogel, M. Conson, and G. Cordasco, “Behavioral Sentiment Analysis of Depressive States,” in *2020 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, Mariehamn, Finland: IEEE, Sep. 2020, pp. 209–214. doi: 10.1109/CogInfoCom50765.2020.9237856.

[6] F. L. Wells, “Motor retardation as a manic-depressive symptom,” *AJP*, vol. 66, no. 1, pp. 1–52, Jul. 1909, doi: 10.1176/ajp.66.1.1.

[7] D. Bennabi, P. Vandel, C. Papaxanthis, T. Pozzo, and E. Haffen, “Psychomotor retardation in depression: a systematic review of diagnostic, pathophysiologic, and therapeutic implications,” *Biomed Res Int*, vol. 2013, p. 158746, 2013, doi: 10.1155/2013/158746.

[8] G. Raimo *et al.*, “Handwriting and Drawing for Depression Detection: A Preliminary Study,” in *Applied Intelligence and Informatics*, Cham: Springer Nature Switzerland, 2022, pp. 320–332. doi: 10.1007/978-3-031-24801-6_23.

[9] L. Zhang, Y. Fan, J. Jiang, Y. Li, and W. Zhang, “Adolescent Depression Detection Model Based on Multimodal Data of Interview Audio and Text,” *Int J Neural Syst*, vol. 32, no. 11, p. 2250045, Nov. 2022, doi: 10.1142/S0129065722500459.

[10] R. Mergl *et al.*, “Kinematical analysis of handwriting movements in depressed patients,” *Acta Psychiatr Scand*, vol. 109, no. 5, pp. 383–391, May 2004, doi: 10.1046/j.1600-0447.2003.00262.x.

[11] K. Stollznow, “Graphology,” in *Language Myths, Mysteries and Magic*, K. Stollznow, Ed., London: Palgrave Macmillan UK, 2014, pp. 219–226. doi: 10.1057/9781137404862_22.

[12] K. Chaudhari and A. Thakkar, “Survey on handwriting-based personality trait identification,” *Expert Systems with Applications*, vol. 124, pp. 282–308, Jun. 2019, doi: 10.1016/j.eswa.2019.01.028.

[13] R. Klimoski and R. Anat, “Inferring personal qualities through handwriting analysis,” *Journal of Occupational Psychology*, vol. 56, pp. 191–202, Aug. 2011, doi: 10.1111/j.2044-8325.1983.tb00127.x.

[14] J. S. Buyukdura, S. M. McClintock, and P. E. Croarkin, “Psychomotor retardation in depression: Biological underpinnings, measurement, and treatment,” *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, vol. 35, no. 2, pp. 395–409, Mar. 2011, doi: 10.1016/j.pnpbp.2010.10.019.

[15] N. Ivančević, “Kinematic analysis of handwriting in neurological, psychiatric and neurodevelopmental disorders of childhood and adolescence,” PhD dissertation, Biomedical engineering and technologies, University of Belgrade, Belgrade, Serbia, 2021.

[16] G. Cordasco, F. Scibelli, M. Faundez-Zanuy, L. Likforman-Sulem, and A. Esposito, “Handwriting and Drawing Features for Detecting Negative Moods,” in *Quantifying and Processing Biomedical and Behavioral Signals*, A. Esposito, M. Faundez-Zanuy, F. C. Morabito, and E. Pasero, Eds., Cham: Springer International Publishing, 2019, pp. 73–86. doi: 10.1007/978-3-319-95095-2_7.

[17] A. Schröter, R. Mergl, K. Bürger, H. Hampel, H.-J. Möller, and U. Hegerl, “Kinematic analysis of handwriting movements in patients with Alzheimer’s disease, mild cognitive impairment, depression and healthy subjects,” *Dement Geriatr Cogn Disord*, vol. 15, no. 3, pp. 132–142, 2003, doi: 10.1159/000068484.

[18] S. Rosenblum, P. Werner, T. Dekel, I. Gurevitz, and J. Heinik, “Handwriting process variables among elderly people with mild Major Depressive Disorder: a preliminary study,” *Aging Clin Exp Res*, vol. 22, no. 2, pp. 141–147, Apr. 2010, doi: 10.1007/BF03324787.

[19] A. U. Rahman and Z. Halim, “Identifying dominant emotional state using handwriting and drawing samples by fusing features,” *Appl Intell*, vol. 53, no. 3, pp. 2798–2814, Feb. 2023, doi: 10.1007/s10489-022-03552-x.

[20] L. Likforman-Sulem, A. Esposito, M. Faundez-Zanuy, S. Clémençon, and G. Cordasco, “EMOTHAW: A Novel Database for Emotional State Recognition From Handwriting and Drawing,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 2, pp. 273–284, Apr. 2017, doi: 10.1109/THMS.2016.2635441.

[21] C. Greco *et al.*, “Discriminative Power of Handwriting and Drawing Features in Depression,” *Int. J. Neur. Syst.*, vol. 34, no. 02, p. 2350069, Feb. 2024, doi: 10.1142/S0129065723500697.

[22] V. Džepina *et al.*, “Detection of Psychomotor Retardation in Youth Depression: A Machine Learning Approach to Kinematic Analysis of Handwriting,” *Applied Sciences*, vol. 15, no. 14, Art. no. 14, Jan. 2025, doi: 10.3390/app15147634.

[23] Y. Saeys, I. Inza, and P. Larrañaga, “A review of feature selection techniques in bioinformatics,” *Bioinformatics (Oxford, England)*, vol. 23, no. 19, Oct. 2007, doi: 10.1093/bioinformatics/btm344.

[24] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, “Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products,” *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 83–90, Sep. 2006, doi: 10.1016/j.chemolab.2006.01.007.

[25] P. A. Devijver, *Pattern recognition : a statistical approach*. Englewood Cliffs, N.J. : Prentice/Hall International, 1982. Accessed: Jul. 26, 2025. [Online]. Available: http://archive.org/details/patternrecogniti00unse

[26] S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” Nov. 25, 2017, *arXiv*: arXiv:1705.07874. doi: 10.48550/arXiv.1705.07874.

[27] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005, doi: 10.1109/TPAMI.2005.159.

[28] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*, Third edition. in Wiley series in probability and statistics. Hoboken, New Jersey: Wiley, 2013.

[29] N. Ivančević *et al.*, “Does handedness matter? Writing and tracing kinematic analysis in healthy adults,” *Psihologija*, vol. 52, pp. 14–14, Jan. 2019, doi: 10.2298/PSI181229014I.

[30] V. Miler-Jerković, V. Kojić, and M. B. Popović, “An Information and Reliability Analysis of handwriting Kinematics,” presented at the 2nd International Conference on Electrical, Electronic and Computing Engineering IcETRAN, Silver Lake, Serbia, Jun. 2015, pp. 1–4.

[31] M. S. Bartlett and R. H. Fowler, “Properties of sufficiency and statistical tests,” *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences*, vol. 160, no. 901, pp. 268–282, Jan. 1937, doi: 10.1098/rspa.1937.0109.

[32] H. B. Mann and D. R. Whitney, “On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other,” *The Annals*

of Mathematical Statistics, vol. 18, no. 1, pp. 50–60, Mar. 1947, doi: 10.1214/aoms/1177730491.

[33] C. L. Cross and W. W. Daniel, “Analysis of Variance,” in *Biostatistics: A Foundation for Analysis in the Health Sciences, 11th Edition*, 11th ed., USA: Wiley, 2018, pp. 267–354.

[34] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene Selection for Cancer Classification using Support Vector Machines,” *Machine Learning*, vol. 46, no. 1, pp. 389–422, Jan. 2002, doi: 10.1023/A:1012487302797.

[35] C. Sobin and H. Sackeim, “Psychomotor symptoms of depression,” *The American journal of psychiatry*, vol. 154, pp. 4–17, Feb. 1997, doi: 10.1176/ajp.154.1.4.