2018 International Conference on Computational Science and Computational Intelligence (CSCI)

Optimizing support vector machine analysis in low density biological data sets

Pablo Rivas[1], Sharon Moore[2], Urszula T. Iwaniec[3], Russell T. Turner[3], Kathy Grant[4], and Erich Baker[2§]

***Abstract*— We explore the effectiveness of Support Vector Machines (SVM) for classification in a sparse data set. Non-human primate models are utilized to analyze Alcohol Use Disorders (AUDs); however, the resulting data have a limited sample size. The challenge of low sample numbers and low replicates are explored using a variety of optimization strategies for feature extraction, including correlation, entropy, density, linear support vector machines for regression (SVR), backward SVR, and forward SVR. We investigate these approaches against the backdrop of the relationship between alcohol consumption and tibial bone mineral density. The results indicate that machine learning (ML) can effectively be used in cases of low and diverse biological data sets. The best relevance feature ranking strategies are correlation, SVR forward, and SVR backward.**

I. INTRODUCTION

The rapid adoption of Machine Learning (ML) approaches to address issues in life science that do not respond well to classic statistical methods opens tremendous opportunities in data analytics. However, certain types of data sets are not optimally shaped to support soft learning due to high variance, low density, high dimensionality, and/or small sample size.

The interdependent characteristics that underlay Alcohol Use Disorders (AUD) spans numerous biological and behavioral disciplines and require deep understanding of influencing factors, such as sex, income, genetics, behavior, and co-morbidity with other disorders [1]. In human subjects these individual characteristics have uneven prevalence, intermittent responses, and low density, and represent the inherent difficulties with establishing robust longitudinal models [2].

Animal models have been used to compensate for these deficiencies, including the construction of a non-human primate (NHP) macaque model of oral alcohol self-administration. This model has the benefit of approximating human physiology and behavior. In addition, recent work has identified categorical levels of drinking severity, *i.e.*, Low Drinkers (LD), Binge Drinkers (BD), Heavy Drinkers (HD), and Very Heaver Drinkers (VHD) [3], that are a reflection of AUD severity in humans. This is particularly true with respect to HD and VHD categories since they are associated with problems of dependence and other brain pathology [4], [5], [6]. Unfortunately, one consequence of the NHP model is that it has low throughput, which impacts sample sizes and the effectiveness of ML-mediated analysis.

*This work is supported by NIAAA grant AA019431 and AA026289, and by the National Council for Science and Technology (CONACyT), Mexico, under grant 193324/303732 provided to PRP.

[1]School of Computer Science and Mathematics, Marist College, Poughkeepsie, NY, USA

[2]Department of Computer Science, Baylor University, Waco, TX, USA

[3]College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA

[4]Oregon National Primate Research Center, OHSU, Portland, OR, USA

§Corresponding author: Erich_Baker@Baylor.edu

II. CLASSIFIER CHOICE IN SPARSE, UNDER POWERED DATA SETS

While a range of ML applications have been used to address non-linear, binary, and multi-class classification problems in life science, there is a gap in the literature about the robustness of ML approaches in complex traits with limited sample sizes. Indeed, we have successfully used ML classification to identify strong behavioral contributors to drinking categories, but only after aggregating numerous animal cohorts [7]. There continues to be unknown effects of low replicate number in high degree feature space given behavioral data. We therefore examine the use of support vector machines for regression as a control model for prediction [8]. We use correlation, entropy, density, and forward-backward selection of features as indicators of the relationship between features and their relevance in modeling bone damage in primates [9]. This includes a ranking of feature selection methods to further identify which features are more predictive. In addition, we have chosen a single clinically relevant parameter to interrogate: bone mineral density (BMD). The study of the relationship between alcohol consumption patterns and BMD is an important initial step for understanding the effects of alcohol on bone.

In this paper we demonstrate that SVM can be employed to effectively classify parameters in low-density space. Whether the data is abundant or scarce, the mining and modeling of biological data is non-trivial because there is a risk of finding under-determined solutions or ill-posed models. This work illustrates potential solutions to feature selection under these constraints, and, consequently, demonstrate that the amount of alcohol consumed in the first two hour period of ethanol open access is the largest determining factor of BMD in monkeys intoxicated early in life.

III. METHODS

A. Primate Subjects

Postmortem tibial BMD was studied in 68 monkeys from the Oregon National Primate Research Center (ONPRC): 14 females and 54 males, representing Cynomolgus and Rhesus monkeys. Each animal was assigned a drinking category based on drinking behavior, see Table I [3].

978-1-7281-1360-9/18/$31.00 ©2018 IEEE
DOI 10.1109/CSCI.2018.00262

TABLE I

SUMMARY OF MONKEY COHORTS.

Cohort Name	Total	Females	Males	BD	LD	HD	VHD
INIA Cyno 2	12	0	12	0	10	0	1
INIA Cyno 9	11	0	11	0	6	2	0
INIA Cyno 8	3	3	0	0	0	0	0
INIA Rhesus 7a	8	0	8	1	3	2	2
INIA Rhesus 7b	5	0	5	1	3	1	0
INIA Rhesus 4	10	0	10	4	5	1	0
INIA Rhesus 5	8	0	8	1	0	3	4
INIA Rhesus 6a	6	6	0	0	0	0	6
INIA Rhesus 6b	5	5	0	0	3	1	1
Total	68	14	54	7	30	10	14

B. Ethanol Self-Administration

All animals followed previously established procedures for *schedule-induced polydipsia* [10], [11]. After an induction period where all monkeys drank to levels that saturated metabolic capacity elevating their Blood Ethanol Content (BEC) over 50 mg/dl, equivalent to the 30$^{\text{th}}$ session of 1.5 g/kg ethanol, animals had concurrent access to 4% w/v ethanol and a 22 h/d *open-access* to water and food in the form of 1g of banana flavored pellets (Noyes), provided at least three times a day during meals, and at least two hours apart. The open-access phase comprised 12 months.

C. Drinking Features for Analysis

Five attributes are analyzed to explore optimal SVM conditions for sparse data sets: drinking category, sex, age at first intoxication [12], [13], maximum bout volume, and ethanol consumption during induction. These produce a total of 14 features, see Table II. For example, from the maximum bout volume occurring during the first 120 minutes of daily open-access, five features are extracted: average maximum bout volume (μ), standard deviation of the maximum bout volume (σ), median of the maximum bout volume, maximum of the maximum bout volume ($\max$), and minimum of the maximum bout volume ($\min$). This amount is also measured as a fraction of what the animals drink. From this data six features are extracted: average percentage of ethanol consumed during induction (μ), median percentage of ethanol consumed during induction, total sum of the percentage of ethanol ingested during induction (Σ), standard deviation of the percentage of ethanol ingested during induction (σ), minimum percentage of ethanol consumed during induction ($\min$), and maximum percentage of ethanol consumed during induction ($\max$).

D. Estimation of Importance and Feature Ranking

Feature ranking is an approach to feature selection that reduces the risk of poor performance and high complexity [9]. Three methods are used to evaluate features based on individual relevance ranking: Pearson correlation, entropy, and density. Three methods are also used to evaluate combined relevance rankings based on the well-known support vector machines for regression (SVRs) [8]: linear SVR, backward SVR, and SVR forward.

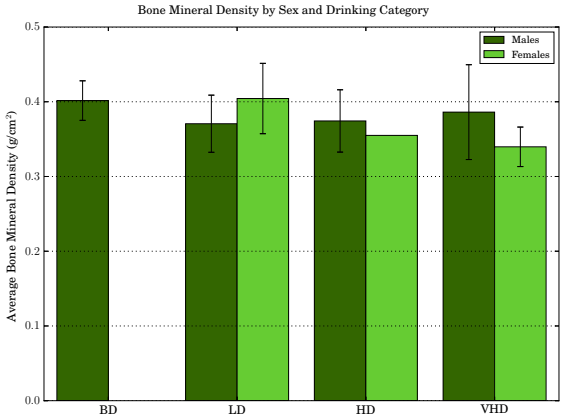

Fig. 1. *Tibial bone mineral density by drinking category and sex.* Drinking categories are displayed along the horizontal axis: low drinker (LD), binge drinker (BD), heavy drinker (HD), and very heavy drinker (VHD). No females are reported as BDs. There is no significant difference between drinking categories or sex.

1) Individual Relevance Ranking using Entropy: Entropy is a measure of the amount of information that is contained in a random variable or feature [14]. If a feature is categorical, *e.g.*, drinking category or sex, we expect entropy based on the distribution of those variable. *Sex* represents a binary category and low entropy, indicating high predictability. On the other hand, the entropy of *drinking category* is more evenly distributed and therefore has higher entropy and lower predictability. Note that calculating entropy requires no knowledge of the target variable Y, which makes this process faster to compute, in comparison with Pearson correlation coefficient, but of the same overall complexity: $\mathcal{O}(mn)$.

2) Individual Relevance Ranking using Density: High density features are highly correlated with many other variables, although not necessarily redundant [9], and may be calculated using the differences of Pearson correlation coefficients. This is important in the study of complex trait disorders, such as AUD, where numerous factors impact outcome [10], [3], [15], [7]. High density features may be calculated as a cost of increased complexity, $\mathcal{O}(m^2n)$, but using computed averages and centered vectors leads to a complexity of $\mathcal{O}(mn)$ in the average case.

3) Combined Relevance Ranking: SVRs have been extensively developed to reduce the high complexity of learning algorithms when kernel mappings of data are used [8] and in problems related to classification [16], [17], [18]. Similarly, they can be used for regression tasks with minor adjustments. The complexity of finding a solution to the SVR optimization problem, using the SMO algorithm with a linear kernel, is $\mathcal{O}(m^2n)$, without considering the cost of the grid search.

E. Feature Selection

We explore two major strategies for ranking sets of features: forward selection and backward selection [9]. Backward selection considers the set of all features first and removes features that produce the smallest contribution. It is

also computationally expensive, pushing the complexity of SVR to $\mathcal{O}(m^3 n)$ due to the kernel function. And since the process of having sets of $n-1, n-2, \ldots, 1$ features selected corresponds to a series that converges to $n(n-1)/2 = \mathcal{O}(n^2)$, then the total overall complexity becomes $\mathcal{O}(m^3 n^2)$. This complexity does not consider the cost of grid search and leave-one out cross validation.

The forward selection of features begins with an empty set and adds individual features one by one, selecting the feature that provided the best score. This process is fast at the beginning since it starts with an empty set, however, it could miss holistic relationships among features, and is computationally expensive, with a complexity of $\mathcal{O}(m^3 n)$.

F. Computational Resources for Analysis

All computations were executed on Monkey Alcohol and Tissue Research Resource (MATRR) [19] servers (four Intel Xeon E5620 processors at 2.4 GHz, with 47 GB of RAM and 1.7 TB storage). Statistical analysis and data processing was completed using Python's Pandas, NumPy, and SciPy packages. For all ML algorithms we used Python's Scikit-Learn package.

IV. RESULTS AND DISCUSSION

The identification of features associated with behavioral drinking patterns in a NHP open access drinking model are used to explore the effectiveness of SVM and associated feature selection models in low data space. Given several cohorts of monkeys and data in the MATRR, the identification of features that impact BMD are explored using a variety of methods.

A. Tibial Bone Mineral Density

Average BMD of all samples is found to be $\mu = 0.376$ g/cm^2, with a standard deviation of $\sigma = 0.049$g/cm^2. Figure 1 depicts BMD broken down by gender and drinking category, and suggests no apparent BMD differences between drinking categories. Sex differences are also within the standard deviation. There is likewise no statistical correlation between BMD and other examined single features.

B. Feature Ranking

Each of the fourteen available features in our data set are ranked using each of the ranking methodologies, see Table II. However, given the small sample size, average rank may be misleading. To establish a more robust ranking approach, a Friedman test is performed [20]. The Friedman statistic is determined to be $\chi_F^2 = 17.3143$ with $p = 0.1853$, with a level of significance of $\alpha = 0.05$. Furthermore, the critical difference δ_{CD} using the Nemenji's test was calculated to determine if two features are significantly different if the corresponding average ranks differ [21]. The value of the critical difference for a confidence level of $\alpha = 0.01$ corresponds to $\delta_{CD} = 1.82$. Using the δ_{CD} value we observe that the top four features are not significantly different from each other, but are significantly better than the rest of the features, suggesting that the most contributing features are

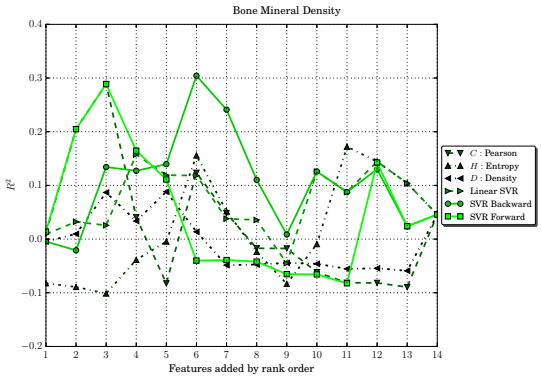

Fig. 2. *Performance of feature sets using the R^2 score.* Pearson correlation coefficient, SVR backward, and SVR forward score the highest among all of the methods.

Drinking Category, *Sex*, *Age at First Intoxication*, μ *of Maximum Bout Volume*, σ *of Maximum Bout Volume*.

Feature Testing

Relevant features are tested based on size-increasing sets according to rank. The input to the algorithm is the data set $\mathcal{D}$ and a vector $\mathbf{r} = [r_1, r_2, \ldots, r_n]$, where r_1 is the index of the best ranked feature and r_n is the index of the worst ranked feature. Note, however, that $\mathbf{r}$ is different for each ranking methodology. *E.g.*, consider the column entitled "Pearson" in Table II, the corresponding $\mathbf{r}$ would be: $\mathbf{r} = [3, 2, 6, 5, 4, \ldots, 1]$

We perform a traditional grid search to find the best set of hyper-parameters $\{C, \gamma, \nu\}$ that produce the best 10-fold cross validation R^2 score. The search is conducted in the logarithmic spaces $C \in \{2^{-5}, 2^{-4}, \ldots, 2^{15}\}$ and $\gamma \in \{2^{-15}, 2^{-14}, \ldots, 2^{3}\}$; and the linear space $\nu \in \{0.05, 0.10, \ldots, 0.95\}$. Then, with the best set of hyper-parameters, $\nu-$SVRs are trained with a leave-one-out (LOO) cross validation strategy, which is proven to provide more accurate error estimates in smaller data sets [22].

Results for all of the relevant feature ranking methods are depicted in Figure 2, which plots all error vectors starting with one feature. Note that since the first feature added, *i.e.*, the best, is distinct for every method, the initial score varies. However, all converge to the same score once all fourteen features are considered together. Ideally, we want the set of features that achieves the highest score with the smallest number of features. SVR forward and Pearson correlation coefficient achieve their highest score with only three features, while SVR backward performs highest with six features.

In order to determine which features appear the most frequently, we analyze which features are in the set that produced the highest score for each method. Table III presents a summary of which features are frequently part of the best set of features. The table also presents a weighted frequency count using the rank of each method. The rank was determined as the best overall score of a given method regardless of the number of features. From Figure 2 it can be

TABLE II

RANKING OF INFORMATIVE FEATURES. EACH COLUMN SHOWS A RANKING METHODOLOGY. THE CRITICAL DIFFERENCE, $\delta_{CD} = 1.82$ WITH $\alpha = 0.01$, SUGGESTS THAT THE TOP FIVE RANKED FEATURES ARE SIGNIFICANTLY BETTER THAN THE REST.

$\mathcal{F}$	Feature	*Pearson*	*Entropy*	*Density*	Linear SVR	SVR Backward	SVR Forward	Avg. Rank
1	Drinking Category	14	11	4	8	6	12	9.16
2	Sex	2	13	5	13	13	1	7.83
3	Age at First Intoxication	1	6	2	7	7	2	4.16
4	μ of Maximum Bout Vol.	5	3	14	1	5	4	5.33
5	σ of Maximum Bout Vol.	4	2	12	3	3	11	5.83
6	Median of Maximum Bout Vol.	3	4	10	2	4	3	4.33
7	max of Maximum Bout Vol.	11	8	7	10	8	9	8.83
8	min of Maximum Bout Vol.	7	5	3	12	14	14	9.16
9	μ % of EtOH During Ind.	13	10	13	6	1	13	9.33
10	Median % of EtOH During Ind.	9	12	8	5	9	10	8.83
11	Σ % of EtOH During Ind.	6	7	9	4	2	8	6.00
12	σ % of EtOH During Ind.	10	1	11	9	10	6	7.83
13	min % of EtOH During Ind.	8	9	6	11	11	5	8.33
14	max % of EtOH During Ind.	12	14	1	14	12	7	10.00

TABLE III

ANALYSIS OF SETS OF FEATURES PRODUCING THE BEST SCORE. COLUMN DATA IS FORMATTED AS b/r, WHERE $b = 1$ IF FEATURE IS PART OF THE BEST SCORING SET; r IS THE METHOD RANK.

$\mathcal{F}$	Feature	*C*	*H*	*D*	Lin. SVR	SVR Bwd.	SVR Fwd.	Freq. ($w_{\text{Freq.}}$)
1	Drinking Category		1/4	1/6		1/1		3 (1.42)
2	Sex	1/2.5		1/6			1/2.5	3 (0.97)
3	Age at First Intoxication	1/2.5	1/4	1/6			1/2.5	4 (1.22)
4	μ of Maximum Bout Vol.		1/4		1/5	1/1		3 (1.45)
5	σ of Maximum Bout Vol.		1/4		1/5	1/1		3 (1.45)
6	Median of Maximum Bout Vol.	1/2.5	1/4		1/5	1/1	1/2.5	5 (2.25)
7	max of Maximum Bout Vol.		1/4					1 (0.25)
8	min of Maximum Bout Vol.		1/4	1/6				2 (0.42)
9	μ % of EtOH During Ind.		1/4			1/1		2 (1.25)
10	Median % of EtOH During Ind.							
11	Σ % of EtOH During Ind.		1/4		1/5	1/1		3 (1.45)
12	σ % of EtOH During Ind.		1/4					1 (0.25)
13	min % of EtOH During Ind.		1/4					1 (0.25)
14	max % of EtOH During Ind.			1/6				1 (0.17)

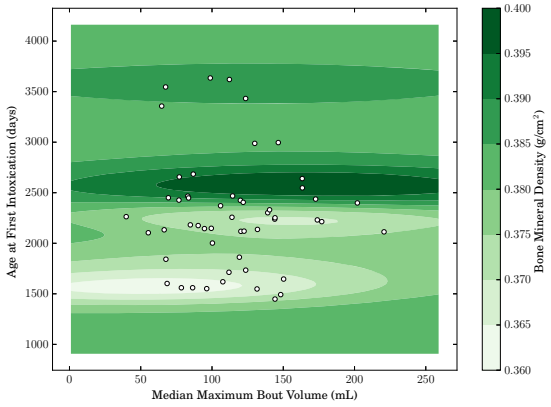

Fig. 3. *BMD values using Median Maximum Bout Volume and Age at First Intoxication as predictors.* Predicted BMD suggests that young monkeys are at higher risk of low BMDs if their *Median Maximum Bout Volume* is in the range of 10 to 150 mL.

determined that the best score corresponds to SVM backward and the worst corresponds to Density.

Careful inspection of Table III reveals that, considering the frequency count of the features, the two features that are frequently part of the best sets of features are *Median of Maximum Bout Volume* and *Age at First Intoxication*. On the other hand, if we consider the weighted frequency count which utilizes the rank, we observe that the top features are *Median of Maximum Bout Volume* with a $w_{\text{Freq.}} = 2.25$ and three ties with $w_{\text{Freq.}} = 1.45$.

Figure 3 depicts how two features, namely *Median Maximum Bout Volume* and *Age at First Intoxication*, interact with each other in predicting BMD. The figure suggests that lower BMD is associated with a median maximum bout volume between 10 and 150 mL for young monkeys whose age is between 1400 and 1700 days. The risk of having low BMD is less for older monkeys of ages greater than 2400 days. These two features do not appear to have a linear relationship in a two-dimensional plane; however, using the kernel method in SVRs, the possibility of having a linear relationship in a higher-dimensional space is often assumed.

A similar analysis is shown in Figure 4. In this case the two features analyzed are *Median Maximum Bout Volume* and *Total Sum of the Percentage of Ethanol Consumed During Induction*. The relationship between these two features appears to be quasi-linear, suggesting a more evident dependence or correlation. From the figure, we can observe that if the *Median Maximum Bout Volume* is less than 100 mL then there is a higher risk of low BMD for almost any sum of the percentage of ethanol, while the risk decreases in the opposite direction.

V. CONCLUSIONS

Several predictive features for bone mineral density were identified using relevance feature ranking techniques. These attributes are both categorical and quantitative, and include *Median of Maximum Bout Volume*, *μ of Maximum Bout Volume*, *σ of Maximum Bout Volume*, *Age at First Intoxication*, and *Total Sum of the Percentage of Ethanol Consumed During Induction*

We also analyzed the complexity of the feature ranking algorithms and observed that the Pearson correlation coefficient is significantly less expensive and performs well for up to three features; after that, Pearson does not consider the

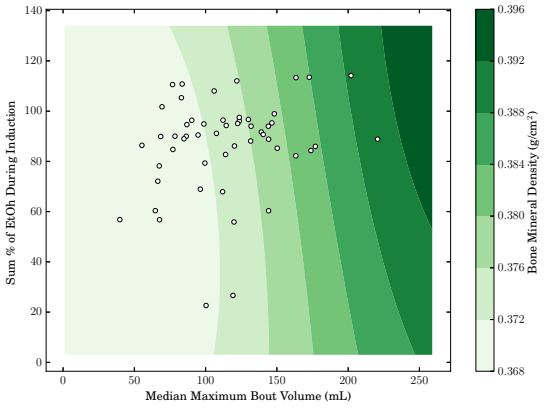

Fig. 4. *BMD values using Median Maximum Bout Volume and Total Sum of the percentage of Ethanol During Induction.* Results suggest that Median Maximum Bout Volume provides a higher correlation to BMD than the percentage of Ethanol intake during Induction.

interdependence of the data. The SVR Backward selection strategy was the best method, since it considers all of the features at once and their relationship as a group; however, it is more expensive than Pearson's correlation coefficient.

Encountering low sample sizes in certain biological data sets is inevitable, and careful application of optimization techniques can ensure that these sets maintain their utility. This approach also demonstrates an alternative to sentiment analysis [23].

REFERENCES

[1] B. F. Grant, R. B. Goldstein, T. D. Saha, S. P. Chou, J. Jung, H. Zhang, R. P. Pickering, W. J. Ruan, S. M. Smith, B. Huang *et al.*, "Epidemiology of dsm-5 alcohol use disorder: Results from the national epidemiologic survey on alcohol and related conditions iii," *JAMA psychiatry*, vol. 72, no. 8, pp. 757–766, 2015.

[2] B. F. Grant, F. S. Stinson, D. A. Dawson, S. P. Chou, M. C. Dufour, W. Compton, R. P. Pickering, and K. Kaplan, "Prevalence and co-occurrence of substance use disorders and independentmood and anxiety disorders: Results from the national epidemiologic survey on alcohol and relatedconditions," *Archives of general psychiatry*, vol. 61, no. 8, pp. 807–816, 2004.

[3] E. J. Baker, J. Farro, S. Gonzales, C. Helms, and K. A. Grant, "Chronic alcohol self-administration in monkeys shows long-term quantity/frequency categorical stability," *Alcoholism: Clinical and Experimental Research*, vol. 38, no. 11, pp. 2835–2843, 2014.

[4] C. D. Kroenke, T. Rohlfing, B. Park, E. V. Sullivan, A. Pfefferbaum, and K. A. Grant, "Monkeys that voluntarily and chronically drink alcohol damage their brains: a longitudinal mri study," *Neuropsychopharmacology*, vol. 39, no. 4, pp. 823–830, 2014.

[5] G. Chen, V. C. Cuzon Carlson, J. Wang, A. Beck, A. Heinz, D. Ron, D. M. Lovinger, and K. J. Buck, "Striatal involvement in human alcoholism and alcohol consumption, and withdrawal in animal models," *Alcoholism: Clinical and Experimental Research*, vol. 35, no. 10, pp. 1739–1748, 2011.

[6] C. A. Siciliano, E. S. Calipari, V. C. C. Carlson, C. M. Helms, D. M. Lovinger, K. A. Grant, and S. R. Jones, "Voluntary ethanol intake predicts κ-opioid receptor supersensitivity and regionally distinct dopaminergic adaptations in macaques," *The Journal of Neuroscience*, vol. 35, no. 15, pp. 5959–5968, 2015.

[7] B. E. J., W. N. A.R., S. Alex, R. P. Pablo, M. Sharon, G. Steven, and G. K. A., "Identifying future drinkers: Behavioral analysis of monkeys initiating drinking to intoxication is predictive of future drinking classification," *Alcoholism: Clinical and Experimental Research*, vol. 41, no. 3, pp. 626–636. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/acer.13327

[8] P. Rivas-Perea, J. Cota-Ruiz, and J. G. Rosiles, "An algorithm for training a large scale support vector machine for regression based on linear programming and decomposition methods," *Pattern Recognition Letters*, vol. 2013, no. 34, pp. 439–451, 2013.

[9] I. Guyon and A. Ekusseeff, "An introduction to feature extraction," in *Feature extraction: foundations and applications*, I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds. New York: Springer, 2008, vol. 207, pp. 1–25.

[10] K. A. Grant, X. Leng, H. L. Green, K. T. Szeliga, L. S. Rogers, and S. W. Gonzales, "Drinking typography established by scheduled induction predicts chronic heavy drinking in a monkey model of ethanol self-administration," *Alcoholism: Clinical and Experimental Research*, vol. 32, no. 10, pp. 1824–1838, 2008.

[11] J. Vivian, H. Green, J. Young, L. Majerksy, B. Thomas, C. Shively, J. Tobin, M. Nader, and K. Grant, "Induction and maintenance of ethanol self-administration in cynomolgus monkeys (macaca fascicularis): Long-term characterization of sex and individual differences," *Alcoholism: Clinical and Experimental Research*, vol. 25, no. 8, pp. 1087–1097, 2001.

[12] T.-K. Li, B. G. Hewitt, and B. F. Grant, "The alcohol dependence syndrome, 30 years later: a commentary," *Addiction*, vol. 102, no. 10, pp. 1522–1530, 2007.

[13] D. A. Dawson, R. B. Goldstein, S. Patricia Chou, W. June Ruan, and B. F. Grant, "Age at first drink and the first incidence of adult-onset dsm-iv alcohol use disorders," *Alcoholism: Clinical and Experimental Research*, vol. 32, no. 12, pp. 2149–2160, 2008.

[14] K. Torkkola, "Information-theoretic methods," in *Feature extraction: foundations and applications*, I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds. New York: Springer, 2008, vol. 207, pp. 167–185.

[15] A. Salo, E. Baker, K. Grant, and P. Rivas-Perea, "Identifying future drinkers: Behavioral analysis of monkeys initiating drinking to intoxication is predictive of future drinking classification," *Nat Genet*, vol. 13, pp. 266–267, 2015.

[16] Y.-W. Chen and C.-J. Lin, "Combininb svms with various feature selection strategies," in *Feature extraction: foundations and applications*, I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, Eds. New York: Springer, 2008, vol. 207, pp. 315–323.

[17] A. Rakotomamonjy, "Variable selection using svm-based criteria," *Journal of Machine Learning Research*, vol. 3, pp. 1357–1370, 2003.

[18] Y.-W. Chang and C.-J. Lin, "Feature ranking using linear svm," *Causation and Prediction Challenge Challenges in Machine Learning*, vol. 2, p. 47, 2008.

[19] J. B. Daunais, A. T. Davenport, C. M. Helms, S. W. Gonzales, S. E. Hemby, D. P. Friedman, J. P. Farro, E. J. Baker, and K. A. Grant, "Monkey alcohol tissue research resource: banking tissues for alcohol research," *Alcoholism: Clinical and Experimental Research*, vol. 38, no. 7, pp. 1973–1981, 2014.

[20] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.

[21] S. García, A. Fernández, J. Luengo, and F. Herrera, "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power," *Information Sciences*, vol. 180, no. 10, pp. 2044–2064, 2010.

[22] T.-T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, 2015.

[23] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, ser. EMNLP '02. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002, pp. 79–86. [Online]. Available: https://doi.org/10.3115/1118693.1118704