Semi-Supervised Machine Learning Enables the Robust Detection of Multireference Character at Low Cost

Chenru Duan[1,2], Fang Liu[1], Aditya Nandy[1,2], and Heather J. Kulik[1,*]

[1]*Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139*

[2]*Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139*

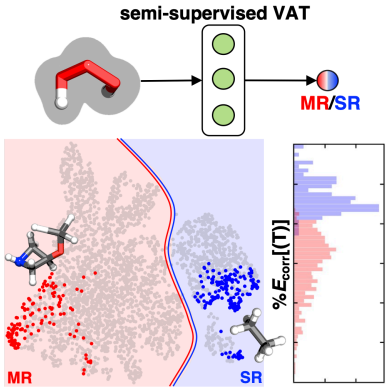
AUTHOR INFORMATION

Corresponding Author

*email: hjkulik@mit.edu, phone: 617-253-4584

ABSTRACT: Multireference (MR) diagnostics are common tools for identifying strongly correlated electronic structure that makes single reference (SR) methods (e.g., density functional theory or DFT) insufficient for accurate property prediction. However, MR diagnostics typically require computationally demanding correlated wavefunction theory (WFT) calculations, and diagnostics often disagree or fail to predict MR effects on properties. To overcome these challenges, we introduce a semi-supervised machine learning (ML) approach with virtual adversarial training (VAT) of an MR classifier using 15 WFT and DFT MR diagnostics as inputs. In semi-supervised learning, only the most extreme SR or MR points are labeled, and the remaining point labels are learned. The resulting VAT model outperforms the alternatives, as quantified by the distinct property distributions of SR- and MR-classified molecules. To reduce the cost of generating inputs to the VAT model, we leverage the VAT model's robustness to noisy inputs by replacing WFT MR diagnostics with regression predictions in a MR decision engine workflow that preserves excellent performance. We demonstrate the transferability of our approach to larger molecules and those with distinct chemical composition from the training set. This MR decision engine demonstrates promise as a low-cost, high-accuracy approach to the automatic detection of strong correlation for predictive high-throughput screening.

TOC GRAPHICS



High-throughput computational chemistry has emerged as an essential tool, especially in the generation of data for chemical discovery[1-6] and for machine learning[7-9] model training[10-16]. Nevertheless, most such efforts[1-6, 17-23] involve black-box, single-reference (SR) DFT or correlated wavefunction theory (WFT). Although progress has been made[24-27] in the automation of parameter (e.g., active space) selection in multireference (MR) methods, MR methods typically require expert knowledge in their application. It would be desirable to have automated tools[9, 18-19, 28-31] that can identify molecules in high-throughput workflows[9] that have significant MR character[32-44] and therefore must be studied with MR methods, or alternatively can identify if an SR approach is suitable.

We recently assembled[45] a dataset of 15 MR diagnostics for 3,165 small organic molecules (*AD-3165*) in equilibrium and randomly or maximally distorted (i.e., both stretched and compressed) geometries[45-46]. The MR diagnostics were evaluated with both DFT[33, 40, 42-44, 47] and WFT (i.e., MP2[40, 48-49], CASSCF[34-37, 40-41], and CCSD[34, 38-39, 50], Table 1). For the 11 diagnostics with literature-recommended cutoffs[34-42, 50-51], classification of MR versus SR character showed significant disagreement (Table 1 and Supporting Information Tables S1–S2). Most CCSD and CAS diagnostics predicted a significant fraction (> 1/3 with $C_0^2 < 0.9$) of molecules to be MR, but few DFT or MP2 diagnostics did (1% with n_{HOMO}[MP2] < 1.9, Figure 1). As could be expected[40, 47, 51], few diagnostics correlated with each other, and the best agreement was observed between diagnostics that probed similar quantities (i.e., occupations or the total atomization energy, TAE[40, 42, 50]) rather than those that used the same level of theory.[45]

Table 1. Summary of MR diagnostics grouped by type and method used

Type	**Method**	**Diagnostic**	**Description**
TAE	DFT	B_1[42], A_{25}[PBE][40]	Differences in TAE with Hartree–Fock exchange fraction
TAE	CCSD(T)	%TAE[(T)][50]	Differences in TAE from (T) term in CCSD(T)

excitations	CCSD(T)	T_1[34], D_1[38], and D_2[39]	Average and maximum singles amplitude and doubles amplitude
occupations	DFT	I_{ND}[PBE][43-44], r_{ND}[PBE][47], I_{ND}[B3LYP][43-44], r_{ND}[B3LYP][47],	Finite-temperature DFT orbital occupations
occupations	MP2	n_{HOMO}[MP2][40, 48], n_{LUMO}[MP2][40, 48]	MP2 natural orbital occupations
occupations	CASSCF	n_{HOMO}[CAS][40-41], n_{LUMO}[CAS][40-41], C_0^2 [34-37]	CAS natural orbital occupations and leading weight

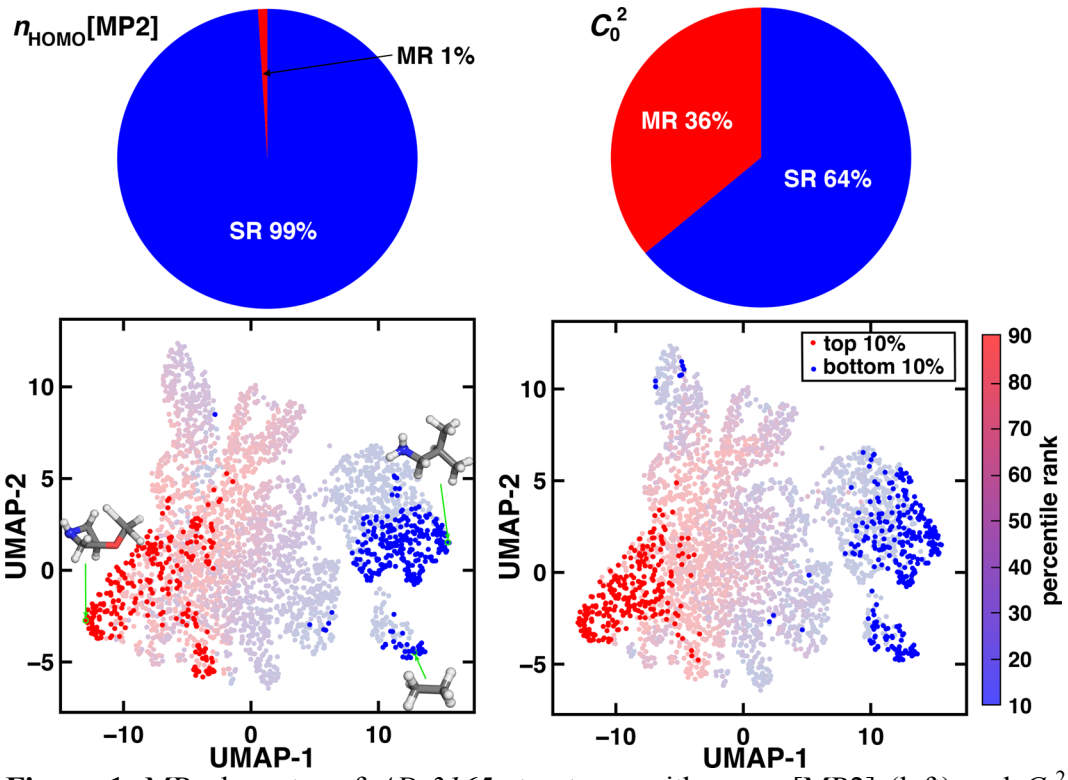


Figure 1. MR character of *AD-3165* structures with n_{HOMO}[MP2] (left) and C_0^2 (right). (Top) Percentage of structures classified as single reference (SR, blue) or multi-reference (MR, red) according to the diagnostic cutoffs. (Bottom) Uniform manifold approximation and projection (UMAP)[52] of all 15 MR diagnostics with the structures in the top 10% (red) and bottom 10% (blue) of the relevant MR diagnostic (n_{HOMO}[MP2] at left and C_0^2 at right) shown as solid red and blue circles, respectively. The remaining data is shown as translucent and colored according to the colorbar at right. A representative MR structure (3-methoxyazetidine) and two SR structures (equilibrium ethane and isobutylamine) are shown inset on the left UMAP plot. Atoms in structures are colored as follows: carbon in gray, hydrogen in white, nitrogen in blue, and oxygen in red.

To ascertain which diagnostics were better at diagnosing MR effects[47, 53], we quantified[45] how well they predicted the difference in recovery of the correlation energy between CCSD and CCSD(T):

$$\%E_{\mathrm{corr}}[(\mathrm{T})] = 100 \times \frac{E[\mathrm{CCSD}] - E[\mathrm{HF}]}{E[\mathrm{CCSD(T)}] - E[\mathrm{HF}]} \quad (1)$$

which we selected for its good agreement with differences between CCSD and CCSDT correlation energies and weak size dependence (Supporting Information Figure S1). Nearly all WFT-based MR diagnostics were better predictors of $\%E_{\mathrm{corr}}[(\mathrm{T})]$ than any of the computationally affordable DFT-based diagnostics, even for the WFT-based diagnostics that were not suitable for cutoff-based MR classification.[45] Although MR diagnostics often disagree, non-linear dimensionality reduction[52] confirmed regions of consensus[45] for strongly SR molecules such as saturated alkanes in equilibrium or MR stretched (e.g., 3-methoxyazetidine) or conjugated molecules (Figure 1).

Consensus among diagnostics suggests that machine learning (ML) models trained on a range of MR diagnostics should be able to overcome present limitations for the automated detection of MR effects (i.e., low $\%E_{\mathrm{corr}}[(\mathrm{T})]$). In choosing an ML approach to apply to this task, supervised learning is not suitable because unambiguous assignment (i.e., MR or SR) is only possible for a small fraction of structures (Figure 1). Conversely, unsupervised learning (e.g., clustering) is expected to struggle[54-55] to uncover the MR versus SR boundary. Here, we develop a fully automated, low-cost, and transferable ML decision engine for MR character classification using semi-supervised learning[56]. As its name suggests, semi-supervised learning requires a fraction of labeled data (i.e., as in supervised learning) to aid illumination of the underlying distribution (i.e., as in unsupervised learning) for the remaining data. We employ virtual adversarial training[57] (VAT) as our semi-supervised learning algorithm, a technique that has demonstrated best-in-class performance in image classification[58] by reducing sensitivity to feature space perturbations[59]. Because VAT only requires augmentation of the supervised learning loss function with an unsupervised term, it can be employed with standard artificial

neural network (ANN) architectures (here, a fully-connected, ANN classifier, see Computational Details and Supporting Information Text S1).

This application of semi-supervised learning to chemistry exploits the robustness of VAT to noisy inputs and demonstrates good performance on a combination of DFT-based and ML-predicted[45] WFT-based diagnostics that ensure modest computational cost. Because VAT model training requires only a fraction of training data (here, molecular geometries) to have assigned labels (here, MR vs SR) for the supervised loss term, we are able to leverage the general consensus among MR diagnostics for the most extreme cases. We compute all 15 MR diagnostics and label structures as MR or SR only when they are in the top or bottom 10% for more than half ($\geqslant$8) of MR diagnostics over the ranges in the training data (Figure 1 and see Computational Details and Supporting Information Figure S2).

To evaluate VAT model performance, we compare to two alternatives: i) a conventional cutoff approach and ii) unsupervised learning with clustering. For the cutoffs, we use a previously recommended combination[51] of four CCSD(T)-based diagnostics, i.e., $D_1 > 0.05$, $T_1 > 0.02$, $D_2 > 0.18$, and %TAE[(T)] > 5% that we augment with a recommended CASSCF leading weight cutoff[34-37, 51], $C_0^2 < 0.9$ (Table 1). We classify a geometry as MR if any of the five diagnostics exceeds cutoff. The clustering approach uses all MR diagnostics to generate one SR and one MR cluster for the binary classification task (see Computational Details). We compare the classification of *AD-3165* test set geometries from these three approaches to the computed $\%E_{corr}$[(T)], which is not provided as input to the models. For a model to be predictive, it should classify input geometries with low $\%E_{corr}$[(T)] as MR, those with high $\%E_{corr}$[(T)] as SR, and transition smoothly from MR to SR classification with increasing $\%E_{corr}$[(T)].

The VAT model's classification of the *AD-3165* test set distinguishes well between

low %E_{corr}[(T)], MR geometries and high %E_{corr}[(T)], SR geometries, transitioning at around 96% (Figure 2). The ca. 120 structures that the VAT classifies as SR are consistent with intuition, i.e., primarily equilibrium or randomly distorted molecules (i.e., only one is in a maximum-energy structure) with relatively saturated bonds (e.g., cyclopropane in Figure 2). The cutoff approach is qualitatively consistent with the VAT model, but its transition region is wider and less smooth (i.e., more structures with similar %E_{corr}[(T)] are classified as both SR and MR, Figure 2). Still, cutoffs misclassify[45] molecules we would expect to be labeled MR such as N_2 or propyne as well as distorted geometries of conjugated molecules (i.e., pyrrole, methanimidamide, or 1-ethylaziridine) because their diagnostics approach but do not exceed recommended cutoff values (Supporting Information Table S3).

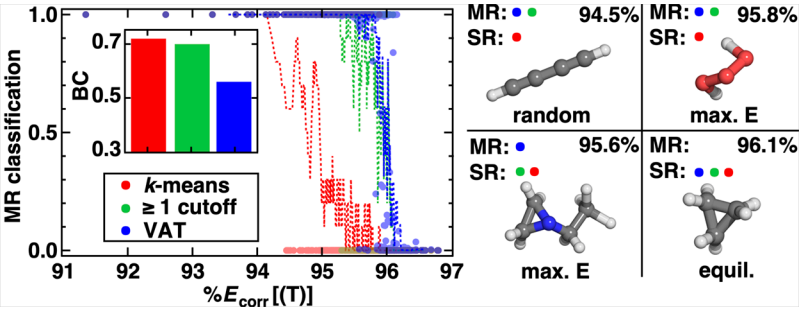


Figure 2. (left) MR classification (0 for SR and 1 for MR) versus %E_{corr}[(T)] for each *AD-3165* test prediction by *k*-means (red translucent circles), the cutoff model (green translucent circles), or VAT model (blue translucent circles) along with a 10-point moving average colored the same as individual symbols, shown inset legend. The Bhattacharyya coefficient (BC) for each model is shown in the inset bar graph for *k*-means in red, cutoff in green, and VAT in blue. (right) Representative structures and their type (maximum energy, max. E; random; or equilibrium, equil.) shown at bottom from *AD-3165* along with the model classification (SR or MR) for each model type (colored as in the left inset legend) along with the %E_{corr}[(T)] for the structure shown at top right. The top left quadrant shows a randomly distorted 1,3-butadiyne, the bottom left quadrant shows a max. E 1-ethylaziridine, top right quadrant shows a max. E trioxidane, and the bottom right quadrant shows an equilibrium cyclopropane. Atoms in structures are colored as follows: carbon in gray, hydrogen in white, nitrogen in blue, and oxygen in red.

Unlike cutoffs or the VAT, unsupervised learning models, i.e., *k*-means and agglomerative clustering (AC), do not transition from MR to SR smoothly with %E_{corr}[(T)] (Figure 2 and Supporting Information Figure S3). The clustering models label too many

structures as SR, including structures that are expected to have MR character (e.g., stretched 1,3-butadiyne in Figure 2). Thus, the *k*-means model only achieves consistent agreement with the VAT or cutoff approach in cases of extreme, unambiguous SR character (e.g., cyclopropane in Figure 2). At the same time, the few structures the *k*-means model has labeled as MR have a large range of $\%E_{corr}[(T)]$ values, indicating the underlying data structure has not been learned (Supporting Information Figure S4).

Because there is no *a priori* cutoff value of $\%E_{corr}[(T)]$ that corresponds to a ground-truth MR or SR assignment against which we can judge model classification, alternate means are used to quantify model performance. The best-performing models should have maximally distinct distributions of $\%E_{corr}[(T)]$ values in their SR vs MR classes (Supporting Information Figure S4). As a measure of the amount of overlap between statistical distributions in two classes, we use the Bhattacharyya coefficient (BC), which we evaluate assuming normal distributions[60]:

$$BC(\mathrm{SR},\mathrm{MR})=\exp\left(-\frac{1}{4}\left(\ln\left(\frac{\sigma_{\mathrm{SR}}^2}{\sigma_{\mathrm{MR}}^2}+\frac{\sigma_{\mathrm{MR}}^2}{\sigma_{\mathrm{SR}}^2}+2\right)+\frac{\left(\mu_{\mathrm{SR}}-\mu_{\mathrm{MR}}\right)^2}{\sigma_{\mathrm{SR}}^2+\sigma_{\mathrm{MR}}^2}\right)\right) \quad (2)$$

Here, σ^2 and μ are the variance and mean of the SR or MR $\%E_{corr}[(T)]$ distributions, and lower BC values indicate greater dissimilarity (Supporting Information Text S2 and Table S4). The VAT model has a much lower BC of 0.55 in comparison to both unsupervised models (*k*-means: 0.72 AC: 0.76), consistent with qualitative observations (Figure 2 and Supporting Information Figures S3–S4 and Tables S4–S5). Although the cutoff approach yields a smoother SR-to-MR transition than clustering, its BC of 0.70 is significantly higher than the VAT model (Figure 2). Attempts to improve the cutoff approach by requiring multiple diagnostics to exceed cutoffs, as has sometimes been recommended[39, 51, 61], reduce the cutoff model's BC slightly (0.67) but do not recover the VAT model performance (Supporting Information Tables S5–S10 and Figure

S3). If we choose a $\%E_{corr}[(T)]$ cutoff to distinguish ground-truth MR or SR labels for conventional assessment of classifier performance, superior VAT model performance is also observed (Supporting Information Figure S5).

Although the VAT model improves upon the cutoff approach, it still requires the calculation of computationally demanding diagnostics (e.g., C_0^2 from CASSCF). We exploit the VAT's insensitivity to feature space perturbations to reduce computational cost by replacing the most expensive (i.e., WFT) MR diagnostics with ML approximations. We recently trained kernel ridge regression (KRR) models to predict WFT-based diagnostics from a combination of six DFT-based diagnostics and size-independent, transferable geometric features[45] (Coulomb-decay revised autocorrelation functions[45, 62-65] or CD-RACs) with low *AD-3165* test set errors (see Computational Details and Supporting Information Text S3 and Figure S6 and Tables S1, S11).

In our streamlined MR "decision engine" workflow, we calculate DFT MR diagnostics, combine them with CD-RACs to obtain KRR-predicted WFT diagnostics, and use these two sets of diagnostics as input to the VAT model (Figure 3). Given the good performance of the individual KRR models, predictions from the MR decision engine are nearly identical (99% unchanged SR/MR assignments) to those made by the original VAT model (Figure 3 and Supporting Information Figures S7–S8). Of the six points reclassified (SR-to-MR: 4 and MR-to-SR: 2) from the full VAT model to the MR decision engine (i.e., with KRR-predicted WFT diagnostics), all are geometries at the SR-to-MR transition ($\%E_{corr}[(T)]$ = 95.6-96.1%) that typically contain strained rings (e.g., 2-ethylaziridine, and 2,3-epoxybutane), including two with intermediate (i.e., 0.25-0.75) VAT model scores (Figure 3 and Supporting Information Figure S9 and Table S12). While for these cases some individual diagnostics (i.e., MP2 or CAS) have relatively high KRR model errors, most others (i.e., CCSD) are intermediate (Supporting

Information Table S13). Conversely, the largest KRR errors are observed for highly distorted geometries (e.g., methoxycyclobutane) far from the SR/MR transition region, making the MR decision engine insensitive to these larger KRR model errors (Figure 3).

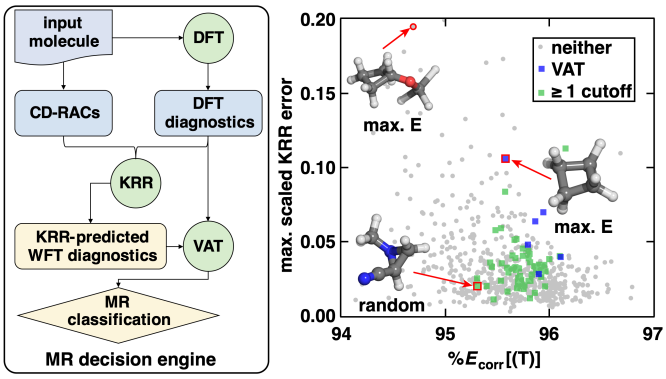


Figure 3. (left) Flowchart of the MR decision engine that combines DFT-calculated diagnostics with CD-RACs as input to KRR models to predict WFT MR diagnostics, which are used alongside DFT-calculated diagnostics as input into the VAT model for MR classification. (right) The maximum scaled KRR errors with respect to calculated values among all KRR-predicted diagnostics versus %E_{corr}[(T)] for each structure in the *AD-3165* test set for structures reassigned with KRR diagnostics by the VAT (6 points, blue translucent squares), cutoff model (58 points, green translucent squares), or neither (569 points, gray translucent circles). Scaled errors refer to errors relative to rescaling each diagnostic from 0 to 1 over the *AD-3165* set. Representative structures are shown in inset with their geometry type: cyclobutane (VAT), methoxycyclobutane (neither), and 1-methylaziridine-2-carbonitrile (cutoff). Atoms in structures are colored as follows: carbon in gray, hydrogen in white, nitrogen in blue, and oxygen in red.

Comparisons to using KRR predictions in the cutoff approach highlight the benefits of the VAT because a small regression error in the KRR-predicted diagnostic can correspond to a large classification error near the recommended cutoff (Supporting Information Figure S10). In the cutoff approach, KRR-predicted diagnostics reclassify 58 structures (i.e., 9%, SR-to-MR: 24 and MR-to-SR: 34) from the calculated diagnostics result, nearly 10× that of the MR decision engine (Figure 3 and Supporting Information Tables S8 and S14–S15). The majority of reclassifications arise for molecules at intermediate (95.3-96.2%) %E_{corr}[(T)] values from modest prediction errors of diagnostics (e.g., D_2 predicted: 0.178 vs calculated: 0.182 for a molecule with a distorted three-membered ring) that cross the recommended cutoff boundary (Figure 3 and

Supporting Information Tables S14–S15 and Figure S11). Quantitatively, the BC of the MR decision engine is nearly unchanged from the VAT, and the insensitivity of the VAT to noisy inputs maintains and widens the MR decision engine's superior performance to a cutoff approach (Supporting Information Figures S8 and S12 and Table S16).

To validate the transferability of our MR decision engine for obtaining MR classification at DFT cost, we curated two additional test datasets of geometries that are chemically distinct from *AD-3165*. *PS-401* consists of equilibrium geometries with heteroatoms (i.e., P or S) added through isovalent substitution (i.e., N or O) of *AD-3165* molecules. *LG-8934* contains larger (i.e., six to eight heavy atom) molecules generated following the same protocol[45] as *AD-3165* (Figure 4 and see Computational Details). We apply the VAT, KRR, and combined MR decision engine trained only on the *AD-3165* training set to these diverse test sets to quantify transferability across chemical composition and size, respectively.

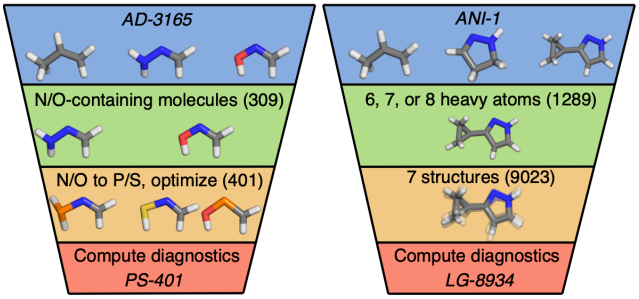


Figure 4. Approach for generation of the *PS-401* (left) and *LG-8934* (right) datasets. For *PS-401*, this approach consists first of selecting equilibrium N/O-containing molecules in *AD-3165*, substituting only one of either an N or O atom with P or S and re-optimizing the geometry, leading to two new molecules when multiple N or O atoms are present, as shown, and computing diagnostics and $\%E_{corr}[(T)]$. For *LG-8934*, the approach follows the protocol of *AD-3165* established in Ref. 1: random selection of 1289 unique molecules of ≥ 6 heavy atoms that are absent from *AD-3165*, selecting seven geometries for each unique molecule (i.e., equilibrium, maximum energy, and five randomly selected), and computing diagnostics and $\%E_{corr}[(T)]$. Representative structures shown in inset are colored as follows: hydrogen in white, carbon in gray, nitrogen in blue, oxygen in red, phosphorus in orange, and sulfur in yellow.

We computed all 15 MR WFT and DFT diagnostics on *PS-401* for inputs to the VAT

model (Figure 4 and see Computational Details). The VAT model performs well on the *PS-401* set, distinguishing high %E_{corr}[(T)] molecules as SR, again smoothly transitioning between 95–96% to MR at lower %E_{corr}[(T)] (Supporting Information Figure S13). This excellent performance is quantified by a VAT model BC value of 0.50 on *PS-401*, which outperforms the alternative cutoff-based or *k*-means models in a manner consistent with observations on *AD-3165* (Supporting Information Table S16).

We next evaluated whether KRR models, which were trained on a combination of the CD-RACs geometric representation with DFT MR diagnostics, performed well on *PS-401* or *LG-8934*.[45] CD-RACs were designed with transferability in mind[16, 30, 66], with reduced size-dependence and explicit encoding of isovalent atomic properties instead of direct incorporation of nuclear charge.[45] As a result, KRR model prediction errors of WFT MR diagnostics on *PS-401* or *LG-8934* are only around 2× *AD-3165* test set errors (Supporting Information Table S17 and Figures S14–S15). Alternative geometric representations with greater size- and nuclear-charge-dependence (i.e., the Coulomb matrix[67]) have poorer transferability (Supporting Information Table S17).

Given the promising results for the KRR models and the robust nature of the VAT models, we can expect good performance in the combined MR decision engine (see Figure 3 and Supporting Information Figure S12). With the MR decision engine, a small number (i.e., 16 or 4%) of *PS-401* molecules are reclassified from their VAT-model assignments (MR-to-SR: 6 and SR-to-MR: 10), thus achieving a good BC (0.51) comparable to the VAT model (Supporting Information Tables S16 and S18–S19 and Figure S16). Broadly, across the *LG-8934*, *PS-401*, and *AD-3165* sets, we observe a consistent, smooth transition from SR to MR at similar %E_{corr}[(T)] values and width of the transition region (Figure 5 and Supporting

Information Table S20 and Figure S17). Comparable performance on P- vs S-containing subsets of *PS-401* as well as *LG-8934* structures grouped by the number of heavy atoms demonstrates the relatively weak dependence of model performance on chemical composition and size (Supporting Information Figures S18–S19 and Tables S21–S22). Quantitatively, SR and MR %E_{corr}[(T)] distributions in all three sets are distinguished by low BC values of 0.50–0.55 that are superior to alternative *k*-means or cutoff approaches (Figure 5 and Supporting Information Table S16 and Figure S20).

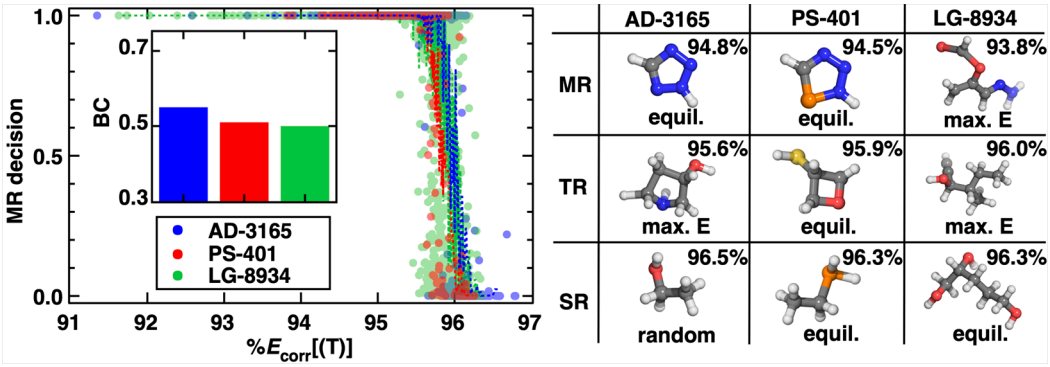


Figure 5. (left) MR classification (0 for SR and 1 for MR) obtained by the MR decision engine vs %E_{corr}[(T)] for structures in *AD-3165* (blue translucent circles), *PS-401* (red translucent circles), and *LG-8934* (green translucent circles) along with a 10-point moving average colored in the same manner, as shown in inset legend. The BC for each set is shown in the inset bar graph with *AD-3165* in blue, *PS-401* in red, and *LG-8934* in green. (right) Representative structures with their type (maximum energy, max. E; random; or equilibrium, equil.) shown at bottom and their %E_{corr}[(T)] value shown at top right. Representative structures are grouped by row with their classification (MR, transition region (TR), or SR) and by data set in columns (*AD-3165*, *PS-401*, and *LG-8934*). Representative structures are colored as follows: hydrogen in white, carbon in gray, nitrogen in blue, oxygen in red, phosphorus in orange, and sulfur in yellow.

Examining representative molecule classifications by the MR decision engine across the three data sets provides chemical insight into the origins of model transferability. Broadly, unsaturated molecules in equilibrium (*AD-3165*: 1H-tetrazole and *PS-401*: the triazaphosphole analogue) and stretched (*LG-8934*: $C_4H_6N_2O_2$) geometries have comparably low %E_{corr}[(T)] (i.e., 93-94%) across sets and are always classified as MR by the decision engine (Figure 5 and Supporting Information Table S23). Conversely, saturated molecules in equilibrium or weakly

distorted geometries (*AD-3165*: ethanol, *PS-401*: ethylphosphine, and *LG-8934*: 1,2,5-pentanetriol) are consistently classified as SR and have comparably high %E_{corr}[(T)] (96.3-96.5%) values (Figure 5 and Supporting Information Table S23). Molecules that span the SR-to-MR transition tend to have intermediate characteristics between these two extremes (Figure 5 and Supporting Information Table S23). Notably, all *PS-401* molecules are P- or S-substituted analogues of *AD-3165* molecules, and this substitution consistently decreases %E_{corr}[(T)] (oxirane: 96.1% vs thiirane: 95.5%) values (Supporting Information Table S24). The MR decision engine captures this effect, reclassifying thiirane and 17 other *PS-401* molecules as MR where oxirane or other *AD-3165* molecules were SR (Figure 5 and Supporting Information Figure S21 and Table S24). Despite good observed transferability, model retraining will likely be necessary when extending to classes of materials (e.g., transition metal complexes) where the relationship between MR diagnostic values and MR effects are qualitatively distinct.

In summary, we have demonstrated a low-cost approach to predicting MR character that overcomes limitations in predictive accuracy and consensus among MR diagnostics. We introduced a semi-supervised VAT approach to training an ANN MR classifier using MR diagnostics as inputs and labeling only the extreme SR or MR points for which class assignment was unambiguous. This VAT model outperformed available alternatives, including unsupervised learning and any traditional cutoff-based approach. We showed that the inherent robustness of the VAT to noisy inputs enabled us to reduce computational cost. We exploited this benefit to replace the calculated WFT MR diagnostics with predictions of these quantities from ML models trained on small organic molecules using a combination of DFT-based and geometric descriptors. The combined MR decision engine workflow preserved performance of the original VAT model, unlike more sensitive cutoff-based approaches. Because the inputs to the MR decision engine are

size-insensitive, it achieves good transferability to molecules larger than those in the training set, as well as those containing distinct heteroatoms (i.e., P or S). This MR decision engine provides a promising avenue for incorporating the automatic detection of strong correlation into high-throughput screening workflows with low computational cost.

Computational Details.

Datasets. We use the *AD-3165* data set curated in Ref. 45, which consists of 3,165 geometries from the ANI-1 data set[46] with up to six C, N, or O heavy atoms in seven geometries, the equilibrium, highest energy (max. E), and five randomly selected structures. This set[45] includes six DFT-based and nine WFT-based MR diagnostics along with $\%E_{corr}[(T)]$ for all *AD-3165* geometries, computed with MultirefPredict[68] and electronic structure codes[57, 69-74] (Table 1 and Supporting Information Tables S1–S2 and Figure S22). We also generated the *PS-401* dataset by randomly substituting a single N or O atom to isovalent P or S, respectively for all 401 compatible molecules (211 N-containing and 190 O-containing) in *AD-3165*[45] (Supporting Information Table S25). To maintain consistency with *AD-3165* structures[46], we optimized these structures with the same DFT functional (i.e., ωB97X[75]) and basis set[76] (i.e., 6-31G*) as in the original work[46] using TeraChem[72-73] with the L-BFGS algorithm in translation rotation internal coordinates[77] with all other defaults applied. We followed the *AD-3165* protocol[45] to compute 15 MR diagnostics (i.e., both DFT-based and WFT-based) and $\%E_{corr}[(T)]$ (Supporting Information Table S25 and Figures S23–S25).

We curated the *LG-8934* data set by adapting the protocol for generating *AD-3165* in Ref. 45 to larger molecules (i.e., six to eight heavy atoms), which were not present in *AD-3165*. As in *AD-3165*, seven geometries were selected including the equilibrium, highest energy, and five random for each of 1,289 unique chemical compositions (Supporting Information Table S26).

For *LG-8934*, we computed the six DFT-based and both MP2 MR diagnostics along with $\%E_{corr}$[(T)] (Table 1 and Supporting Information Figures S26–S29). From a theoretical dataset size of 9,023, a small number (89, ca. 1%) of DFT-based diagnostics could not be computed, leading to a final set of 8,934 geometries (Supporting Information Table S27).

ML models. We employ KRR models trained in Ref. 45 on a combination of DFT-based diagnostics and CD-RACs geometric descriptors[9, 45, 62-65] to predict WFT-based diagnostics (Supporting Information Text S3). Consistent with observations on *AD-3165*[45], the DFT diagnostics alone cannot predict $\%E_{corr}$[(T)] in *PS-401* or *LG-8934* (Supporting Information Figure S30). In this work, VAT ANN classifier models were built with PyTorch[78-79], and hyperparameters were tuned using Hyperopt[80] (Supporting Information Tables S28 and S29). We use the same 80% train/20% test set partitions of *AD-3165* as in the prior KRR models.[45] We selected the VAT model that gives the lowest unsupervised loss on a validation set that is 20% of *AD-3165* training data (i.e., 16% overall, Supporting Information Figures S31–S32). As model inputs, we label only the training data structures in the bottom or top 10 percentile over a majority (i.e., ≥ 8 of 15) of MR diagnostics, which corresponds to 379 (181 MR and 198 SR) of 2,532 *AD-3165* training points. This value was selected by trial and error, and the VAT model is relatively insensitive to this choice (Supporting Information Figure S32 and Tables S30–S31). VAT scores are assigned SR (i.e., 0) if they are below 0.5 and MR (i.e., 1) if ≥ 0.5. The *k*-means and AC models are applied on all 15 MR diagnostics, as implemented in scikit-learn. After inspection, the clusters are assigned SR or MR based on their extreme points. Increasing the number of clusters and assigning them as SR or MR to minimize the BC value on $\%E_{corr}$[(T)] values can improve performance but at the cost of requiring manual inspection and reassignment of the additional clusters (Supporting Information Table S32). UMAP[52] was carried out on 15

MR diagnostics for *AD-3165* in Ref. 45 and adapted for Figure 1.

ASSOCIATED CONTENT

Supporting Information. The following files are available free of charge.

Overview of calculation approach for 15 MR diagnostics; recommended cutoffs for diagnostics; system size dependence of $\%E_{corr}[(T)]$; description of VAT algorithm; example of labeling procedure; VAT MR classifications; histogram of $\%E_{corr}[(T)]$; details of Bhattacharyya coefficients; statistics for MR and SR $\%E_{corr}[(T)]$ distributions; Bhattacharyya coefficients, confusion matrix, and ROC curves for different methods; description of CD-RACs; KRR errors analysis on *AD-3165* test set; VAT classification and MR and SR $\%E_{corr}[(T)]$ distributions using KRR-predicted diagnostics; KRR error and $\%E_{corr}[(T)]$ analysis on reclassified structures on *AD-3165* test set; summary of Bhattacharyya coefficients; agreement of VAT MR classification with noisy inputs; KRR errors on *PS-401* and *LG-8934*; KRR error and $\%E_{corr}[(T)]$ analysis on reclassified structures on *PS*-401; Bhattacharyya coefficients, Fermi–Dirac fitting, MR decisions versus $\%E_{corr}[(T)]$, and MR and SR $\%E_{corr}[(T)]$ distributions on different datasets; detailed information for main text structures; detailed information and $\%E_{corr}[(T)]$ analysis on pairs of geometries in *PS-401* and *AD-3165* where the MR decisions are changed after element substitution; histogram of all diagnostics on all datasets; list of excluded structures from *LG-8934*; MR diagnostics vs $\%E_{corr}[(T)]$ for *PS-401* and *LG-8934*; range of hyperparameters allowed; best hyperparameters for VAT models; unsupervised loss vs model accuracy; agreement of VAT models on different labeling thresholds; additional *k*-means cluster model BC values. (PDF)

Total energies, correlation energies, MR diagnostics, KRR-predicted diagnostics, and MR classification of different methods of all molecules in *AD-3165*, *PS-401*, and *LG-8934*; list of molecules eliminated during *LG-8934* set curation; list of molecules in *PS-401* that do not have corresponding structures in *AD-3165*; the VAT model trained on *AD-3165* training set; geometries of all molecules in *AD-3165*, *PS-401*, and *LG-8934* (ZIP)

AUTHOR INFORMATION

Notes

The authors declare no competing financial interests.

ACKNOWLEDGMENT

The authors acknowledge primary support by the Office of Naval Research under grant numbers N00014-17-1-2956, N00014-18-1-2434, and N00014-20-1-2150 (to C.D., A.N., and H.J.K). F.L. was partially supported by the Department of Energy under grant number DE-SC0018096 and a MolSSI fellowship (grant no. ACI-1547580). This work made use of Department of Defense HPCMP computing resources. This work was also carried out in part using computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. H.J.K. holds a Career Award at the Scientific Interface from the Burroughs Wellcome Fund and an AAAS Marion Milligan Mason Award, which supported this work. The authors thank Adam H. Steeves for providing a critical reading of the manuscript.

REFERENCES

(1) Shu, Y.; Levine, B. G. Simulated Evolution of Fluorophores for Light Emitting Diodes. *J. Chem. Phys.* **2015,** *142*, 104104.
(2) Gomez-Bombarelli, R.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Duvenaud, D.; Maclaurin, D.; Blood-Forsythe, M. A.; Chae, H. S.; Einzinger, M.; Ha, D. G.; Wu, T.; et al. Design of Efficient Molecular Organic Light-Emitting Diodes by a High-Throughput Virtual Screening and Experimental Approach. *Nat. Mater.* **2016,** *15*, 1120-+.
(3) Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R. Efficient Computational Screening of Organic Polymer Photovoltaics. *J. Phys. Chem. Lett.* **2013,** *4*, 1613-1623.
(4) Vogiatzis, K. D.; Polynski, M. V.; Kirkland, J. K.; Townsend, J.; Hashemi, A.; Liu, C.; Pidko, E. A. Computational Approach to Molecular Catalysis by 3d Transition Metals: Challenges and Opportunities. *Chem. Rev.* **2018,** *119*, 2453-2523.
(5) Janet, J. P.; Chan, L.; Kulik, H. J. Accelerating Chemical Discovery with Machine Learning: Simulated Evolution of Spin Crossover Complexes with an Artificial Neural Network. *J. Phys. Chem. Lett.* **2018,** *9*, 1064-1071.
(6) Janet, J. P.; Ramesh, S.; Duan, C.; Kulik, H. J. Accurate Multi-Objective Design in a Space of Millions of Transition Metal Complexes with Neural-Network-Driven Efficient Global Optimization. *ACS Cent. Sci.* **2020,** *6*, 513-524.
(7) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; Von Lilienfeld, O. A. Quantum Chemistry Structures and Properties of 134 Kilo Molecules. *Sci. Data* **2014,** *1*, 140022.
(8) Gossett, E.; Toher, C.; Oses, C.; Isayev, O.; Legrain, F.; Rose, F.; Zurek, E.; Carrete, J.; Mingo, N.; Tropsha, A. AFLOW-ML: A RESTful API for Machine-Learning Predictions of Materials Properties. *Comput. Mater. Sci.* **2018,** *152*, 134-145.

(9) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Ind. Eng. Chem. Res.* **2018,** *57*, 13973-13986.
(10) Collins, C. R.; Gordon, G. J.; von Lilienfeld, O. A.; Yaron, D. J. Constant Size Descriptors for Accurate Machine Learning Models of Molecular Properties. *J. Chem. Phys.* **2018,** *148*, 241718.
(11) Bleiziffer, P.; Schaller, K.; Riniker, S. Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations. *J. Chem. Inf. Model.* **2018,** *58*, 579-590.
(12) De, S.; Bartok, A. P.; Csanyi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space. *Phys. Chem. Chem. Phys.* **2016,** *18*, 13754-13769.
(13) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials. *npj Comput. Mater.* **2016,** *2*, 16028.
(14) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating Materials Property Predictions Using Machine Learning. *Sci. Rep.* **2013,** *3*, 2810.
(15) Meyer, B.; Sawatlon, B.; Heinen, S.; von Lilienfeld, O. A.; Corminboeuf, C. Machine Learning Meets Volcano Plots: Computational Discovery of Cross-Coupling Catalysts. *Chem. Sci.* **2018,** *9*, 7069-7077.
(16) Welborn, M.; Cheng, L.; Miller, T. F. Transferability in Machine Learning for Electronic Structure via the Molecular Orbital Basis. *J. Chem. Theory Comput.* **2018,** *14*, 4772-4779.
(17) Curtarolo, S.; Hart, G. L.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O. The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **2013,** *12*, 191-201.
(18) Curtarolo, S.; Setyawan, W.; Hart, G. L.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.; Wang, S.; Xue, J.; Yang, K.; Levy, O. AFLOW: An Automatic Framework for High-Throughput Materials Discovery. *Comput. Mater. Sci.* **2012,** *58*, 218-226.
(19) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013,** *68*, 314-319.
(20) Nørskov, J. K.; Bligaard, T. The Catalyst Genome. *Angew. Chem., Int. Edit.* **2013,** *52*, 776-777.
(21) Han, S.; Huang, Y.; Watanabe, T.; Dai, Y.; Walton, K. S.; Nair, S.; Sholl, D. S.; Meredith, J. C. High-Throughput Screening of Metal–Organic Frameworks for CO2 Separation. *ACS Comb. Sci.* **2012,** *14*, 263-267.
(22) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal–Organic Frameworks. *Nat. Chem.* **2012,** *4*, 83-89.
(23) Witman, M.; Ling, S.; Anderson, S.; Tong, L.; Stylianou, K. C.; Slater, B.; Smit, B.; Haranczyk, M. In Silico Design and Screening of Hypothetical MOF-74 Analogs and Their Experimental Synthesis. *Chem. Sci.* **2016,** *7*, 6263-6272.
(24) Stein, C. J.; Reiher, M. Automated Selection of Active Orbital Spaces. *J. Chem. Theory Comput.* **2016,** *12*, 1760-1771.
(25) Jeong, W.; Stoneburner, S. J.; King, D.; Li, R.; Walker, A.; Lindh, R.; Gagliardi, L. Automation of Active Space Selection for Multireference Methods via Machine Learning on Chemical Bond Dissociation. *J. Chem. Theory Comput.* **2020,** *16*, 2389-2399.
(26) Veryazov, V.; Malmqvist, P. Å.; Roos, B. O. How to Select Active Space for Multiconfigurational Quantum Chemistry? *Int. J. Quantum Chem.* **2011,** *111*, 3329-3338.

(27) Schriber, J. B.; Evangelista, F. A. Communication: An Adaptive Configuration Interaction Approach for Strongly Correlated Electrons with Tunable Accuracy. *J. Chem. Phys.* **2016,** *144*, 161106.
(28) Ioannidis, E. I.; Gani, T. Z. H.; Kulik, H. J. molSimplify: A Toolkit for Automating Discovery in Inorganic Chemistry. *J. Comput. Chem.* **2016,** *37*, 2106-2117.
(29) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminf.* **2011,** *3*, 33.
(30) Duan, C.; Janet, J. P.; Liu, F.; Nandy, A.; Kulik, H. J. Learning from Failure: Predicting Electronic Structure Calculation Outcomes with Machine Learning Models. *J. Chem. Theory Comput.* **2019,** *15*, 2331-2345.
(31) McAnanama-Brereton, S.; Waller, M. P. Rational Density Functional Selection Using Game Theory. *J. Chem. Inf. Model.* **2017,** *58*, 61-67.
(32) Süß, D.; Huber, S. E.; Mauracher, A. On the Impact of Multi-Reference Character of Small Transition Metal Compounds on Their Bond Dissociation Energies. *J. Chem. Phys.* **2020,** *152*, 114104.
(33) Grimme, S.; Hansen, A. A Practicable Real-Space Measure and Visualization of Static Electron-Correlation Effects. *Angew. Chem., Int. Edit.* **2015,** *54*, 12308-12313.
(34) Lee, T. J.; Taylor, P. R. A Diagnostic for Determining the Quality of Single-Reference Electron Correlation Methods. *Int. J. Quantum Chem.* **1989**, 199-207.
(35) Sears, J. S.; Sherrill, C. D. Assessing the Performance of Density Functional Theory for the Electronic Structure of Metal-Salens: The d(2)-Metals. *J. Phys. Chem. A* **2008,** *112*, 6741-6752.
(36) Sears, J. S.; Sherrill, C. D. Assessing the Performance of Density Functional Theory for the Electronic Structure of Metal-Salens: The 3d(0)-Metals. *J. Phys. Chem. A* **2008,** *112*, 3466-3477.
(37) Langhoff, S. R.; Davidson, E. R. Configuration Interaction Calculations on the Nitrogen Molecule. *Int. J. Quantum Chem.* **1974,** *8*, 61-72.
(38) Janssen, C. L.; Nielsen, I. M. B. New Diagnostics for Coupled-Cluster and Moller-Plesset Perturbation Theory. *Chem. Phys. Lett.* **1998,** *290*, 423-430.
(39) Nielsen, I. M. B.; Janssen, C. L. Double-Substitution-Based Diagnostics for Coupled-Cluster and Moller-Plesset Perturbation Theory. *Chem. Phys. Lett.* **1999,** *310*, 568-576.
(40) Fogueri, U. R.; Kozuch, S.; Karton, A.; Martin, J. M. L. A Simple DFT-Based Diagnostic for Nondynamical Correlation. *Theor. Chem. Acc.* **2013,** *132*, 1291.
(41) Tishchenko, O.; Zheng, J. J.; Truhlar, D. G. Multireference Model Chemistries for Thermochemical Kinetics. *J. Chem. Theory Comput.* **2008,** *4*, 1208-1219.
(42) Schultz, N. E.; Zhao, Y.; Truhlar, D. G. Density Functionals for Inorganometallic and Organometallic Chemistry. *J. Phys. Chem. A* **2005,** *109*, 11127-11143.
(43) Ramos-Cordoba, E.; Salvador, P.; Matito, E. Separation of Dynamic and Nondynamic Correlation. *Phys. Chem. Chem. Phys.* **2016,** *18*, 24015-24023.
(44) Ramos-Cordoba, E.; Matito, E. Local Descriptors of Dynamic and Nondynamic Correlation. *J. Chem. Theory Comput.* **2017,** *13*, 2705-2711.
(45) Duan, C.; Liu, F.; Nandy, A.; Kulik, H. J. Data-Driven Approaches Can Overcome the Cost– Accuracy Trade-Off in Multireference Diagnostics. *J. Chem. Theory Comput.* **2020,** https://doi.org/10.1021/acs.jctc.0c00358.
(46) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1, a Data Set of 20 Million Calculated Off-Equilibrium Conformations for Organic Molecules. *Sci. Data* **2017,** *4*, 170193.

(47) Kesharwani, M. K.; Sylvetsky, N.; Kohn, A.; Tew, D. P.; Martin, J. M. L. Do CCSD and Approximate CCSD-F12 Variants Converge to the Same Basis Set Limits? The Case of Atomization Energies. *J. Chem. Phys.* **2018,** *149*, 154109.
(48) Jensen, H. J. A.; Jorgensen, P.; Ågren, H.; Olsen, J. Second‐Order Moller–Plesset Perturbation Theory as a Configuration and Orbital Generator in Multiconfiguration Self‐Consistent Field Calculations. *J. Chem. Phys.* **1988,** *88*, 3834-3839.
(49) Moller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934,** *46*, 0618-0622.
(50) Karton, A.; Daon, S.; Martin, J. M. L. W4-11: A High-Confidence Benchmark Dataset for Computational Thermochemistry Derived from First-Principles W4 Data. *Chem. Phys. Lett.* **2011,** *510*, 165-178.
(51) Jiang, W. Y.; DeYonker, N. J.; Wilson, A. K. Multireference Character for 3d Transition-Metal-Containing Molecules. *J. Chem. Theory Comput.* **2012,** *8*, 460-468.
(52) McInnes, L.; Healy, J.; Melville, J. Umap: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426* **2018**.
(53) Sprague, M. K.; Irikura, K. K. Quantitative Estimation of Uncertainties from Wavefunction Diagnostics. In *Thom H. Dunning, Jr.*; Springer: 2015; pp 307-318.
(54) Priebe, C. E.; Park, Y.; Vogelstein, J. T.; Conroy, J. M.; Lyzinski, V.; Tang, M.; Athreya, A.; Cape, J.; Bridgeford, E. On a Two-Truths Phenomenon in Spectral Graph Clustering. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116*, 5995-6000.
(55) Ceriotti, M. Unsupervised Machine Learning in Atomistic Simulations, between Predictions and Understanding. *J. Chem. Phys.* **2019,** *150*, 150901.
(56) Chapelle, O.; Scholkopf, B.; Zien, A. *Semi-Supervised Learning*. MIT Press: Cambridge, Mass., 2006; p x, 508 p.
(57) Miyato, T.; Maeda, S. I.; Koyama, M.; Ishii, S. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2019,** *41*, 1979-1993.
(58) Oliver, A.; Odena, A.; Raffel, C.; Cubuk, E. D.; Goodfellow, I. J. Realistic Evaluation of Deep Semi-Supervised Learning Algorithms. In *Advances in Neural Information Processing Systems*, 2018.
(59) Goodfellow, I.; Shlens, J.; Szegedy, C. Explaining and Harnessing Adversarial Examples. In *International Conference on Learning Representations*, 2015.
(60) Coleman, G. B.; Andrews, H. C. Image Segmentation by Clustering. *Proceedings of the IEEE* **1979,** *67*, 773-785.
(61) Lee, T. J. Comparison of the T1 and D1 Diagnostics for Electronic Structure Theory: A New Definition for the Open-Shell D1 Diagnostic. *Chem. Phys. Lett.* **2003,** *372*, 362-367.
(62) Janet, J. P.; Kulik, H. J. Resolving Transition Metal Chemical Space: Feature Selection for Machine Learning and Structure-Property Relationships. *J. Phys. Chem. A* **2017,** *121*, 8939-8954.
(63) Janet, J. P.; Gani, T. Z. H.; Steeves, A. H.; Ioannidis, E. I.; Kulik, H. J. Leveraging Cheminformatics Strategies for Inorganic Discovery: Application to Redox Potential Design. *Ind. Eng. Chem. Res.* **2017,** *56*, 4898-4910.
(64) Broto, P.; Moreau, G.; Vandycke, C. Molecular Structures: Perception, Autocorrelation Descriptor and Sar Studies: System of Atomic Contributions for the Calculation of the N-Octanol/Water Partition Coefficients. *Eur. J. Med. Chem.* **1984,** *19*, 71-78.

(65) Virshup, A. M.; Contreras-García, J.; Wipf, P.; Yang, W.; Beratan, D. N. Stochastic Voyages into Uncharted Chemical Space Produce a Representative Library of All Possible Drug-Like Compounds. *J. Am. Chem. Soc.* **2013,** *135*, 7296-7303.
(66) Cheng, L. X.; Welborn, M.; Christensen, A. S.; Miller, T. F. A Universal Density Matrix Functional from Molecular Orbital-Based Machine Learning: Transferability across Organic Molecules. *J. Chem. Phys.* **2019,** *150*, 131103.
(67) Rupp, M.; Tkatchenko, A.; Muller, K. R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012,** *108*, 058301.
(68) Liu, F.; Duan, C.; Kulik, H. J. Multirefpredict. https://github.com/hjkgrp/MultirefPredict (accessed May 23, 2020).
(69) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; et al. Psi4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017,** *13*, 3185-3197.
(70) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012,** *2*, 73-78.
(71) Neese, F. Software Update: The ORCA Program System, Version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018,** *8*, e1327.
(72) Petachem LLC. Petachem. http://www.petachem.com (accessed May 23, 2020).
(73) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009,** *5*, 2619-2628.
(74) MolSSI. QCEngine. https://github.com/MolSSI/QCEngine (accessed May 23, 2020).
(75) Chai, J. D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008,** *128*, 084106.
(76) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods .9. Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971,** *54*, 724-+.
(77) Wang, L.-P.; Song, C. Geometry Optimization Made Simple with Translation and Rotation Coordinates. *J. Chem. Phys.* **2016,** *144*, 214108.
(78) Pytorch. https://pytorch.org (accessed May 23, 2020).
(79) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc.: 2019; pp 8024--8035.
(80) Bergstra, J. C., D. D.; Yamins, D. Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms *Proceedings of the 12th Python in science conference* **2013**, 13-20.