

Deciphering Hydrogen Bond Features to Develop Machine Learning Models for Predicting Non-ionic Deep Eutectic Solvents

Usman L. Abbas*¹ | Yuxuan Zhang² | Tapia, Joseph¹ | Md, Selim³ | Chen, Jin³ | Jian Shi² | Qing Shao*¹

¹Department of Chemical and Materials Engineering, University of Kentucky, Lexington, Kentucky 40506.

²Department of Biosystems and Agricultural Engineering, University of Kentucky, Lexington, Kentucky 40506.

³Institute for Biomedical Informatics, Department of Computer science, University of Kentucky, Lexington, Kentucky 40506.

* Corresponding authors.

Email: usman.abbas@uky.edu, qshao@uky.edu.

Abstract

Non-ionic deep eutectic solvents (DESs) have emerged as designer solvents for applications such as catalysis, extraction, and carbon capture and conversion. A major challenge is the lack of an efficient tool to discover DES candidates. Currently, the search relies heavily on the researchers' intuition or a trial-and-error process, which leads to a low success rate or bypassing of promising candidates. Recognizing the central role that hydrogen bonds play in the DES formation, this work aims to decipher the hydrogen bond features for DESs and develop machine learning models to predict the potential of a system to be DES based on the hydrogen bond-based descriptors. We first analyze the hydrogen bond properties of 38 known DES and 111 known non-DES systems using their molecular dynamics simulation trajectories. The analysis reveals two features for DES compared to non-DES: (a) the imbalance between the numbers of the two intra-component hydrogen bonds, and (b) more and stronger inter-component hydrogen bonds. Based on the analysis results, we developed 30 machine learning models using ten algorithms and three types of hydrogen bond-based descriptors. The model performance is first benchmarked using their average and minimal ROC-AUC values. We also analyze the importance of individual features in the models and the results are consistent with the simulation-based statistical analysis. Finally, we validate the prediction ability of the models using the experimental results of 34 systems. The Extra Trees outperforms the others in the validation with an ROC-AUC of 0.88. Our work iterates the importance of hydrogen bond in DES formation and shows the potential of machine learning in discovering new DESs.

1. Introduction

Deep eutectic solvents (DESs) are liquid mixtures of hydrogen bond acceptors (HBAs) and donors (HBDs) with tunable properties[1-13]. DESs have gained attention from researchers as sustainable solvents in a number of applications including carbon capture[2, 14-16], pharmaceuticals[9, 14, 15, 17-23], material synthesis[8, 19, 24], electrochemistry[9, 14, 25-37], decontamination[17, 18] and extractions[6, 8, 24, 38-41] where the solvent can be recovered[42] and potentially reused[43]. Non-ionic DESs show several desirable properties including biodegradability, high conductivity, low volatility, and low toxicity as compared to conventional solvents [2, 8, 38, 43, 44]. Non-ionic DESs, popularly classified as type V DESs[4, 14-16, 38], can be made with natural compounds and have characteristics such as low viscosity that make them particularly suitable for industrial applications such as liquid-liquid extractions and carbon nanomaterial production[12, 38, 45].

A challenge for DES-related research is to discover more DESs so the community can possess a large pool to search for the ones with the desired properties. Numerous experimental and computational studies have shown the importance of hydrogen bonds (HBs) for DESs[1, 2, 4, 9, 14, 15, 39, 42, 46]. Farias et al.[43] carried out an experimental study to understand the role of HBDs of DESs in aqueous biphasic systems. They concluded that HBDs with high relative hydrophilicity are mainly adjuvants in biphasic systems, HBDs with moderate hydrophilicity control the formation of the biphasic systems, and HBDs with low hydrophilicity (high hydrophobicity) form aqueous biphasic systems, with the HBA acting as adjuvants in such systems. Abranches et al.[1] investigated the suitability of betaine, due to its polarity imbalance, as a universal HBA in the formation of DESs. Their study used a combination of experiments and density functional theory calculations and concluded that betaine is a suitable choice for making natural DESs due to its non-selective nature, low cost, and low toxicity. All these fundamental studies highlight the importance of HBs in DES formation and properties. This important role indicates that the HB-based descriptors could serve as suitable inputs to discover new DESs.

Machine learning (ML) models have been used to predict DES physicochemical and thermophysical properties[17, 19, 46-51]. The review of Benworth et al.[15] covered the studies that developed quantitative structure- property relationship models for predicting DES properties[6, 15]. Halder et al.[51] used a cheminformatics approach to find out which structural

attributes of DESs are required to get accurate predictions of densities for industrial applications. They combined the top performing models to make consensus predictions of densities. They concluded that a consensus modelling approach could be used to obtain high accuracy estimates of novel DES densities using features like number of HBs, lipophilicity, polarizability, and the van der Waals surface area. Dietz et al.[6] used Perturbed-Chain Statistical Association Fluid Theory (PC-SAFT) modeling to predict the liquid-liquid equilibrium and solid-liquid equilibrium of mixtures of hydrophobic DES with water or hydroxymethyl furfural. Their results showed such an approach to be adequate for predicting phase behavior of hydrophobic DES mixtures. Abdollahzadeh et al.[19] compared 7 machine learning algorithms to estimate the densities of 149 DESs. Their results showed that the least squares support vector regression had the highest accuracy and performed 74.5% better than the best results obtained via empirical correlations. Zamora et al.[16] compared the suitability of 5 different ML algorithms, trained on experimental data, to accurately predict the densities and viscosities of type V DESs. The study concluded that support vector machines performed best at predicting densities, and gaussian process regression models did best at predicting viscosities. Xu et al.[50] used gradient boosting models to obtain good prediction accuracies on DES viscosities. The model showed nice prediction when trained and tested on experimental and simulation data. These studies highlight the potential of combining ML and molecular simulations to predict the properties of DESs.

MD simulation has become a useful tool for determining descriptors as inputs for machine learning models[14, 24, 46]. We hypothesize that HB properties could be used to predict the formation of DESs. However, it is not trivial to determine which HB properties could be used. Our previous work[52] found that the non-ionic DESs can be broadly classified into three groups based on the ratio of intra- and inter-component HB numbers. Such observations inspire us to explore the possibility of developing machine learning models using HB-based descriptors.

The database plays a central role in any machine learning model development. We curated a library of 38 known DESs and 111 non-DESs from the literature. We will use this database to conduct statical analysis on molecular simulation data, to further construct training and testing datasets for model development. Finally, we will utilize the experimental results for 34 systems to validate our model performance. Due to the size of the database, this paper will focus on traditional machine learning algorithms. This work will utilize 10 machine learning algorithms. However, it

is noted that the deep learning algorithms have emerged as a promising technique for designing materials. The rest of this paper is structured as follows: Section 2 will display the computational details. Section 3 will present results and discussion and section 4 will present our conclusions.

2. Methodology

2.1 The list of DES and non-DES systems.

Tables S1-S8 show the details of the 183 systems simulated in this study. They included 38 known DES and 111 known non-DES systems from literature and 34 systems for the experimental validation set. The 38 known DES and 111 known non-DES systems are determined based on the experimental results of van Osch et al.[53, 54]. Only the non-ionic DESs from their list are considered in our study. We also excluded the DESs that did not have all the three types of HBs (A-A, B-B and A-B). Compounds are represented using three letters e.g., DEA represents Decanoic acid. The systems are labeled using the three letters of their compounds and the molar ratio e.g., DEA-MEN11 represents a 1:1 mixture of decanoic acid and menthol. Table S17-19 list the abbreviations used for chemical compounds.

2.2 Molecular simulations

2.2.A Molecular models

The OPLSAA/M force field[55] was used to describe the molecules in this study. The nonbonded and bonded parameters in the systems were assigned using the OPLSAA/M force field because this force field can adequately model the behavior of organic molecules. The force field parameters were generated using the Ligpargen[56] web server.

2.2.B Simulation detail

The simulation systems were created by inserting specific numbers (depending on the molar ratio) of the chosen organic molecules randomly in a cubic box. Figure 1 shows a snapshot, generated using VMD[57], of the Thy-Men11 system.

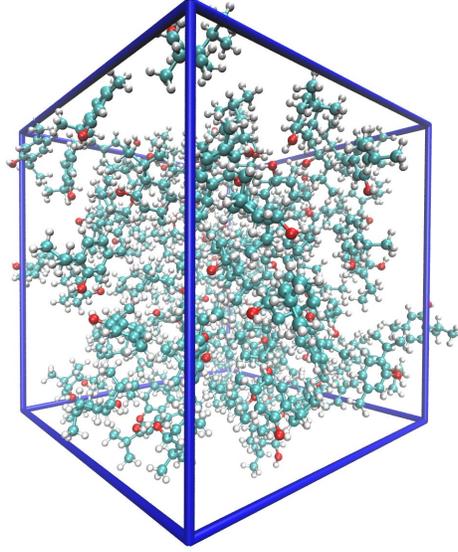


Figure 1. A snapshot of the equilibrated Thy-Men11 system. The molecules are display in CPK model. The color of atoms: C: cyan, O: red, and H: white.

For each system, the simulation process comprises three steps: (a) an energy minimization to remove any atomic overlaps, (b) a 50-ns isobaric-isothermal (NPT, P=1 atm, T=295 K) ensemble MD simulation to enable the system reach thermodynamic equilibrium and (c) a 10-ns canonical (NVT, T=295 K) ensemble MD simulation to collect the data at a frequency of 10 ps. In step (b), the MD simulation uses the Berendsen[58] method to control the system pressure while the velocity rescaling method is used to control the system temperature.

The short and long range nonbonded interactions in the OPLS-AA/M force field are calculated using the Lennard-Jones 12-6 and Coulomb potential, respectively using equation 1.

$$E = \sum_i \sum_{j<i} \left\{ \frac{1}{4\pi\epsilon_0} \frac{q_i q_j e^2}{r_{ij}} + 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right\} \quad (1)$$

where r_{ij} is the distance between atoms i and j ; q_i and q_j are the partial charges of atoms i and j respectively, ϵ_0 is the free space permittivity, ϵ_{ij} and σ_{ij} are energetic and geometric parameters respectively. The particle mesh Ewald[59] (PME) sum is used to calculate long-range potentials, and the LINCS algorithm[60] is used to constrain bonds involving hydrogen atoms. All energy minimization and MD simulations were conducted using GROMACS 2021.2.[61]

2.2. Hydrogen bond analysis

We characterize the HBs using the criteria developed by the Chandler group [62]: (1) the distance between the O(donor) and O(acceptor) $\leq 0.35\text{nm}$; (2) the O(acceptor)-H(donor)-O(donor) angle $\leq 30^\circ$.

We calculated the HB lifetime in two steps: (a) Calculate the correlation function $C(t)$ as shown in equation 3:

$$C(t) = \frac{\langle N_{HB}(t) \rangle}{\langle N_{HB}(0) \rangle} \quad (2)$$

where $\langle N_{HB}(0) \rangle$ is the ensemble average of the number of HBs at the initial status, and $\langle N_{HB}(t) \rangle$ is the ensemble average of the number of HBs still existing at time=t. The HBs are counted even if they break intermittently, based on Rappaport's definition[63].

(b) Calculate the lifetime τ by numerically integrating the $C(t)$ curves.

2.3 Machine learning models

The literature-based database contains more non-DES systems than DES ones. This imbalance in data distribution may cause bias in model performance. To attenuate this artificial effect, we curate a database containing 38 DES and 38 non-DES for the machine learning model development. The 38 non-DES is randomly selected from the 111 ones. We further split this database into the training set (30 DES and 30 non-DES) and testing set (8 DES and 8 non-DES). We use fixed seeds when sampling from the DES and non-DES set to ensure all models are evaluated on the same dataset slices. All models are further validated with experimentally verified DESs and non-DESs, as described in section 3.

We trained machine learning models using 10 algorithms. The algorithm implementations from scikit-learn[64, 65] are: (1) Decision tree, (2) logistic regression, (3) AdaBoost, (4) Gradient Boost, (5) Extra trees forest, (6) random forest, (7) K-nearest neighbors, and (8) support vector machine. The (9) XGBoost and (10) XGBoost random forest are from the XGBoost package[66].

Hyperparameter optimization was done using scikit-learn's grid search. Each model's performance was measured via repeated k-fold (6 folds, 10 repeats) cross-validation with the ROC-AUC metric. The architecture with the highest ROC-AUC during optimization is judged to be the best. For each model, further training and testing will only be conducted on the architecture with the best architecture.

This work considers three types of input features: (a) HB numbers alone, (b) HB lifetimes alone, and (c) HB numbers combined with lifetimes. All HB input features, generated from MD simulations, are shown in Figures S5-S8. The total number of the trained models is 30. The model hyperparameters are shown in Figures S20-22.

Here is a list of python packages used to conduct this work: Python (version 3.10.8), scikit-learn[65] (version 1.2.0), pandas[67] (version 1.5.2), Numpy[68] (version 1.22.3), matplotlib[69] (version 3.6.2), scipy[70] (version 1.7.3) and XGBoost[66] (version 1.7.3). All machine learning work is done on an 8th Gen Intel core i7-8750H processor.

2.4 Experiment

To validate the trained models, a list of solvent formulas was tested to determine whether they can form DESs or not. The systems are prepared by mixing the two components at a specific molar ratio (e.g., 1:1, 1:2, and 2:1) and followed with constant stirring and heating to ensure fully mixing. The two compounds were first weighted based on their molar ratio and transferred into a glass bottom sequentially. The compounds were then premixed using a glass rod, after which a magnetic stirrer was added, and the glass bottle was sealed. Subsequently, the bottle was placed in an oil bath and heated to 80-120 °C on a heating plate at a stirring speed of 500 RPM for 1 h. After the heating process, the mixture was air-cooled to room temperature and kept in a desiccator for 24 h. The sample that remained in a liquid form with no crystals within the 24 h was considered a DES candidate. We observe that some systems turn out to be DES-like at beginning but form solid phase after several days. These systems are not used in this work.

3. Results and discussion

3.1 Statistical analysis of hydrogen bond features

3.1.1 Hydrogen bond number features

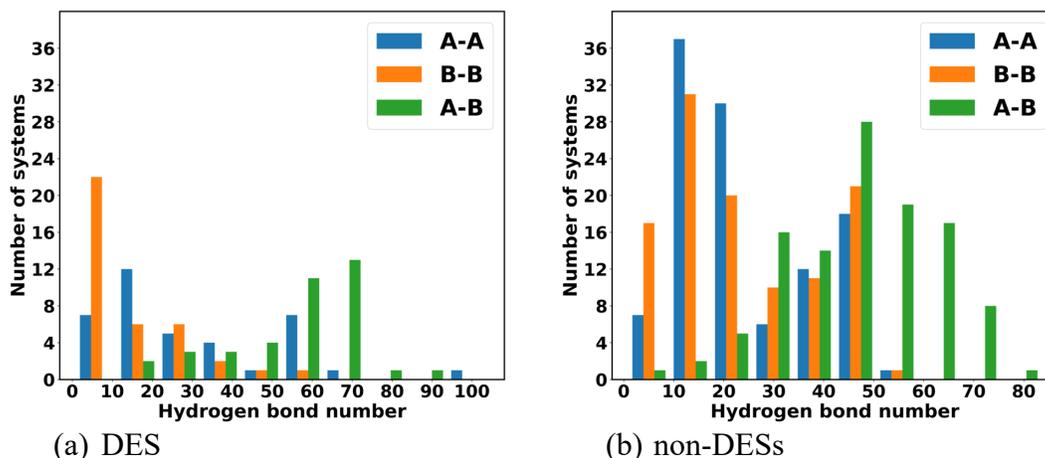


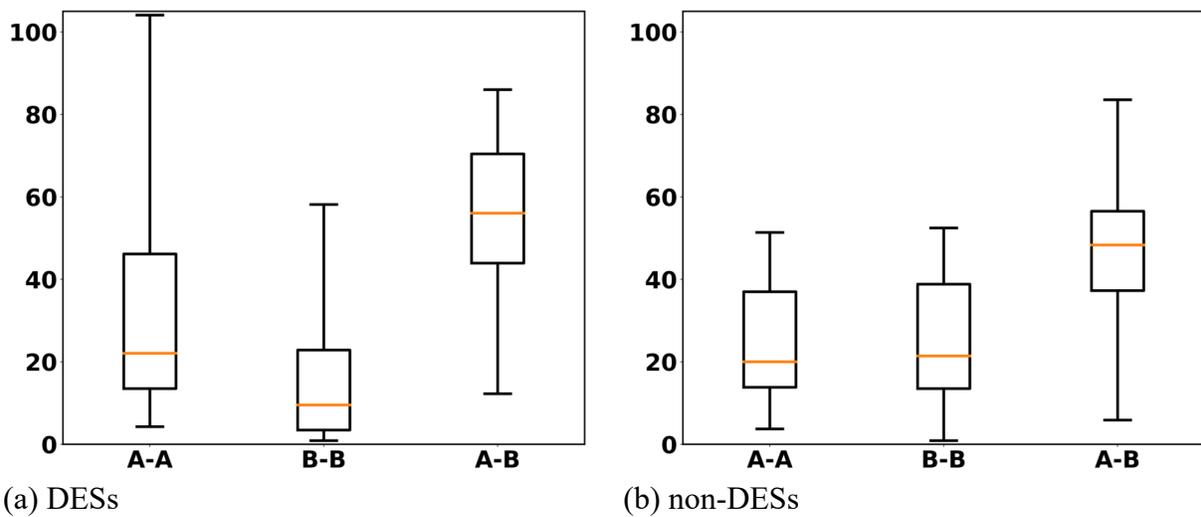
Figure 2.2 Distribution of average inter- (A-B) and intra- (A-A and B-B) component HB numbers. (a) DES and (b) non-DES.

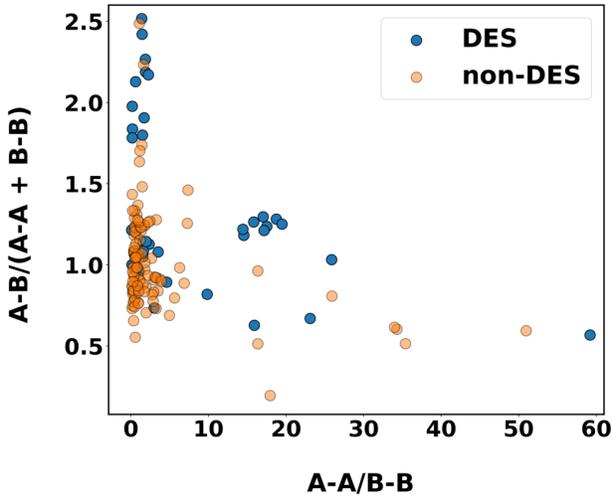
We first analyze the distribution of actual inter-component and intra-component HB numbers for DES and non-DES systems. Figure 2 shows the distribution of the 38 DES and 111 non-DES systems based on their average inter- and intra-component HB numbers. There are no distinct differences regarding the two pattern distributions in Figure 2. As shown in Figure 2a, the intra-component HB numbers (A-A and B-B) for DESs skew to the left, indicating that most of the DESs in our dataset have average HB numbers less than 20. The B-B HB numbers are concentrated on the lower end of the spectrum compared to the A-A HBs. The inter-component HB numbers skew to the right, suggesting that most of the DESs have higher inter-component HB numbers compared to intra-component HB numbers. As shown in Figure 2b, the intra-component HB numbers for non-DESs also skew to the left. In addition, most of the inter-component HB numbers are skewed to the right. Thus, the actual number of intra- and inter-component HBs may not be a suitable HB feature to differentiate DES and non-DES systems.

However, some distinct pattern appears when we plot the average inter- and intra-component HB numbers as boxplots for DES and non-DES systems. As shown in Figure 3a-b, the DES systems present a large difference between the median values for the two intra-component HB numbers (A-A vs. B-B) compared to the non-DES systems. In addition, the inter-component HBs in the DES systems display a median value of 56.07, 61 – 83% higher than the median values

of the two intra-component HBs. For the non-DES systems, the A-B HBs only present a median value of 48.36, 55 – 59% higher than A-A and B-B respectively. Even when the intra-component HB numbers, A-A and B-B, are summed up, the inter-component HB number, A-B, is greater. Such differences in the median implies that the ratio of the two intra-component and the inter-/intra- may serve as features for system classification.

The plot of the A-A/B-B and A-B/(A-A+B-B) in Figure 3c further confirms our hypothesis. The ratio of inter-component to intra-component HB numbers is well above 1.5 for some DESs. On average, the inter-component HB numbers are 35% more than the total intra-component HB numbers. Finally, we looked at the ratio of the intra-component bonds to get more insight into the magnitudes of their differences. On average, the number of A-A HBs is about 700% greater than the B-B HBs. This difference does not exist in non-DESs. The average intra-component HB numbers, A-A and B-B, for non-DES components are roughly the same (24.31 and 24.08 respectively). The median HB numbers for A-A and B-B are also similar in non-DESs with 20.01 and 21.41 respectively. This suggests that there is no dominant intra-component HB in non-DESs. This is also seen in Figure 3c as most of the intra-component HB number ratios cluster around 1.0, with few outliers. The average inter-component (A-B) HB numbers are close to twice (1.89 – 1.96) that of the average intra-component (A-A and B-B) HBs respectively. The ratios of intra-component HBs in non-DESs are also closer across all percentiles, with a median of 1.01 compared to 2.00 for DESs.

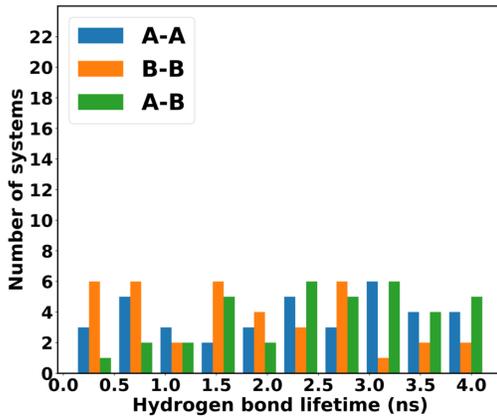




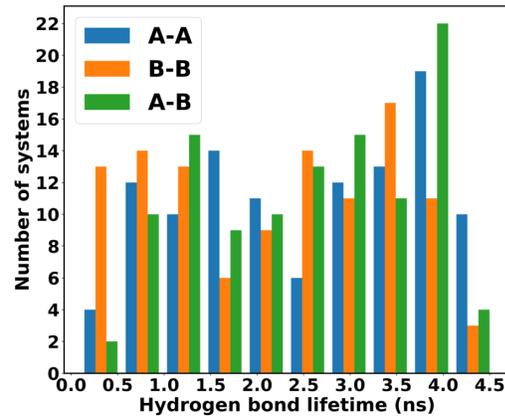
(c)

Figure 3.3 HB number features for DESs and non-DESs. (a) and (b) show the distributions of average HB numbers. (c) shows the ratio of inter-component to intra-component HB numbers vs ratio of intra-component HB numbers.

3.1.2 Hydrogen bond lifetimes



(a) DESs



(b) non-DESs

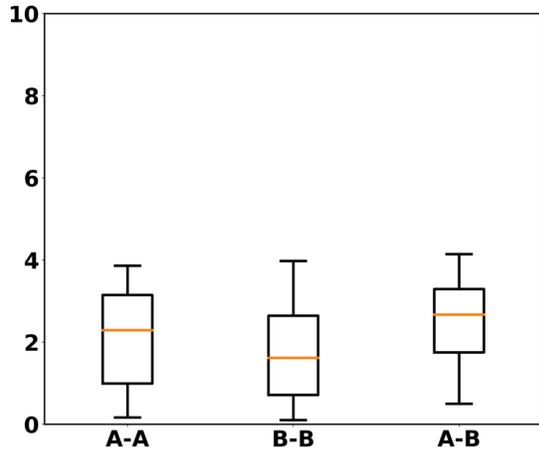
Figure 4. 4 Distribution of average inter- (A-B) and intra-component (A-A and B-B) HB lifetimes. (a) DES and (b) non-DES.

We also analyzed the distribution of the inter- and intra- component HB lifetimes of the 38 DESs and 111 non-DESs. There are no distinct patterns emerging from Figure 4. In Figure 4a, one of the intra-component HB lifetime bonds (A-A) is concentrated at 2.0 – 4.0 ns while the B-B lifetime is concentrated at 0.25 – 2.5 ns for DESs. The inter-component HB lifetimes, A-B, appear to skew to the right, and last longer than the intra-component lifetimes.

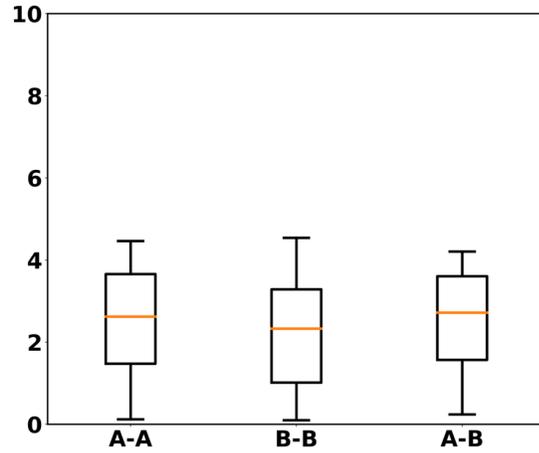
From Figure 4b, we can observe that one of the intra-component HB lifetimes for non-DESs dominates in different bins but there is no clear trend e.g., B-B dominates at lifetimes less than 1.25 ns but A-A dominates at lifetimes greater than 3.0 ns. In each bin, the inter-component lifetimes (A-B) appear to be more dominant than one of the intra-component bonds in lifetimes but are similar to the other intra-component bond. The lack of a clear pattern means actual intra- and inter- HB lifetime features alone might not be enough to differentiate DES and non-DES systems.

Some differences emerge when we plot the intra- and inter- HB lifetime distributions as boxplots. From Figure 5a, the DESs present a small difference between the median values for the inter-component (A-B) and one of the intra-component (A-A) lifetimes; the difference is wider between the median values of A-B and the other intra-component (B-B) lifetimes. The A-B lifetimes present a median of 2.67, which is 14% and 39% greater than the A-A and B-B lifetimes respectively. From Figure 5b, the non-DESs present a smaller difference between the median values for the inter-component and intra-component lifetime values. The A-B lifetimes present a median of 2.72, which is only 3.6% and 14% greater than the A-A and B-B lifetimes respectively. These differences indicate that the ratios of inter- to intra- component lifetimes could be more useful as features than the actual lifetimes.

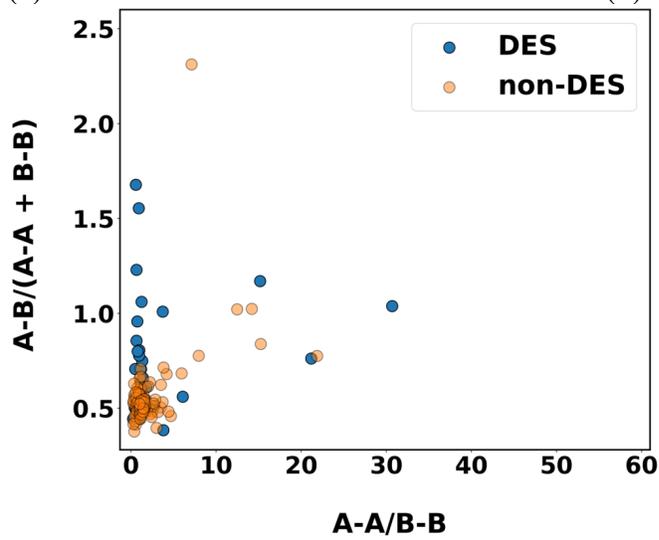
The plot of $A-A/B-B$ vs $A-B/(A-A + B-B)$ in Figure 5c confirms this hypothesis. The A-A median lifetimes last about 7% longer than B-B lifetimes compared to 13% for non-DESs. Even though there are more inter-component HBs than intra-component HBs, the intra-component HBs last longer. The median value of $A-B/(A-A + B-B)$ lifetimes is 0.63 for DESs and 0.53 for non-DESs. The ratio of inter-component to intra-component HBs in DESs varies from 0.5 to 2.0 while most of the non-DESs have ratios of inter-component to intra-component lifetimes clustered around 0.5. Similar to the HB numbers, ratios of HB lifetimes might be more useful as features than the actual lifetime values.



(a) DESs



(b) non-DESs



(c)

Figure 5.5 HB lifetime features for DESs and non-DESs. (a) and (b) show the distributions of average HB lifetimes. (c) shows the ratio of inter-component to intra-component HB lifetimes vs ratio of intra-component HB lifetimes.

3.2 Model development

We trained 30 models with 10 algorithms (logistic regression, random forest, decision tree, extra tree forest, KNN, SVC, Ada boost, gradient boost, and XGBoost random forest, XGBoost) and three types of input features (HB number, HB lifetime and a combination of HB number and lifetime features) to predict if a system could be DES. We trained each model for 100 rounds and calculated the average Receiver operating characteristic-Area under the curve (ROC-AUC) values from 100 iterations. For each round, we randomly sampled 38 (30 for training, 8 for testing) entries each from the DES and non-DES dataset. To ensure a fair comparison, each model is trained and tested with the same samples from the DES and non-DES dataset.

ROC is a probability curve and AUC represents the degree or measure of separability. It shows how much the model is capable of distinguishing between classes. The higher the AUC, the better the model is at predicting DES classes as DES and non-DES classes as non-DES. We ranked the models using two criteria: (1) average ROC-AUC score, and (2) minimum ROC-AUC score.

Table 1. Average ROC-AUC values of the 30 models. Best value is in bold.

S/N	Algorithm	lifetime	number	number + lifetime
1	Logistic regression	0.68	0.78	0.77
2	Decision Tree	0.63	0.74	0.68
3	Gradient Boost	0.66	0.78	0.76
4	AdaBoost	0.70	0.78	0.75
5	Random Forest	0.69	0.81	0.79
6	Extra Trees Forest	0.70	0.80	0.78
7	Support Vector Classifier	0.64	0.77	0.77
8	K-Nearest Neighbors	0.63	0.77	0.77
9	XGBoost	0.67	0.81	0.77
10	XGBoost Random Forest	0.62	0.82	0.79

With an average ROC-AUC score of 0.70, the AdaBoost and Extra Trees classifiers are tied for the best performing models when trained with HB lifetime features. When trained with HB number features, XGBoost-Random Forest and XGBoost are the two top performing models, with an average ROC-AUC of 0.82 and 0.81 respectively. When HB numbers and lifetimes

features are combined, the top performing models are the Random Forest and the XGBoost-Random Forest classifiers with both having an average ROC-AUC of 0.79. Overall, the top performing models are the XGBoost-Random Forest and Extra Trees based on the average and minimum ROC-AUC values respectively.

Table 2. Minimum ROC-AUC scores for the 30 models

S/N	Algorithm	lifetime	number	number + lifetime
1	Logistic regression	0.25	0.55	0.50
2	Decision Tree	0.30	0.50	0.45
3	Gradient Boost	0.25	0.50	0.40
4	AdaBoost	0.30	0.40	0.38
5	Random Forest	0.30	0.45	0.45
6	Extra Trees Forest	0.30	0.70	0.55
7	Support Vector Classifier	0.15	0.10	0.10
8	K-Nearest Neighbors	0.30	0.45	0.45
9	XGBoost	0.20	0.50	0.45
10	XGBoost Random Forest	0.20	0.55	0.45

The minimum ROC-AUC score in 100 training iterations could also be used to evaluate the performance of a model. Table 2 lists the minimum ROC-AUC score for the 30 models. Nine algorithms display the lowest minimum ROC-AUC scores when trained with HB lifetime features alone. Such observations indicates that the HB lifetime alone might not be sufficient to develop a machine learning model for classifying DES systems. Interestingly, all the algorithms present the highest minimum ROC-AUC scores when trained with HB numbers alone. Across all categories, the Extra Trees classifier has the highest minimum ROC-AUC score of 0.70 when trained with HB numbers. All the algorithms also recorded their highest overall ROC-AUC scores when trained with HB numbers alone.

Some algorithms are among the top performers regardless of the criteria used for model selection. For models trained with HB numbers, the top performing model is the Extra Trees based on minimum ROC-AUC and it is only slightly behind the XGBoost-RF when judged by average ROC-AUC score. For models trained with HB lifetime features, the Extra Trees and the AdaBoost are the top performers using either average ROC-AUC scores or highest minimum ROC-AUC

scores. For models trained with combined HB number and lifetimes features, the Extra Trees classifier is the top performer using highest minimum ROC-AUC score or average ROC-AUC score. However, it should be noted that stellar performance observed during training does not necessarily translate into excellence in the validation stage, as will be seen in the next section.

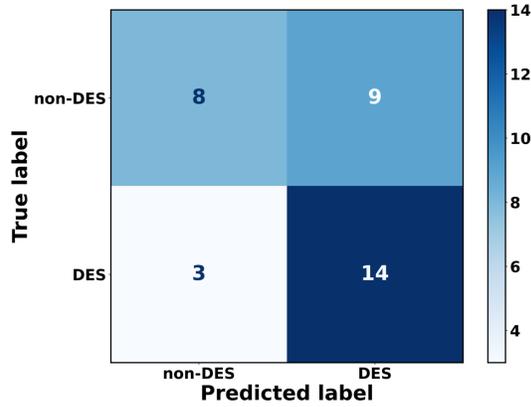
3.3 Model validation with experimental results

We validate the 30 trained models were using experimental results of 34 experimental results (17 non-DESSs and 17 DESSs) The results are presented in Table 3.

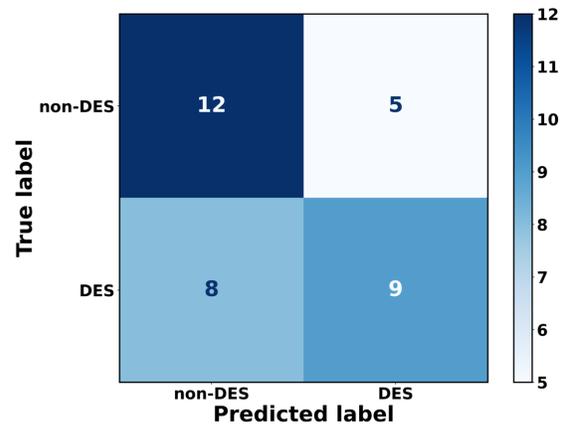
Table 3. ROC-AUC values of the trained models when tested with validation data. Top performing model under each feature type has its ROC-AUC value in bold.

S/N	Algorithm	lifetime	number	number + lifetime
1	Logistic regression	0.65	0.66	0.84
2	Decision Tree	0.52	0.69	0.65
3	Gradient Boost	0.57	0.77	0.81
4	AdaBoost	0.61	0.74	0.66
5	Random Forest	0.54	0.76	0.79
6	Extra Trees Forest	0.65	0.79	0.88
7	Support Vector Classifier	0.56	0.80	0.80
8	K-Nearest Neighbors	0.47	0.53	0.57
9	XGBoost	0.61	0.65	0.74
10	XGBoost Random Forest	0.68	0.69	0.79

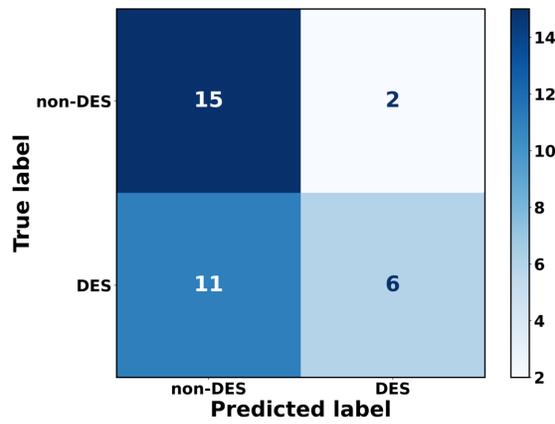
For models trained with HB lifetime features, the XGBoost Random Forest, Logistic regression and Extra Trees were the top performers with ROC-AUC values of 0.68, 0.65 and 0.65 respectively during validation. Support Vector, Extra Trees, and Gradient Boost were the top performers with ROC-AUC values of 0.80, 0.79 and 0.77 respectively when models were trained with HB number features. The Extra Trees, Logistic regression, and Gradient Boost were the top performing models with ROC-AUC of 0.88, 0.84 and 0.81 respectively for models trained with both HB numbers and lifetimes.



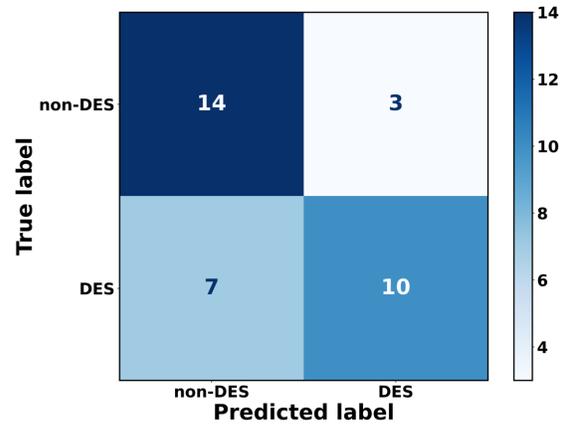
(a) XGBoost-RF



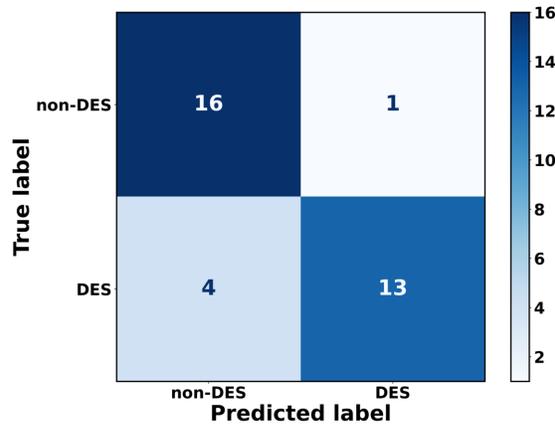
(b) Logistic Regression



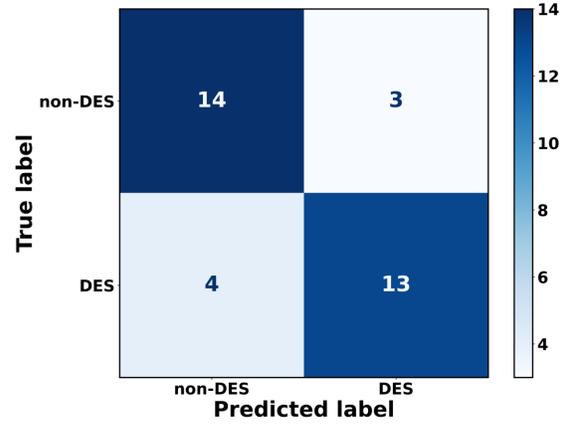
(c) Support Vector



(d) Extra Trees



(e) Extra Trees



(f) Logistic regression

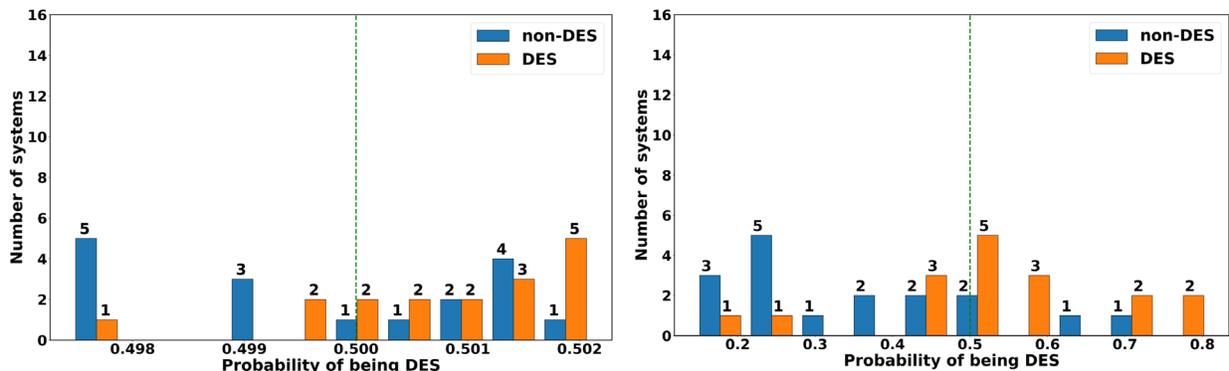
Figure 7. Confusion matrices for the top performing models during validation. (a) and (b) are for HB lifetime features (c) and (d) are for HB number features. (e) and (f) are for combined HB numbers and lifetimes features.

Figure 7 presents confusion matrices for the top performing models under each of the 3 input feature categories during validation. Confusion matrices present true positives, true negatives, false positives, and false negatives for each model's predictions. In this case, DESs are positives while non-DESs are negatives. Sensitivity measures how many DESs were correctly predicted to be DESs while specificity measures how many non-DESs were correctly predicted to be non-DESs by a model. Some models are better at predicting DESs (high sensitivity) while some are better at predicting non-DESs (high specificity).

XGBoost-RF is the top performing algorithm for models trained with HB lifetime features. It performs best at predicting which systems are DESs, as shown with its high sensitivity of 0.82 but it is not good at predicting which systems are non-DESs (low specificity of 0.47). For models trained with HB number features, the support vector is the top performing algorithm. It has a specificity of 0.88, which means it performs best at predicting which systems are non-DESs. Its low sensitivity of 0.35 means it is not good at predicting DESs.

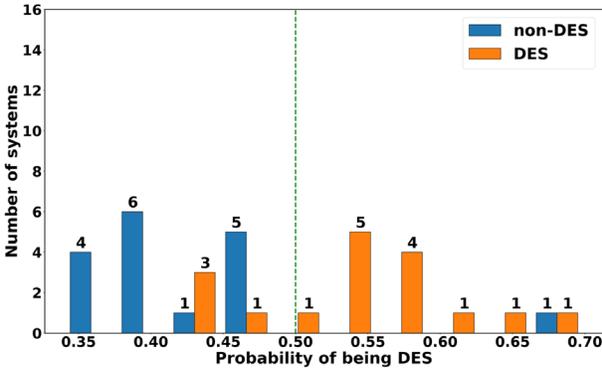
When models are trained with combined HB lifetime and number features as inputs, the Extra Trees model performs best. It has a sensitivity of 0.76, indicating it is among the top performers at predicting which systems are DESs. It has a specificity of 0.94, indicating it is the top performer at predicting which systems are non-DESs. Relative to the top performing models in other input feature categories, the Extra Trees algorithm is the best overall at predicting both DESs and non-DESs. The confusion matrices for all models are shown in S18-20.

Prediction probabilities



(a) XGBoost-RF

(b) Support vector

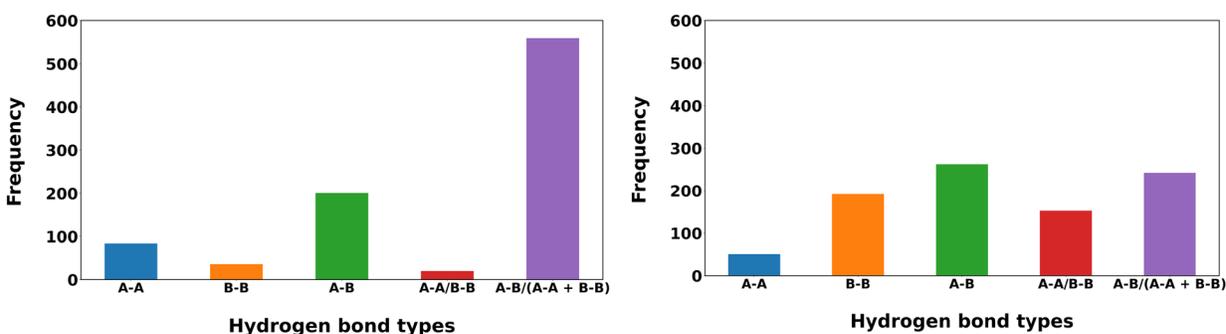


(c) Extra Trees

Figure 8. Distribution of prediction probabilities for the top performing models during validation. (a) is for HB lifetime features, (b) is for HB number features. (c) is for combined HB numbers and lifetimes features. The number of systems within each bin is indicated on the bars. The vertical dashed line indicates the classification threshold. Perfect model will have all non-DESs on the left and DESs on the right of the vertical dashed line.

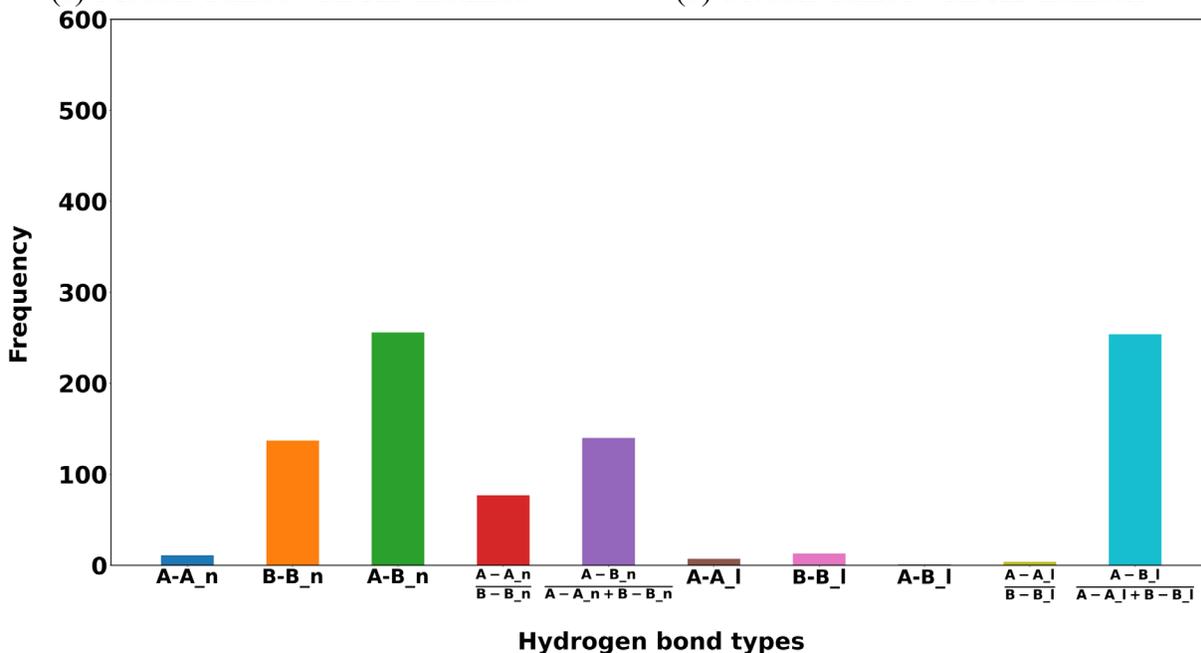
Prediction probabilities are useful indicators of how well each model separates DESs and non-DESs. An ideal model would have all its non-DES predictions with probability of being DES < 0.5 , and its DES predictions with probability of being DES > 0.5 . Figure 8 presents the distribution of prediction probabilities for the best models during validation. It can be seen from the probabilities in figure 8a that the predictions of the XGBoost-RF are closely distributed around 0.49 to 0.51 suggesting there is not much separation for models trained with lifetime features. Notably, all of the XGBoost-RF's 14 DES predictions made with confidence > 0.5 were correct. The separation improves in Figure 8b with probabilities distributed around 0.46 to 0.54, suggesting that HB number features help models detect non-DESs relatively better than lifetimes alone. This is backed up by the observation that all 15 non-DES predictions made by the support vector model with probability of being DES < 0.5 were correct. The probabilities are distributed between 0.30 and 0.70 in Figure 8c, indicating better confidence in the Extra Tree model's predictions when HB number and lifetime features are combined. The Extra Tree model shows better separation in its classifications and is relatively more confident in its non-DES predictions, and this is backed up by its specificity of 0.94 (Figure 7e). The prediction probabilities for all the other models are shown in Figures S21-23.

It is useful to have some insight into which input features carry the most weight when the ML models are making predictions. Figure 9 shows how the models ranked the importance of input features. Models that were trained with HB lifetime features alone overwhelmingly ranked the ratio of inter- to intra- species HB lifetime as the most important feature for predictions, followed by the inter-component HB lifetime. When trained with HB numbers features alone, the models ranked the inter-component HB numbers as the most important feature, but it should be noted that the ratio of inter- to intra- species HB numbers is not far behind in second place. When numbers and lifetimes were combined, the trained models ranked the inter-component HB numbers as the most important feature, closely followed by the ratio of inter- to intra- species HB lifetimes.



(a) Models trained with HB lifetimes.

(b) Models trained with HB numbers.



(c) Models trained with both HB numbers and lifetimes.

Figure 9. Important features during training iterations for all models. In (c), “_n” and “_l” denote HB number and lifetime features respectively.

Conclusion

We analyzed the HB features of 38 known DES and 111 known non-DES systems using MD simulation trajectories. The statistical analysis of inter- and intra-component HB numbers and HB lifetimes revealed two types of HB features for DESs: An imbalance between the two intra-component HB numbers in DES, and more and stronger inter-component HBs. We then developed 30 machine learning models by training 10 algorithms on 3 types of input features. We validated the models using 17 DESs and 17 non-DESs that have been experimentally verified. Using two criteria of highest average and highest minimum ROC-AUC scores, we found the logistic regression, gradient boost, support vector and Extra Trees to be among the top performers when using the HB lifetime, number, as well as combined lifetime and number features respectively. Extra Trees classifier was the top performing model overall with an average ROC-AUC of 0.88 when the HB numbers and lifetimes were combined. Intuitively, it makes sense that models would perform better when fed information about the population of HB numbers as well as how long those HBs last. All models ranked the inter-component as well as the ratio of inter- to intra-component HB numbers and lifetimes as the most important features for predicting a system to be DES or not. DESs are promising solvents that hold huge potential. Due to the sheer size of the candidate pool, it is important to have models that can accurately predict which compounds will or will not form DESs when mixed. Our work sheds light on which compounds are likely to form DESs but does not say what their physicochemical properties are likely to be. In the future, more work needs to be done to be able to predict which compounds will form DESs with application-specific properties.

REFERENCES

- [1] D.O. Abranches, L.P. Silva, M.A.R. Martins, S.P. Pinho, J.A.P. Coutinho, Understanding the Formation of Deep Eutectic Solvents: Betaine as a Universal Hydrogen Bond Acceptor, *ChemSusChem* 13(18) (2020) 4916-4921.
- [2] N.M. Stephens, E.A. Smith, Structure of Deep Eutectic Solvents (DESs): What We Know, What We Want to Know, and Why We Need to Know It, *Langmuir* 38(46) (2022) 14017-14024.

- [3] A.T. Celebi, N. Dawass, O.A. Moulτος, T.J.H. Vlught, How sensitive are physical properties of choline chloride-urea mixtures to composition changes: Molecular dynamics simulations and Kirkwood-Buff theory, *J Chem Phys* 154(18) (2021) 184502.
- [4] D.O. Abranches, J.A.P. Coutinho, Type V deep eutectic solvents: Design and applications, *Current Opinion in Green and Sustainable Chemistry* 35 (2022).
- [5] R. Alcalde, A. Gutiérrez, M. Atilhan, S. Aparicio, An experimental and theoretical investigation of the physicochemical properties on choline chloride – Lactic acid based natural deep eutectic solvent (NADES), *Journal of Molecular Liquids* 290(110916) (2019).
- [6] C.H.J.T. Dietz, A. Erve, M.C. Kroon, M. van Sint Annaland, F. Gallucci, C. Held, Thermodynamic properties of hydrophobic deep eutectic solvents and solubility of water and HMF in them: Measurements and PC-SAFT modeling, *Fluid Phase Equilibria* 489 (2019) 75-82.
- [7] C. Florindo, L.C. Branco, I.M. Marrucho, Development of hydrophobic deep eutectic solvents for extraction of pesticides from aqueous environments, *Fluid Phase Equilibria* 448 (2017) 135-142.
- [8] H. Kivela, M. Salomaki, P. Vainikka, E. Makila, F. Poletti, S. Ruggeri, F. Terzi, J. Lukkari, Effect of Water on a Hydrophobic Deep Eutectic Solvent, *J Phys Chem B* 126(2) (2022) 513-527.
- [9] A. Kovacs, E.C. Neyts, I. Cornet, M. Wijnants, P. Billen, Modeling the Physicochemical Properties of Natural Deep Eutectic Solvents, *ChemSusChem* (2020).
- [10] T. Křížek, M. Bursová, R. Horsley, M. Kuchař, P. Tůma, R. Čabala, T. Hložek, Menthol-based hydrophobic deep eutectic solvents: Towards greener and efficient extraction of phytocannabinoids, *Journal of Cleaner Production* 193 (2018) 391-396.
- [11] K. Li, Y. Jin, D. Jung, K. Park, H. Kim, J. Lee, In situ formation of thymol-based hydrophobic deep eutectic solvents: Application to antibiotics analysis in surface water based on liquid-liquid microextraction followed by liquid chromatography, *J Chromatogr A* 1614 (2020) 460730.
- [12] M. Lukaczynska-Anderson, M.H. Mamme, A. Ceglia, K. Van den Bergh, J. De Strycker, F. De Proft, H. Terryn, J. Ustarroz, The role of hydrogen bond donor and water content on the electrochemical reduction of Ni(2+) from solvents - an experimental and modelling study, *Phys Chem Chem Phys* 22(28) (2020) 16125-16135.

- [13] M.A.R. Martins, L.P. Silva, N. Schaeffer, D.O. Abranches, G.J. Maximo, S.P. Pinho, J.A.P. Coutinho, Greener Terpene–Terpene Eutectic Mixtures as Hydrophobic Solvents, *ACS Sustainable Chemistry & Engineering* 7(20) (2019) 17414-17423.
- [14] D. Tolmachev, N. Lukasheva, R. Ramazanov, V. Nazarychev, N. Borzdun, I. Volgin, M. Andreeva, A. Glova, S. Melnikova, A. Dobrovskiy, S.A. Silber, S. Larin, R.M. de Souza, M.C.C. Ribeiro, S. Lyulin, M. Karttunen, Computer Simulations of Deep Eutectic Solvents: Challenges, Solutions, and Perspectives, *Int J Mol Sci* 23(2) (2022).
- [15] B.B. Hansen, S. Spittle, B. Chen, D. Poe, Y. Zhang, J.M. Klein, A. Horton, L. Adhikari, T. Zelovich, B.W. Doherty, B. Gurkan, E.J. Maginn, A. Ragauskas, M. Dadmun, T.A. Zawodzinski, G.A. Baker, M.E. Tuckerman, R.F. Savinell, J.R. Sangoro, Deep Eutectic Solvents: A Review of Fundamentals and Applications, *Chem Rev* 121(3) (2021) 1232-1285.
- [16] L. Zamora, C. Benito, A. Gutierrez, R. Alcalde, N. Alomari, A.A. Bodour, M. Atilhan, S. Aparicio, Nanostructuring and macroscopic behavior of type V deep eutectic solvents based on monoterpenoids, *Phys Chem Chem Phys* 24(1) (2021) 512-531.
- [17] F. Bergua, M. Castro, C. Lafuente, M. Artal, Thymol+1-menthol eutectic mixtures: Thermophysical properties and possible applications as decontaminants, *Journal of Molecular Liquids* 368 (2022).
- [18] F. Bergua, M. Castro, J. Muñoz-Embid, C. Lafuente, M. Artal, L-menthol-based eutectic solvents: Characterization and application in the removal of drugs from water, *Journal of Molecular Liquids* 352 (2022).
- [19] M. Abdollahzadeh, M. Khosravi, B. Hajipour Khire Masjidi, A. Samimi Behbahan, A. Bagherzadeh, A. Shahkar, F. Tat Shahdost, Estimating the density of deep eutectic solvents applying supervised machine learning techniques, *Sci Rep* 12(1) (2022) 4954.
- [20] Y. Dai, G.J. Witkamp, R. Verpoorte, Y.H. Choi, Tailoring properties of natural deep eutectic solvents with water to facilitate their applications, *Food Chem* 187 (2015) 14-9.
- [21] A. Gutierrez, S. Aparicio, M. Atilhan, Design of arginine-based therapeutic deep eutectic solvents as drug solubilization vehicles for active pharmaceutical ingredients, *Phys Chem Chem Phys* 21(20) (2019) 10621-10634.
- [22] A. Gutierrez, M. Atilhan, S. Aparicio, A theoretical study on lidocaine solubility in deep eutectic solvents, *Phys Chem Chem Phys* 20(43) (2018) 27464-27473.

- [23] M.H. Zainal-Abidin, M. Hayyan, G.C. Ngoh, W.F. Wong, C.Y. Looi, Emerging frontiers of deep eutectic solvents in drug discovery and drug delivery systems, *J Control Release* 316 (2019) 168-195.
- [24] X. Zhong, C. Velez, O. Acevedo, Partial Charges Optimized by Genetic Algorithms for Deep Eutectic Solvent Simulations, *J Chem Theory Comput* 17(5) (2021) 3078-3087.
- [25] N. Chaabene, K. Ngo, M. Turmine, V. Vivier, New hydrophobic deep eutectic solvent for electrochemical applications, *Journal of Molecular Liquids* 319 (2020).
- [26] T. Hanada, M. Goto, Synergistic Deep Eutectic Solvents for Lithium Extraction, *ACS Sustainable Chemistry & Engineering* 9(5) (2021) 2152-2160.
- [27] L. Yurramendi, J. Hidalgo, A. Siriwardana, A Sustainable Process for the Recovery of Valuable Metals from Spent Lithium Ion Batteries by Deep Eutectic Solvents Leaching, *International Conference on Raw Materials and Circular Economy*, 2022.
- [28] K. Du, E.H. Ang, X. Wu, Y. Liu, Progresses in Sustainable Recycling Technology of Spent Lithium - Ion Batteries, *Energy & Environmental Materials* 5(4) (2022) 1012-1036.
- [29] J. Neumann, M. Petranikova, M. Meeus, J.D. Gamarra, R. Younesi, M. Winter, S. Nowak, Recycling of Lithium - Ion Batteries—Current State of the Art, Circular Economy, and Next Generation Recycling, *Advanced Energy Materials* 12(17) (2022).
- [30] S. Tang, M. Zhang, M. Guo, A Novel Deep-Eutectic Solvent with Strong Coordination Ability and Low Viscosity for Efficient Extraction of Valuable Metals from Spent Lithium-Ion Batteries, *ACS Sustainable Chemistry & Engineering* 10(2) (2022) 975-985.
- [31] J. Zhang, M. Wenzel, J. Steup, G. Schaper, F. Hennersdorf, H. Du, S. Zheng, L.F. Lindoy, J.J. Weigand, 4-Phosphoryl Pyrazolones for Highly Selective Lithium Separation from Alkali Metal Ions, *Chemistry* 28(1) (2022) e202103640.
- [32] Y. Chen, Y. Wang, Y. Bai, Y. Duan, B. Zhang, C. Liu, X. Sun, M. Feng, T. Mu, Significant Improvement in Dissolving Lithium-Ion Battery Cathodes Using Novel Deep Eutectic Solvents at Low Temperature, *ACS Sustainable Chemistry & Engineering* 9(38) (2021) 12940-12948.
- [33] K. Wang, T. Hu, P. Shi, Y. Min, J. Wu, Q. Xu, Efficient Recovery of Value Metals from Spent Lithium-Ion Batteries by Combining Deep Eutectic Solvents and Coextraction, *ACS Sustainable Chemistry & Engineering* 10(3) (2021) 1149-1159.
- [34] G. Zante, M. Boltoeva, Review on Hydrometallurgical Recovery of Metals with Deep Eutectic Solvents, *Sustainable Chemistry* 1(3) (2020) 238-255.

- [35] L. Chen, Y. Chao, X. Li, G. Zhou, Q. Lu, M. Hua, H. Li, X. Ni, P. Wu, W. Zhu, Engineering a tandem leaching system for the highly selective recycling of valuable metals from spent Li-ion batteries, *Green Chemistry* 23(5) (2021) 2177-2184.
- [36] M.K. Tran, M.-T.F. Rodrigues, K. Kato, G. Babu, P.M. Ajayan, Deep eutectic solvents for cathode recycling of Li-ion batteries, *Nature Energy* 4(4) (2019) 339-345.
- [37] S. Wang, Z. Zhang, Z. Lu, Z. Xu, A novel method for screening deep eutectic solvent to recycle the cathode of Li-ion batteries, *Green Chemistry* 22(14) (2020) 4473-4482.
- [38] N. Aguilar, R. Barros, J. Antonio Tamayo-Ramos, S. Martel, A. Bol, M. Atilhan, S. Aparicio, Carbon nanomaterials with Thymol + Menthol Type V natural deep eutectic solvent: From surface properties to nano-Venturi effect through nanopores, *Journal of Molecular Liquids* 368 (2022).
- [39] M. Tiecco, F. Cappellini, F. Nicoletti, T. Del Giacco, R. Germani, P. Di Profio, Role of the hydrogen bond donor component for a proper development of novel hydrophobic deep eutectic solvents, *Journal of Molecular Liquids* 281 (2019) 423-430.
- [40] M.H. Zainal-Abidin, M. Hayyan, W.F. Wong, Hydrophobic deep eutectic solvents: Current progress and future directions, *Journal of Industrial and Engineering Chemistry* 97 (2021) 142-162.
- [41] R. Paul, A. Mitra, S. Paul, Phase separation property of a hydrophobic deep eutectic solvent-water binary mixture: A molecular dynamics simulation study, *J Chem Phys* 154(24) (2021) 244504.
- [42] P. Makoś, E. Słupek, J. Gębicki, Extractive detoxification of feedstocks for the production of biofuels using new hydrophobic deep eutectic solvents – Experimental and theoretical studies, *Journal of Molecular Liquids* 308 (2020).
- [43] F.O. Farias, J.F.B. Pereira, J.A.P. Coutinho, L. Igarashi-Mafra, M.R. Mafra, Understanding the role of the hydrogen bond donor of the deep eutectic solvents in the formation of the aqueous biphasic systems, *Fluid Phase Equilibria* 503 (2020).
- [44] P. Vainikka, S. Thallmair, P.C.T. Souza, S.J. Marrink, Martini 3 Coarse-Grained Model for Type III Deep Eutectic Solvents: Thermodynamic, Structural, and Extraction Properties, *ACS Sustainable Chemistry & Engineering* 9(51) (2021) 17338-17350.
- [45] M. Atilhan, S. Aparicio, Molecular dynamics simulations of mixed deep eutectic solvents and their interaction with nanomaterials, *Journal of Molecular Liquids* 283 (2019) 147-154.

- [46] I.I.I. Alkhatib, D. Bahamon, F. Llovel, M.R.M. Abu-Zahra, L.F. Vega, Perspectives and guidelines on thermodynamic modelling of deep eutectic solvents, *Journal of Molecular Liquids* 298(112183) (2020).
- [47] I. Adeyemi, M.R.M. Abu-Zahra, I.M. AlNashef, Physicochemical properties of alkanolamine-choline chloride deep eutectic solvents: Measurements, group contribution and artificial intelligence prediction techniques, *Journal of Molecular Liquids* 256 (2018) 581-590.
- [48] K. Shahbaz, F.S.G. Bagh, F.S. Mjalli, I.M. AlNashef, M.A. Hashim, Prediction of refractive index and density of deep eutectic solvents using atomic contributions, *Fluid Phase Equilibria* 354 (2013) 304-311.
- [49] F.S.G. Bagh, K. Shahbaz, F.S. Mjalli, I.M. AlNashef, M.A. Hashim, Electrical conductivity of ammonium and phosphonium based deep eutectic solvents: Measurements and artificial intelligence-based prediction, *Fluid Phase Equilibria* 356 (2013) 30-37.
- [50] X. Xu, J. Range, G. Gygli, J. Pleiss, Analysis of Thermophysical Properties of Deep Eutectic Solvents by Data Integration, *Journal of Chemical & Engineering Data* 65(3) (2019) 1172-1179.
- [51] A.K. Halder, R. Haghbakhsh, I.V. Voroshylova, A.R.C. Duarte, M. Cordeiro, Density of Deep Eutectic Solvents: The Path Forward Cheminformatics-Driven Reliable Predictions for Mixtures, *Molecules* 26(19) (2021).
- [52] U.L. Abbas, Q. Qiao, M.T. Nguyen, J. Shi, Q. Shao, Molecular Dynamics Simulations of Heterogeneous Hydrogen Bond Environment in Hydrophobic Deep Eutectic Solvents, *AIChE Journal* (2021).
- [53] D.J.G.P. van Osch, C.H.J.T. Dietz, S.E.E. Warrag, M.C. Kroon, The Curious Case of Hydrophobic Deep Eutectic Solvents: A Story on the Discovery, Design, and Applications, *ACS Sustainable Chemistry & Engineering* (2020).
- [54] D.J.G.P. van Osch, C.H.J.T. Dietz, J. van Spronsen, M.C. Kroon, F. Gallucci, M. van Sint Annaland, R. Tuinier, A Search for Natural Hydrophobic Deep Eutectic Solvents Based on Natural Components, *ACS Sustainable Chemistry & Engineering* 7(3) (2019) 2933-2942.
- [55] M.J. Robertson, J. Tirado-Rives, W.L. Jorgensen, Improved Peptide and Protein Torsional Energetics with the OPLSAA Force Field, *J Chem Theory Comput* 11(7) (2015) 3499-509.

- [56] L.S. Dodda, I. Cabeza de Vaca, J. Tirado-Rives, W.L. Jorgensen, LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands, *Nucleic Acids Res* 45(W1) (2017) W331-W336.
- [57] W.H.A.D.K. Schulten, VMD: Visual Molecular Dynamics, *Journal of Molecular Graphics* 14(1) (1996) 33-38.
- [58] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, *The Journal of Chemical Physics* 81(8) (1984) 3684-3690.
- [59] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems, *The Journal of Chemical Physics* 98(12) (1993) 10089-10092.
- [60] H.B. Berk Hess, Herman J. C. Berendsen,, J.G.E.M. Fraaije, LINCS: A Linear Constraint Solver for Molecular Simulations, *Journal of Computational Chemistry* 18(12) (1997) 1463-1472.
- [61] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1-2 (2015) 19-25.
- [62] A. Luzar, D. Chandler, Hydrogen-bond kinetics in liquid water, *Nature* 379(6560) (1996) 55-57.
- [63] A. Luzar, Resolving the hydrogen bond dynamics conundrum, *The Journal of Chemical Physics* 113(23) (2000) 10663-10675.
- [64] G.L. Lars Buitinck, Mathieu Blondel, Fabian Pedregosa,, O.G. Andreas C. Muller, Vlad Niculae, Peter Prettenhofer,, J.G. Alexandre Gramfort, Robert Layton, Jake Vanderplas,, B.H. Arnaud Joly, and Gael Varoquaux, API design for machine learning software: experiences from the scikit-learn project, *European Conference on Machine Learning and Principles and Practices of Knowledge Discovery in Databases* (2013) (2013) 108-122.
- [65] F.a.V. Pedregosa, G. and Gramfort, A. and Michel, V., B.a.G. and Thirion, O. and Blondel, M. and Prettenhofer, P., R.a.D. and Weiss, V. and Vanderplas, J. and Passos, A. and, D.a.B. Cournapeau, M. and Perrot, M. and Duchesnay, E., Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825-2830.

- [66] T.a.G. Chen, Carlos, XGBoost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining ACM, San Francisco, California, USA, 2016, pp. 785–794.
- [67] W. McKinney, Data Structures for Statistical Computing in Python, in: S.v.d.W.a.J. Millman (Ed.) Proceedings of the 9th Python in Science Conference, 2010, pp. 56-61.
- [68] C.R. Harris, K.J. Millman, S.J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N.J. Smith, R. Kern, M. Picus, S. Hoyer, M.H. van Kerkwijk, M. Brett, A. Haldane, J.F. Del Rio, M. Wiebe, P. Peterson, P. Gerard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T.E. Oliphant, Array programming with NumPy, *Nature* 585(7825) (2020) 357-362.
- [69] J.D. Hunter, Matplotlib: A 2D graphics environment, *Computing in Science & Engineering* 9 (2007) 90--95.
- [70] P. Virtanen, R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, I. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, C. SciPy, SciPy 1.0: fundamental algorithms for scientific computing in Python, *Nat Methods* 17(3) (2020) 261-272.