Efficient Machine-Learning-Based New Tools to Design Eutectic Mixtures and Predict Their Viscosity

Stella Christodoulou, Camille Cousseau, Emmanuelle Limanton, Lorris Toucouere, Fabienne Gauffre, Béatrice Legouin, Laurent Maron, Ludovic Paquin, Romuald Poteau

Efficient machine learning-based new tools to design eutectic mixtures and predict their viscosity

Stella Christodoulou[1] ‡, Camille Cousseau[2] ‡, Emmanuelle Limanton[2], Lorris Toucouere[1], Fabienne Gauffre[2], Béatrice Legouin[2], Laurent Maron[1], Ludovic Paquin[2] and Romuald Poteau[1]***

[1]Laboratoire de Physique et Chimie des Nano-Objets (LPCNO, INSA-UPS-CNRS UMR 5215), Université de Toulouse, 135 avenue de Rangueil, 31077 Toulouse Cedex, France

[2]University of Rennes, CNRS, ISCR-UMR6226, F-35000 Rennes, France

* Email: ludovic.paquin@univ-rennes.fr ; Phone: +33 2 23 23 69 83

** Email: romuald.poteau@univ-tlse3.fr ; Phone: +33 5 61 55 96 64

KEYWORDS eutectic mixture, data mining, regression, classification, viscosity, XAI

ABSTRACT

The development of models that accurately predict the formation of Eutectic Mixtures (EM) – including the well-known Deep Eutectic Solvents (DES) – and their viscosity is imperative to save time in synthesizing new solvents. We developed reliable machine learning-based classifiers able to discern between eutectic and non-eutectic (non-EM) mixtures and regressors able to predict the viscosity of EM. A new experimental dataset of 219 EM, 384 non-EM and

1450 viscosity points at different temperatures and water contents is provided and used to challenge several models, defined both by an algorithm and by descriptors. The top performing EM/non-EM classifier yields accuracy of 92% and the best regressor achieves viscosity predictions with a Mean Absolute Error (MAE) of 2.2 mPa.s; the extrapolation capabilities of the latter were assessed on additional measurements at temperatures and water contents outside the range of the training dataset, revealing good accuracy at low viscosities. The SHapley Additive exPlanations (SHAP) algorithm was employed in several models as an eXplainable Artificial Intelligence (XAI) technique to quantify input feature contributions to model output. These results represent a significant step forward in developing robust and highly accurate models for determining eutectic mixtures and their viscosity.

Introduction

Deep eutectic solvents (DES) have emerged as promising environmentally friendly alternatives to conventional organic solvents. Since their discovery by Abbott et al. in 2003, DES have been applied in a wide range of fields.[1] Whether in catalysis[2,3] or biocatalysis[4], analytical chemistry[5], organic synthesis[6], electrochemistry[7], extraction, medicine or CO_2 capture[8], they have found their place among conventional solvents.

DES are composed of at least two compounds associated by weak interactions, such as hydrogen bonds or electrostatic interactions. This combination enables the mixture to be liquid at a lower temperature than if it behaved as an ideal mixture. The so-called "eutectic point" is the temperature and composition at which the mixture is liquid. In the case of DES, this eutectic point is significantly "deeper" than the ideal eutectic point, *i.e*. at much lower temperature.[9,10] DES have a number of advantages: as well as being easy to prepare, they have physico-

chemical properties that can be modulated for a specific application according to their composition, such as viscosity, pH, polarity, density or conductivity.[11]

Given the multitude of possible mixtures, it is essential to be able to rationalize the design of new DES, and to predict their influential properties according to their intended application. For instance, one major drawback of DES lies in their generally high viscosity, a characteristic that poses practical hurdles in their utilization for industrial applications.[12,13] It is reported that the high viscosity of DES can be attributed to a notable hydrogen bonding network, as well as to large molecular size, small void volume or electrostatic or van der Waals interactions.[14] The viscosity of DES is known to also be influenced by other factors, including the temperature, the water content and the molar ratio. Moreover, great variability can be observed in experimental measurements of a given DES under the same conditions; for instance, for Choline chloride-Urea (1:2) at 30°C, viscosities close to 527 mPa.s are observed by Dou et al. (527.0 mPa.s), Yadav and Pandey (527.3 mPa.s) as well as Shekaari et al. (527.1 mPa.s), while Xie et al. reported a viscosity of 953.7 mPa.s.[15–18] Hence, special attention should be given to outliers or other possible inconsistencies within the dataset. Given the complexity of the hydrogen bonding network, the non-ideal behavior of DES and the number of parameters affecting their physical properties, predicting the viscosity of DES is a very challenging task. Hence, the demand for accurate predictive models has become paramount. Several models can be found in the current literature using either experimental viscosity data[19,20] or thermodynamic properties, such as the COSMO-RS (COnductor-like Screening MOdel for Real Solvents) model.[21–24] However, the COSMO-RS model has so far mainly been applied to the prediction of Solid-Liquid Equilibria (SLE) of DES based on choline chloride, and has certain limitations, which will be discussed later in more detail.[21–24] Machine learning (ML) models have also been in the spotlight for the prediction of the viscosities of DES.[25,26] There is still room for evaluating new models and descriptors on the basis of a set of experimental data obtained in similar

conditions and in a relatively large chemical space. The availability of predictive models for the feasibility and the viscosity of DES taking the water content into account is limited, as already underlined by Bakhtyari et al.[19], primarily due to the lack of this information in datasets. ML models have been developed for the prediction of other properties of DES, for instance Ayres et al. developed an extreme gradient boosting model to predict the melting point of DES and evaluated several sets of descriptors for this task. Their top performing model was trained over the molecular fingerprints (ECFP) of the components of DES, the molar ratios and a set of other chemical descriptors.[27]

In this study, we focus on eutectic mixtures (EM) in the broad sense, that means mixtures that are liquid and homogeneous at working temperature and do not necessarily exhibit a "*deep*" eutectic point. We report a new dataset of 1450 viscosity points of EM in a wide range of temperatures, molar ratios and water contents, alongside additional data representing a total of 219 EM and 384 non-eutectic mixtures (non-EM), *i.e.* mixtures of compounds that have not liquefied completely or at all, all collected through our experimental investigations and measurements. Data analysis techniques were used to uncover patterns and correlations between the data and the property of the mixture (EM or not). The dataset was then used with machine learning classification models aiming to discern between EM and non-EM and with regression models established for viscosity predictions. Several structural descriptors were evaluated for both classification and regression tasks. EXplainable Artificial Intelligence (XAI) techniques were employed in order to interpret and rationalize the predictions. It is both a way to open the AI black box, to verify that a seemingly good behavior of an algorithm is chemically grounded and to investigate the system from an AI perspective.

Experimental Procedures

Chemicals

All compounds were purchased from Alfa Aesar. More details are available in ESI.

Mixtures Preparation

EM and non-EM were prepared by mixing and heating the components from 65°C to 80°C until homogeneous liquids were obtained for EM; in cases where the mixture melted only partially or not at all, heating was stopped after several hours. In some cases, a known quantity of water was added to the mixture; water was not considered as a third component of the mixture. Non-EM were not stored, while EM were cooled to room temperature and stored at the same temperature. They have been assigned internal laboratory identifiers, in the form "OC-xx-xx" or "DIV-xx-xx".

Water Content

The water content of the eutectic mixtures was determined in mass percentage using a Karl Fischer 831 KF Coulometer titrator from Metrohm. The percentage given is the average of 3 measurements.

Viscosity

Viscosity measurements were performed with an Anton Paar MCR301 rheometer using parallel plate geometry. The upper plane has a diameter of 75 mm (reference PP75/P2 1288) and the surface of the wall of the plates in contact with the samples was roughened to prevent slipping. Temperature was controlled by a Peltier system. Measurements were made at a fixed shear rate of 4 s^{-1}. Temperature variation ranged from 25°C to 75°C for initial database measurements; for extrapolation tests, temperature variation ranged from 20°C to 90°C.

Modeling Procedures

Comparison of Existing Methods

Due to their ease of preparation, eutectic mixtures have been developed and applied for many years with little fundamental knowledge and molecular understanding.[28] Especially since there is an infinite number of possible combinations of compounds for the formation of new EM, it has become crucial to be able to define upstream whether a mixture of compounds can become an EM and to know its physico-chemical properties, in order to determine whether it will be interesting for the desired application. Quite a few studies have been carried out in this field, and we propose a brief overview in order to better understand the interest in developing new predictive models in this article.

Models based on thermodynamics and quantum chemistry

Several thermodynamic models have been used in the literature as predictive tools, mainly to identify new eutectic mixtures. These include the UNIFAC, PC-SAFT and COSMO-SAC methods.[29–31] The best known is undoubtedly COSMO-RS. Developed by Klamt, COSMO-RS is a quantum chemistry-based method which can be used to predict the chemical potential of pure compounds or mixtures.[32] Using this method, it is possible to predict the SLE diagram of a mixture, useful for eutectic solvents screening. The prediction of SLE diagrams highlights a window at which the mixture is liquid; it can therefore be used to determine the eutectic temperature T_e, or eutectic point, and the associated ratio.[22] Several studies have been carried out for the evaluation of this method as a predictive tool for the SLE of binary eutectic solvents, and to our knowledge have focused mainly on the association of a quaternary ammonium, choline chloride (ChCl) with a hydrogen bond donor (HBD).[21–24] Although some results are

promising, these predictions by COSMO-RS have optimization constraints; according to Song et al., depending on the different conformers employed and parameterizations of the software used for the SLE calculation the SLE diagram of some mixtures cannot be predicted.[22] Overall, the main limitation of thermodynamic models is that these methods require knowledge of certain physico-chemical properties such as the melting temperature or enthalpy of fusion of the compounds, which are not always available in the literature or measurable experimentally. For example, some compounds, such as amino acids, degrade before reaching their melting point.[21]

For viscosity prediction, COSMO-RS is mainly used to generate σ-profiles of molecules, which allow the molecular structure to be described and then included in a model.[25,33,34] The direct viscosity prediction function in the software is not commonly used for EM as far as we are aware. Mutalib et al. compared predicted and experimental values in the case of ionic liquid (IL) derivatives, solvents described by some as the predecessors of DES. It was found that COSMO-RS significantly underestimated the viscosity of these IL derivatives.[33]

Whether predicting the formation of eutectic mixtures or their viscosity, COSMO-RS studies rarely take the mixture's water content into account.

Models based on experimental data

Other models exist, based on experimental data. Most of them can be used to predict the physico-chemical properties of eutectic solvents, such as their viscosity. For instance, Al-Dawsary et al. employ the Arrhenius and Vogel-Fulcher-Tamman (VFT) equations to fit experimental viscosity data to temperature for 70 DES and obtained fitting parameters for each of the considered DES used to estimate viscosities.[20] Bakhtyari et al. propose a global model for the estimation of the viscosities of DES over wide ranges of temperatures using critical temperature, critical pressure and a reference viscosity point as input variables; however, these

values are not always easily available.[19] The authors also optimized a DES-specific VFT model.

On the machine learning side, Roosta et al. employed the Least Square Support Vector Machine (LSSVM)[35] and Multilayer Perceptron Artificial Neural Network (MLPANN)[36] algorithms coupled with the Group Contribution (GC) method[37], temperature and molar ratios, on a databank of 2533 data points collected from the literature.[38] Mohan et al. developed several machine learning models to predict viscosities of DES, using a dataset of 4949 data points at various temperatures and molar ratios, and they employed COSMO-RS σ-profiles as training features.[26] Cat-Boost was their best-performing model. While these models yield highly satisfactory predictions, they do not consider the water content of the solvent, a parameter that has a significant impact on the viscosity.[17,39] Benguerba et al. proposed a Quantitative Structure Property Relationship (QSPR) model using multilinear regression (MLR) and Artificial Neural Network (ANN) methods applied to 108 experimental viscosity measurements of DES at different temperatures using σ-profiles derived from COSMO-RS as descriptors.[25] The ANN model outputs satisfactory predictions on the validation set, however its applicability is limited to amine-based DES. Shi et al. used a dataset of 994 experimental viscosities in different temperatures, along with basic properties, Morgan fingerprints and water content as input features for Support Vector Regression (SVR)[40,41], Random Forests (RF)[42,43], Neural Network (NN)[44] and XGBoost[45] algorithms for viscosity predictions.[46] The latter exhibited the best performance. More models for viscosity prediction can be found in the literature; some of these have enabled Odegova et al. to develop an open-access predictive tool, DESignSolvents. Note that this tool does not take into account the water content of mixtures in its predictions.[47–49]

Other models

Fewer models can predict the formation of eutectic solvents. Abbas et al. analyzed the hydrogen bonding features of known DES and non-Deep Eutectic Solvents (non-DES) systems using Molecular Dynamics (MD) trajectories and developed several machine learning models aiming to discern between the two classes, using input features based on intra-component and inter-component hydrogen bonds. However, their models have only been validated on 17 DES and 17 non-DES.[50]

In summary, although predictive models for the design of new EM or for estimating their viscosity are being developed, those that take the water content of EM into account are scarce.

Development of New Machine Learning Models

We therefore present nine classification models employing NN, Support Vector Machines (SVM) and RF (**Figure 1**). These algorithms are amongst the most common classification algorithms, and they achieved satisfactory accuracy in the work of Abbas et al.[50] In addition, they differ significantly and employing them for the same task covers different algorithmic perspectives. NN, SVM and RF have also been successfully employed for other classification tasks in materials science.[51,52] The dataset used with the classifiers comprises 219 EM and 384 non-EM. Three structural descriptors were evaluated with each algorithm: Morgan fingerprints[53], Bag of Bonds (BoB)[54] and Bond-Angle-Torsion (BAT)[55], with and without performing Principal Component Analysis (PCA) for dimensionality reduction. We also present two new regression models employing NN to predict the viscosities of EM, for which we evaluate two structural descriptors: Morgan fingerprints without PCA dimensionality reduction, Morgan fingerprints upon PCA dimensionality reduction, BAT upon PCA dimensionality reduction. The dataset used with the regressors comprises 1450 experimental viscosity measurements.

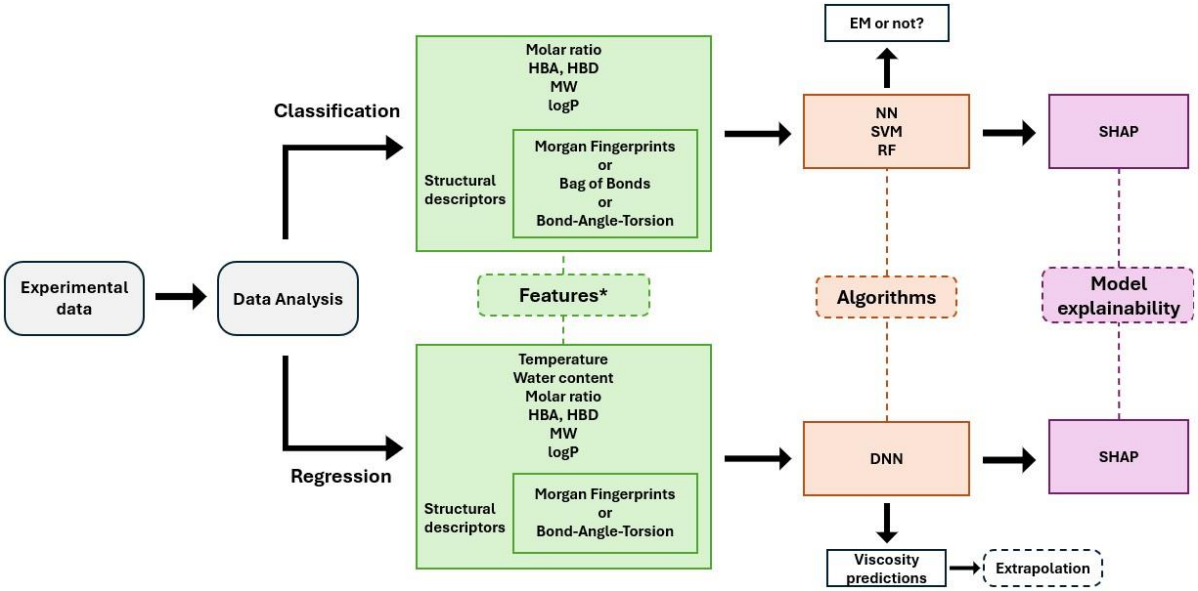


Figure 1: Workflow used for the development of classification (EM/non-EM) and regression (viscosity) models.

***Models were assessed with and without PCA for dimensionality reduction on structural descriptors. Regression employing BAT was only assessed upon PCA dimensionality reduction**

Most machine learning algorithms applied to chemistry, cheminformatics and materials science require to translate 2D or 3D molecular structures into numerical representations. Yet, the training and the accuracy of a ML algorithm is crucially dependent on the choice of data representation in a machine-readable form, also called "featurization". Whatever the method chosen, highly problem-specific, the resulting representation must be a unique identifier, invariant with respect to rotation operations, translation in space and permutation of the ordering of the atoms. The SMILES, Morgan fingerprints, BoB and BAT used in the present study are descriptors often encountered in ML applications to chemistry. Morgan fingerprints are a commonly used descriptor for the prediction of properties of DES, with a recent variant aiming to take into account the number of atom groups[56], a possibility that has not been

thoroughly explored in the present paper. In this work, we used the ECFP implementation of the Morgan fingerprints with radius=3 and nbits=2048 per molecule. As far as we are aware, BoB and BAT have not been used in the context of eutectic mixtures yet. Our work can serve as a first example of the potential of such descriptors for this specific task. A nice summary can be found in the work of Raghunathan and Priyakumar.[57] These descriptors are schematized in **Figure 2**. To describe the electron density distribution in molecules in a more realistic way, the original version of BoB and BAT descriptors has been modified such that the generated matrix is calculated on the natural charges, also known as natural population charges, of the involved atoms instead of their atomic numbers. More details about BoB and BAT descriptors can be found in the original articles.[54,55] Natural charges were obtained through Density Functional Theory (DFT) calculations (**see section S2 in ESI for computational details**).

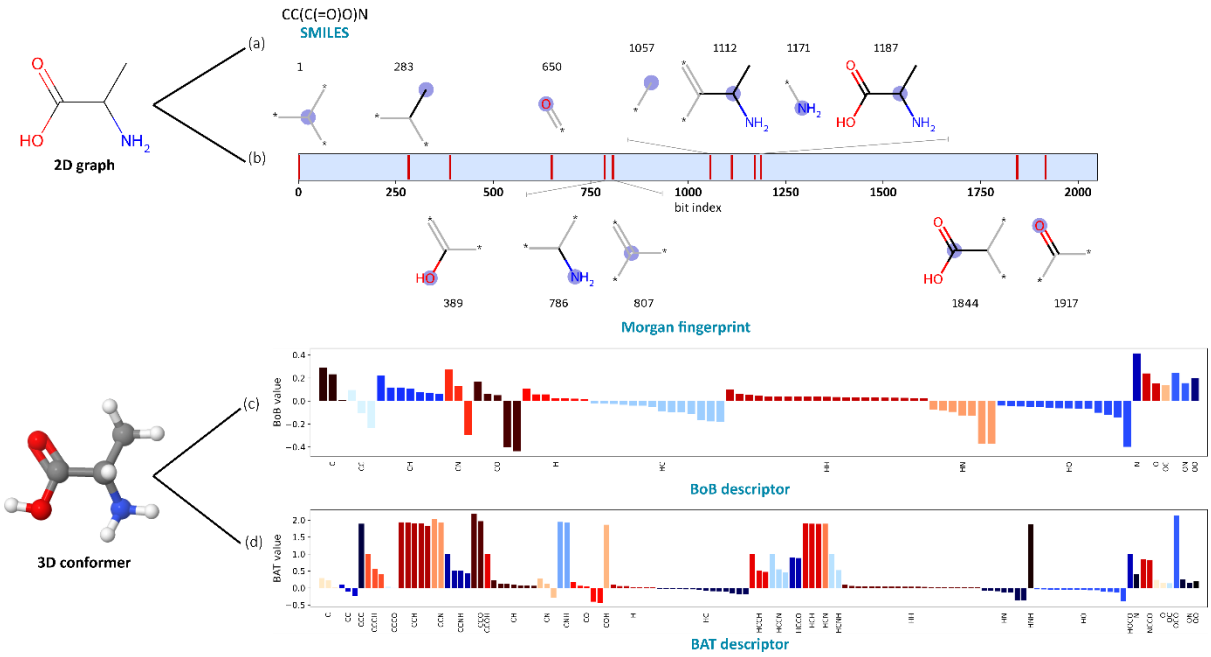


Figure 2: Molecular featurization of alanine. (a) SMILES is a textual representation that encodes the molecular structure as a string, that can be further processed into binary features; (b) Morgan fingerprints are binary vectors where each non-zero bit represents a specific substructure within the molecule; (c) The Bag of Bonds (BoB) representation is an extension of the Bag-of-Words concept used in natural language processing. It is based

on the types and counts of bonds it contains; (d) The Bond-Angle-Torsion (BAT) representation is expected to improve the BoB featurization by adding direct and torsional angles contributions. The BoB and BAT packages automatically apply a process called zero-padding, which ensures that all descriptors have a consistent length, by adding zeros to shorter descriptors as needed. This allows for uniform descriptor length across all compounds. While SMILES and Morgan fingerprints can be inferred from a 2D graph, BoB and BAT are calculated from a 3D representation, usually obtained by DFT, or by X-ray measurements

PCA dimensionality reduction was employed to reduce the dimensionality of the structural descriptors, while retaining the essential information. This enhances the model's efficiency and speed.[58] In addition to the aforementioned descriptors, molar ratios, number of hydrogen bond donor (HBD) and acceptor (HBA) sites, molecular weights (MW) and logP (**see Computational details and Figure S6-S1 in ESI**) of each component were given as input features to all predictive models. Temperatures and water contents were also incorporated as input features for the regressors. It is worth noting that the possible hydrogen bonding interactions between water molecules and EM components are not taken into account explicitly in this work. In fact, quantifying the impact of such interactions on viscosity would require methods such as molecular dynamics, that add complexity in terms of model development and computational time. As mentioned in the work of Herschlag and Pinney, hydrogen bonding interactions are affected by a multitude of parameters.[59] For instance, information on pKa could be an interesting additional descriptor, however these data are not always easily available. The structural descriptors before PCA dimensionality reduction were described as the concatenation of the structural descriptors for molecule A and molecule B, where molecule A is defined as the component with the highest logP. Hyperparameters have been carefully fine-

tuned (**see sections S9 and S11 in ESI for details**). Descriptors have been standardized as needed using StandardScaler provided by scikit-learn.[60] The regressors' performance was monitored using Mean Absolute Error (MAE), Mean Squared Error (MSE) and R^2 metrics during the training process, where MAE was the loss function. The accuracy of the classifiers was monitored during the training process. Each model was trained on 80% of the dataset and tested on the remaining 20%. The generalizability of all models was assessed through k-fold cross-validation (k=5) and the regressors were further evaluated on a separate dataset of unseen data. The extrapolative ability of the best regressor was also evaluated on new experimental measurements, beyond the temperature and water content ranges of the training dataset. Finally, the SHapley Additive exPlanations (SHAP, https://shap.readthedocs.io/en/latest/index.html) algorithm was employed as model explainability technique, aiming to provide insights on feature importance for model output.[61] SHAP values have been preferred over feature selection. Although the latter focuses on optimizing model performance and on reducing complexity, it may remove features that, while not highly correlated with the target variable, contribute to subtle information when combined with other features. All models were developed using Keras (v2.14.0, available on https://keras.io) and TensorFlow (software available on https://www.tensorflow.org) libraries, both implemented in Python programming language.[62–64]

Results and Discussion

Data Analysis

Data analysis is a crucial prerequisite in ML and is an integral part of the ML workflow. It involves preparing the data, identifying patterns and cleaning the data from possible outliers. In this section, similarity heatmaps and boxplots were employed to investigate correlations

between various dataset features and the target property. The aim is to highlight relevant descriptors for our predictive models and to provide an overview of the EM and non-EM from a data perspective. Heatmap in **Figure 3** depicts the molecular similarity between 20 randomly chosen EM and 20 randomly chosen non-EM in 1:1 component ratio, as computed using the Dice similarity algorithm (**see section S5 in ESI for equation**) with Morgan fingerprints.[53] A detailed heatmap with similarity values is available in the **ESI (Figure S7-S1)**.

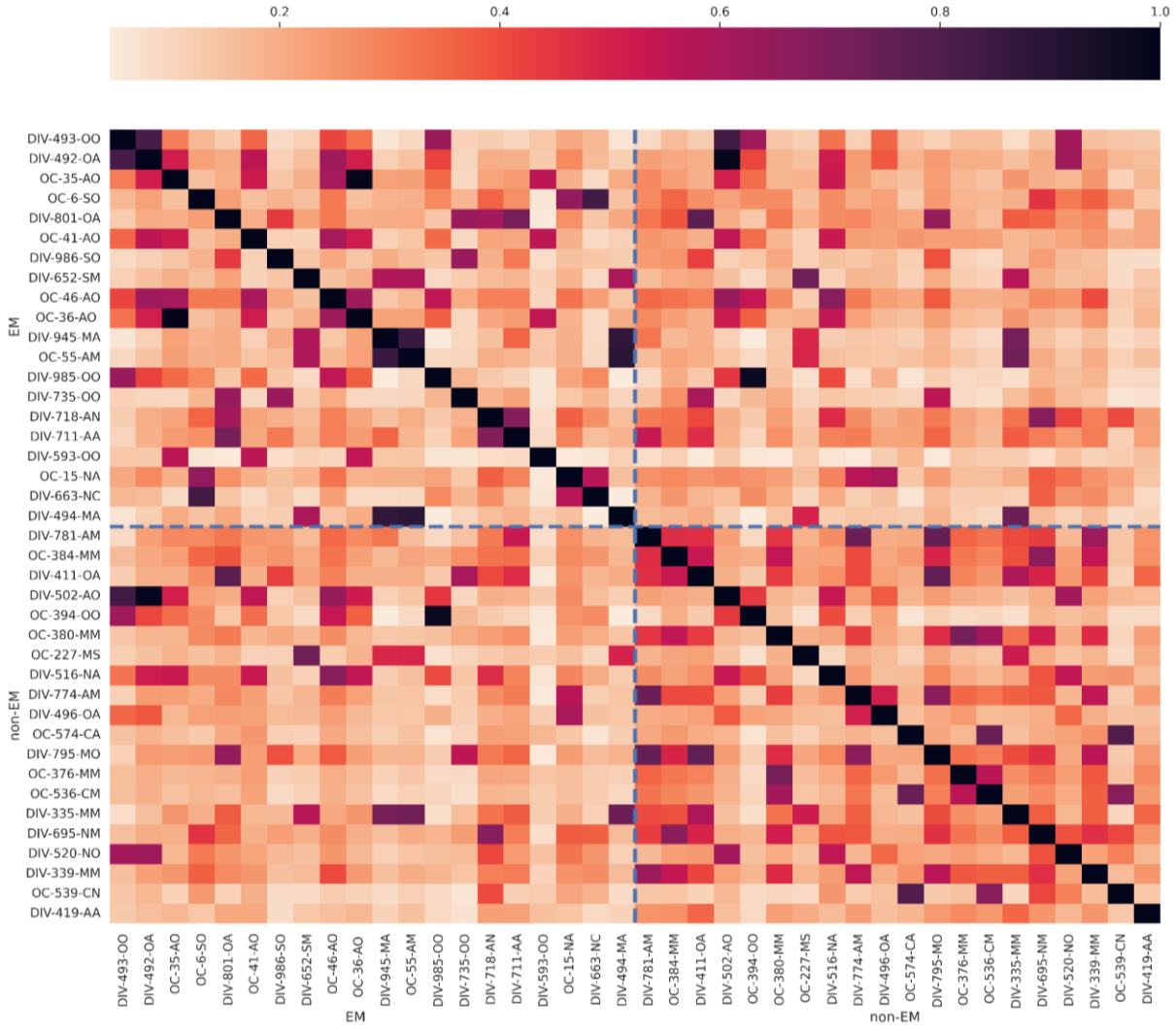


Figure 3: Heatmap of structural similarities between 20 randomly chosen EM and 20 randomly chosen non-EM in ratio 1:1 calculated using Dice Similarity Coefficient on Morgan fingerprints. The darker the color, the more structurally similar the mixtures. Blue dashed lines separate EM and non-EM. Labels correspond to identifiers assigned by

the authors to each mixture. Similarity values and details are reported in ESI (Figure S7-S1 and Table S7-S2)

Computed similarities do not reveal any significant patterns between either EM, or EM and non-EM. For instance, DIV-502-AO (Phenoxyethanol-Decanoic acid), a non-EM, and DIV-492-OA (Phenoxyethanol-Octanoic acid), an EM, are 99% similar in terms of molecular structure. The difference lies in the length of the acid's alkyl chain (**Figure 4a**).

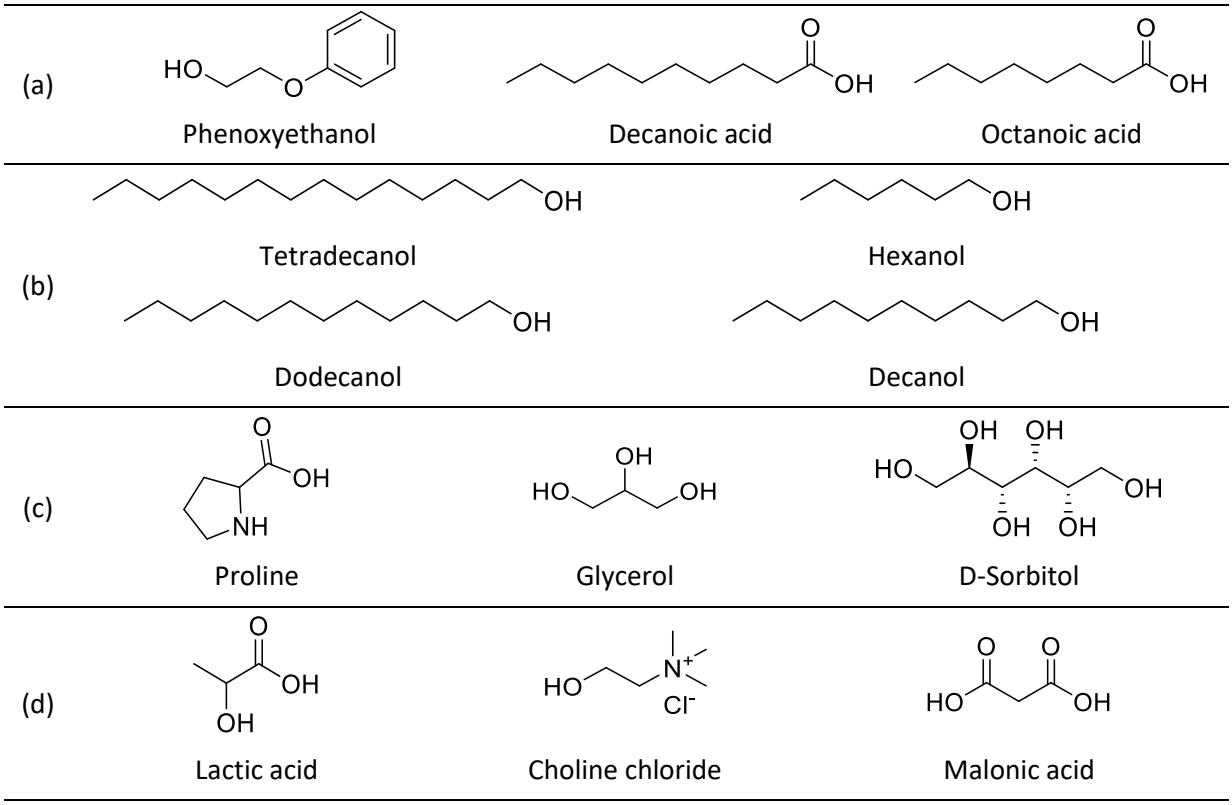


Figure 4: Chemical structures of some compounds involved in EM and non-EM. (a) DIV-502-AO and DIV-492-OA; (b) OC-394-OO and DIV-985-OO; (c) OC-228-MO and OC-229-MO; (d) OC-07-AO and DIV-13-NA

Likewise, the non-EM OC-394-OO (Tetradecanol-Hexanol) and the EM DIV-985-OO (Dodecanol-Decanol) present a computed similarity of 97%. DIV-985-OO has a total of two

more carbons than OC-394-OO (**Figure 4b**). The different distributions of carbons between the compounds of each mixture seems to influence the formation of the EM, since in the first case the mixture did not become liquid and homogeneous, unlike the second mixture.

The boxplots in **Figure 5** depict the distribution of the number of hydrogen bonding acceptor and donor sites for each component of EM and non-EM on the whole dataset. In **Figure 5a**, the median values indicate that the number of HBA sites of 50% of the first components of non-EM is within the range of 5-7, whereas 50% of EM exhibit numbers of HBA sites of the first component ranging from 3 to 7. Similarly, 50% of the numbers of HBA sites of the second component of non-EM ranges from 4 to7, while for EM they range within 5-11 (**Figure 5a**). Additionally, 50% of the first components of non-EM have numbers of HBD sites between 3 and 8, whereas the corresponding numbers for EM are between 2 and 6 (**Figure 5b**). Considering these points, the overall distribution of the number of HBA and HBD sites across the components of the mixtures is different for EM and non-EM. Moreover, non-EM generally have higher mean molecular weights (**Figure 5c**). It has been reported in the literature that this parameter has an impact on the viscosity of EM; Al-Dawsari et al. demonstrate an increase in the viscosity with an increase in the molecular weight of the salt involved in the mixture.[20] These findings suggest that the number of HBA and HBD sites, and molecular weights may serve as suitable input features for our classifiers. It will be interesting to evaluate, thanks to XAI, chemically agnostic by definition, if the best models will agree with these observations.

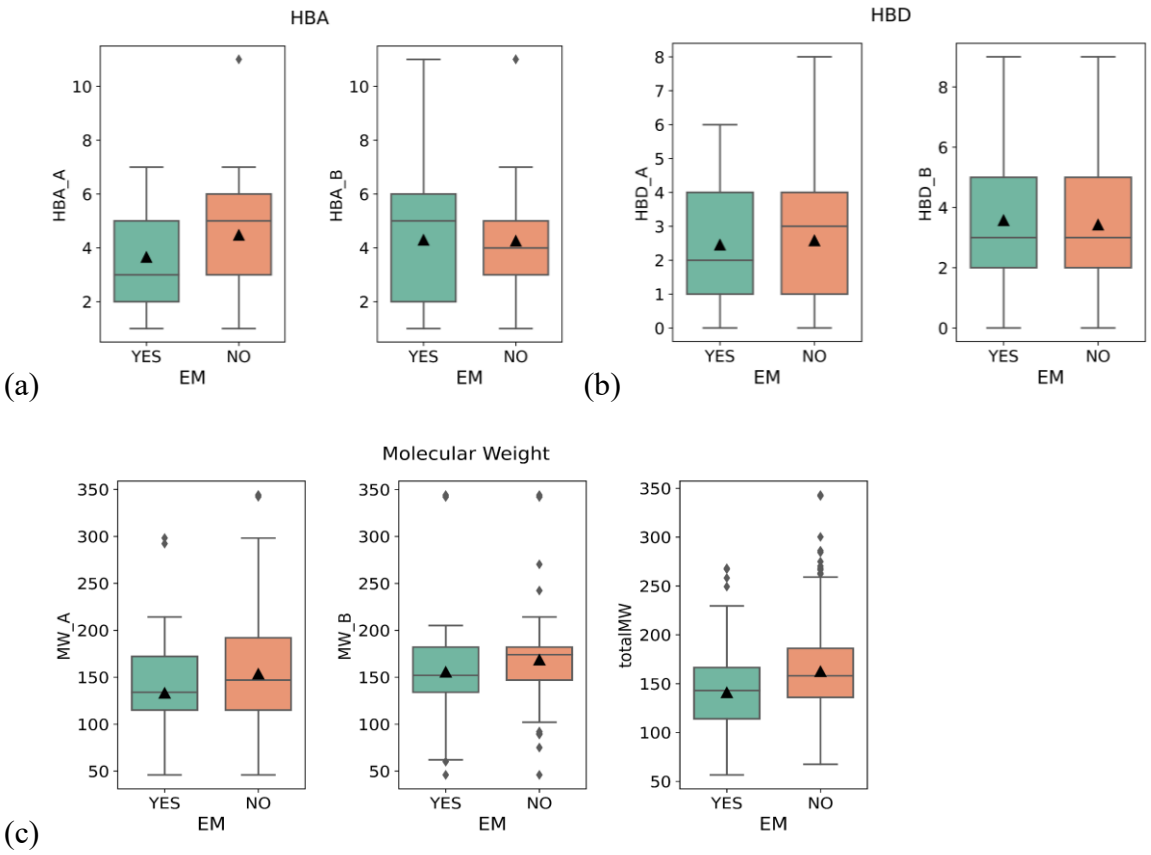


Figure 5: Boxplots representing the distribution of (a) HBA sites of the two components of the mixture in the case of a eutectic, or not, mixture; (b) HBD sites of the two components of the mixture in the case of a eutectic, or not, mixture; (c) Molecular weight of each of the two components of the mixture and the ratio weighted total MW in the case of a eutectic or not mixture ("YES" label: EM; "NO" label: non-EM)

A possible simple chemical similarity, EM vs non-EM relationship, has already been discarded (**see Figure 3)**, shedding light on the necessity to train a model able to account for multifactorial properties. The same evaluation was carried out for the viscosity of a selection of EM. A possible molecular similarity of EM in 1:1 component ratio, as depicted in the heatmap in **Figure 6**, does not exhibit correlations between molecular structure and viscosities (**Table 1**).

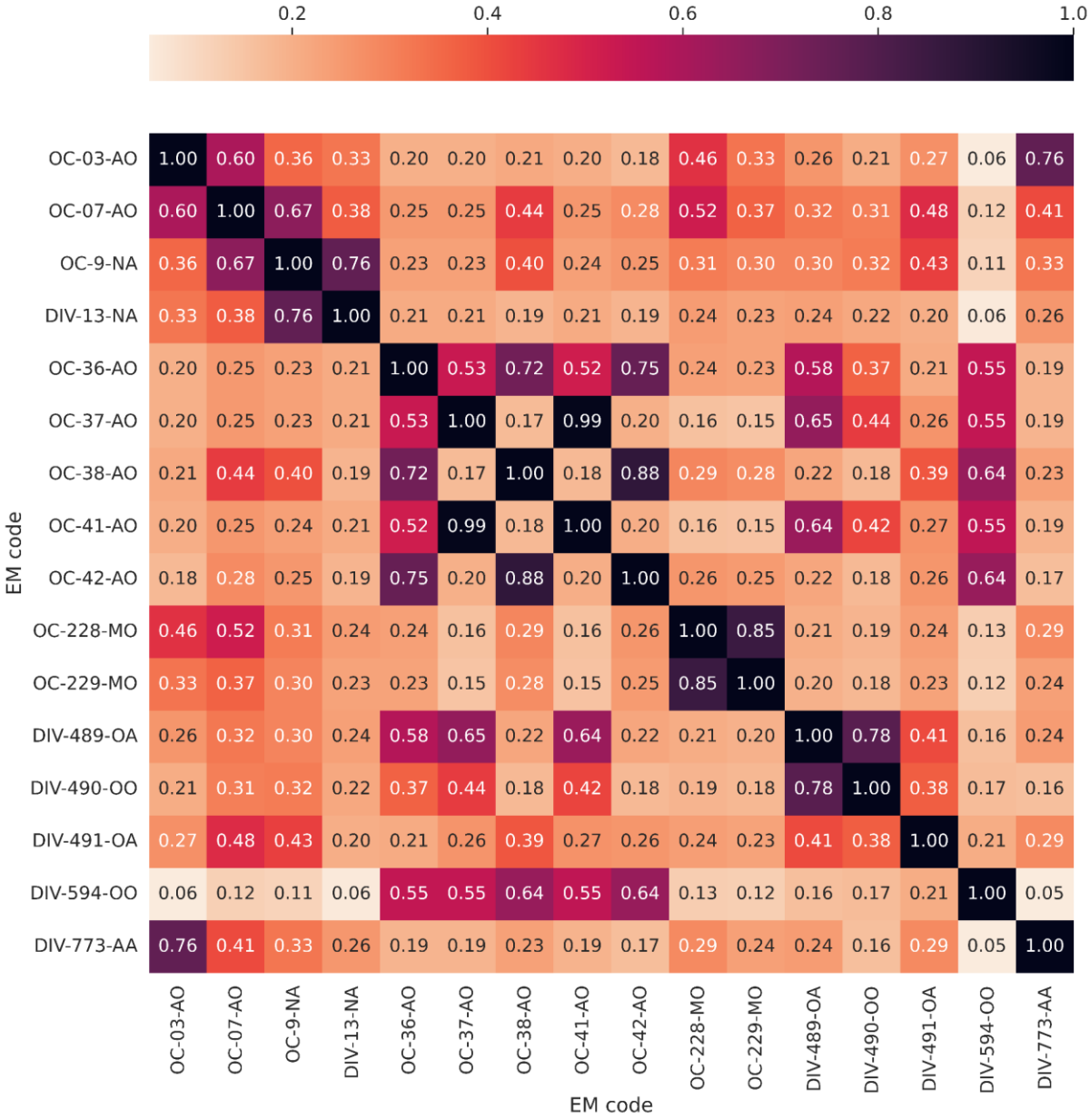


Figure 6: Molecular similarities between EM at 1:1 ratio (associated viscosities in Table 1, with more details in ESI, Table S7-S3)

Table 1: Viscosity at 25°C of the EM involved in the heatmap in Figure 6. The complete table is available in ESI (Table S7-S3)

EM number	Composition	Water content (% by mass)	Viscosity η at 25°C (mPa.s)
OC-07-AO	Lactic acid-Glycerol	30.0	19.3

DIV-13-NA	Choline chloride-Malonic acid	30.0	10.3
OC-37-AO	Thymol-Decanoic acid	0.8	79.4
OC-41-AO	Thymol-Octanoic acid	1.0	9.6
OC-228-MO	Proline-Glycerol	17.5	246.0
OC-229-MO	Proline-Sorbitol	18.0	1578.1

For instance, OC-41-AO (Thymol-Octanoic acid) and OC-37-AO (Thymol-Decanoic acid), both in water content ⩽1%, show a calculated similarity of 99% and measured viscosities of 9.6 and 79.4 mPa.s, respectively. The increase in the viscosity with the length of the alkyl chain of a DES compound has already been reported by Chen et al.[65] They revealed that the viscosities of alkylammonium bromide-based DES increased with the length of the salt's alkyl chain: Ethylammonium bromide-Glycerol (307 cP at 20°C) < Propylammonium bromide-Glycerol (398 cP at 20°C) < Butylammonium bromide-Glycerol (421 cP at 20°C).[65] The solvents OC-228-MO (Proline-Glycerol) and OC-229-MO (Proline-Sorbitol), with a water content of 17.5% and 18% respectively, are 85% similar, with respective viscosities of 246.0 and 1578.1 mPa.s. The structural difference between glycerol and sorbitol is that sorbitol has three more carbons and three more hydroxyl functions than glycerol (**Figure 4c**). This may explain the major difference in viscosity. Indeed, the hydroxyl functions are known to make hydrogen bonds, thus generating a more interconnected network of hydrogen bonds, which induces a higher viscosity. Grayson et al. carried out studies aiming to quantify the relationship between viscosity and the number of hydroxyl functional groups in organic solvents.[66] Their findings suggest an increase in the viscosity of organic compounds upon addition of a hydroxyl functional group to a carbon backbone, following a linear relationship.

In contrast, OC-07-AO (Lactic acid-Glycerol) and DIV-13-NA (Choline chloride-Malonic acid), both in water contents of 30%, demonstrate closer viscosity values of 19.3 and 10.3

mPa.s respectively, despite having a lower similarity of 38% as can be seen from the component structures of each of these two EM (**Figure 4d**).

Overall, molecular similarity alone does not seem to be sufficient to discern between EM and non-EM, nor to be correlated to the viscosity of EM. This can be partially explained by the inability of molecular similarity to quantify the number of times a fragment is present in a molecule, as it only considers the number of features that are shared between two molecules.

In addition to the observations made above, it is important to note that a minor difference between two molecules can have a major impact on properties such as melting temperature. For instance, the anion associated with tetrabutylammonium has a strong influence: while tetrabutylammonium chloride has a melting temperature of 72.85°C, tetrabutylammonium bromide melts at 121.85°C.[67,68] This is further evidence that molecular similarity calculations are a limited technique for addressing our problem.

Experimental Findings

From these selected descriptors and the starting dataset, the models will learn trends and behaviors. Correlations between the viscosity of eutectic mixtures and certain parameters are easily observable and known in the literature; some are studied in this article.

The general tendency for EM in our dataset is the decrease of viscosity when temperature rises. Examples of eutectic mixtures with different compositions but the same water content and ratio are shown in **Figure 7a**.

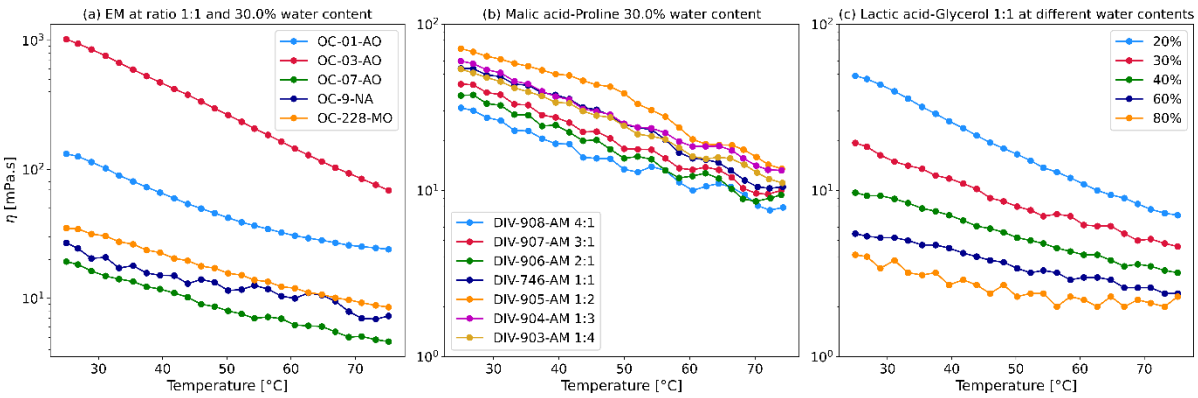


Figure 7: (a) Viscosity η (mPa.s) of several EM. The water content is 30% by mass for each and the ratio is 1:1. Values are reported in Table S8-S1 (ESI); (b) Viscosity η (mPa.s) of Malic acid-Proline EM at different ratios. The water content is 30% by mass for each. Values are reported in Table S8-S2 (ESI); (c) Viscosity η (mPa.s) of OC-07-AO (Lactic acid-Glycerol 1:1) EM at different water contents. Values are reported in Table S8-S3 (ESI)

Eutectic mixture OC-03-AO composed of Citric acid-Glycerol (red line in **Figure 7a**) has a high viscosity at 25°C (1020.0 mPa.s) but it drops drastically as temperature rises to 75°C (68.8 mPa.s). Some EM are less viscous at working temperature, like OC-07-AO (Lactic acid-Glycerol) which has a viscosity of 19.3 mPa.s at 25°C. The decrease in viscosity with increasing temperature can also be observed for other types of EM, based on quaternary ammonium salts: at 25°C, the viscosity of OC-228-MO (ChCl-Lactic acid) was 26.9 mPa.s, while it was 7.3 mPa.s at 75°C. This observation is not unknown in the literature; previous studies have demonstrated a direct correlation between temperature and viscosity. Adeyemi et al. report decreased viscosities with rising temperature, attributing this observation to the enhanced mobility of the molecules induced by increasing temperature, resulting in weakened binding energies.[69] Similar observations have been reported in numerous studies such as those by Zuo et al., Yadav et al. and Gajardo-Parra et al.[39,70,71]

Additionally, the molar ratio of DES components has been shown to impact the viscosities. For instance, Adeyemi et al. report increasing viscosities of ChCl-Monoethanolamine DES with increasing ChCl molar ratio, however little to no influence of the molar ratio was observed for ChCl-Diethanolamine DES.[69] The authors attribute these findings to internal changes of the hydrogen bonding network between the components of the DES. Mjalli and Naser, and Shah et al. also report increase in viscosity with increasing salt mole fraction in ChCl based solvents.[47,72] In this work, the case of the Malic acid-Proline mixture was studied at different molar ratios (**Figure 7b**).

Although increasing the ratio of malic acid to proline seems to have a downward impact on viscosity, there is no observable trend when the proline ratio is increased.

Furthermore, the viscosities of EM are remarkably affected by the water content (**Figure 7c**).

Although the definition of a eutectic mixture becomes questionable when the water content exceeds 40–50%, this does not prevent us from observing that the more water present in the mixture, the lower the viscosity.[73] At 25°C, OC-07-AO, composed of lactic acid and glycerol in a 1:1 ratio, has a viscosity of 48.9 mPa.s with a water content of 20%, while at 80% water its viscosity is 4.1 mPa.s. This significant decrease in viscosity has been widely reported when the water content of a DES is increased; for instance, Sarmad et al. report decrease in viscosity from 716.6 to 20.4 mPa.s at 25°C upon addition of 0.11 mol of water to Benzyltrimethylammonium chloride-Glycerol (1:2).[74]

Finally, the logP of the mixture components were compared with respect to their viscosities. The logP is the partition coefficient between octanol and water; it indicates the hydrophobicity of a compound. The higher the logP, the more hydrophobic the compound. LogP values of part of the components involved in our classification and regression datasets are available in **ESI** (**Figure S6-S1**). A study of three hydrophobic EM based on decanoic acid revealed that the

lower the logP of the associated compound, the lower the viscosity (**Figure 8**). Conversely, for the two hydrophilic proline-based EM studied, an increase in viscosity implied a lower logP for the associated compound. A trend that can be observed, however, is that blends with compounds having a very low logP have a very high viscosity, even with a higher percentage of water. For example, the mixture OC-228-MO (Proline-Glycerol 17.5% water) has a viscosity of 246.0 mPa.s, while OC-37-AO (Thymol-Decanoic acid 0.8% water) has a viscosity of 79.4 mPa.s.

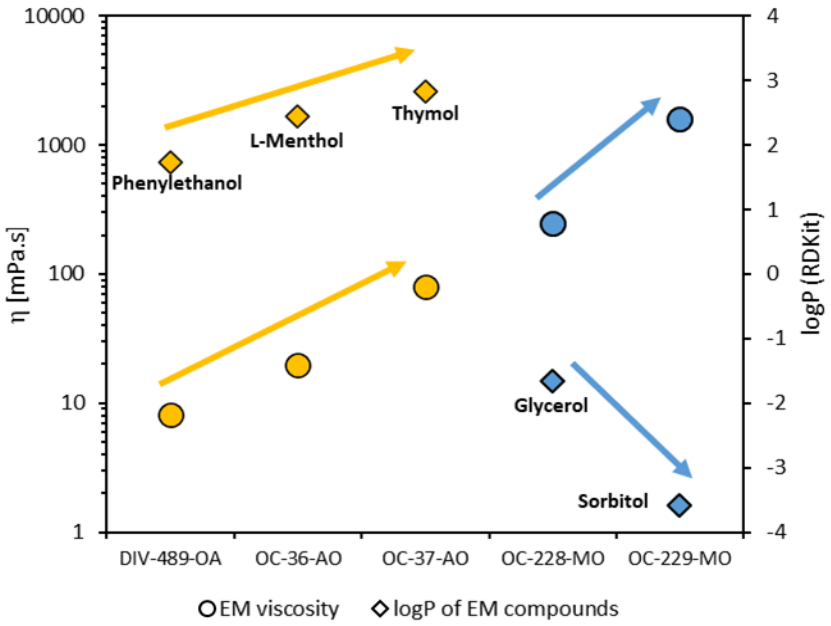


Figure 8: Comparison of the viscosities of eutectic mixtures as a function of the logP of their components. Yellow, hydrophobic EM: compounds associated with decanoic acid (logP = 3.21); Blue, Hydrophilic EM: compounds associated with proline (logP = -0.18). Values are reported in Figure S1-S6 and Table S7-S3 (ESI)

Data analysis and observation of experimental results from viscosity measurements were used to understand the trends and correlations that the models will learn from. It is important to note that limited experimental viscosities are available for some mixtures. Surfaces in **Figure 9** show the range of temperatures and water contents explored experimentally in the cases of

Malic acid-Adonitol and Lactic acid-Glycerol. For Malic acid-Adonitol (**Figure 9a**) viscosities were measured only at water contents of 24.1% and 31.3%, while for Lactic acid-Glycerol (**Figure 9b**), water content ranges from 10.0% to 80.0%. As a result, the model must be able to learn enough from EM for which a wide range of data is available (a wide range of water contents or temperatures, for example) to be able to determine a trend; it will, then, be able to extrapolate and predict viscosities for EM with a limited range of data, such as Malic Acid-Adonitol. The new models developed on this basis for the prediction of the formation of eutectic mixtures and their viscosity are presented below.

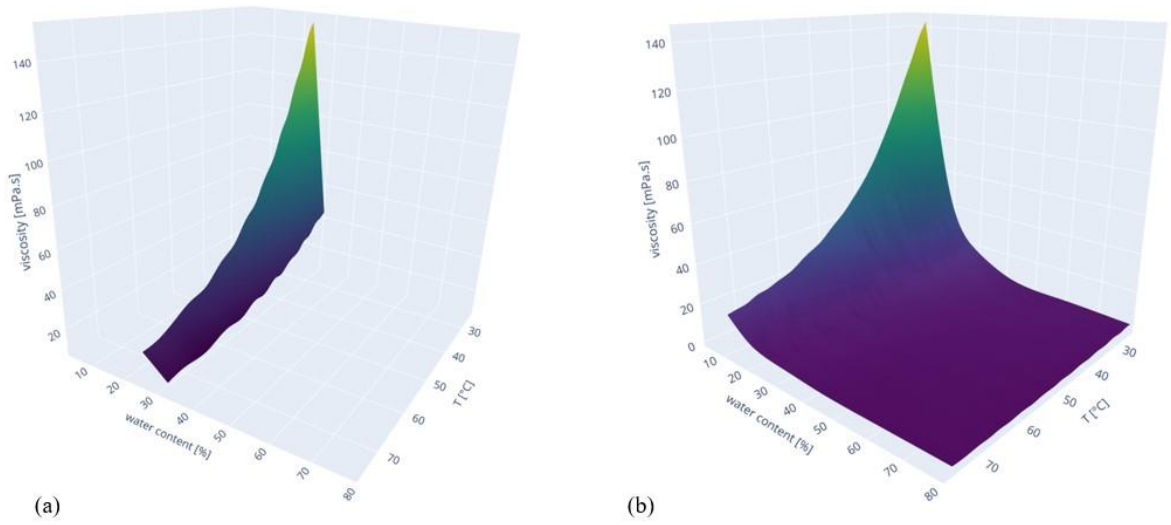


Figure 9: Three-dimensional surface plots illustrating experimental viscosities as a function of temperature and water content for EM: (a) Malic acid-Adonitol (DIV-861-AO); (b) Lactic acid-Glycerol (OC-07-AO). Surface plots were obtained by cubic interpolation

Classification between EM and non-EM

Several classification models have been compared in order to define which one is the best at predicting which mixture of compounds can give rise to a eutectic or a non-eutectic mixture. The accuracies of the three algorithms employing each of the three sets of descriptors with and without PCA dimensionality reduction, on the same test set, are reported in **Figure 10a**. In the case of PCA, the final dataset after dimensionality reduction comprises 80 columns for the BAT structural descriptor instead of the initial 6718 columns (**see section S12 in ESI for Cumulative explained variance ratio by principal components**).

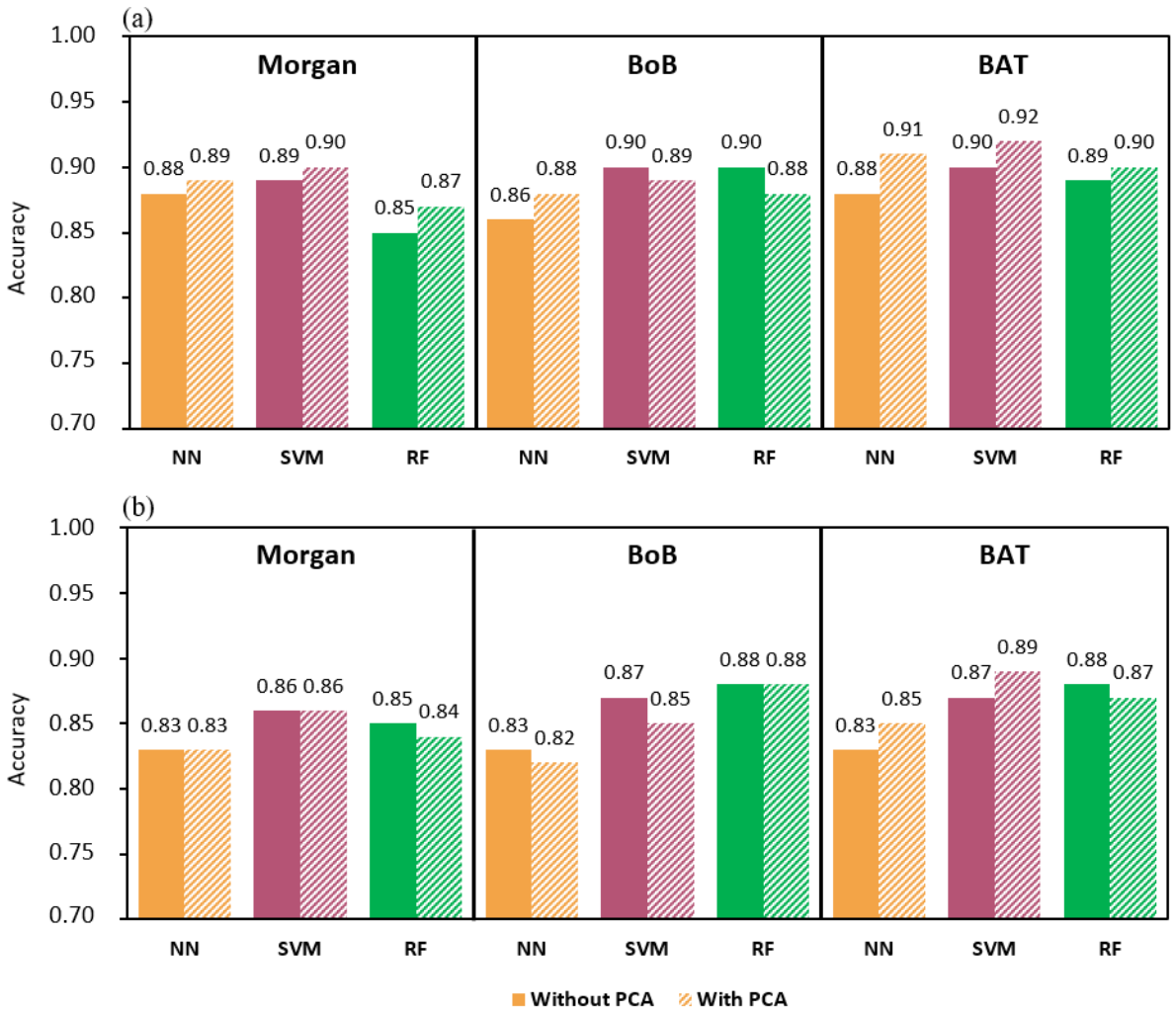


Figure 10: (a) Accuracies achieved on the same test set for the different classification models with and without PCA. The model with the highest accuracy is considered the top-performing classifier for a given dataset; (b) Cross-validation test set mean accuracies for the different classification models with and without PCA. Cross-validation is used to determine the generalizability and robustness of the models. (NN: neural network, SVM: support vector machine, RF: random forest)

All models show accuracy of at least 85%. As the differences are small, only the best and least accurate models are studied in greater detail. SVM-BAT-PCA model demonstrates the highest accuracy, achieving 92%, whereas RF-Morgan exhibits the lowest, although satisfactory, accuracy of 85% on the same test set. The mean cross-validation accuracies for all classification models have been calculated to assess the generalizability of the models. The corresponding values are reported in **Figure 10b**.

Here again SVM-BAT-PCA yields the highest mean cross-validation accuracy (89%). Neural networks exhibit the least satisfactory cross-validation mean accuracies with NN-BoB-PCA displaying the lowest accuracy of 82%. Models trained on BAT descriptors result in higher overall cross-validation performance compared to BoB and Morgan descriptors. Such behavior may be due to a better representation of molecular structures when BAT is employed, as it takes into account more structural parameters than BoB, while bit collision (**section S10 in ESI**) in Morgan fingerprints may reduce model performance due to inconsistencies in the encoded information. Model evaluation on the same test set suggests that PCA dimensionality reduction results in small overall improvement in accuracy, with SVM-BoB and RF-BoB being the only exceptions. To quantify the statistical significance of those differences we performed a paired t-test in terms of cross-validation accuracies. To be consistent, we compared the three algorithms (NN, SVM, RF) with the same descriptor upon PCA, this means we compared SVM-BAT-PCA (mean CV accuracy=89%), RF-BAT-PCA (mean CV accuracy=87%) and

NN-BAT-PCA (mean CV accuracy=85%). The differences in accuracy between the two top performing models are statistically insignificant (p-value>0.05). In contrast the differences in accuracy between the most and the least accurate models are statistically significant (p-value<0.05).

Highest overall feature contributions to SVM-BAT-PCA EM probability predictions are illustrated in **Figure 11**. SHAP, the XAI method used in the present article, suggests that higher values of total molecular weight, and of descriptors 1 and 3 tend to decrease the probability of EM formation, whereas higher values for descriptors 6 and 7 tend to increase this probability. Descriptors 1, 3, 6 and 7 result from PCA dimensionality reduction and correspond to the structure contributions in **Figure 12**.

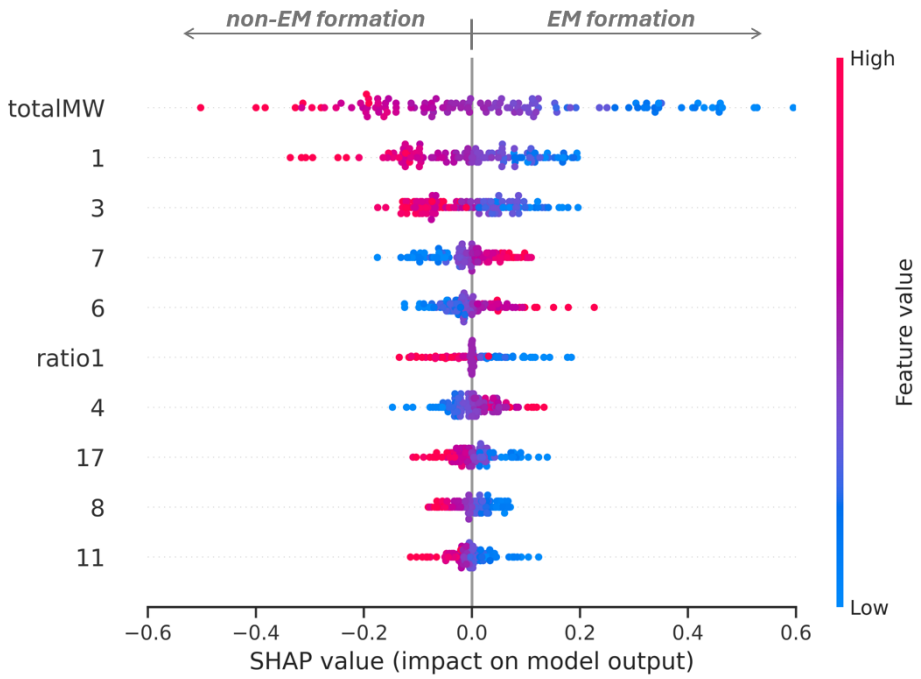


Figure 11: Contributions of the most influential descriptors to EM formation for the classification model SVM-BAT-PCA. 1, 3, 7, 6, 4, 17, 8 and 11 are resulting structural descriptors from PCA dimensionality reduction. SHAP value > 0, in favor of EM

formation; SHAP value < 0, against EM formation. In red, high value of the descriptor; In blue, low value of the descriptor

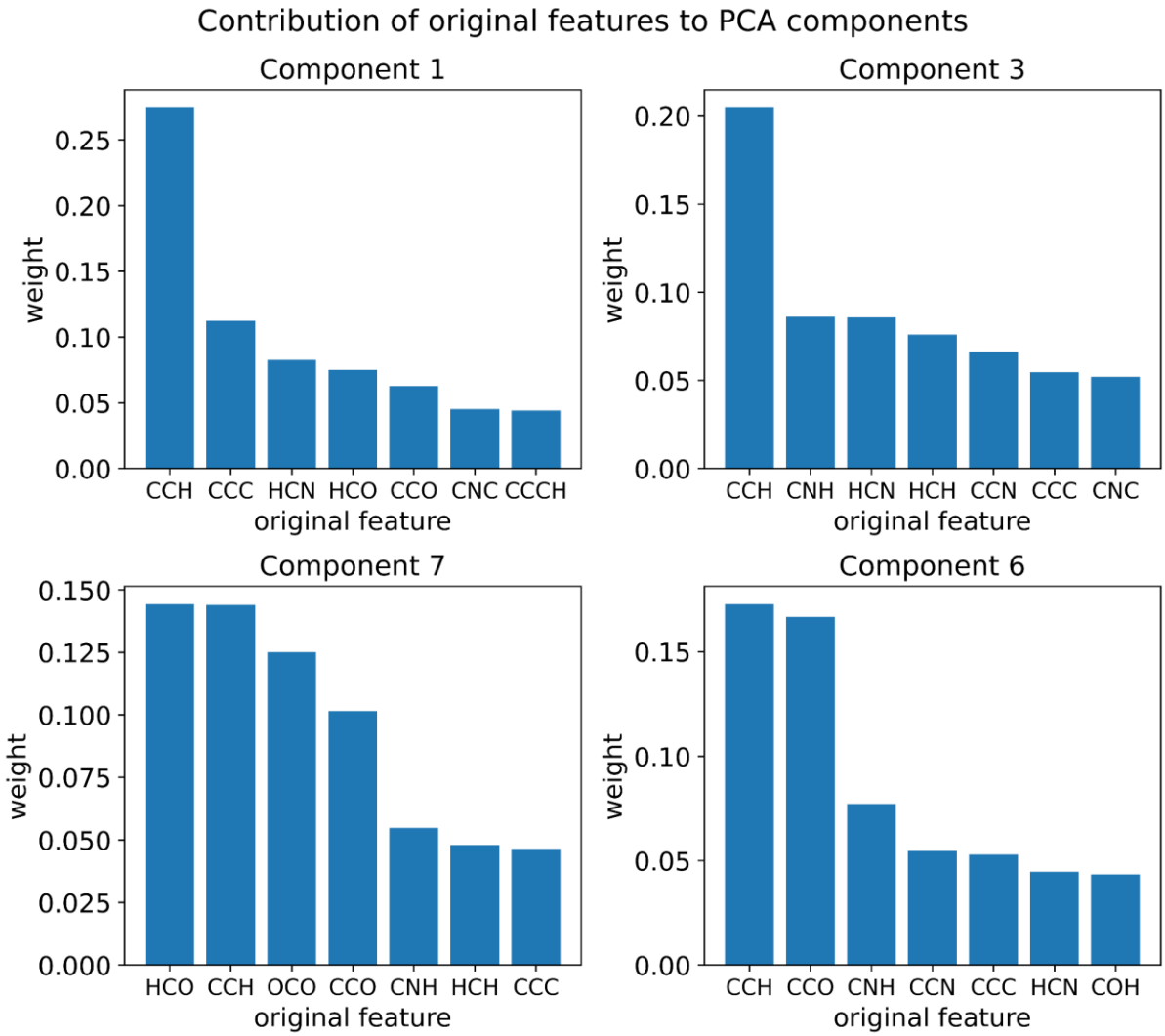


Figure 12: Contribution of original features to BAT-PCA components 1, 3, 7 and 6 (order given by Shapley values reported in Figure 11). Weights were calculated as the sum of the squares of the initial feature contributions associated to similar molecular substructures

The higher molecular weight may be linked to a longer alkyl chain and, in systems where components contain alkyl chains, it has been observed upstream that a longer alkyl chain can hinder the formation of a eutectic mixture. This is in line with the other two of the three most

influential descriptors (1 and 3), where the majority contribution comes from CCH and CCC groups, that usually describe alkyl chains. On the other hand, the favorable impact of descriptors 7 and 6 is due to the significant contribution of the H-bond donor HCO and H-bond acceptor CCO groups. The latter describes hydrophilic interactions, while the presence of alkyl chains can only give rise to hydrophobic interactions.

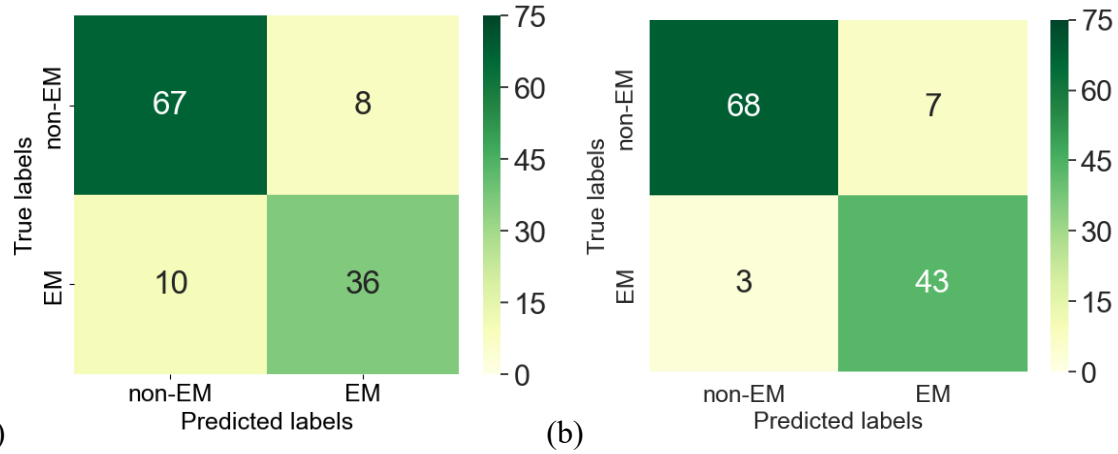


Figure 13: Confusion matrices for the same test set. (a) RF-Morgan; (b) SVM-BAT-PCA. The color represents the number of mixtures in each true/false positive/negative square (also indicated in numbers)

Next, we focus on the performance of the RF-Morgan model. It misclassifies 18 out of 121 samples within the test set (**Figure 13a**). Of these misclassifications, SVM-BAT-PCA correctly classifies 11. It is worth noting that both models misclassify the same 7 samples, among others. They are listed in **Table 2** and thanks to the feature contributions given by the XAI-SHAP algorithm, it is possible to evaluate the origin of such misclassification (**Figure 14**).

Table 2: EM and non-EM misclassified by both RF-Morgan and SVM-BAT-PCA classification models

Mixture number	Mol_A	Mol_B	Ratio	Real	Predicted
OC-177-AO	Hexadecanol	Dodecanoic acid	1:1	non-EM	EM
OC-196-OO	Decanol	Panthenol	1:1	EM	non-EM
DIV-360-MM	Glutamine	Arginine	4:1	non-EM	EM
DIV-671-MO	Proline	Sorbitol	1:2	non-EM	EM
DIV-792-AA	Pyroglutamic acid	Malic acid	9:1	EM	non-EM
DIV-874-AO	Malic acid	Arabitol	1:9	non-EM	EM
DIV-955-MA	Proline	Citric acid	1:2	non-EM	EM

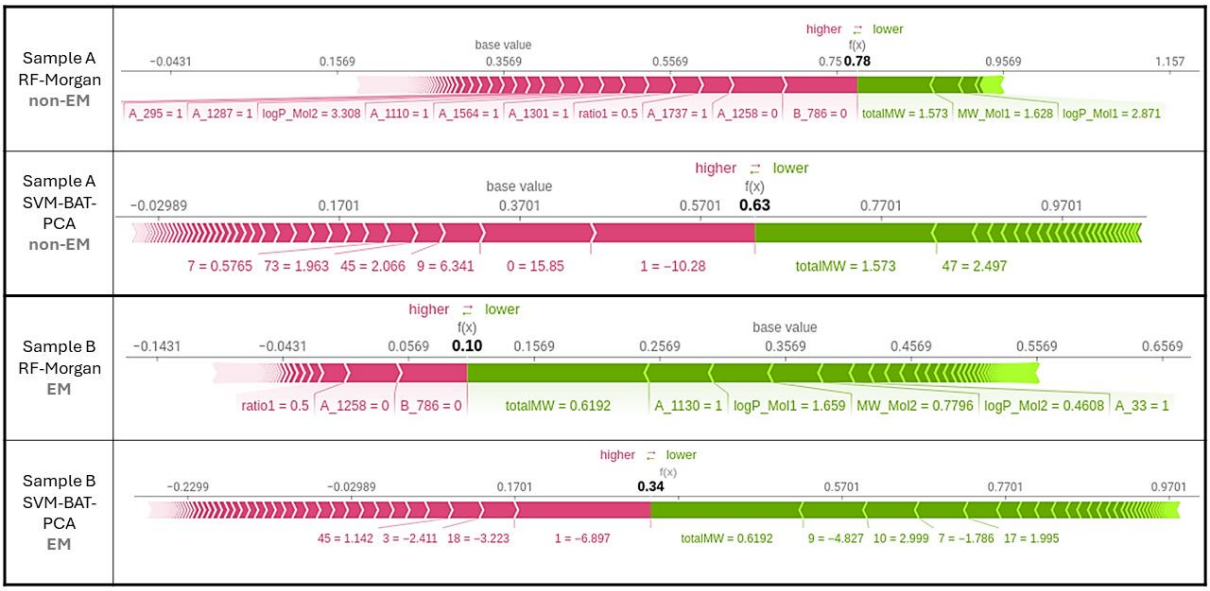


Figure 14: SHAP force plots on feature contributions to the classification of two mixtures: sample A (OC-177-AO) and sample B (OC-196-OO). The standardized values of totalMW are reported in the figure. Numbers in bold correspond to the probability of formation of an EM assigned by the classifiers to each sample; probabilities < 0.5 indicate

prediction of formation of a non-EM, while probabilities > 0.5 indicate prediction of formation of an EM. One should refer to the SHAP force plot documentation for an in-depth understanding of these plots. As an example, B_786=0 means that the absence of the corresponding group in sample A increases the EM probability.

In the case of RF-Morgan, the labeling of Sample A (Hexadecanol-Dodecanoic acid 1:1), a non-EM, and the Sample B (Decanol-Panthenol 1:1), an EM, are positively influenced by the 787^{th} bit (descriptor B_786) and the 1259^{th} bit (descriptor A_1258) for samples A and B (**Figure 14, Figure 18**). The model indicates that the absence of these chemical structures featuring an amine group as an HBA site, would favor the formation of an EM. This result is not in line with previous observations and could be at the origin of the lower accuracy of the RF-Morgan model. Both mixtures exhibit a totalMW descriptor in the upper quartile of the boxplots reported in **Figure 5**. Together with the SHAP summary plot reported in **Figure 11**, this explains why this descriptor lowers the probability of EM formation for both mixtures. Unfortunately, no other descriptors counterbalance this effect for Sample B, thus yielding a wrong probability.

Even when both classifiers fail to label Sample A correctly, the SVM-BAT-PCA model exhibits a lesser degree of error by assigning a lower probability (0.63) to the wrong label compared to RF-Morgan (0.78) for Sample A. Likewise, SVM-BAT-PCA demonstrates lower confidence (0.34) in the wrong label in the case of Sample B compared to RF-Morgan (0.10).

Having determined SVM-BAT-PCA as the top-performing classifier and highlighted the molecular weight, HBA/HBD sites and CCH, CCC groups as the most impactful descriptors, we now focus on regression algorithms developed for viscosity prediction.

Regression

The second challenge is to be able to anticipate whether EM physico-chemical properties are suitable for the desired application, in particular viscosity. To this end, new models employing neural networks were developed and two structural descriptors were compared: Morgan fingerprints, Morgan fingerprints upon PCA dimensionality reduction, BAT upon PCA dimensionality reduction; in case of PCA, the dataset comprises 50 columns for the Morgan fingerprints after dimensionality reduction instead of the initial 4096 columns (**see section S12 in ESI for Cumulative explained variance ratio by principal components**). Utilization of the complete set of Morgan fingerprints (Morgan-Full), yields a MAE of 2.2 mPa.s and a R^2 of 0.99 in predicting the viscosity values of the test set, whereas employing PCA for dimensionality reduction on the Morgan fingerprint descriptors (Morgan-PCA) yields a MAE of 3.5 mPa.s and R^2 of 0.99. Additionally, employing PCA on the BAT descriptor (BAT-PCA) slightly decreases the test set accuracy, achieving a MAE of 4.1 and an R^2 of 0.99 on the test set. Cross-validation was performed to assess the performance and generalizability of the predictive models (**see section S14 in ESI**).

We further assessed the generalizability of the regressors by evaluating their predictive performance on a separate dataset consisting of previously unseen data, hereafter named Challenging set.

Morgan-Full yields predictions with MAE of 6.5 mPa.s and R^2 of 0.99, Morgan-PCA achieves MAE of 5.3 mPa.s and R^2 of 0.99 and BAT-PCA achieves MAE of 6.8 mPa.s and R^2 of 0.99, on this new dataset. PCA dimensionality reduction of Morgan fingerprints improved the predictive accuracy of the regression model by 1.2 mPa.s on the challenging set compared to the full descriptor. No improvement was achieved when PCA was employed on the BAT descriptor, although the results remained satisfactory. **Figure 15** depicts the models' predictions plotted against the experimental viscosities.

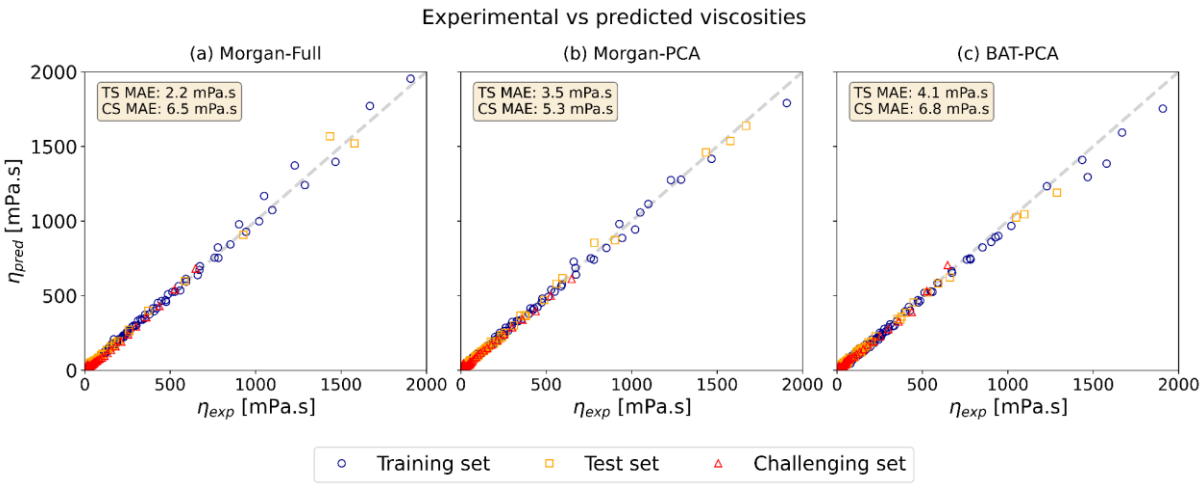


Figure 15: Predicted versus experimental viscosities. (a) Morgan-Full; (b) Morgan-PCA; (c) BAT-PCA. TS, Test set; CS, Challenging set. R^2=0.99 for the 3 different models, for the test set (randomly selected) and the challenging set (same challenging set). The cross validation reported in the ESI (section S14) has been performed with the same random seed.

By looking at Figure 15 we notice an abundance of data at low viscosities, while data are relatively scarce at high viscosities. To ensure that this data imbalance is not impacting neither leading to an overestimation of the model's performance, we calculated the test set MAE of each of the three regressors for viscosities greater than 1000 mPa.s: 95.4 mPa.s (Morgan-Full), 32.3 mPa.s (Morgan-PCA) and 95.7 mPa (BAT-PCA). The corresponding test set mean relative errors are 6.5%, 2.1% and 6.2% respectively. These values are satisfactory, and they suggest that the regressors learn trends from data-rich regions and output good predictions at regions where data are scarce. This evaluation is necessary to monitor the reliability of the models when data are disproportionately distributed.

As the differences in model performance were not significant, the best-performing model on the test set, the Morgan-Full model, was chosen for extrapolation testing. The extrapolation capabilities of the model were evaluated for six mixtures. Its performance was examined on

temperatures and water contents beyond the range over which it was trained for these mixtures. The predictions are depicted in **Figure 16**. Values are reported in **Table S11-S4** (**ESI**).

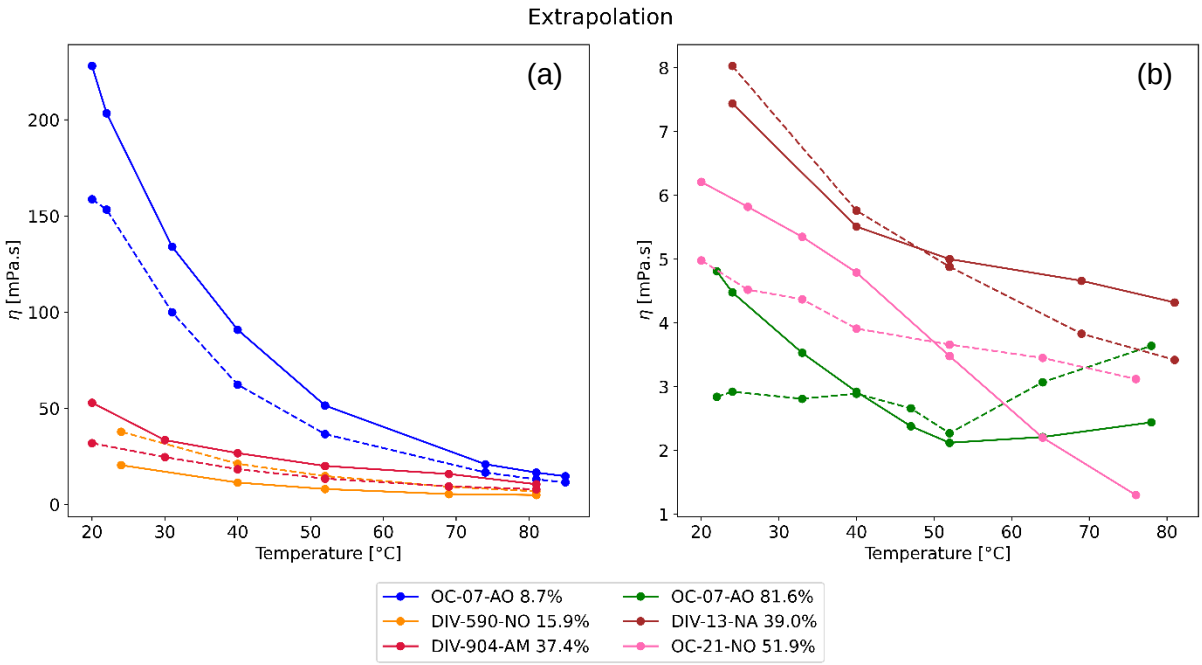
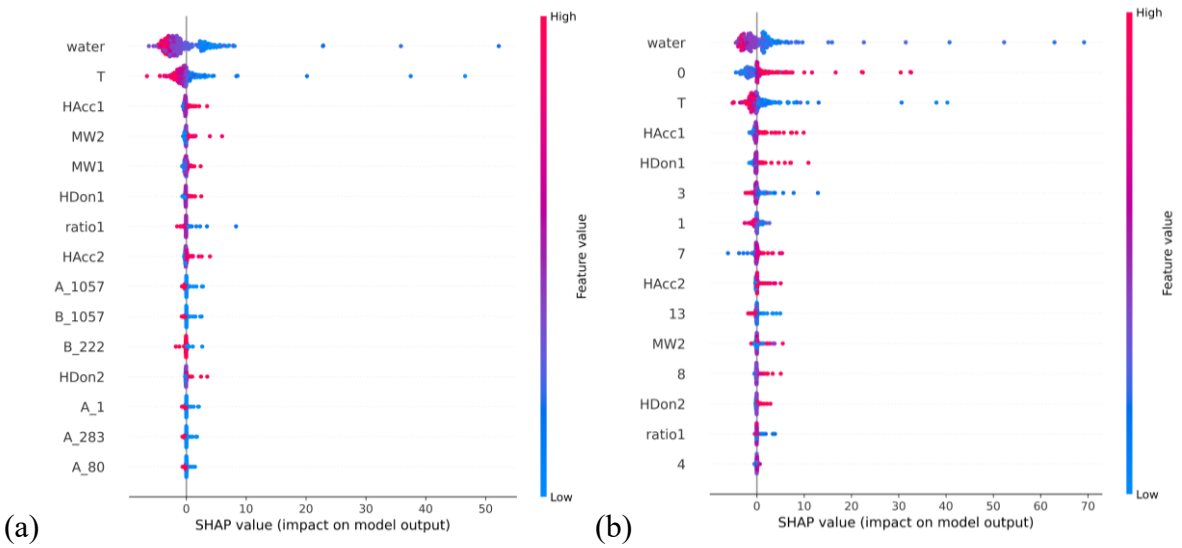


Figure 16: Evaluation of the extrapolation performance of the Morgan-Full regression model on viscosity prediction. (a) Viscosities greater than 10 mPa.s; (b) Viscosities below 10 mPa.s. Solid curves: predicted values; Dashed curves: experimental values. Values are reported in Table S11-S4 (ESI)

In the three cases with the lowest water content, the error increases with experimental viscosity. For the mixtures DIV-590-NO (Glycerol-Choline chloride 1:2, 15.9% water) at 24°C and DIV-904-AM (Proline-DL-Malic acid 3:1, 37.4% water) at 20°C, the errors reach 17.4 and 21.0 mPa.s respectively. For OC-07-AO (Lactic acid-Glycerol 1:1, 8.7% water, blue curve), the largest absolute error is 69.3 mPa.s. In the case of lower viscosity predictions, below 10 mPa.s, model predictions are much closer to reality, with smaller absolute errors, ranging from 0 to 2.0 mPa.s. Overall, the regressor can be considered more reliable in cases of temperature and water content extrapolations that are expected to reduce the viscosity.

Figures 17a and 17b depict the 15 features that contribute the most to the predictions given by Morgan-Full and Morgan-PCA models, respectively.



(a) (b)

Figure 17: Contributions of the most influential descriptors to EM viscosity for the regression models a, Morgan-Full; b, Morgan PCA. 0, 3, 1, 7, 13, 8, and 4 are resulting structural descriptors from PCA dimensionality reduction. SHAP value > 0, in favor of increasing viscosity; SHAP value < 0, in favor of decreasing viscosity. In red, high value of the descriptor; In blue, low value of the descriptor

As expected, SHAP reveals strong dependence of the predictions on water content and temperature conditions; lower water contents and temperatures have a positive impact on model output, which equals an increase in viscosity. Higher molecular weights and numbers of hydrogen bonding sites also exhibit a positive impact on the viscosity for the two models. This is in line with studies attributing high viscosities to a strong HB network decreasing the mobility of free species in the mixture. There is also an influence of the "ratio1". This descriptor corresponds to the fraction of the component with the highest logP. The higher a compound's logP, the more hydrophobic it is, and vice versa. Thus, to observe in **Figures 17a and 17b** that the lower the ratio1, the higher the viscosity means that the lower the fraction of the more

hydrophobic compound, the higher the viscosity. The main structural descriptors involved in each model's output are shown in **Figure 18**.

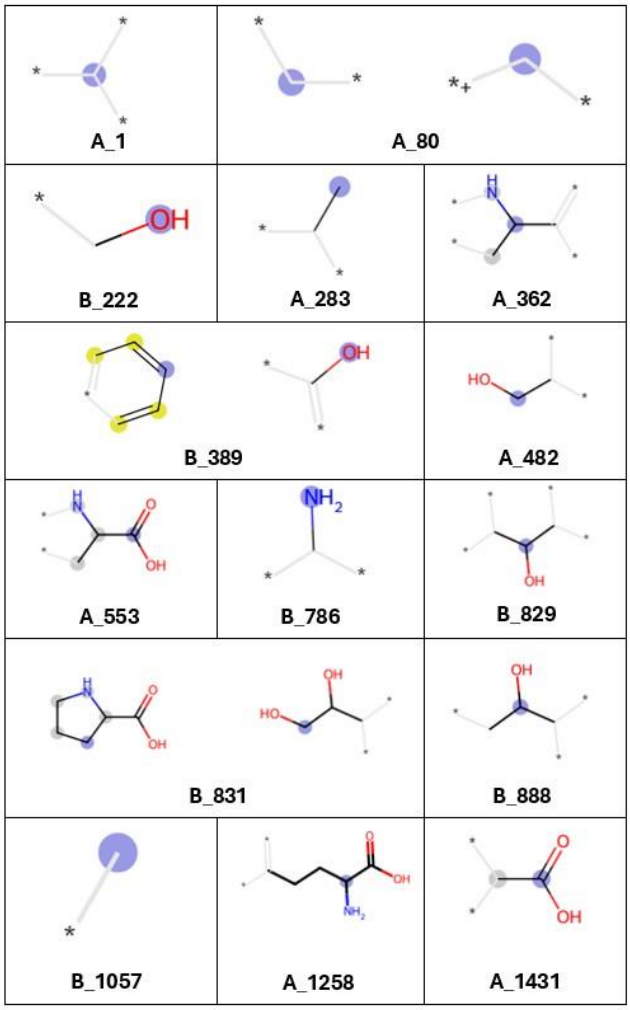


Figure 18: Morgan fingerprints substructures contributing to predictions as shown by SHAP for regression models. In the case of Morgan-PCA regressor, the substructures contributing the most to most influential principal components in order of importance are: A_362, A_1431, A_553 for component '0'; B_482, B_222, B_389 for component '1'; B_888, B_831, B_829 for component '3'. A_1057 is the same as B_1057. Blue: the central atom in the environment; yellow: aromatic atoms; gray: aliphatic ring atoms

Bit collision (**see ESI**) in some Morgan fingerprints, namely A_80, B_389, B_831, overshadows the true impact of specific features; however, this limitation does not seem to impede the overall accuracy of the regressors. It can be seen that the weaker the structural descriptors involved in **Figure 17a**, the higher the viscosity. This correlates with the fact that the more hydrophobic an EM is, the less viscous it is. In the Morgan-PCA model, higher values of structural descriptors 0 and 7 have an upward impact on viscosity. The major contribution to descriptor 0 comes from H-bonding structures (**Figure 17b**). Conversely, lower values of descriptors 1 and 3 tend to increase the viscosity of eutectic mixtures. Although the molecular structures contributing most to these two descriptors are composed of hydroxyl groups, these are linked to multi-carbon chains.

When applied on a subset of the test set containing only hydrophobic solvents, the SHAP explainer revealed contributions of hydrophobic substructures, such as 4- and 7-carbon alkyl chains, among HBA/HBD and logP descriptors (**see section S15 in ESI**). This analysis suggests that the model is able not only to capture trends for both hydrophobic and hydrophilic mixtures, but also to use the proper descriptors to make accurate predictions.

Among the regressors developed, Morgan-Full and Morgan-PCA exhibited equal efficiency, while temperature and water content were identified among the most influential descriptors across both cases; Morgan-Full demonstrated satisfactory extrapolation capabilities, especially for low viscosities. SHAP analyses show that the models output accurate predictions for the right reasons.

Conclusions

A new experimental dataset including 219 EM and 384 non-EM, along with 1450 viscosity points, is reported in this work, and experimental findings are in agreement with the literature; as expected, the viscosity of EM depends on the temperature, molar ratio and water content of

the mixture. Data analysis based on structural similarity was not sufficient to predict whether an EM would form, nor to highlight a potential correlation between the molecular structure of EM and their viscosities.

Nine classification models, based on 2D or 3D structural descriptors and ANN, SVM and RF ML algorithms, were developed to predict the potential formation of a eutectic mixture with and without PCA for dimensionality reduction. Classification models show better accuracy with the BAT 3D structural descriptor rather than with BoB or 2D Morgan fingerprints. Dimensionality reduction improved the accuracy of 7 out of the 9 models tested. The SVM-BAT-PCA model exhibited the best generalization performance during cross-validation. Furthermore, to the best of our knowledge, the dataset presented and used in this study for classification is the largest of its kind in the existing literature used for this purpose.

The regressors were evaluated on a test set and on a set of unseen data (challenging set). Morgan-Full, where “Full” implies the use of the complete set of the Morgan structural descriptor without PCA dimensionality reduction, was the top performing regressor on the test set, while Morgan-PCA performed best on the challenging set. Overall, all regressors yielded satisfactory predictions and can be reliable tools for viscosity predictions of new solvents. In several cases, the extrapolative capabilities of the Morgan-Full regressor were assessed on water contents and temperatures beyond the range of the training set data of those mixtures and yielded accurate predictions especially at low viscosities. Apart from their good performance, all models employ descriptors that are obtained easily and without significant computational cost. Moreover, the experimental viscosities we provide compose the largest dataset used for viscosity prediction, that includes data on the water content; the effect of this parameter on the viscosity is lacking in predictive models within the current literature.

The SHAP algorithm, employed to quantify the global and individual contributions of the input features to the models' output for both classification and regression tasks, actually serves other critical purposes. It provides insights on model reliability and allows to draw meaningful correlations between features and the target property. It has shown that the molecular structure is not the only parameter influencing the formation of an EM and its viscosity, highlighting the key role of temperature and water content. It sometimes might seem to re-explain known information, but this provides evidence for the trustworthiness of the models. It can help evaluate the origin of errors, providing information about possible weaknesses of the model, which is crucial for improving model performance, as part of our machine learning workflow.

It is important to note that the descriptor that demonstrated the poorest performance in regression, i.e. BAT, achieved the highest accuracy in classification, highlighting the interest of distinct evaluation of structural descriptors for each task.

Finally, further improvements of the predictions are possible by incorporating even larger experimental datasets including a greater diversity of chemical compounds. This would enhance the accuracy and generalizability of our models. The replacement of the binary Morgan fingerprints used in the present study by count-based Morgan fingerprints, alongside optimizing the model's hyperparameters, could be a worthwhile approach to slightly improve accuracy. Future research could also explore the potential benefits of combining several structural descriptors, as well as the impact of dimensionality reduction and feature selection within this combined framework. However, in their current state, the models developed in this work can be used as powerful tools for the design of new EM; prior to any experiment, the chemical space of possible mixtures can be narrowed down to potential EM candidates, whose viscosity can be predicted by the best regressor.

ASSOCIATED CONTENT

Supporting Information (.xlsx)

- **S1: Chemical details**
- **S2: Computational details**
- **S3: Classification dataset**
- **S4: Regression dataset**
- **S5: Dice similarity**
- **S6: logP:**

Figure S6-S1. Comparison of logP values as given by ChemDraw, pubchem and RDKit for 34 components

- **S7: Data Analysis:**

Figure S7-S1. Detailed heatmap of structural similarities between EM and non-EM at ratio 1:1 calculated using Dice Similarity Coefficient on Morgan fingerprints

Table S7-S2. Composition of the EM and non-EM involved in the heatmap in Figure 3 and Figure S7-S1. Ratio 1:1

Table S7-S3. Viscosity η (mPa.s) at 25°C of the EM involved in the heatmap in Figure 6

- **S8: Experimental Findings:**

Table S8-S1. Viscosity η (mPa.s) of several EM involved in Figure 7a. Ratio 1:1, water content 30%

Table S8-S2. Viscosity η (mPa.s) of Malic acid-Proline (MalA-P) EM at different ratios involved in Figure 7b. Water content 30%

Table S8-S3. Viscosity η (mPa.s) of OC-07-AO (Lactic acid-Glycerol 1:1) EM at different water contents involved in Figure 7c

- **S9: Classifier:**

Table S9-S1. SVM-BAT-PCA optimal hyperparameters

- **S10: Morgan Fingerprints: bit collision - examples:**

Figure S10-S1. Examples of bit collision in Morgan fingerprints

- **S11: Regressors:**

Table S11-S1. 'Morgan-Full' neural network architecture

Table S11-S2. 'Morgan-PCA' neural network architecture

Table S11-S3. 'BAT-PCA' neural network architecture

Table S11-S4. Evaluation of the extrapolation performance of regression models on viscosity η prediction

- **S12: PCA: Cumulative Explained Variance Ratio:**

Figure S12-S1. Cumulative explained variance ratio by principal components – BAT descriptor for classification

Figure S12-S2. Cumulative explained variance ratio by principal components – Morgan fingerprints for regression

- **S13: Distribution of values of descriptors totalMW and 1 involved in classification:**

Figure S13-S1. Distribution of scaled totalMW values within the classification training set

Figure S13-S2. Distribution of PCA component 1 values within the classification training set

- **S14: Cross-validation for regressors:**

Figure S14-S1. MAE of k-fold cross-validation for the Morgan-Full regressor

Figure S14-S2. MAE of k-fold cross-validation for the Morgan-PCA regressor

Figure S14-S3. MAE of k-fold cross-validation for the BAT-PCA regressor

- **S15: Viscosity predictions of hydrophobic EM and corresponding SHAP plot:**

Figure S15-S1. Experimental versus predicted viscosities of a subset of the test set containing only hydrophobic EM.

Figure S15-S2. SHAP summary plot of the top 10 most impactful descriptors for the prediction of viscosity of hydrophobic EM

- **S16: Comparison between training dataset and challenging dataset:**

Figure S16-S1. Viscosities of training set and challenging set mixtures

Table S16-S2. Statistics of the training set in regression

Table S16-S3. Statistics of the challenging set in regression

Figure S16-S4. Heatmap of structural similarities between training set and challenging set mixtures

Table S16-S5. Mixtures in challenging set

AUTHOR INFORMATION

Corresponding Author

*Email: ludovic.paquin@univ-rennes.fr ; Phone: +33 2 23 23 69 83

** Email: romuald.poteau@univ-tlse3.fr ; Phone: +33 5 61 55 96 64

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

Funding Sources

Daniel Langlois Foundation and French Ministry of Higher education and Research for PhD funding

ACKNOWLEDGMENTS

We would like to thank Dr. Arnaud Saint-Jalmes and Florian Scholkopf for allowing us to carry out viscosity measurements on their rheometer, as well as for their expertise and the exchanges we had concerning our results. Thanks to Alexy Brunel for answering our rheology questions during the writing of this article. Some experimental results were obtained as part of the Océactif project. We would like to thank Agrimer for their collaboration on this project. S.C. thanks Stella Eva Tsiapali for insightful discussions about neural networks. S.C., L.M. and R.P. thank the CALMIP computer center for generous allocation of computational time (P0611 project).

ABBREVIATIONS

AARD, Average absolute relative deviation; AI, Artificial Intelligence; ANN, Artificial neural network; BAT, Bond-Angle-Torsion; BoB, Bag of Bonds; COSMO-RS, COnductor-like Screening MOdel for Real Solvents; DES, Deep eutectic solvent; DFT, Density functional theory; DNN, Deep neural network; EM, Eutectic mixture; GC, Group Contribution; HBA, Hydrogen bond acceptor; HBD, Hydrogen bond donor; LSSVM, Least square support vector machine; MAE, Mean absolute error; MD, Molecular dynamics; MLPANN, Multilayer perceptron artificial neural network; MSE, Mean squared error; MW, Molecular weight; NN, Neural Network; Non-DES, Non-deep eutectic solvents; Non-EM, non-eutectic mixture; PCA, Principal component analysis; QSPR, Quantitative structure property relationships; RF, Random forests; SHAP, Shapley additive explanations; SLE, Solid-liquid equilibria; SVM, Support vector machine; SVR, Support vector regression; VFT, Vogel-Fulcher-Tammann; XAI, Explainable artificial intelligence

REFERENCES

(1) Abbott, A. P.; Capper, G.; Davies, D. L.; Rasheed, R. K.; Tambyrajah, V. Novel Solvent Properties of Choline Chloride/Urea Mixtures. *Chem. Commun.* **2003**, No. 1, 70–71. https://doi.org/10.1039/b210714g

(2) Azizi, N.; Batebi, E. Highly Efficient Deep Eutectic Solvent Catalyzed Ring Opening of Epoxides. *Catal. Sci. Technol.* **2012**, *2* (12), 2445–2448. https://doi.org/10.1039/C2CY20456H

(3) Ni, Y.; Bi, Z.; Su, H.; Yan, L. Deep Eutectic Solvent (DES) as Both Solvent and Catalyst for Oxidation of Furfural to Maleic Acid and Fumaric Acid. *Green Chem.* **2019**, *21* (5), 1075–1079. https://doi.org/10.1039/C8GC04022B

(4) Pätzold, M.; Siebenhaller, S.; Kara, S.; Liese, A.; Syldatk, C.; Holtmann, D. Deep Eutectic Solvents as Efficient Solvents in Biocatalysis. *Trends Biotechnol.* **2019**, *37* (9), 943–959. https://doi.org/10.1016/j.tibtech.2019.03.007

(5) Shishov, A.; Bulatov, A.; Locatelli, M.; Carradori, S.; Andruch, V. Application of Deep Eutectic Solvents in Analytical Chemistry. A Review. *Microchem. J.* **2017**, *135*, 33–38. https://doi.org/10.1016/j.microc.2017.07.015

(6) Handy, S. Deep Eutectic Solvents in Organic Synthesis. In *Ionic Liquids - Current State of the Art*; Handy, S., Ed.; InTech, **2015**. https://doi.org/10.5772/59254

(7) Abbott, A. P. Deep Eutectic Solvents and Their Application in Electrochemistry. *Curr. Opin. Green Sustain. Chem.* **2022**, *36*, 100649. https://doi.org/10.1016/j.cogsc.2022.100649

(8) Fourmentin, S., Costa Gomes, M., Lichtfouse, E. *Deep Eutectic Solvents for Medicine, Gas Solubilization and Extraction of Natural Substances*; Springer International Publishing: Cham, **2021**, 103-129 and 131-155. https://doi.org/10.1007/978-3-030-53069-3

(9) Kollau, L. J. B. M.; Vis, M.; Van Den Bruinhorst, A.; Tuinier, R.; De With, G. Entropy Models for the Description of the Solid–Liquid Regime of Deep Eutectic Solutions. *J. Mol. Liq.* **2020**, *302*, 112155. https://doi.org/10.1016/j.molliq.2019.112155

(10) Martins, M. A. R.; Pinho, S. P.; Coutinho, J. A. P. Insights into the Nature of Eutectic and Deep Eutectic Mixtures. *J. Solut. Chem.* **2019**, *48* (7), 962–982. https://doi.org/10.1007/s10953-018-0793-1

(11) Prabhune, A.; Dey, R. Green and Sustainable Solvents of the Future: Deep Eutectic Solvents. *J. Mol. Liq.* **2023**, *379*, 121676. https://doi.org/10.1016/j.molliq.2023.121676

(12) Aissaoui, T.; AlNashef, I. M.; Qureshi, U. A.; Benguerba, Y. Potential Applications of Deep Eutectic Solvents in Natural Gas Sweetening for CO_2 Capture. *Rev. Chem. Eng.* **2017**, *33* (6), 523–550. https://doi.org/10.1515/revce-2016-0013

(13) Omar, K. A.; Sadeghi, R. Physicochemical Properties of Deep Eutectic Solvents: A Review. *J. Mol. Liq.* **2022**, *360*, 119524. https://doi.org/10.1016/j.molliq.2022.119524

(14) Zhang, Q.; De Oliveira Vigier, K.; Royer, S.; Jérôme, F. Deep Eutectic Solvents: Syntheses, Properties and Applications. *Chem. Soc. Rev.* **2012**, *41* (21), 7108. https://doi.org/10.1039/c2cs35178a

(15) Dou, W.; Yu, J.; Wang, X. Effect of Ethanol on the Density and Viscosity of Choline Chloride/Urea Eutectic System. *J. Mol. Liq.* **2023**, *382*, 121923. https://doi.org/10.1016/j.molliq.2023.121923

(16) Yadav, A.; Pandey, S. Densities and Viscosities of (Choline Chloride + Urea) Deep Eutectic Solvent and Its Aqueous Mixtures in the Temperature Range 293.15 K to 363.15 K. *J. Chem. Eng. Data* **2014**, *59* (7), 2221–2229. https://doi.org/10.1021/je5001796

(17) Xie, Y.; Dong, H.; Zhang, S.; Lu, X.; Ji, X. Effect of Water on the Density, Viscosity, and CO_2 Solubility in Choline Chloride/Urea. *J. Chem. Eng. Data* **2014**, *59* (11), 3344–3352. https://doi.org/10.1021/je500320c

(18) Shekaari, H.; Zafarani-Moattar, M. T.; Mohammadi, B. Thermophysical Characterization of Aqueous Deep Eutectic Solvent (Choline Chloride/Urea) Solutions in Full Ranges of Concentration at T = (293.15–323.15) K. *J. Mol. Liq.* **2017**, *243*, 451–461. https://doi.org/10.1016/j.molliq.2017.08.051

(19) Bakhtyari, A.; Haghbakhsh, R.; Duarte, A. R. C.; Raeissi, S. A Simple Model for the Viscosities of Deep Eutectic Solvents. *Fluid Phase Equilibria* **2020**, *521*, 112662. https://doi.org/10.1016/j.fluid.2020.112662

(20) Al-Dawsari, J. N.; Bessadok-Jemai, A.; Wazeer, I.; Mokraoui, S.; AlMansour, M. A.; Hadj-Kali, M. K. Fitting of Experimental Viscosity to Temperature Data for Deep Eutectic Solvents. *J. Mol. Liq.* **2020**, *310*, 113127. https://doi.org/10.1016/j.molliq.2020.113127

(21) Song, Z.; Hu, X.; Wu, H.; Mei, M.; Linke, S.; Zhou, T.; Qi, Z.; Sundmacher, K. Systematic Screening of Deep Eutectic Solvents as Sustainable Separation Media Exemplified by the CO_2 Capture Process. *ACS Sustain. Chem. Eng.* **2020**, *8* (23), 8741–8751. https://doi.org/10.1021/acssuschemeng.0c02490

(22) Song, Z.; Wang, J.; Sundmacher, K. Evaluation of COSMO-RS for Solid–Liquid Equilibria Prediction of Binary Eutectic Solvent Systems. *Green Energy Environ.* **2021**, *6* (3), 371–379. https://doi.org/10.1016/j.gee.2020.11.020

(23) Silva, L. P.; Fernandez, L.; Conceição, J. H. F.; Martins, M. A. R.; Sosa, A.; Ortega, J.; Pinho, S. P.; Coutinho, J. A. P. Design and Characterization of Sugar-Based Deep Eutectic Solvents Using Conductor-like Screening Model for Real Solvents. *ACS Sustain. Chem. Eng.* **2018**, *6* (8), 10724–10734. https://doi.org/10.1021/acssuschemeng.8b02042

(24) Abranches, D. O.; Larriba, M.; Silva, L. P.; Melle-Franco, M.; Palomar, J. F.; Pinho, S. P.; Coutinho, J. A. P. Using COSMO-RS to Design Choline Chloride Pharmaceutical Eutectic Solvents. *Fluid Phase Equilibria* **2019**, *497*, 71–78. https://doi.org/10.1016/j.fluid.2019.06.005

(25) Benguerba, Y.; Alnashef, I. M.; Erto, A.; Balsamo, M.; Ernst, B. A Quantitative Prediction of the Viscosity of Amine Based DESs Using σ-Profile Molecular Descriptors. *J. Mol. Struct.* **2019**, *1184*, 357–363. https://doi.org/10.1016/j.molstruc.2019.02.052

(26) Mohan, M.; Jetti, K. D.; Smith, M. D.; Demerdash, O. N.; Kidder, M. K.; Smith, J. C. Accurate Machine Learning for Predicting the Viscosities of Deep Eutectic Solvents. *J. Chem. Theory Comput.* **2024**. https://doi.org/10.1021/acs.jctc.3c01163

(27) Ayres, L. B.; Bandara, M.; McMillen, C. D.; Pennington, W. T.; Garcia, C. D. eutXG: A Machine-Learning Model to Understand and Predict the Melting Point of Novel X-Bonded

Deep Eutectic Solvents. *ACS Sustain. Chem. Eng.* **2024**. https://doi.org/10.1021/acssuschemeng.4c02844

(28) Afonso, J.; Mezzetta, A.; Marrucho, I. M.; Guazzelli, L. History Repeats Itself Again: Will the Mistakes of the Past for ILs Be Repeated for DESs? From Being Considered Ionic Liquids to Becoming Their Alternative: The Unbalanced Turn of Deep Eutectic Solvents. *Green Chem.* **2023**, *25* (1), 59–105. https://doi.org/10.1039/D2GC03198A

(29) Song, Z.; Chen, J.; Qin, H.; Qi, Z.; Sundmacher, K. Extending UNIFAC Models for Solid-Liquid Equilibria Prediction and Design of Eutectic Solvent Systems. *Chem. Eng. Sci.* **2023**, *281*, 119097. https://doi.org/10.1016/j.ces.2023.119097

(30) Pontes, P. V. A.; Crespo, E. A.; Martins, M. A. R.; Silva, L. P.; Neves, C. M. S. S.; Maximo, G. J.; Hubinger, M. D.; Batista, E. A. C.; Pinho, S. P.; Coutinho, J. A. P.; Sadowski, G.; Held, C. Measurement and PC-SAFT Modeling of Solid-Liquid Equilibrium of Deep Eutectic Solvents of Quaternary Ammonium Chlorides and Carboxylic Acids. *Fluid Phase Equilibria* **2017**, *448*, 69–80. https://doi.org/10.1016/j.fluid.2017.04.007

(31) Peng, D.; Alhadid, A.; Minceva, M. Assessment of COSMO-SAC Predictions for Solid–Liquid Equilibrium in Binary Eutectic Systems. *Ind. Eng. Chem. Res.* **2022**, *61* (35), 13256–13264. https://doi.org/10.1021/acs.iecr.2c00856

(32) Klamt, A. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224–2235. https://doi.org/10.1021/j100007a062

(33) Mutalib, N. F. A.; Bustam, M. A.; Wirzal, M. D. H.; Idris, A. A Prediction for the Conversion Performance of H2S to Elemental Sulfur in an Ionic-Liquid-Incorporated Transition Metal Using COSMO-RS. *Chemistry* **2022**, *4* (3), 811–826. https://doi.org/10.3390/chemistry4030058

(34) Lemaoui, T.; Darwish, A. S.; Attoui, A.; Abu Hatab, F.; Hammoudi, N. E. H.; Benguerba, Y.; Vega, L. F.; Alnashef, I. M. Predicting the Density and Viscosity of Hydrophobic Eutectic Solvents: Towards the Development of Sustainable Solvents. *Green Chem.* **2020**, *22* (23), 8511–8530. https://doi.org/10.1039/D0GC03077E

(35) Suykens, J. A. K.; Vandewalle, J. Least Squares Support Vector Machine Classifiers. *Neural Process. Lett.* **1999**, *9* (3), 293–300. https://doi.org/10.1023/A:1018628609742

(36) Almeida, L. B. Multilayer Perceptrons. In *Handbook of Neural Computation*; CRC Press, **1996**.

(37) Joback, K. G.; Reid, R. C. Estimation of Pure-Component Properties from Group-Contributions. *Chem. Eng. Commun.* **1987**, *57* (1–6), 233–243. https://doi.org/10.1080/00986448708960487

(38) Roosta, A.; Haghbakhsh, R.; Rita C. Duarte, A.; Raeissi, S. Deep Eutectic Solvent Viscosity Prediction by Hybrid Machine Learning and Group Contribution. *J. Mol. Liq.* **2023**, *388*, 122747. https://doi.org/10.1016/j.molliq.2023.122747

(39) Zuo, Y.; Xuan, Y.; Wang, Y.; Tong, J. Effect of Water on the Physical Properties of Quaternary Ammonium Salt-(DL)-Lactic Acid Deep Eutectic Solvents. *J. Chem. Eng. Data* **2024**. https://doi.org/10.1021/acs.jced.3c00781

(40) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20* (3), 273–297. https://doi.org/10.1007/BF00994018

(41) Vapnik, V. N. The Support Vector Method. In *Artificial Neural Networks — ICANN'97*; Gerstner, W., Germond, A., Hasler, M., Nicoud, J.-D., Eds.; Springer: Berlin, Heidelberg, **1997**; pp 261–271. https://doi.org/10.1007/BFb0020166

(42) Joshi, A. V. *Machine Learning and Artificial Intelligence*; Springer International Publishing: Cham, **2020**. https://doi.org/10.1007/978-3-030-26622-6

(43) Ho, T. K. Random Decision Forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*; **1995**; Vol. 1, pp 278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994

(44) McCulloch, W. S.; Pitts, W. A Logical Calculus of the Ideas Immanent in Nervous Activity. *Bull. Math. Biophys.* **1943**, *5* (4), 115–133. https://doi.org/10.1007/BF02478259

(45) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; KDD '16; Association for Computing Machinery: New York, NY, USA, **2016**; pp 785–794. https://doi.org/10.1145/2939672.2939785

(46) Shi, D.; Zhou, F.; Mu, W.; Ling, C.; Mu, T.; Yu, G.; Li, R. Deep Insights into the Viscosity of Deep Eutectic Solvents by an XGBoost-Based Model plus SHapley Additive exPlanation. *Phys. Chem. Chem. Phys.* **2022**, *24* (42), 26029–26036. https://doi.org/10.1039/D2CP03423A

(47) Mjalli, F. S.; Naser, J. Viscosity Model for Choline Chloride-Based Deep Eutectic Solvents. *Asia-Pac. J. Chem. Eng.* **2015**, *10* (2), 273–281. https://doi.org/10.1002/apj.1873

(48) Haghbakhsh, R.; Parvaneh, K.; Raeissi, S.; Shariati, A. A General Viscosity Model for Deep Eutectic Solvents: The Free Volume Theory Coupled with Association Equations of State. *Fluid Phase Equilibria* **2018**, *470*, 193–202. https://doi.org/10.1016/j.fluid.2017.08.024

(49) Odegova, V.; Lavrinenko, A.; Rakhmanov, T.; Sysuev, G.; Dmitrenko, A.; Vinogradov, V. DESignSolvents: An Open Platform for the Search and Prediction of the Physicochemical Properties of Deep Eutectic Solvents. *Green Chem.* **2024**, *26* (7), 3958–3967. https://doi.org/10.1039/D3GC04533A

(50) Abbas, U.; Zhang, Y.; Tapia, J.; Md, S.; Chen, J.; Shi, J.; Shao, Q. Machine-Learning Assisted Design of Deep Eutectic Solvents Based on Uncovered Hydrogen Bond Patterns. *Eng.* **2024**, *39*, 74-83. https://doi.org/10.1016/j.eng.2023.10.020

(51) Dardzinska, A.; Zdrodowska, M. Classification Algorithms in the Material Science and Engineering Data Mining Techniques. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *770* (1), 012096. https://doi.org/10.1088/1757-899X/770/1/012096

(52) Casas, A.; Rodríguez-Llorente, D.; Rodríguez-Llorente, G.; García, J.; Larriba, M. Machine Learning Screening Tools for the Prediction of Extraction Yields of Pharmaceutical Compounds from Wastewaters. *J. Water Process Eng.* **2024**, *62*, 105379. https://doi.org/10.1016/j.jwpe.2024.105379

(53) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* **2010**, *50* (5), 742–754. https://doi.org/10.1021/ci100050t

(54) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; von Lilienfeld, O. A.; Müller, K.-R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6* (12), 2326–2331. https://doi.org/10.1021/acs.jpclett.5b00831

(55) Huang, B.; von Lilienfeld, O. A. Communication: Understanding Molecular Representations in Machine Learning: The Role of Uniqueness and Target Similarity. *J. Chem. Phys.* **2016**, *145* (16), 161102. https://doi.org/10.1063/1.4964627

(56) Zhong, S.; Guan, X. Count-Based Morgan Fingerprint: A More Efficient and Interpretable Molecular Representation in Developing Machine Learning-Based Predictive Regression Models for Water Contaminants' Activities and Properties. *Environ. Sci. Technol.* **2023**, *57* (46), 18193–18202. https://doi.org/10.1021/acs.est.3c02198

(57) Raghunathan, S.; Priyakumar, U. D. Molecular Representations for Machine Learning Applications in Chemistry. *Int. J. Quantum Chem.* **2022**, *122* (7), e26870. https://doi.org/10.1002/qua.26870

(58) Zhou, Z.-H. *Machine Learning*; Springer Nature, **2021**. https://doi.org/10.1007/978-981-15-1967-3

(59) Herschlag, D.; Pinney, M. M. Hydrogen Bonds: Simple after All? *Biochemistry* **2018**, *57* (24), 3338–3352. https://doi.org/10.1021/acs.biochem.8b00217

(60) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, No. 12, 2825–2830.

(61) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. arXiv **2017**. https://doi.org/10.48550/ARXIV.1705.07874

(62) Joseph, F. J. J.; Nonsiri, S.; Monsakul, A. Keras and TensorFlow: A Hands-on Experience. In *Advanced Deep Learning for Engineers and Scientists: A Practical Approach*; Prakash, K. B., Kannan, R., Alexander, S. A., Kanagachidambaresan, G. R., Eds.; Springer International Publishing: Cham, **2021**; pp 85–111. https://doi.org/10.1007/978-3-030-66519-7

(63) Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mane, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viegas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *Google Res.* **2015**, 1–19.

(64) Chollet, F. Keras, **2015**. https://github.com/keras-team/keras (accessed 2024-06-26).

(65) Chen, Z.; Ludwig, M.; Warr, G. G.; Atkin, R. Effect of Cation Alkyl Chain Length on Surface Forces and Physical Properties in Deep Eutectic Solvents. *J. Colloid Interface Sci.* **2017**, *494*, 373–379. https://doi.org/10.1016/j.jcis.2017.01.109

(66) Grayson, J. W.; Song, M.; Evoy, E.; Upshur, M. A.; Ebrahimi, M.; Geiger, F. M.; Thomson, R. J.; Bertram, A. K. The Effect of Adding Hydroxyl Functional Groups and

Increasing Molar Mass on the Viscosity of Organics Relevant to Secondary Organic Aerosols. August 4, **2016**. https://doi.org/10.5194/acp-2016-672

(67) Costa, T.; Sanchez-Vicente, Y.; Yang, Z.; A. Stevens, L.; S. Dias, F. de; Pereira, S.-C. C. Thermophysical Properties of Tetrabutylammonium Chloride, Paraffin and Fatty Acids for Thermal Energy Applications. *RSC Adv.* **2024**, *14* (36), 26246–26258. https://doi.org/10.1039/D4RA03782K

(68) Burns, J. A.; Verrall, R. E. Thermodynamics of Tetraalkyl- and Bis-Tetraalkylammonium Bromides: II. Heat Capacities of Solid State from 273 to 373 K. *Thermochim. Acta* **1974**, *9* (3), 277–288. https://doi.org/10.1016/0040-6031(74)80008-0

(69) Adeyemi, I.; Abu-Zahra, M. R. M.; AlNashef, I. M. Physicochemical Properties of Alkanolamine-Choline Chloride Deep Eutectic Solvents: Measurements, Group Contribution and Artificial Intelligence Prediction Techniques. *J. Mol. Liq.* **2018**, *256*, 581–590. https://doi.org/10.1016/j.molliq.2018.02.085

(70) Yadav, A.; Trivedi, S.; Rai, R.; Pandey, S. Densities and Dynamic Viscosities of (Choline chloride + Glycerol) Deep Eutectic Solvent and Its Aqueous Mixtures in the Temperature Range (283.15–363.15) K. *Fluid Phase Equilibria* **2014**, *367*, 135–142. https://doi.org/10.1016/j.fluid.2014.01.028

(71) Gajardo-Parra, N. F.; Cotroneo-Figueroa, V. P.; Aravena, P.; Vesovic, V.; Canales, R. I. Viscosity of Choline Chloride-Based Deep Eutectic Solvents: Experiments and Modeling. *J. Chem. Eng. Data* **2020**, *65* (11), 5581–5592. https://doi.org/10.1021/acs.jced.0c00715

(72) Shah, D.; Mjalli, F. S. Effect of Water on the Thermo-Physical Properties of Reline: An Experimental and Molecular Simulation Based Approach. *Phys Chem Chem Phys* **2014**, *16* (43), 23900–23907. https://doi.org/10.1039/C4CP02600D

(73) Zhekenov, T.; Toksanbayev, N.; Kazakbayeva, Z.; Shah, D.; Mjalli, F. S. Formation of Type III Deep Eutectic Solvents and Effect of Water on Their Intermolecular Interactions. *Fluid Phase Equilibria* **2017**, *441*, 43–48. https://doi.org/10.1016/j.fluid.2017.01.022

(74) Sarmad, S.; Xie, Y.; Mikkola, J.-P.; Ji, X. Screening of Deep Eutectic Solvents (DESs) as Green CO_2 Sorbents: From Solubility to Viscosity. *New J. Chem.* **2017**, *41* (1), 290–301. https://doi.org/10.1039/C6NJ03140D

SYNOPSIS

Machine learning models enhance green solvent discovery and yield accurate property predictions, minimizing material wasted on trial-and-error laboratory experiments.

GRAPHICAL ABSTRACT