Digital Discovery

PAPER

Check for updates

Understanding impact sensitivity of energetic molecules by supervised machine learning

Heather M. Quayle, [a] Karthik Mohan, [b] Sohan Seth, [b] Colin R. Pulham [a] and Carole A. Morrison *[a]

Machine learning models have been developed to rationalise correlations between molecular structure and sensitivity to initiation by mechanical impact for a data set of 485 energetic molecules. The models use readily obtainable features derived from SMILES strings to classify structures, first by a binary split to differentiate between primary and secondary energetic material behaviour, and by subsequent boundary divisions to create up to five impact sensitivity classes. The best accuracy score was 0.79, which was obtained for the binary classifier random forest model. Feature importance and SHAP analysis showed that the features most likely to categorise a molecule with a high impact sensitivity were a high oxygen balance and a high molecular flexibility. The outcome of this study gives easily interpretable information on how the structure of a molecule can be tailored to design energetic materials with desired impact sensitivity properties. Included model codes also allow users to predict the sensitivity classes of any additional molecular structures from a SMILES string.

Introduction

Energetic materials (EMs, explosives, propellants and pyrotechnics) are substances that contain large amounts of stored chemical energy that can be quickly released upon initiation.[1] Their sensitivity to mechanical stimuli, *i.e.*, the impact of a falling weight, can be quantified as impact sensitivity (IS), and is an important safety metric for EMs. For this reason, substantial efforts are devoted to measuring this property, which is typically achieved using a fall-hammer test,[2] such as the BAM apparatus.[3] IS values are quoted as either an h_{50} value, which represents the height in centimetres at which a weight of known mass dropped onto the sample will induce initiation 50% of the time, or alternatively as an E_{50} value, where the data is recast in units of Joules. It is well known that this test comes with inherent subjectivity, as well as inconsistency, due to dependencies on variables such as crystal size and morphology, sample purity, humidity and temperature.[2,4,5] While valuable research is ongoing to improve data reproducibility,[6–8] one further limitation exists for the fall-hammer test, namely the maximum height from which the weight can reasonably be dropped. This leads to issues with data resolution, as many insensitive compounds are simply recorded as having an IS value of greater than 40 J, meaning opportunities to differentiate between compounds that fall into this category are lost.

Many efforts have also been undertaken to establish relationships between IS and both the molecular and crystalline structures of EMs. Studies have reported correlations with the amount of free space per molecule in its corresponding crystal lattice,[9,10] the electronic band gap of the EM,[11] bond dissociation energies,[12] charge distributions within molecules,[13] and activation energy of thermal decomposition.[14] More computationally expensive methods include models which are rooted in mechanochemistry *via* the vibrational up-pumping model[15–18] such as those by Ye, Bernstein and Michalchuk.[19–24] However, although this approach has demonstrated success in predicting trends in IS for a broad range of crystalline EMs, the process is computationally intensive, which naturally limits the number of compounds that can be processed in this way. This means that the opportunity to 'learn' what features are most likely to influence IS is limited. An obvious strategy to address this is to shift the focus from the solid-state structure onto the molecular structure. Although this means that any correlations with crystal structure properties, *e.g.*, polymorphism, will be lost, a substantial increase in the size of the data set offers opportunities to apply machine learning (ML) methods to identify aspects of molecular structure that correlate with IS.

Recent work has shown that ML models can be successfully trained to predict several important properties of EMs, including formation energies and crystal densities.[25–27] However, IS presents challenges for training ML models because of the relatively small size of the data set,[28,29] as well as the inherent limitations of experimental measurements as outlined above. Several ML models for IS have been reported, but these have typically been restricted to one or two functional

aSchool of Chemistry and EaStCHEM Research School, University of Edinburgh, The King's Buildings, David Brewster Road, EH9 3FJ, UK. E-mail: c.morrison@ed.ac.uk
bSchool of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh, EH8 9AB, UK

groups or structure types[26,30–34] which limits their general applicability to other types of EMs. A recent report has drawn on a larger and broader data set,[35] although it included both molecular and ionic compounds which means any conclusions drawn will likely be complex, as salt or co-crystal formation are known to markedly alter IS values of EMs.[36] Other ML models have been explored that are rooted in the kinetics of the decomposition process behind energetic initiation,[37,38] or which use EM-property-based features calculated using semi-empirical or quantum mechanical calculations.[39–41] While these models have achieved high accuracies, the complex nature of the features employed means that it is harder to rationalise how these features translate to the structure of the molecule, and therefore how the insights learned can be directly used to design new molecules with desired IS values. Therefore, methods involving prediction of IS using features that can be extracted purely from the 2D structure of the molecule, i.e., a SMILES string,[42] or even just the molecular formula, allow for simpler translation into structure design.[43–48]

From this short overview it is clear that opportunities exist to apply ML models to capture the correlations between molecular structure and measured IS values, but the available data presents challenges. While transfer learning is being applied to address the issues associated with limited data,[49,50] this still leaves the issue of data reliability. To address this, we have opted to explore classification models, where molecules are assigned to a particular category based on a range of IS values. In this way we not only tackle the problems associated with numerical variation, but we can also include the reports that simply state IS values above the common 40 J threshold value.

Herein we report our efforts to derive a classification ML model to link impact sensitivity to molecular structure by undertaking the following steps:

(1) Creating a substantial IS data set for molecular EMs from the available literature, taking care to include a broad range of functional groups and structural motifs. Salts and co-crystals are excluded from this data set for the reasons outlined above.

(2) Training ML models based on classification methods (specifically, linear support vector classification (SVC), logistic regression, random forest (RF) and Light Gradient Boosting Machine (LightGBM)) to group EMs into classes with similar reported IS values. As previously described, this can address some of the experimental limitations of the data set. We start with a simple binary classifier model, i.e., setting one sensitive/insensitive boundary, and extend the methodology to encompass up to five classes.

(3) Defining model features that are quick and straightforward to obtain (i.e., no quantum mechanically computed data). These should be already known to be important features for EM design, be inspired by insights gained in our previous work on a mechanochemistry-based vibrational up-pumping model or be generally important structural features that influence molecular design.

(4) Analysing how the features highlighted by the ML model can be applied to direct the design of new molecules with desired IS values.

Computational methodology

An extensive literature search was undertaken to create a large drop-weight sensitivity data set for molecular-based EMs (485 compounds, see SI). This includes some (ca. 40%), of the previous data set reported by Storm which contains a total of 279 compounds.[51] This new data set encompasses all the main molecular EM structure types (azides, ca. 6%, tetrazoles 17%, pyrazoles 17%, triazoles 16%, nitrate esters 7%, aromatics 22%, and other ring/cage compounds 14% – note that most compounds fall into multiple classes, see SI), and encompasses a wide spread of IS behaviour. Only solid molecular crystals were included, i.e., polymers, liquids, salts, hydrates and co-crystals are excluded. The data were not sorted or differentiated by the instrument used to measure experimental IS. Other large data sets, such as those by Muravyev et al.[52] and Marrs et al.[39] were not used when compiling this data set but are acknowledged as significant contributions to the availability of IS data and data on other EM properties.

We then grouped compounds into the classifications defined in Table 1. Final class boundary decisions were made to create approximately equal class sizes; this was particularly pertinent for the tertiary, quaternary and quinary classes. Altering the class boundaries was not explored as this would create an uneven distribution of data in each class; this imbalance would likely lead to unreliable outcomes.

Our choices of model features are summarised in Fig. 1. Some were selected based on criteria that are already known to be important for general EM performance. One such feature is oxygen balance (OB; a metric that defines whether a molecule contains sufficient oxygen to completely convert all carbon present to CO_2 during oxidation). Another feature is the number of weak bond linkages, described as trigger bonds (TBs).[53] In this work we have defined two classes of trigger bonds – the widely accepted R–NO_2 (where R = C, N, or O), which are denoted as class I trigger bonds, together with other weak bonds (C–N)$_{aliphatic}$, (C–O)$_{aliphatic}$, $C_{aliphatic}$–$N_{aromatic}$, N–NH_2 and C–P, which we term class II trigger bonds. These have recently been highlighted as being just as weak as the class I trigger bonds.[24] We also include several features that take inspiration from insights learned from the vibrational up-pumping model, which has highlighted that a correlation exists between predicted impact sensitivity and the number of so-called doorway modes;[18,24] these are low-energy vibrational modes typically describing angle bends and torsional motions, together with some bond-stretching character. A high density of doorway modes is essentially indicative of molecular flexibility. For our ML study this has been translated into the Kier Molecular Flexibility (KMF) index (a measure of a molecule's flexibility based on its size, atom types and degree of one-bond and two-bond connectivity information),[24,54] along with the number of rotatable bonds. We also include a count of the total number of rings, as well as differentiating between the number of aromatic and other rings, which essentially define stiffer structures, as well as a ratio of hydrogen bond donor groups to hydrogen bond acceptor groups. This defines the molecular intra- and inter-

Table 1 Classification boundaries for impact sensitivity measurements. The number of entries in each class is given in square brackets

Classification	Class				
	0	1	2	3	4
Binary	IS ≤ 8 J [212]	IS > 8 J [273]	—	—	—
Tertiary	IS < 6 J [162]	6 ≤ IS < 20 J [162]	IS > 20 J [161]	—	—
Quaternary	IS ≤ 4 J [129]	4 < IS ≤ 10 J [131]	10 < IS ≤ 30 J [117]	IS > 30 J [108]	—
Quinary	IS ≤ 3 J [100]	3 < IS ≤ 8 J[112]	8 < IS ≤ 18 J [95]	18 < IS < 40 J [93]	IS ≥ 40 J [85]

molecular bonding potential, and is also indicative of molecular flexibility, since a high potential for forming hydrogen bonding interactions will likely restrict the molecular conformation. We also note that the vibrational up-pumping model has previously highlighted that IS correlates very strongly with the number of trigger bonds.[16,24] Our final set of features take inspiration from the design approach commonly adopted by synthetic chemists, and includes molecular weight (MW), empirical formula (defined by the number of hetero, carbon, oxygen and nitrogen atoms), and the decisions on functional group placement (e.g. whether to position an –NO_2, –NH_2, –NH, –OH, or –CH_3 group adjacent to another –NO_2 group). A similar approach of including structural fragments has been shown to be advantageous in earlier ML models,[55,56] and all substituent groups included here are commonly found in some combination in energetic molecules. Additionally, interactions between adjacent groups could lead to intra- and inter-molecular interactions, affecting the sensitivity as described above. We also include the number of azide (–N_3) functional groups present. Finally, we include three cheminformatics features that define fundamental properties related to electron distribution and surface area. These are (i) the topological polar surface area (TPSA), a measure of the total polar area on a molecule,[57] (ii) VSA_EState8, which is related to the ease with which two atoms can interact (due to electronegativity difference and physical distance) and is therefore associated with hydrogen bond interactions,[58] and (iii) SMR_VSA5, which is related to molecular polarisability and therefore also interactions.[59] These features also represent a straightforward way to approximate the effects of the electrostatic surface potential, which has previously been observed to correlate with IS for a small sample of nitroaromatics and nitroheterocycles.[13,60] Feature correlation analysis was carried out prior to model training to detect and remove any highly correlated features (see Fig. S1 in the SI). As a result of this first step, five features were highlighted as highly correlated, and were therefore removed from the initial feature set. The affected features were the (i) total number of rings, (ii) number of rotatable bonds, (iii) molecular weight, (iv) number of oxygen atoms and (v) number of heteroatoms were removed. This is reflected in Fig. 1.

Each molecular structure was parsed in the form of a SMILES string,[42] while the Python3 (ref. 61) library RDKit[62] and SMARTS queries[63] were used to extract the data for the selected features. The SMARTS parsing script was based on published scripts by Rein et al.[32] Scikit-learn[64] and LightGBM[65] were used for data pre-processing and model implementation. During pre-processing, the continuous features were log-transformed and scaled for modelling. Classification machine learning models linear support vector classifier (SVC), logistic regression (LogReg), random forests (RF) and Light Gradient Boosting Machine (LightGBM) were implemented. These were chosen to test a range of well-established linear, non-linear and tree-based ensemble models to balance interpretability and accuracy.

The input data were randomly split such that 75% of the data was used for training the models and 25% was used for testing. For models implemented in scikit-learn, hyperparameters were tuned using 5-fold Grid Search Cross Validation. Bayesian Optimisation (via HyperOpt) was used for hyperparameter tuning in LightGBM.[66] These optimised hyperparameters were used to train all four ML models for each of the four classification tasks. Model outcomes were assessed via precision and recall values. Analysis of feature importance for all four classification tasks was performed for the most accurate models using SHapley Additive exPlanations (SHAP) analysis.[67,68]
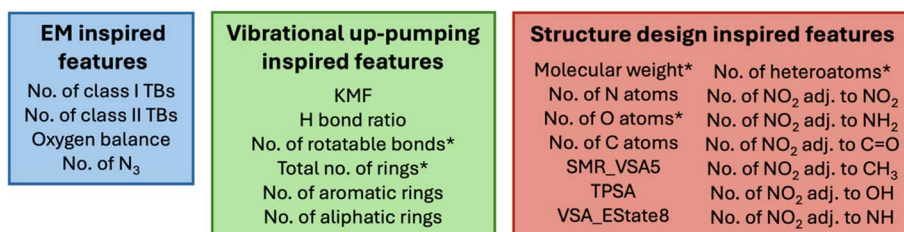
Fig. 1 Initial features selected for ML models. Those marked with an asterisk were removed before model training due to high correlation with other features.

Results and discussion

The best outcomes obtained for training and testing data sets for the four classifications are shown in Table 2. Scores for all models are available in the SI. The class assignment threshold was 0.5 probability for binary classification (*i.e.*, if the model prediction for a molecule is > 0.5, we consider that as class 1, or otherwise class 0) and based on highest predicted probability for the multi-class classification.

The most visible and expected outcome is that the highest accuracy is achieved for the binary classification model. This is to be expected because having only two classification groups means that there is more data for each class to be trained on and there are fewer possible outcomes to predict. From an experimental perspective this is the most important boundary division, as an IS value of about 8 J marks the approximate differentiator between primary and secondary EM behaviour. As the number of classification groups increases, it follows that training and outcome prediction will be affected as the data set is divided into increasingly smaller classes. The RF model performs best in three out of the four classification tasks, which is not an unexpected result, since Random Forests are known to perform well in cases of non-linear, complex relationships.

The outcomes from the models reported in Table 2 are presented as confusion matrices in Fig. 2. For the binary classification task (Fig. 2a) this shows that 33 (out of 47) molecules are correctly predicted as true class 0 (*i.e.* IS ≤ 8 J; this is a recall rate of 70%), while 63 (out of 75) molecules are correctly predicted as true class 1 (*i.e.* IS > 8 J; 84% recall). This slightly skewed performance could arise due to a little more of the data set (and therefore more of the training data set) being assigned to class 1 (see Table 1). For the tertiary classification task (Fig. 2b), the recall rates for true class 0–2 prediction are 74%, 48%, and 74%, respectively, indicating now that the middle-ranking class (where IS falls in the range 6–20 J) is the most challenging to predict. This trend continues for the quaternary and quinary data sets (Fig. 2c and d), where recall accuracy increases at either end of the sensitivity spectrum. For the quaternary classification task, the recall rates for true class 0 and 1 are 56% and 62%, respectively, dropping to 34% for true class 2 assignment, before improving back up to 55% for the true class 3. For the quinary classification task, the overall performance accuracy is lowest, at 0.48, but the recall rate for true class 0 (*i.e.* compounds with IS ≤ 3 J) is 52%, and true class 1 (also sensitive compounds with IS in the range 3–8 J) is 65%, showing best ability for prediction of most sensitive compounds, important for the safety aspect of the prediction.

Another metric to report the predictions for the ML models is the precision scores. This defines, for example, what proportion of the molecules in a given model predicted to be class 0 are true class 0. The confusion matrices (Fig. 2) show that precision and recall scores follow similar trends for most of the models.

It is important to note that impact sensitivity classification data is ordinal, *i.e.*, we know that class 0 is more sensitive than class 1, which is more sensitive than class 2, *etc.* However, this is not accounted for by the models, which will treat the data as though it is nominal, *i.e.* as if it has no implicit ordering. Additionally, since we know that there is an experimental error on all IS data (which is not always reported, and so is not accounted for here), a molecule assigned to class 2, for example, could actually be class 1. This becomes more of an issue as we move to higher numbers of categories that span smaller numerical ranges of IS values. We see that the majority of wrongly classed predictions are assigned to neighbouring categories of the true class, which suggests that the models are nonetheless doing well at interpreting the order of the classes.

Next, we extracted the feature importances from each model; the outcome for the quaternary RF model is shown in Fig. 3 and is broadly representative of the outcomes obtained for all the classification models (see SI). This highlights that several of the features that were already known to be important for EM design (blue bars), and those that were inspired by the vibrational up-pumping model (green bars), were ranked highly in comparison to the structurally-inspired features (red bars). We note that the high ranking of oxygen balance has been previously observed in other ML models for IS prediction.[39,41] Although the ordering of feature importance does change between the four classification tasks, the list of the most important features is consistent. Some slight change in ordering is to be expected due to the differing decision boundaries for the four different tasks.

In addition to simply identifying which features correlate with IS, the central aim of this work was to create a model which could inform molecular design of more- or less-sensitive EMs in a straightforward way, by using readily accessible features that correlate with molecular structure. The next step was therefore to analyse how the most important structural parameters affect sensitivity, and how these could be used as a design tool for new EMs with predicted IS behaviour. This can be performed using SHAP analysis, with the binary classification model being the most intuitive model to interpret. The analysis ranks the features, from most important (top) to least (bottom), in terms

Table 2 Best models (as judged by highest test set accuracy scores) for each of the four classification models

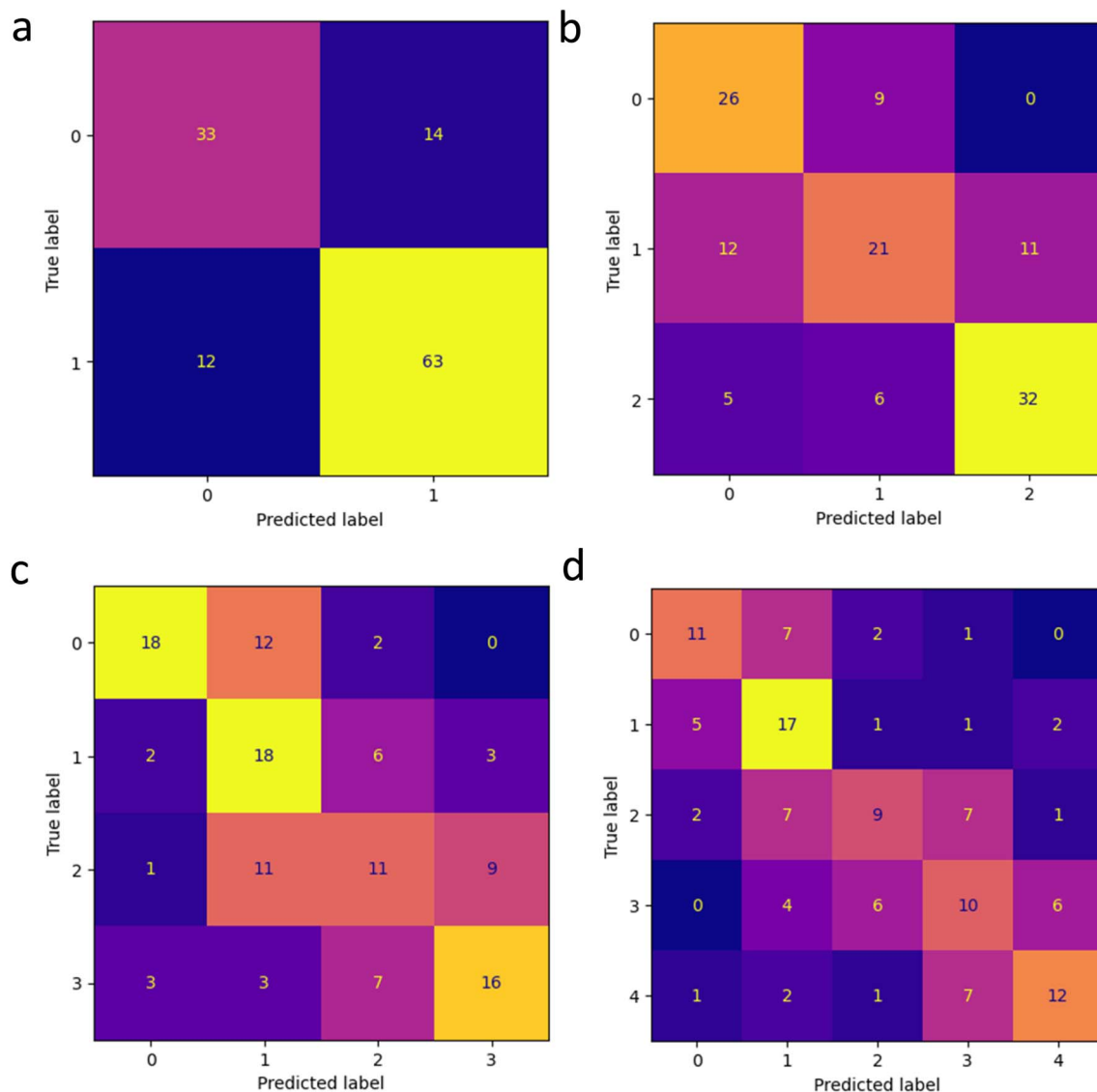| Classification	Best model	Accuracy score		Macro averaged precision	Macro averaged recall
		Train	Test		
Binary	RF	0.95	0.79	0.78	0.77
Tertiary	RF	0.80	0.65	0.64	0.65
Quaternary	RF	0.99	0.52	0.54	0.52
Quinary	LogReg	0.52	0.48	0.49	0.49

Fig. 2 Confusion matrices for the most accurate ML model for each classification task, showing only testing data: (a) binary RF, (b) tertiary RF, (c) quaternary RF, (d) quinary LogReg.

of whether a high (red) or low (blue) numerical value of each parameter will consign a given molecule to class 0 (positive SHAP values) or class 1 (negative SHAP values) (see Fig. 4). Thus, analysis of the colour distribution of the features given in the top right-hand side of the plot gives the most important information to categorise a molecule as a sensitive EM (IS ≤ 8 J). It is important to note that most of the features are scaled in the pre-processing step to have values between 0 and 1. The exceptions to this are the number of C atoms and the six functional group relative position parameters. Therefore, whilst the effect these features have on the model output is correctly accounted for, the impact of the number of C atoms on the result may be slightly overestimated.

A number of observations become readily apparent. Firstly, a high oxygen balance is more likely to assign a given EM to be impact sensitive, a correlation that has been shown previously.[46,47,69] Secondly, features that code for high molecular

flexibility also correlate with increased impact sensitivity. Specifically, this is represented by a low ratio of hydrogen bonds (i.e. molecules are less likely to constrained by hydrogen bonding interactions), high KMF values, a low number of aromatic rings in the structure, and VSA_EState8 being low.[24] The low values of VSA_EState8 and SMR_VSA5 correlating with higher sensitivity is somewhat in contrast with previous observations documenting the relationship between electrostatic surface potential and impact sensitivity.[13] However, it should be noted that the previous work used a dataset constrained to nitroaromatics only, which only account for 22% of our data set, so the relationship may be different when more structural variety is considered, as it is in this study. Thirdly, for a more sensitive molecule, the number of class I and II trigger bonds should be high. Increased impact sensitivity also correlates with the number of carbon atoms being low, in agreement with the oxygen balance being high. Molecules with a high number of
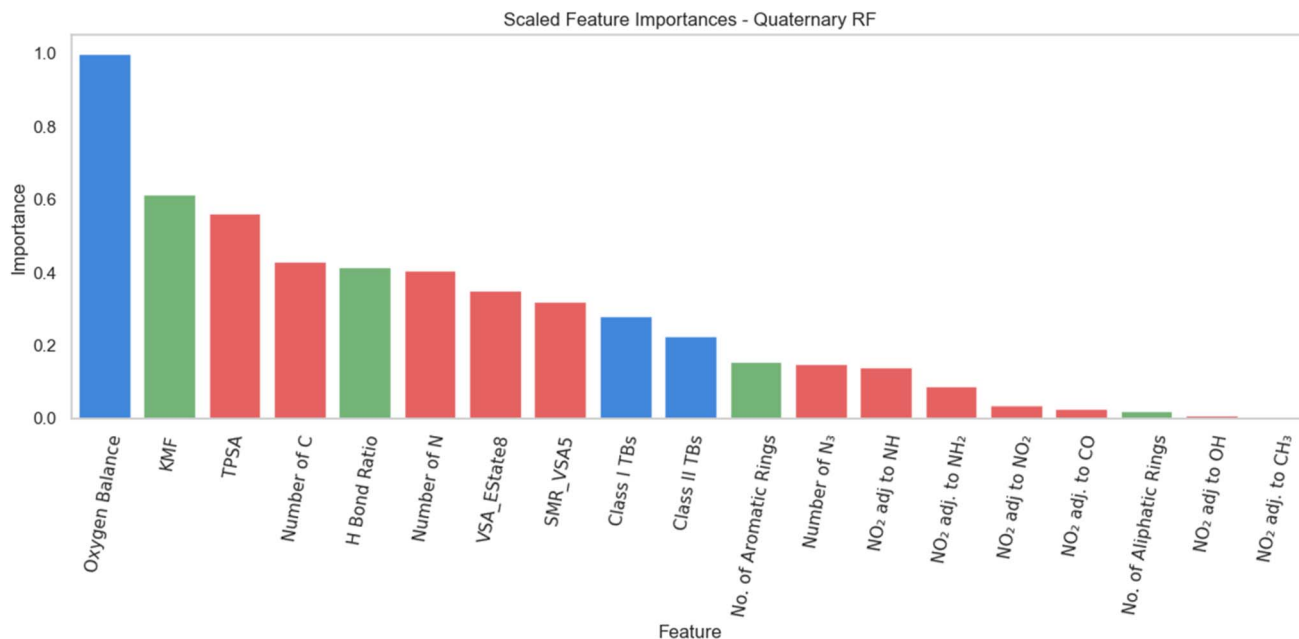
Fig. 3 Feature importance (normalised to the most important feature) for the quaternary RF model, based on splits (the number of times the feature is used in the model). The bars are coloured according to the feature groupings shown in Fig. 1.

azide functional groups are more likely to be categorised as highly impact sensitive, although this may reflect a bias in the data set, since 28 out of 30 azides in our data set are very impact sensitive. Finally, the features that relate to functional group placement have a very weak effect on classification. This is an important finding for the design of energetic molecules, as it suggests that, at least for the 485 molecules in the data set explored here, we are not able to identify a correlation between placement of these particular functional groups and IS. This may be symptomatic of the size of the data set employed, or SMILES strings being too simplistic to capture molecular geometry features.

Since this work is employs a ML model, rather than a physical model, any reasoning as to why some features are more important than others is ultimately speculation. The exceptions to this are the features that relate to previous work performed on a physical model (vibrational up-pumping) which permitted structure/property relationships to be explored, albeit on a substantially smaller data set (33 compounds)[24] and oxygen balance which, as discussed above, has been shown to be an important feature in multiple previous models. Mathieu proposed that this could be due to initiation depending on the ability of oxygen-containing groups to fuel conversion of energetic molecules to decomposition products, and therefore formulated a link between the proportion of oxygen in a molecule and its ease of initiation.[70] Features coding for high flexibility (such as a low hydrogen bond ratio, high KMF and a low number of rings), and a high number of trigger bonds giving more sensitive molecules is in agreement with the general findings from the mechanically-induced impact sensitivity model.[24] This permits a physical interpretation for these correlations that more flexible molecules have more vibrational

modes of appropriate frequencies to more efficiently channel the impact energy up to the higher frequency modes that excite weak chemical bonds.

This work is the first attempt to impose a predictive classification system on impact sensitivity data, and we believe that this straightforward approach, particularly with respect to the binary classification scheme, will be beneficial to experimentalists who seek guidance on molecular design features that will likely result in primary or secondary energetic behaviour. While regression and classification models are fundamentally different approaches it is, of course, possible to binarise the continuous output from a regression model to try and offer a comparison. The data set provided by Matheiu offers this possibility, as they provide numerical predictions alongside the experimental values.[38] Applying the binary dividing value of 8 J to the output from their Mod7P model resulted in 78% of the structures being correctly assigned as primary energetics, and 90% as secondary energetics (test set data size: 217 structures). These recall rates are better than ours (70% and 84% for primary and secondary, respectively), but we note that this is a surface analysis only. More in-depth analysis would be required to convert an existing regression model into a categorisation model, and both would need to be run with the same data set to assess whether the model we have proposed here is formally more or less accurate than previously published regression models, and whether classification offers improved predictions over regression.

Our models and data set are open access. In order to maximise their utilisation beyond training and testing, we also include a prediction functionality, which provides a simple route to predict the sensitivity class for an additional molecule outside of our data set from a user-provided SMILES string. This
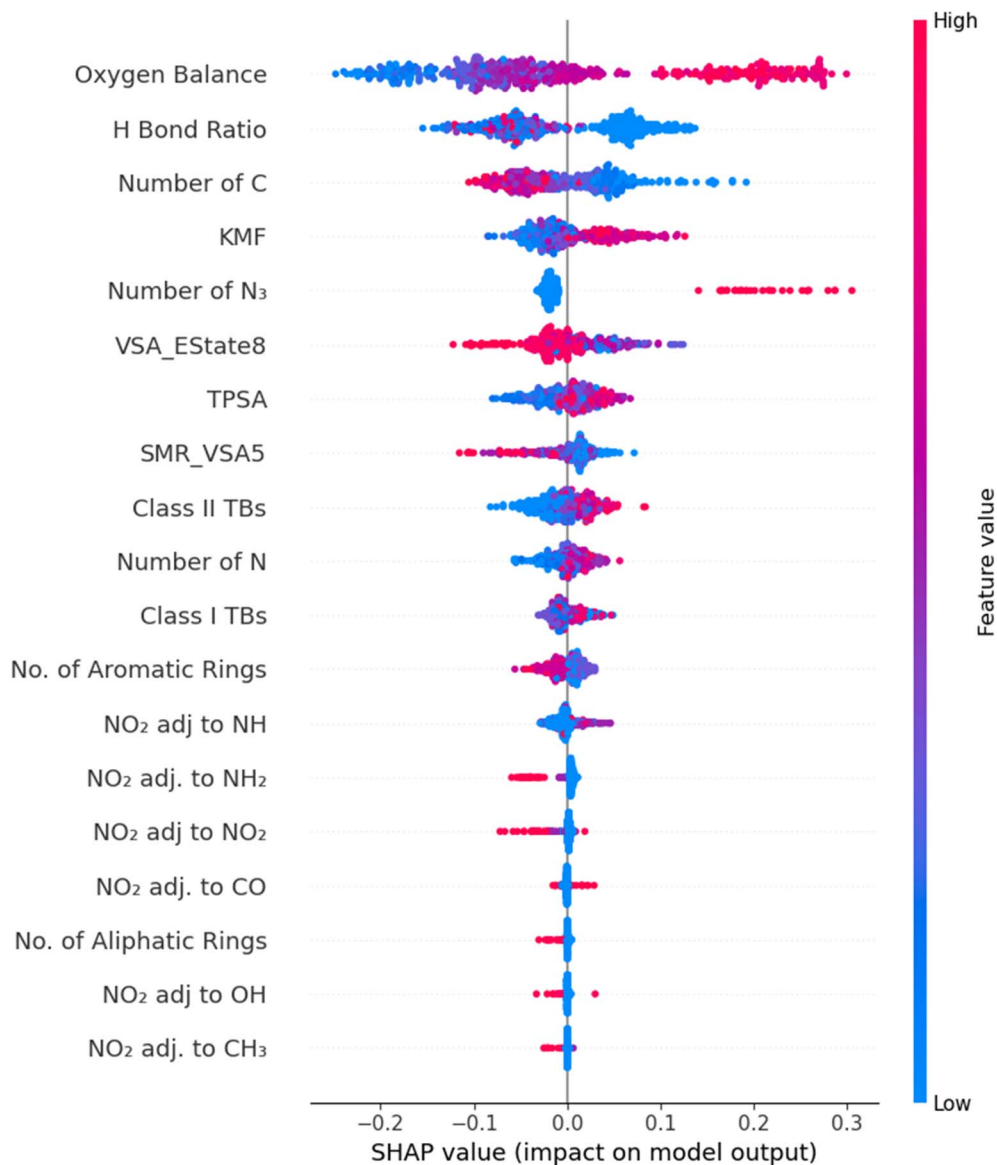
Fig. 4 SHAP analysis for the whole data set RF binary classification model. More positive SHAP values direct more strongly towards class 0 (IS ≤ 8 J) status, while the blue/red colour bar indicates how the numerical value of the feature directs towards this class.

additional functionality is available for the four models outlined in Table 2.

Conclusion

In this work, machine learning models have been developed to rationalise correlations between molecular structure and impact sensitivity. This was achieved using one of the largest energetic molecule data sets constructed to date (485 structures). The model requires only readily obtainable features derived from a SMILES string (no QM-calculated parameters), informed by insights gained from our previous work on the vibrational up-pumping model for impact initiation, or which are generally known to be important structural features in energetic molecular design. The work also addresses how to account for some of the limitations in experimental IS data,

particularly where published values show numerical variation or simply report IS values above 40 J, by implementing classification ML models to group the compounds, first by a binary split, and then by further dividing the compounds to create up to five separate groupings. For the test data, the best accuracy score for the binary classification model was 0.79, which fell to 0.48 as the number of classification categories rose to five. This result was to be expected, as the higher number of classification groupings means that the number of data points available for model training in each class falls, and the number of possible outcomes to predict rises. From a molecular design perspective, the binary classifier is the most important, as an IS value of *ca.* 8 J (chosen as the differentiator in our ML model) represents the approximate boundary divider between primary and secondary EM behaviour. Feature importance and SHAP analysis were conducted to investigate how the features highlighted by the

model could be directly applied to design new molecules with tailored IS values. For the binary classifier model, we have shown that a more sensitive molecule will likely have a higher oxygen balance and have a more flexible structure. This model, working with simply the molecular structure, does not account for crystal structure factors, such as polymorphism, or macroscopic factors including defects, particle size or hotspot formation. However, the insights gained from this study offers a straightforward tool with readily relatable information for chemists to design new EMs with desired IS responses.

Author contributions

HMQ collated the data, and with KM wrote and ran the ML models, with supervision from SS and CAM. All authors contributed to the central concepts of the paper. HMQ and CAM wrote the original draft, which all authors reviewed and edited.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data set is included as a spreadsheet in the supplementary information (SI) (with structures, IS values and references) and as a sheet of SMILES strings on the GitHub site. The molecular impact sensitivity database and associated ML models are available to download from GitHub at: **https://github.com/carolemorris1/ML_for_impact_sensitivity_prediction**. This content is associated with **https://doi.org/10.5281/zenodo.17157808**. Supplementary information is available. See DOI: **https://doi.org/10.1039/d5dd00357a**.

Acknowledgements

References

1 W. Zhu, in *Molecular Modeling of the Sensitivities of Energetic Materials*, ed. D. Mathieu, Elsevier, 2022, pp. 291–345.

2 M. Sućeska, *Test Methods for Explosives*, Springer New York, New York, NY, 1995.

3 Bundesanstalt für Materialforschung und -prüfung, **https://www.bam.de**, accessed 19 September 2024.

4 V. J. Bellitto and M. I. Melnik, Surface defects and their role in the shock sensitivity of cyclotrimethylene-trinitramine, *Appl. Surf. Sci.*, 2010, **256**, 3478–3481.

5 R. W. Armstrong, C. S. Coffey, V. F. DeVost and W. L. Elban, Crystal size dependence for impact initiation of cyclotrimethylenetrinitramine explosive, *J. Appl. Phys.*, 1990, **68**, 979–984.

6 F. W. Marrs, V. W. Manner, A. C. Burch, J. D. Yeager, G. W. Brown, L. M. Kay, R. T. Buckley, C. M. Anderson-Cook and M. J. Cawkwell, Sources of Variation in Drop-Weight Impact Sensitivity Testing of the Explosive Pentaerythritol Tetranitrate, *Ind. Eng. Chem. Res.*, 2021, **60**, 5024–5033.

7 N. V. Muravyev, D. B. Meerov, K. A. Monogarov, E. K. Kosareva, I. N. Melnikov, D. K. Pronkin, L. L. Fershtat, M. S. Klenov, A. V. Kormanov, I. L. Dalinger, I. V. Kuchurov, I. V. Fomenkov and A. N. Pivkina, Impact and Friction Sensitivity of Reactive Chemicals: From Reproducibility Study to Benchmark Data Set for Modeling, *Ind. Eng. Chem. Res.*, 2024, **63**, 6504–6511.

8 D. Christensen, G. P. Novik and E. Unneberg, Estimating sensitivity with the Bruceton method: Setting the record straight, *Propellants, Explos., Pyrotech.*, 2024, **49**, e202400022.

9 M. Pospíšil, P. Vávra, M. C. Concha, J. S. Murray and P. Politzer, A possible crystal volume factor in the impact sensitivities of some energetic compounds, *J Mol Model*, 2010, **16**, 895–901.

10 P. Politzer and J. S. Murray, Impact sensitivity and crystal lattice compressibility/free space, *J Mol Model*, 2014, **20**, 2223.

11 W. Zhu and H. Xiao, First-principles band gap criterion for impact sensitivity of energetic crystals: a review, *Struct. Chem.*, 2010, **21**, 657–665.

12 X. Song, X. Cheng, X. Yang, D. Li and R. Linghu, Correlation between the bond dissociation energies and impact sensitivities in nitramine and polynitro benzoate molecules with polynitro alkyl groupings, *J. Hazard. Mater.*, 2008, **150**, 317–321.

13 P. Politzer, J. Martinez, J. S. Murray, M. C. Concha and A. Toro-Labbé, An electrostatic interaction correction for improved crystal density prediction, *Mol. Phys.*, 2009, **107**, 2095–2101.

14 N. Zohari, M. H. Keshavarz and S. A. Seyedsadjadi, A link between impact sensitivity of energetic compounds and their activation energies of thermal decomposition, *J. Therm. Anal. Calorim.*, 2014, **117**, 423–432.

15 C. S. Coffey and E. T. Toton, A microscopic theory of compressive wave-induced reactions in solid explosives, *J. Chem. Phys.*, 1982, **76**, 949–954.

16 F. J. Zerilli and E. T. Toton, Shock-induced molecular excitation in solids, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1984, **29**, 5891–5902.

17 S. Califano, V. Schettino and N. Neto, *Lattice Dynamics of Molecular Crystals*, Springer Berlin Heidelberg, Berlin, Heidelberg, 1981, vol. 26.

18 D. D. Dlott and M. D. Fayer, Shocked molecular solids: Vibrational up pumping, defect hot spot formation, and the onset of chemistry, *J. Chem. Phys.*, 1990, **92**, 3798–3812.

19 S. Ye, K. Tonokura and M. Koshi, Energy transfer rates and impact sensitivities of crystalline explosives, *Combust. Flame*, 2003, **132**, 240–246.

20 J. Bernstein, Ab initio study of energy transfer rates and impact sensitivities of crystalline explosives, *J. Chem. Phys.*, 2018, **148**, 084502.

21 A. A. L. Michalchuk, P. T. Fincham, P. Portius, C. R. Pulham and C. A. Morrison, A Pathway to the Athermal Impact Initiation of Energetic Azides, *J. Phys. Chem. C*, 2018, **122**, 19395–19408.

22 A. A. L. Michalchuk, M. Trestman, S. Rudić, P. Portius, P. T. Fincham, C. R. Pulham and C. A. Morrison, Predicting the reactivity of energetic materials: an *ab initio* multi-phonon approach, *J Mater Chem A Mater*, 2019, **7**, 19539–19553.

23 A. A. L. Michalchuk, J. Hemingway and C. A. Morrison, Predicting the impact sensitivities of energetic materials through zone-center phonon up-pumping, *J. Chem. Phys.*, 2021, **154**, 064105.

24 J. M. Hemingway, H. M. Quayle, C. Byrne, C. R. Pulham, S. Mondal, A. A. L. Michalchuk and C. A. Morrison, Predicting impact sensitivities for an extended set of energetic materials via the vibrational up-pumping model: molecular-based structure-property relationships identified, *Phys. Chem. Chem. Phys.*, 2025, **27**, 11640–11648.

25 P. Nguyen, D. Loveland, J. T. Kim, P. Karande, A. M. Hiszpanski and T. Y.-J. Han, Predicting Energetics Materials' Crystalline Density from Chemical Structure by Machine Learning, *J. Chem. Inf. Model.*, 2021, **61**, 2147–2158.

26 N. Lease, L. M. Klamborowski, R. Perriot, M. J. Cawkwell and V. W. Manner, Identifying the Molecular Properties that Drive Explosive Sensitivity in a Series of Nitrate Esters, *J. Phys. Chem. Lett.*, 2022, **13**, 9422–9428.

27 J. V. Davis, F. W. Marrs, M. J. Cawkwell and V. W. Manner, Machine Learning Models for High Explosive Crystal Density and Performance, *Chem. Mater.*, 2024, **36**(22), 11109–11118.

28 W.-H. Liu, Q.-J. Liu, M. Zhong, Y.-D. Gan, F.-S. Liu, X.-H. Li and B. Tang, Predicting impact sensitivity of energetic materials: insights from energy transfer of carriers, *Acta Mater.*, 2022, **236**, 118137.

29 X. Wang, Y. He, X. Zhang, M. Hu, W. Zhao, H. Sun, X. Yang, X. Liu and R. Liu, Interpretable-machine-learning-guided discovery of dominant intrinsic factors of sensitivity of high explosives, *Mater Adv*, 2024, **5**, 3921–3928.

30 Q. Deng, J. Hu, L. Wang, Y. Liu, Y. Guo, T. Xu and X. Pu, Probing impact of molecular structure on bulk modulus and impact sensitivity of energetic materials by machine learning methods, *Chemom. Intell. Lab. Syst.*, 2021, **215**, 104331.

31 D. Mathieu, Physics-Based Modeling of Chemical Hazards in a Regulatory Framework: Comparison with Quantitative Structure–Property Relationship (QSPR) Methods for Impact Sensitivities, *Ind. Eng. Chem. Res.*, 2016, **55**, 7569–7577.

32 J. Rein, J. M. Meinhardt, J. L. Hofstra Wahlman, M. S. Sigman and S. Lin, A Physical Organic Approach towards Statistical Modeling of Tetrazole and Azide Decomposition**, *Angew. Chem., Int. Ed.*, 2023, **62**, e202218213.

33 V. Prana, G. Fayet, P. Rotureau and C. Adamo, Development of validated QSPR models for impact sensitivity of nitroaliphatic compounds, *J. Hazard. Mater.*, 2012, **235–236**, 169–177.

34 D. Mathieu and T. Alaime, Predicting Impact Sensitivities of Nitro Compounds on the Basis of a Semi-empirical Rate Constant, *J. Phys. Chem. A*, 2014, **118**, 9720–9726.

35 Z. Zhang, Y. Ma, C. Chen, S. V. Bondarchuk and Y. Liu, A General Model of Impact Sensitivity for Nitrogen-Rich Energetic Materials: A Combined Incremental Theory and Genetic Function Approximation Study, *ChemPhysChem*, 2024, **25**, e202400014.

36 S. R. Kennedy and C. R. Pulham, in *Co-crystals: Preparation, Characterization and Applications*, ed. C. B. Aakeröy and A. S. Sinha, Royal Society of Chemistry, 2018, pp. 231–266.

37 D. Mathieu, Modeling Sensitivities of Energetic Materials using the Python Language and Libraries, *Propellants, Explos., Pyrotech.*, 2020, **45**, 966–973.

38 R. Claveau, J. Glorian and D. Mathieu, Impact sensitivities of energetic materials derived from easy-to-compute *ab initio* rate constants, *Phys. Chem. Chem. Phys.*, 2023, **25**, 10550–10560.

39 F. W. Marrs, J. V. Davis, A. C. Burch, G. W. Brown, N. Lease, P. L. Huestis, M. J. Cawkwell and V. W. Manner, Chemical Descriptors for a Large-Scale Study on Drop-Weight Impact Sensitivity of High Explosives, *J. Chem. Inf. Model.*, 2023, **63**, 753–769.

40 J. C. Duarte, R. D. da Rocha and I. Borges, Which molecular properties determine the impact sensitivity of an explosive? A machine learning quantitative investigation of nitroaromatic explosives, *Phys. Chem. Chem. Phys.*, 2023, **25**, 6877–6890.

41 Q. Wu, X. Wang, B. Yan, S. Luo, X. Zheng, L. Tan and W. Zhu, Prediction of impact sensitivity and electrostatic spark sensitivity for energetic compounds by machine learning and density functional theory, *J. Mater. Sci.*, 2024, **59**, 8894–8910.

42 D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.

43 S. V. Bondarchuk, Z. Zhang, C. Chen, L. Wen, J. Zhang and Y. Liu, Grammar of Impact Sensitivity: An Incremental Theory, *J. Phys. Chem. A*, 2023, **127**, 10506–10516.

44 J. V. Davis, F. W. Marrs, A. C. Burch, N. Lease, M. Cawkwell and V. W. Manner, in *Proceedings of the Conference of the American Physical Society Topical Group on Shock Compression of Condensed Matter*, 2023, p. 280001.

45 D. Mathieu and T. Alaime, Impact sensitivities of energetic materials: Exploring the limitations of a model based only on structural formulas, *J Mol. Graph. Model.*, 2015, **62**, 81–86.

46 M. J. Kamlet, in *Proceedings of the 6th Symposium on Detonation, ONR Report ACR*, vol. 221, 1976, p. 312.

47 M. J. Kamlet and H. G. Adolph, The relationship of Impact Sensitivity with Structure of Organic High Explosives. II. Polynitroaromatic explosives, *Propellants, Explos., Pyrotech.*, 1979, **4**, 30–34.

48 M. Keshavarz and H. Pouretedal, Simple empirical method for prediction of impact sensitivity of selected class of explosives, *J. Hazard. Mater.*, 2005, **124**, 27–33.

49 R. J. Appleton, D. Klinger, B. H. Lee, M. Taylor, S. Kim, S. Blankenship, B. C. Barnes, S. F. Son and A. Strachan, Multi-Task Multi-Fidelity Learning of Properties for Energetic Materials, *Propellants, Explos., Pyrotech.*, 2025, **50**, e202400248.

50 J. L. Lansford, B. C. Barnes, B. M. Rice and K. F. Jensen, Building Chemical Property Models for Energetic Materials from Small Datasets Using a Transfer Learning Approach, *J. Chem. Inf. Model.*, 2022, **62**, 5397–5410.

51 C. B. Storm, J. R. Stine and J. F. Kramer, in *Chemistry and Physics of Energetic Materials*, ed. S. N. Bulusu, Springer Netherlands, Dordrecht, 1990, pp. 605–639.

52 N. V. Muravyev, D. R. Wozniak and D. G. Piercey, Progress and performance of energetic materials: open dataset, tool, and implications for synthesis, *J Mater Chem A Mater*, 2022, **10**, 11054–11073.

53 T. M. Klapötke, in *Chemistry of High-Energy Materials*, De Gruyter, 2017, pp. 127–160.

54 L. B. Kier, An Index of Molecular Flexibility from Kappa Shape Attributes, *Quant. Struct.-Act. Relat.*, 1989, **8**, 221–224.

55 M. H. Keshavarz, Prediction of impact sensitivity of nitroaliphatic, nitroaliphatic containing other functional groups and nitrate explosives, *J. Hazard. Mater.*, 2007, **148**, 648–652.

56 M. H. Keshavarz, A New General Correlation for Predicting Impact Sensitivity of Energetic Compounds, *Propellants, Explos., Pyrotech.*, 2013, **38**, 754–760.

57 P. Ertl, B. Rohde and P. Selzer, Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties, *J. Med. Chem.*, 2000, **43**, 3714–3717.

58 L. B. Kier and L. H. Hall, An Electrotopological-State Index for Atoms in Molecules, *Pharm. Res.*, 1990, **07**, 801–807.

59 S. A. Wildman and G. M. Crippen, Prediction of Physicochemical Parameters by Atomic Contributions, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 868–873.

60 J. S. Murray, P. Lane and P. Politzer, Relationships between impact sensitivities and molecular surface electrostatic potentials of nitroaromatic and nitroheterocyclic molecules, *Mol. Phys.*, 1995, **85**, 1–8.

61 B. M. Rice and J. J. Hare, A Quantum Mechanical Investigation of the Relation between Impact Sensitivity and the Charge Distribution in Energetic Molecules, *J. Phys. Chem. A*, 2002, **106**, 1770–1783.

62 Python3, https://www.python.org, accessed 13 January 2024.

63 RDKit: Open-source cheminformatics, https://www.rdkit.org, accessed 19 November 2023.

64 SMARTS, https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html, accessed 12 November 2023.

65 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, Scikit-learn: Machine Learning in Python, *Journal of Machine Learning Research*, 2011, **12**, 2825–2830.

66 LightGBM, https://github.com/Microsoft/LightGBM, accessed 13 June 2024.

67 P. Refaeilzadeh, L. Tang and H. Liu, in *Encyclopedia of Database Systems*, ed. M. T. Özsu and L. Liu, Springer US, Boston, MA, 2009, pp. 532–538.

68 S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal and S.-I. Lee, From local explanations to global understanding with explainable AI for trees, *Nat. Mach. Intell.*, 2020, **2**, 56–67.

69 S. M. Lundberg and S.-I. Lee, in *Advances in Neural Information Processing Systems*, ed. I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett, 2017, vol. 30.

70 D. Mathieu, Sensitivity of Energetic Materials: Theoretical Relationships to Detonation Performance and Molecular Structure, *Ind. Eng. Chem. Res.*, 2017, **56**, 8191–8201.