Which molecular properties determine the impact sensitivity of an explosive? A machine learning quantitative investigation of nitroaromatic explosives†

Julio Cesar Duarte,[a,b] Romulo Dias da Rocha,[b] Itamar Borges Jr., [a,c,]*

We decomposed density functional theory charge densities of 53 nitroaromatic molecules into atom-centered electric multipoles using the distributed multipole analysis that provides a detailed picture of the molecular electronic structure. Three electric multipoles, $\sum Q_0(NO_2)$ (the charge of the nitro groups), $\sum Q_1(NO_2)$ (the total dipole, i.e., polarization, of the nitro groups), $\sum Q_2(C)$ (the total electron delocalization of the C ring atoms), and the number of explosophore groups ($\#NO_2$) were selected as features for a comprehensive machine learning (ML) investigation. The target property was the impact sensitivity h_{50} (cm) values quantified by drop-weight measurements. After a preliminary screening of 42 ML algorithms, four were selected based on the lowest root mean square errors: Extra Trees, Random Forests, Gradient Boosting, and AdaBoost. The predicted h_{50} values of molecules having very different sensitivities for the four algorithms are in the range 19% - 28% compared to experimental data. The most important properties for predicting h_{50} are the electron delocalization in the ring atoms and the polarization of the nitro groups with averaged weights of 39% and 35%, followed by the charge (16%) and number (10%) of nitro groups. A significant result is how the contribution of these properties to h_{50} depends on its sensitivities: for the most sensitive explosives (h_{50} up to $\sim$ 50 cm), the four properties contribute to reducing h_{50}, and for intermediate ones ($\sim$ 50 cm $\lesssim$ h_{50} $\lesssim$ 100 cm) $\#NO_2$ and $\sum Q_1(NO_2)$ contribute to increasing it and the other two properties to reducing it. For highly insensitive explosives ($h_{50} \gtrsim$ 200 cm), all four properties essentially contribute to increasing it. These results furnish a consistent molecular basis of the sensitivities of known explosives that also can be used for developing safer new ones.

Keywords:

Impact sensitivities of explosives; Machine Learning; Molecular charge (electronic) densities; Drop weight test; Distributed Multipole Analysis (DMA); Extra Trees; Random Forests; Gradient Boosting; AdaBoost

---

[a] *Departamento de Engenharia de Computação, Instituto Militar de Engenharia.*
[b] *Programa de Pós-Graduação em Engenharia de Defesa, Militar de Engenharia.*
[c] *Departamento de Química, Militar de Engenharia.*
Instituto Militar de Engenharia, Praça General Tibúrcio, 80, Urca, Rio de Janeiro (RJ), 22290-270, Brazil: E-mail: itamar@ime.eb.br
† Electronic supplementary information (ESI) available.

1. Introduction

The impact sensitivity of explosives is a critical issue because large amounts of energy can be released quickly due to mechanical impacts. Shock, thermal or electric stimuli also produce the same effect.[1-3] These materials are controllable storage systems of chemical energy that are frequently stored, transported, and broadly used. Therefore, a comprehensive understanding of the sensitivity of an explosive is paramount, for example, for reducing its sensitivity, thus increasing safety.[4-8]

Energetic materials are usually molecular solids made of polyatomic molecules arranged in complicated crystal structures.[9-12] This class of materials, which also includes propellants and pyrotechnics, is most valuable when combining low sensitivity (i.e., they are safer) and high energy content.[4] However, good performance of an energetic material usually corresponds to increased sensitivity.[13, 14] These two apparently incompatible requirements demand a more significant fundamental understanding of the molecular origins of the sensitivity of energetic materials either known or to be developed.

The aforementioned different types of stimuli are helpful (and have practical implications) in quantifying the sensitivity of an energetic material.[8, 15] The impact stimulus, the focus of this work, is frequently quantified by the h_{50} value corresponding to the height h in cm from where a standard weight dropped over a certain amount of explosive will start an explosion 50% of the drops.[16, 17] Therefore, insensitive materials have large h_{50} and vice-versa.

Different types of uncertainties involved in drop-weight measurements of h_{50} include atmospheric conditions, particle size distribution and shape, presence of crystalline defects, humidity, and even the operator technique.[18] Furthermore, the impact sensitivity depends not only on these macroscopic properties and experimental conditions but also on microscopic properties (e.g., bond-breaking activation energies and energy per molecule),[19, 20] mesoscale properties (e.g., strain[21, 22] and phonon energy transfer mechanics[23, 24]). Despite these issues, h_{50} values are very helpful in quantifying the impact sensitivity. Additionally, this plethora of possible influences shows that the prediction of impact sensitivity based on physical properties "remains a challenge".[25]

Considering that the stability of a molecule of an energetic material is related to its sensitivity, the latter stems from the chemical character of the materials,[13] which can be accurately investigated by quantum chemical methods.[26-28] Furthermore, this approach is particularly convenient because the macroscopic property impact sensitivity "is not known to depend directly on intermolecular interactions in bulk",[29] typical hazards of the experimental work are not involved especially on potential new energetic materials,[29] not to say that "imprecise impact sensitivity measurements can provide a false sense of security that might result in fatalities when handling even small quantities of the material".[18, 30] Quantum chemical methods also allows safer screening of a more significant number of potentially greener energetic materials to reduce environmental contamination.[31]

Attempts to correlate molecular properties and impact sensitivities have a long history, beginning in the 1970s. These approaches developed models that included the proportion of oxygen atoms in the explosive molecules,[17] molecular structural parameters such as the number of carbon atoms and molecular masses,[32-34] quantitative structure-property relationships (QSPR) involving structural (e.g., weaker molecular bonds) and energetic (e.g., dissociation energies), among many other properties, known as descriptors in QSPR jargon.[19, 25, 35, 36] Quantum chemical modeling of impact sensitivities, in particular, has quite been quite successful in this regard, especially for a given class category of materials such as nitroaromatics, nitramines, among others. Representative works of this approach include correlating impact sensitivities with molecular properties such as bond strengths and molecular electrostatic potentials,[29, 37-39] whereas others employed Mulliken charge values of nitro groups[15, 40] and examined binding forces using the Wiberg bonding index.[41, 42]

Our own quantum chemical work on the molecular origins of the impact sensitivity has been based on the decomposition of a quantum chemical molecular charge (electronic) density (for instance, computed using density functional theory, DFT or an ab initio method) into atom-centered electric multipoles determined by the distributed multipole analysis (DMA) method.[43-46] This approach provides a detailed and accurate picture of the molecular charge density, which we have been used to investigate the impact sensitivity of different families of explosive molecules[47-53]

and different phenomena related to catalysis.[54-57] We recently reviewed this work on the prediction of impact sensitivities and others based on different theoretical descriptions of molecular charge distributions.[58]

The variety and diverse nature of the properties related to sensitivity indicate that those relationships cannot all represent the fundamental causes of sensitivity because "many of them can be symptomatic".[14] Therefore, those properties correspond to general trends, not correlations,[59, 60] providing valuable estimates of the magnitude of the sensitivities.[61] In this work, we search for these predictive trends by identifying and quantifying molecular properties that affect the sensitivity to impact.[29]

The application of artificial intelligence (AI) and machine learning (ML) techniques to chemistry and materials science, including the prediction of properties of energetic materials, is growing tremendously.[62-67] In quantum chemistry applications, a given training data ML is now a "game changer" for inferring complicated and unknown nonlinear dependencies of molecular properties from fitting functions and making accurate predictions;[68-70] In particular, ML can be used to learn features (properties) of materials properties.[71-73]

Concerning the application of ML techniques to sensitivity, and in particular, to impact sensitivity, there have been sparse works in the last two decades. Most of them used a large number of molecular structure descriptors to fit experimental h_{50} values employing a classical QSPR analysis.[74] Most of these investigations employed QSPR packages to generate and select sometimes over 1000 molecular descriptors[75] that are afterward reduced by statistical techniques.[76] The pioneering work of Nefati et al. employed neural networks to predict h_{50} of 204 molecules of different families selected from originally 39 descriptors (27 purely topological and 12 from semiempirical electronic structure calculations) by a cross-validated (leave one out) standard error criterion,[77] producing much better results compared to those from a linear regression. Choa and co-workers optimized neural networks based on this work and employed over 200 molecules of the same dataset, finding that compositional and topological descriptors gave better results compared to those composed of electronic descriptors, which may be due to their use of the not very accurate semiempirical AM1 method.[78] Keshavarz and Jaafari also employed neural

networks, with 275 molecules comprising the training set and 14 explosive molecules the test set.[79] The descriptors were primarily structural, namely, the number of $N-NO_2$ bonds, the number of α −hydrogens, aromatic and heteroaromatic characters, and the number of each type of atom divided by the molecular weight. The predicted h_{50} values were superior to the five quantum chemical models of Rice and Hare based on generalized interaction property functions (GIPFs) computed from molecular electrostatic potentials.[29] Wang et al. employed neural networks and atom- and group-type electrotopological-state indices (ETSI),[80] which combine the electronic and topological properties of a molecule and take the binding environment into account, to predict impact sensitivities of 156 compounds (training set with 127 molecules and test set with 29) of different chemical families (49 nitroaromatic, 55 nitramine, 40 nitroaliphatic compounds containing other functional groups, 7 nitrate esters, and 5 nitroaliphatics).[81] Their results with neural networks had good agreement with experimental h_{50} values and were superior compared to multi-linear regression and partial least squares modeling.

In another work, Wang and coworkers employed a genetic algorithm for selecting descriptors for the QSPR analysis of 186 explosive molecules using neural networks.[82] Nine topological and quantum descriptors were selected from an initial pool of over 500. Prana et al. investigated the impact sensitivity of 50 nitroaliphatic compounds using about 400 descriptors including quantum chemical ones based on DFT calculations.[83] Linear and multi-linear regressions including internal and external validations provided accurate relationships with experimental h_{50} values. One interesting mechanistic finding is that several descriptors, especially the quantum mechanical ones, were related to the explosophore (functional group that makes a compound explosive) NO_2 group involved in the decomposition of nitro compounds. Out of the four models, the most accurate also involved quantum chemical descriptors. Xu et al. applied ten three-dimensional (3D) descriptors that gather different information from the spatial structure of the molecule and neural networks[74] on 156 molecules, obtaining better results compared to linear regression approaches. Deng and collaborators used sure independence screening and sparsifying operator (SISSO) methods to investigate the correlation between impact sensitivity and bulk modulus B (a mechanical property that quantifies resistance of a material to compression) of 240 nitroaromatics, using 14 molecular structure descriptors and 7 structural parameters (including oxygen balance, crystal density and hydrophilicity), with results of similar accuracy when

compared to neural networks.[76] They found that the oxygen-containing group, hydrophilicity, and some atomic properties are the main contributions to impact sensitivity. A particularly interesting result is that the impact sensitivity sharply decreases with increasing B when B is small, while the opposite trend occurs for materials with relatively large bulk modulus B.

Lansford and coworkers used Chemprop's[84] purely graph-based directed-message passage neural networks (D-MPNNs)[85] to apply a transfer learning approach simultaneously for training a multi-target regression model on a small number of molecules with experimentally measured h_{50} values and related computed properties.[30] They iteratively constructed an initial dataset of 172 million likely energetic molecules from the PNNL (Pacific Northwest National Laboratory) database[86] employing the ML ANI-1ccx force field[87] (in some instances, DFT) to compute different energetic and geometric properties. Direct models using Chemprop's bond features such as aromaticity, atomic mass and bond type were trained only on impact sensitivities, while other models were co-trained with DFT or other computed properties for small datasets. The co-training technique applied to experimental sensitivities improved the model performance of the physics-based Mathieu's modeling.[61] Using an ensemble of models and 14 molecules as test set, they predicted the experimental h_{50} values of PETN and CL-20, and slightly larger h_{50} values of RDX, HMX, and TNAZ within the modeling error. The ordering of the predicted impact sensitivity follows the measured values for HNS, PYX, TNT, and TATB, while the insensitive nature of DNAN and NQ that have inaccurate experimental h_{50} values were found.

A related work was carried out by Wen at al. where they used estimated h_{50} values computed with Mathieu's approach[19, 61, 88] and three more criteria (oxygen fluorine balance, detonation velocity and synthesis difficulty) to screen about 10^5 $CHONF$-containing molecules by combinatorial library design based on fragment-types. The top 10 promising molecules were further examined using DFT/B3LYP resulting in two of them having a large detonation velocity and comparable sensitivity compared with the widely used CL-20 explosive.

The neural networks employed in those works have a powerful learning capacity but display a limited model explainability for identifying the features (in our case, the molecular properties) that affect the target property (here, the h_{50} value).[89] In most cases, the investigations

employ such a large number of descriptors that it becomes almost impossible to identify the real molecular origin of the impact sensitivity. Moreover, while some descriptors have a direct chemical interpretation (e.g., an aromaticity index), others such as WHIM (Weighted Holistic Invariant Molecular), which is based on a principal component analysis of the weighted covariance matrix from atomic cartesian coordinates,[90] though helpful, are not straightforward to interpret. An additional issue concerning the number of descriptors is that the results produced by adding more of them may lead to overfitting and chance correlations.[74]

In an important work of ML applied to energetic materials, different machine learning models and several featurization methods were compared using a small training data set of 87 diverse $CNOHF$ energetic molecules, and a test data of 22 molecules, predicted different explosive properties with small errors, such as detonation pressure, detonation velocity, explosive energy, heat of formation, density, among others.[91] This work is also significant for the present purposes because it showed that a small number of descriptors – in some cases, only one (e.g., oxygen balance) can provide good results for the target feature. Their main conclusion is that feature selection is more important than model selection for making property predictions on a small data set, a conclusion also confirmed by our results.

In this work, we combine high-level quantum chemical data – in our case, DMA site electric multipoles that accurately describe the DFT molecular charge density of molecules, with ML algorithms to identify molecular trends affecting the impact sensitivity of explosives. We investigate the broadly used class of nitroaromatics explosives that usually combine low sensitivity and high performance, hence, for this reason, are the most used type in warlike devices.[92] The nitroaromatic molecules include nitrobenzenes, nibrobenzyls, nitroanilines, and nitrophenols, among other types.

2. Methods

Our original set of 50 nitroaromatic molecules[52] was extended to 53 with the inclusion of 2,4-DNP, TATB and ATNAN – see Scheme 1S of the Electronic Supplementary Information (ESI) for their molecular structures, and Table 1S for the meaning of the acronyms. In all cases, we employed the previously M06-2X/TZVP optimized geometries[41] to compute single point

B3LYP/6-311+G(d) electron densities, which afterward were divided into atom-centered DMA electric multipoles. The properties of each molecule comprised three derived DMA properties described below and the number of nitro (NO_2) groups. The experimental h_{50} values were taken from reference 41.

The DMA of Stone decomposes a molecular charge density into atom-centered electric multipoles.[43-46] It is a rapidly convergent expansion of the charge density that describes it accurately and in great detail. The expansion here employed was truncated at the electric quadrupole term, and its three first terms have a straight chemical interpretation. The monopole (in units of the fundamental charge $e = 1{,}602 \times 10^{-19}$ C) represents a localized charge in an atom, with different atoms in a bond displaying some degree of charge separation. Dipole atom-centered moment vectors (in atomic units of ea_0, where a_0 is the Bohr radius) are represented by a vector pointing from a negative charge to a positive one of the same magnitude that describes an atom-centered charge displacement. Depending on the electronegativities of the involved atoms and the remaining bond environment, bond densities can produce significant atom-centered dipoles. An isolated atom has a perfectly spherical electron cloud, hence no dipole. Finally, the last term of the DMA atom-centered expansion is the electric quadrupole (in units of ea_0^2), which is the first electrostatic moment to include contributions from out-of-plane density. For this reason, it is associated with delocalized π electrons and lone pairs of electrons.[44] Given that the quadrupole moment is a tensor, it is represented here by a number that corresponds to the square root of the sum of all components of the tensor squared.

We illustrate in Fig. 1 the power of the DMA atom-centered decomposition to rationalize the molecular charge density by describing the molecular charge density of DATB (2,4-diamino-1,3,5-trinitrobenzene), one of the investigated molecules. The charges and the quadrupoles of the carbon ring atoms are clearly affected by the bonded group: for the strong electron-withdrawing nitro groups, their charges are slightly negative, while the quadrupoles, indicators of electron delocalization, have $\sim 30\%$ larger values compared to the carbons bonded to the electron donor amines (NH_2) due a resonance effect with NO_2; the charges on the carbons bonded to the amines are positive in contrast. Depending on the nature of the moieties bonded to the other ring carbon atoms, these effects have different magnitudes on the pi-electron structure[93] that are clearly

quantified by the DMA. Atom-centered polarization, given by the dipoles, is only appreciable in the highly electronegative oxygen atoms that have large negative charges and different quadrupoles depending on the type of neighbor hydrogen, either isolated or in an amine group. The nitrogen atoms in the amine groups have more than twice the quadrupole value of the nitrogen atoms in the nitro groups due to the lone pairs of the former; their charges are negative in the amine group while they are positive in the nitro groups. Hence, a very detailed picture of the molecular is furnished by the DMA.

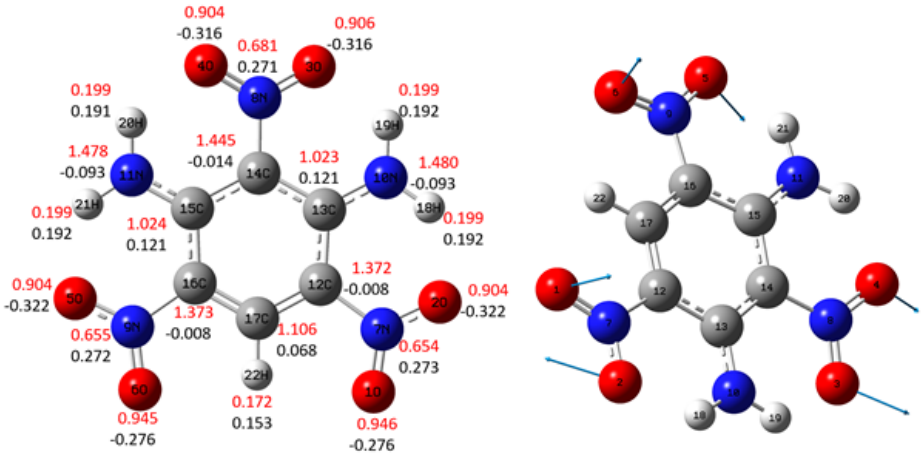

Fig. 1 DMA picture of the DATB (2,4-diamino-1,3,5-trinitrobenzene) molecule. Atomic dipoles are represented by vectors drawn at the corresponding nuclei. Monopole (charge) numbers (black) in units of elementary charge e ($1{,}602 \times 10^{-19}$ C), and quadrupole values (red) in units of ea_0^2 ($= 4.486 \times 10^{-40}$ Cm^2).

The three molecular properties based on DMA used as features in the ML algorithms are described at Table 1. These properties represent what has been considered as relevant molecular properties affecting the impact sensitivity of nitroaromatic properties.[58]

Table 1 The three DMA molecular properties used as features in the machine learning simulations

DMA formula	Property
$\sum Q_0(NO_2)$	Total charge of the nitro(s) group(s)
$\sum Q_1(NO_2)$	Magnitude of the total dipole of the nitro(s) group(s)
$\sum Q_2(C)$	Magnitude of the quadrupole moment of the carbon atoms

2.1 The machine learning algorithms

We now briefly discuss the four selected algorithms. In the following subsection, we present how they were selected and implemented. The input data – the four molecular properties and the target-feature (h_{50}) are shown in Table 2S of the Supporting Information.

The root mean square error (RMSE) value is very used in ML to compare and evaluate regressors (i.e., to compute their errors). Additionally, very low RMSE values on the training set may indicate the overfitting of the model. The RMSE metric is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - \hat{y})^2} \qquad (1)$$

where n is the total number of samples, y is the observed (target) value and $\hat{y}$ is the value estimated by the algorithm (model). In our case, y is the experimental impact sensitivity value h_{50} and $\hat{y}$ is the predicted value (target feature) predicted by each model. As it is well-known, the square of the difference $(y - \hat{y})^2$ penalizes the most significant deviations, especially outliers. The square of the RMSE is the mean square error (MSE) or variance. The mean absolute error (MAE), another measure of the error, is defined in the usual way as $1/n \sum_{i=1}^{n} |y - \hat{y}|$.

Four algorithms out of 42 tested in a preliminary screening were selected (see next section for the selection process). Two algorithms (Extra Trees and Random Forests) are based on Regression Trees, which are a Decision Trees Regressor[94] used for regression tasks with the purpose of predicting output values. The other two, Gradient Boosting[95] and AdaBoost (Adaptive Boosting),[96, 97] are based on the concept of boosting, which is an ensemble method combining several weak learners (base methods) into a strong learner. Boosting methods train predictors (predicted target features) sequentially, each trying to correct its predecessors.

Extra Trees. When growing a tree in the Random Forest algorithm, at each node, only a subset of molecular properties (features or attributes in ML jargon) is considered for the subsequent tree

splitting into other nodes (see Fig. 2). The trees can be made even more random by using train thresholds for each feature instead of searching for the best possible threshold, similarly to the process employed by a regular Decision Tree. These extremely random trees are called an extremely randomized ensemble, known for short as Extra-Trees.[98] The algorithm "trades more bias for a lower-variance",[94] and, in this way, the Extra-Trees is much faster to train compared with the Random Forest because the most consuming task of growing a tree is to find the best possible threshold for each feature at every node, which is not an issue here given the present small size of the training set compared with more usual applications. In the present work, 125 trees were generated employing the *GridSearch* tool and used for prediction.

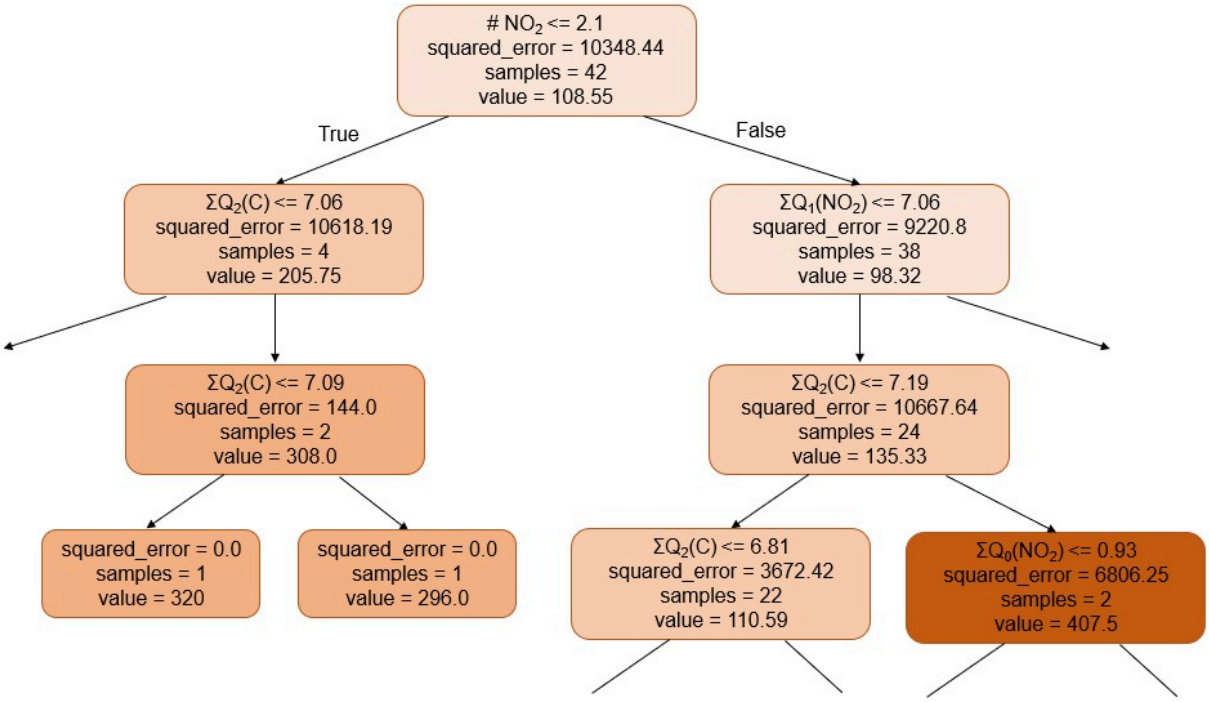

Fig. 2 An excerpt of a regression tree built by the Extra-Trees algorithm. The colors are related to the predicted h_{50} values - the darker colors correspond to larger values and the lighter ones to the smaller values.

Fig. 2 illustrates the growing process of a representative tree in the Extra-Trees algorithm for our problem. The first (root) node starts with a training group of 42 molecules (*samples = 42*), creating a pre-established rule for growing the tree. In this tree, the rule is the number of nitro groups bonded to the ring ($\#\ NO_2$) less than 2.1. In each ramification of the tree, these values are

modified, the MSE is computed, the number of samples (molecules) is reduced, and at the end, a value of h_{50} is computed for a given molecule. The algorithm selects the cuff-off value leading to the lowest MSE. Since Extra Trees is an ensemble algorithm, several trees are created, thus increasing its prediction ability. The final result is the average of the results computed by the 125 trees, although other criteria for selection (voting in ML jargon) can be used (e.g., median or mode).

From the initial node, two more nodes are produced. If $\#\, NO_2 \leq 2.1$, the left branch is formed with 4 new molecules (samples) in this node. If this inequality (i.e., the rule) is false, the right branch is formed with the remaining 38 samples of the training set. Subsequent nodes are formed, resulting in lower MSE values. When a terminal node is reached, the predicted h_{50}value of a molecule given by the algorithm will be the average of the values from each of the 125 trees.

Random Forests. Similarly to Extra Trees, Random Forests[99] is an ensemble algorithm with a significant difference: for determining the initial node, its base algorithm (Regression Tree) selects the best property for further ramifications. Whereas Random Forests in the Scikit-Learn[100] implementation here used employed bootstrap replicas, i.e., it subsamples the input data, Extra Trees uses the whole sample, which may create more variance. Furthermore, Random Forests choose the optimum split, whereas Extra Trees select it randomly. However, after selecting the split points, the two algorithms pick up the best one between all subsets of features. Therefore, Extra Trees adds randomization but still has optimization.

Gradient Boosting. This is a boosting algorithm of an ensemble method combining several weak learners into a strong learner training predictor sequentially.[95] Gradient boosting builds many small models (e.g., decision trees) to sequentially reduce the residual errors from the previous steps by employing a gradient descent optimization algorithm.[95]

AdaBoost (Adaptative Boosting) Regressor. Similarly to Gradient Boosting, AdaBoost is a boosting method because it combines sequentially weak models[96, 97] However, in contrast to Gradient Boosting, it adjusts the instance weights at every interaction. In other words, the predictors are trained sequentially, each one trying to correct the predecessors' output. In this kind

of algorithm, a component of randomness is removed compared to other techniques, considering that it uses the previous predictor to calibrate the next one.

2.2 The machine learning computational approach

The Scikit-Learn Python library[100] was used for the machine learning experiments run in the Google Colab environment.[101] The high-quality quantum-chemical data combined with a piece of structural information (the number of the explosophore nitro groups) comprised the four properties (features) of each molecule.

The dataset was divided into training and testing data using the function *train_test_split* and the *test_size* parameter set as *0.2*. This specifies that 20% of the original set will comprise the test set. Therefore, 42 molecules made up the training set and 11 the test set. The selection of the molecules in the training and test sets was made randomly but controlled by a predetermined seed. In this way, each set in each algorithm has the same molecules when the separation is done. The target feature of all ML simulations was a predicted impact sensitivity h_{50} value for each molecule.

The Lazy Predict tool of Scikit-Learn was employed to carry out a preliminary test of 42 different algorithms using for the training step their default hyperparameter set. From those 42 ML algorithms, the tool selected the best four algorithms according to the lowest RMSE values defined as by Eq. (1). The RMSE of the 42 algorithms are collected in Table 3S of the Supporting Information.

As mentioned in the previous section, the following algorithms were selected by Lazy Predict: Random Forest,[99] Extra Trees,[98] Gradient Boosting,[95] and AdaBoost.[96, 97] The XGBoost algorithm had similar RMSE values but was discarded because it is very similar to Gradient Boosting and more prone to overfitting. To obtain the best hyperparameter set for each of the four selected algorithms, a grid search technique that tests all the possible combinations of predetermined discrete parameter ranges was employed.[102] For this purpose, we used the GridSearchCV class that produces the optimum hyperparameter set for each algorithm, resulting in a more calibrated estimator for the problem.[100] Once the optimum set of hyperparameters of

each algorithm was found, the models were trained using a k-fold cross-validation[103] technique with 5 folds. Hence, the training set was randomly divided into 5 subsets. The remaining 11 (20%) molecules comprised the test set, as already mentioned. This number of folds was used due to the small size of the dataset.

Afterward, the relative importance of the molecular properties (features) - the three DMA multipole values and the number of nitro groups, most relevant for predicting the target feature h_{50} in each case, according to each algorithm, was investigated. For this step, the *feature_importances* tool was used to estimate the importance of each property for a given algorithm during its training step.[104]

For the four algorithms, SHAP (Shapley Additive exPlanations) plots[105] were built to display how much each feature (independent variable used in the regression analysis) – in our case, a molecular property, contributed either positively or negatively to the predicted h_{50} value with respect to the average value of the set. This is a game theoretical approach developed to rationalize the output of any ML model, and it was used recently in a neural network QSPR work on impact sensitivity.[76]

Those plots can be drawn for a given molecule or collectively for the complete test set. Features that contribute to increasing h_{50} values are shown in red, whereas those that contribute to decreasing the predictions are in blue. Therefore, SHAP values allow us to identify relevant patterns in the data.

3. Results and discussion

We previously developed quadratic models for predicting h_{50} values employing the same features used here, namely, three different combinations of atom-centered DMA multipole values and the number of nitro groups.[48, 52] The most extensive set comprised 50 of the present 53 molecules. Simple functions combining 2, 3, or 4 of the present molecular properties were built. Five models were then constructed. These functions, called Γ, were fitted to quadratic expressions of the type $h_{50} = \alpha\Gamma^2 + \beta\Gamma + \gamma$, where h_{50} are the experimental values of the impact sensitivity.

In that work, it was found that the quadrupole of the ring atoms, $\sum Q_2(C)$, a measure of electronic delocalization (aromaticity) of the ring atoms affected by substituent effects, and the total charge of the nitro groups, $\sum Q_0(NO_2)$, a measure of charge localization on the explosophore moiety, were the most important features affecting the prediction of h_{50}. This conclusion was based on the good correlation (adjusted R^2 values greater than 0.90) of the models. However, the contribution weight of each property to the prediction of h_{50} could not be quantified in that work. In contrast, we report here the weight of each one of the features involved, improved predicted h_{50} values and investigate in detail the molecular properties that affect them.

The averaged contribution weight of each feature to a predicted h_{50} value computed by the four algorithms are collected in Table 2 for the test set (11 molecules) and Table 3 for the training set (42 molecules). Although we report the averaged contribution of the values computed by each algorithm, it is important to note that overall, for the test set, the percentage contribution of each feature for a given algorithm is similar. The ordering, thus the importance, of the features is the same, except for a slight inversion in the AdaBoost ordering of the $\sum Q_2(C)$ (37% weight) and $\sum Q_1(NO_2)$ (38%) properties. For the training group, the electron delocalization of the ring atoms ($\sum Q_2(C)$) and the NO_2 total dipole ($\sum Q_1(NO_2)$) are still by far the most dominant properties, but in contrast with the test group, it is followed by the number of the nitro groups ($\#NO_2$) as the third most important and their charges, $\sum Q_0(NO_2)$, as the forth.

Although the four properties contribute appreciably to the predicted h_{50}, two of them are by far the most important, with similar weights both for the training and test groups. The importance of electron delocalization in the aromatic ring for predicting h_{50} was found before by us[48-52] and others.[29, 39, 106] However, the importance of the magnitude of the total dipole moment of the explosophore nitro(s) group(s) was not identified before by us or the literature – its contribution in the test group (35%) is similar to the electron delocalization (39%). This is chemically justified because the dipole moment measures the polarization of the explosophore NO_2 that affects the ring electrons and, accordingly, its $C-NO_2$ bond strengths, thus characteristics of electronic structure of the molecule that are relevant for the sensitivity. We found that the charge of the nitro group contributes far less (16%), although, in other works, they were treated as the sole molecular contributor to h_{50}.[15, 107]

The test group included 11 molecules with a wide range of sensitivity (h_{50}) values, varying from very sensitive (PTETNT, $h_{50} = 25$ cm) to highly insensitive (DATB, $h_{50} = 320$ cm). Table 4 displays the errors of the predicted h_{50} values for each molecule of the test set. The average errors in the predicted values for each algorithm (shown in the last row) are in the range 19% - 28%, which can be considered quite good concerning the unknown uncertainties in the measured h_{50} values and the variety of type of molecules and sensitivities in the test group. In the test group, the AdaBoost algorithm not only has the lowest average error (19%) but presents six molecules possessing very different experimental h_{50} values in the range 25 cm^{-1} – 320 cm^{-1} with errors of the predicted h_{50} values lower than 10%. GradientBoost also shows similarly consistent values, with eight molecules having errors less than 25%. The two algorithms based on Decision Trees, Random Forest, and Extra Trees, have six molecules with errors less than about 20%.

Table 2 The averaged contribution weight of each feature for predicting the impact sensitivity h_{50} values using the four algorithms for the molecules of the **test set**. The average value in the last column is the average of the percent values computed by each algorithm.

Property	TEST GROUP				
	Extra Trees (%)	Random Forest (%)	Gradient Boost (%)	AdaBoost (%)	Average (%)
$\sum Q_2(C)$	37	41	40	37	39
$\sum Q_1(NO_2)$	24	40	36	38	35
$\sum Q_0(NO_2)$	21	13	14	19	16
$\#NO_2$	18	6	10	6	10

Table 3 The averaged contribution weight of each feature for predicting the impact sensitivity h_{50} values using the four algorithms for the molecules of the **training set**. The average value in the last column is the average of the percent values computed by each algorithm.

Property	TRAINING GROUP				
	Extra Trees (%)	Random Forest (%)	Gradient Boost (%)	AdaBoost (%)	Average (%)
$\sum Q_2(C)$	36	40	35	47	40
$\sum Q_1(NO_2)$	21	40	28	35	31
$\#NO_2$	29	13	20	8	17
$\sum Q_0(NO_2)$	14	7	17	10	12

Table 4 Computed errors of each algorithm for the molecules of the test group.

MOLECULE	h_{50} **(cm)**	**AdaBoost (%)**	**Random Forest (%)**	**Gradient Boost (%)**	**Extra Trees (%)**
DATB	320	1.6%	4.1%	0.3%	12.5%
DATNP	112	0.0%	22.3%	8.9%	34.8%
DCLTNAN	75	41.3%	58.7%	65.3%	78.7%
DNAN	220	38.0%	20.5%	13.2%	13.2%
PATETNT	47	6.4%	12.8%	46.8%	19.1%
PTETNT	25	4.0%	24.0%	20.0%	20.0%
TETNB	28	21.4%	39.3%	21.4%	39.3%
TMTNB	110	32.7%	57.3%	53.6%	31.8%
TNBEtOH	68	45.6%	17.6%	0.0%	32.4%
TNT	98	8.2%	1.0%	14.3%	7.1%
DNB	100	10.0%	6.7%	8.7%	14.7%
AVERAGE		**19.0%**	**25.8%**	**23.0%**	**27.6%**

The RMSE values of the test set for each algorithm and the previous results[52] of the same molecules are presented in Fig. 3. In the previous work, the 21 molecules of the training group were selected according to the diversity of structures, groups, and bond types. The remaining 29 comprised the test set, with four molecules (DATB, DCLTNAN, PTETNT, and TNT) common to both test sets. The four ML algorithms have RMSE values of about 30 cm compared to 50 cm of the previous work, a substantial improvement in the prediction of the h_{50}. Despite the overall result of about 30 cm RMSE value for the eleven molecules of the test set, for sensitive molecules with h_{50} values less than 30 cm (e.g., PTETNT and TETNB), good agreement (error between

parentheses) was found for AdaBoost, namely, PTETNTN (4.0%), and TETNB (21.4%); Gradient Boost also gave good results for the same molecules, respectively 20.0% and 21.4%.

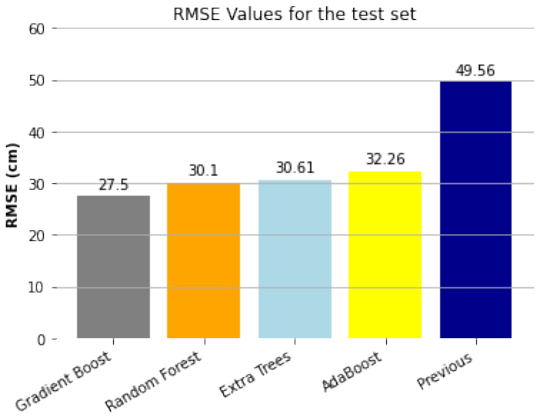

Fig. 3 Root mean square errors (RMSE) in cm in the predicted h_{50} values for the four algorithms and previous work of the molecules of the test set.[52]

$A \ -\sum Q_2(C)$

$B \ -\sum Q_1(NO_2)$

$C \ -\sum Q_0(NO_2)$

$D \ -\ \#NO_2$

Fig. 4. Influence of the features (molecular properties) for each algorithm according to the SHAP analysis for the training group – **Train. G** (42 molecules), and for the Test group – **Test G** (11 molecules). Each dot represents a molecule, and each molecular property (listed in the vertical

axis) includes all molecules of the corresponding set. Note that the horizontal scale may be different in each plot. Red values contribute to increasing the h_{50} value, and blue to reducing it. The darker the color, the larger the contribution – see the scale on the right. The features on the vertical axis are indicated by the letters A, B, C and D according to the correspondence on the top right of the figure.

The SHAP plots in Fig. 4 present the results for each molecule that produced the average information of Tables 3 and 4. The ordering in the vertical axis indicates that the higher the property, the higher is its contribution to the final h_{50}. The horizontal axis indicates the amount of contribution (positive or negative) of a given property to the h_{50} value. The coloring of the plots conveys relevant information concerning the predicted value of h_{50}: for the training group of the four algorithms, in most molecules, electron delocalization ($\sum Q_2(C)$) contributes to reducing the h_{50} value (dots are mostly in blue and have negative values) in contrast with what was found before [52, 106] – however, this is a subtle question, as we discuss below when examining the molecules of the test set individually. The other three properties, especially the number of nitro groups ($\#NO_2$) and their dipole moments ($\sum Q_1(NO_2)$) also in most cases contribute to reducing h_{50}. To examine in detail what is the role of each property in increasing or reducing the predicted h_{50}, we examine the SHAP plots of individual molecules of the test set.

Fig. 5 collects the individual SHAPs of the molecules of the test set computed with the AdaBoost algorithm along with the experimental values of impact sensitivities (h_{50}) collected from low to high values. The SHAP plots obtained with the other three algorithms, which are very similar to AdaBoost's, are not shown.

A clear and very interesting pattern emerges from these plots in Fig. 5. The 11 molecules of the test group can be divided into three groups according to the values of h_{50} and the direction of the contributing properties. For the most sensitive explosives (h_{50} value between parentheses), PTETNT (25 cm), TETNB (28 cm) and PATETNT (47 cm), the four molecular properties contribute to reducing the h_{50} value (i.e., they are in blue). According to AdaBoost results, which are similar to the one obtained with the other algorithms, the polarization of the nitro groups ($\sum Q_1(NO_2)$), followed by the electron delocalization of the ring atoms ($\sum Q_2(C)$), are the two most important properties in this group by far. For the second group comprised of TNBEtOH

(68 cm), DCLTNAN (75 cm), TNT (98 cm), DNB (100 cm), TMTNB (110 cm), and DATNP (112 cm), two properties contribute to increasing h_{50} (i.e., they are in red) and two to reducing it (i.e., they are in blue). The number of nitro groups ($\#NO_2$), and with a more significant contribution, the polarization of the nitro groups, contribute to increasing h_{50}, whereas the other two properties (most importantly $\sum Q_2(C)$, and the charge of the nitro groups, $\sum Q_0(NO_2)$) contribute to lowering it. The exception is DNB, which has the charge of nitro groups instead of their polarization, contributing to an increase h_{50}. In all cases of this intermediate group, the electron delocalization of the ring atoms is the most significant contributor to reducing h_{50} (i.e., to increasing the sensitivity of the explosive).

Finally, for the two most insensitive molecules of the test set, DNAN (220 cm) and DATB (320 cm), there is an extreme case – DATB – in which all properties contribute to increasing h_{50}, whereas for DNAM the polarization of the nitro group contributes to slightly reducing h_{50} and the other three properties significantly contribute to increasing it. In particular, we see the role of the electron delocalization ($\sum Q_2(C)$): this property only contributes to increasing the value of h_{50} for these two highly insensitive molecules.

Fig. 5. AdaBoost SHAP plots for the 11 molecules of the test set. Experimental impact sensitivity values (h_{50}) are also shown. Red properties contribute to increasing the value of h_{50} while blue one contributes to reducing it.

4. Conclusions

We carried out comprehensive machine learning (ML) simulations to investigate the molecular origin of the impact sensitivities of 53 nitroaromatic explosives characterized by h_{50} values. After an initial screening of 42 ML algorithms, the four with the smallest RMSE of the predicted h_{50} values, namely, Extra Trees, Random Forests, Gradient Boosting, and AdaBoost, were further investigated. The algorithms used four features (i.e., properties), and were run with a training set of 42 molecules and a test set comprising 11 molecules involving the extensive range of sensitivities $25\text{ cm} \leq h_{50} \leq 320\text{ cm}$. Three electronic features were atomic-centered DMA electric multipoles that characterize accurately in detail the density functional DFT charge density of the explosive. The fourth was a structural one, the number of the explosophore nitro groups.

Therefore, three electronic molecular properties and a structural one were the input features for the ML regressions of the target feature, the impact sensitivity h_{50} (cm). We found that the electron delocalization of the ring atoms, given by $\sum Q_2(C)$, is a significant contribution to the predicted h_{50} with a weight computed as the average of the four ML algorithms equal to 39%. The importance of this molecular electronic property agrees with previous results and our own, but this is the first time that its contribution has been quantified. The other contributing property with a similar average weight (35%) is the polarization of the nitro group, given by its total dipole $\sum Q_1(NO_2)$ value, a result not found before. The other two examined properties, the charge ($\sum Q_0(NO_2)$) and the number ($\#NO_2$) of the nitro groups, had average weights of 16% and 10% respectively.

By using special plots known as SHAP plots, the contribution of a given feature to increase or decrease the predicted h_{50} value of each molecule was evaluated. For the 11 molecules of the test set, the direction of the contribution depends on their sensitivity. For sensitive molecules ($h_{50} \lesssim 50$ cm), the four features contributed to reducing h_{50}. For the intermediate range $70\text{ cm} \lesssim h_{50} \lesssim 100$ cm, the number of the nitro groups and, more significantly, the polarization of the nitro groups, contributed to increasing h_{50} while the other two contributed to reducing it; in all cases, the electron delocalization of the ring atoms contributed to reducing h_{50}. Finally, for the most insensitive molecules, DNAN ($h_{50} = 220$ cm) and DATB (320 cm), all properties contributed to

increasing h_{50} (i.e., to decrease the sensitivity), except for the polarization of the nitro groups that contributes to reducing slightly h_{50} in DNAM.

By identifying the molecular properties that most affect the impact sensitivity, our work illuminated structure-property relationships of known explosives and may be used as a valuable guide for designing safer new energetic molecules.

Authors contributions

Julio Cesar Duarte – Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing - review & editing
Romulo Dias da Rocha – Data curation, Formal Analysis, Investigation, Software, Visualization,
Itamar Borges Jr.- Conceptualization, Data curation, Formal Analysis, Funding Acquisition, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft; Writing - review & editing

Acknowledgments

I. B. thanks the Brazilian agencies CNPq (Grant numbers 304148/2018-0 and 409447/2018-8) and FAPERJ (Grant number E26/201.197/2021) for support of this research. The authors thank Professor Craig A. Bayse for providing the molecular geometries used in this work.

References

1. L. E. Fried, M. R. Manaa, P. F. Pagoria and R. L. Simpson, *Design and synthesis of energetic materials*, *Ann. Rev. Mater. Res.*, 2001, **31**, 291-321.
2. A. K. Sikder and N. Sikder, *A review of advanced high performance, insensitive and thermally stable energetic materials emerging for military and space applications*, *J. Hazard. Mater.*, 2004, **112**, 1-15.
3. A. Demenay, L. Catoire and A. Osmont, in *Theoretical and Computational Chemistry*, ed. D. Mathieu, Elsevier, 2022, vol. 22, pp. 107-137.
4. E. R. Bernstein, R. W. Shaw, T. B. Brill and D. L. Thompson, *Overviews of Recent Research on Energetic Materials*, 2005.
5. P. M. Politzer, J. S., *Energetic Materials. Part 1. Decomposition, Crystal and Molecular Properties*, Elsevier, Amsterdan, 2003.
6. P. M. Politzer, J. S., *Energetic Materials. Part 2. Detonation, Combustion.*, Elsevier, Amsterdan, 2003.
7. N. Kubota, *Propellants and Explosives: Thermochemical Aspects of Combustions*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2nd edn., 2007.
8. D. Mathieu, *Molecular Modeling of the Sensitivities of Energetic Materials*, Elsevier. . Amsterdam, Netherlands, 2022.
9. M. M. Kuklja, E. V. Stefanovich and A. B. Kunz, *An excitonic mechanism of detonation initiation in explosives*, *J. Chem. Phys.*, 2000, **112**, 3417-3423.
10. A. A. Dippold and T. M. Klapotke, *A Study of Dinitro-bis-1,2,4-triazole-1,1 '-diol and Derivatives: Design of High-Performance Insensitive Energetic Materials by the Introduction of N-Oxides*, *J. Am. Chem. Soc.*, 2013, **135**, 9931-9938.
11. T. M. Klapotke, B. Krumm and A. Widera, *Synthesis and Properties of Tetranitro-Substituted Adamantane Derivatives*, *ChemPlusChem*, 2018, **83**, 61-69.
12. A. Osmont and A. Lefrancois, in *Theoretical and Computational Chemistry*, ed. D. Mathieu, Elsevier, 2022, vol. 22, pp. 3-27.
13. S. Zeman and M. Jungova, *Sensitivity and Performance of Energetic Materials*, *Propellants Explos. Pyrotech.*, 2016, **41**, 426-451.
14. P. Politzer and J. S. Murray, in *Theoretical and Computational Chemistry*, ed. D. Mathieu, Elsevier, 2022, vol. 22, pp. 173-194.
15. G. Li and C. Zhang, *Review of the molecular and crystal correlations on sensitivities of energetic materials*, *J. Hazard. Mater.*, 2020, **398**, 122910.
16. M. J. Kamlet, San Diego, California, 1976.
17. M. J. Kamlet and H. G. Adolph, *Relationship of impact sensitivity with structure of organic high explosives: . polynitroaromatic explosives*, *Propellants and Explosives*, 1979, **4**, 30-34.
18. P. J. Rae and P. M. Dickson, *Some Observations About the Drop-weight Explosive Sensitivity Test*, *Journal of Dynamic Behavior of Materials*, 2020, DOI: 10.1007/s40870-020-00276-2.
19. D. Mathieu and T. Alaime, *Impact sensitivities of energetic materials: Exploring the limitations of a model based only on structural formulas*, *J. Mol. Graph.*, 2015, **62**, 81-86.

20. T. L. Jensen, J. F. Moxnes, E. Unneberg and D. Christensen, *Models for predicting impact sensitivity of energetic materials based on the trigger linkage hypothesis and Arrhenius kinetics*, *J. Mol. Model.*, 2020, **26**, 65.
21. M. M. Kuklja, S. N. Rashkeev, B. T. and J. K., *Shear-strain-induced chemical reactivity of layered molecular crystals*, *Appl. Phys. Lett.*, 2007, **90**, 151913.
22. N. K. Rai, H. S. Udaykumar, S. O., R. N. K., D. A. S., H. D. B. and U. H. S., *Void collapse generated meso-scale energy localization in shocked energetic materials: Non-dimensional parameters, regimes, and criticality of hotspots*, *Phys. Fluids*, 2019, **31**, 016103.
23. K. L. McNesby and C. S. Coffey, *Spectroscopic Determination of Impact Sensitivities of Explosives*, *The Journal of Physical Chemistry B*, 1997, **101**, 3097-3104.
24. S. V. Bondarchuk, in *Theoretical and Computational Chemistry*, ed. D. Mathieu, Elsevier, 2022, vol. 22, pp. 195-213.
25. J. A. Morrill, B. C. Barnes, B. M. Rice and E. F. C. Byrd, in *Theoretical and Computational Chemistry*, ed. D. Mathieu, Elsevier, 2022, vol. 22, pp. 139-156.
26. P. Politzer, J. S. Murray, J. M. Serninario, P. Lane, M. E. Grice and M. C. Concha, *Computational characterization of energetic materials*, *Theochem-J. Mol. Struct.*, 2001, **573**, 1-10.
27. B. M. Rice and E. F. C. Byrd, *Theoretical chemical characterization of energetic materials*, *J. Mater. Res.*, 2006, **21**, 2444-2452.
28. Q. L. Yan and S. Zeman, *Theoretical evaluation of sensitivity and thermal stability for high explosives based on quantum chemistry methods: A brief review*, *International Journal of Quantum Chemistry*, 2013, **113**, 1049-1061.
29. 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.
30. 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, DOI: 10.1021/acs.jcim.2c00841.
31. A. Witze, *Green fuels blast off*, *Nature*, 2013, **500**, 509-510.
32. M. H. Keshavarz, *Simple Relationship for Predicting Impact Sensitivity of Nitroaromatics, Nitramines, and Nitroaliphatics*, *Propellants Explos. Pyrotech.*, 2010, **35**, 175-181.
33. M. H. Keshavarz, *A Simple Way to Predict Heats of Detonation of Energetic Compounds only from Their Molecular Structures*, *Propellants Explos. Pyrotech.*, 2012, **37**, 93-99.
34. M. H. Keshavarz, *A New General Correlation for Predicting Impact Sensitivity of Energetic Compounds*, *Propellants Explos. Pyrotech.*, 2013, **38**, 754-760.
35. D. Mathieu, *Toward a physically based quantitative modeling of impact sensitivities*, *J. Phys. Chem. A*, 2013, **117**, 2253-2259.
36. D. Mathieu, *Sensitivity of Energetic Materials: Theoretical Relationships to Detonation Performance and Molecular Structure*, *Ind. Eng. Chem. Res.*, 2017, **56**, 8191-8201.
37. P. Politzer, J. S. Murray, P. Lane, P. Sjoberg and H. G. Adolph, *Shock-sensitivity relationships for nitramines and nitroaliphatics*, *Chem. Phys. Lett.*, 1991, **181**, 78-82.
38. 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.

39. J. S. Murray, P. Lane and P. Politzer, *Effects of strongly electron-attracting components on molecular surface electrostatic potentials: application to predicting impact sensitivities of energetic molecules*, *Mol. Phys.*, 1998, **93**, 187-194.
40. C. Y. Zhang, *Review of the establishment of nitro group charge method and its applications*, *J. Hazard. Mater.*, 2009, **161**, 21-28.
41. A. L. Shoaf and C. A. Bayse, *Trigger bond analysis of nitroaromatic energetic materials using wiberg bond indices*, *Journal of Computational Chemistry*, 2018, **39**, 1236-1248.
42. C. A. Bayse and M. Jaffar, *Bonding analysis of the effect of strain on trigger bonds in organic-cage energetic materials*, *Theor. Chem. Acc.*, 2020, **139**, 11.
43. A. J. Stone, *Distributed multipole analysis, or how to describe a molecular charge-distribution*, *Chem. Phys. Lett.*, 1981, **83**, 233-239.
44. A. J. Stone and M. Alderton, *Distributed multipole analysis - methods and applications*, *Mol. Phys.*, 1985, **56**, 1047-1064.
45. A. J. Stone, *The Theory of Intermolecular Forces*, Clarendon Press, Oxford, 1997.
46. A. J. Stone, *Distributed multipole analysis: Stability for large basis sets*, *J. Chem. Theory Comput.*, 2005, **1**, 1128-1132.
47. I. Borges, *Conformations and charge distributions of diazocyclopropanes*, *International Journal of Quantum Chemistry*, 2008, **108**, 2615-2622.
48. G. Anders and I. Borges, *Topological analysis of the molecular charge density and impact sensitivy models of energetic molecules*, *J. Phys. Chem. A*, 2011, **115**, 9055-9068.
49. T. Giannerini and I. Borges, *Molecular Electronic Topology and Fragmentation Onset via Charge Partition Methods and Nuclear Fukui Functions: 1,1-Diamino-2,2-dinitroethylene*, *J. Braz. Chem. Soc.*, 2015, **26**, 851-859.
50. R. S. S. de Oliveira and I. Borges, *Correlation between molecular charge densities and sensitivity of nitrogen-rich heterocyclic nitroazole derivative explosives*, *J. Mol. Model.*, 2019, **25**, 314.
51. M. A. S. Oliveira and I. Borges, *On the molecular origin of the sensitivity to impact of cyclic nitramines*, *International Journal of Quantum Chemistry*, 2019, **119**, 14.
52. R. Siqueira Soldaini Oliveira and I. Borges Jr., *Correlation Between Molecular Charge Properties and Impact Sensitivity of Explosives: Nitrobenzene Derivatives*, *Propellants, Explosives, Pyrotechnics*, 2021, **46**, 309-321.
53. M. A. S. Oliveira, R. S. S. Oliveira and I. Borges, *Quantifying bond strengths via a Coulombic force model: application to the impact sensitivity of nitrobenzene, nitrogen-rich nitroazole, and non-aromatic nitramine molecules*, *J. Mol. Model.*, 2021, **27**, 69.
54. I. Borges, A. M. Silva, A. P. Aguiar, L. E. P. Borges, J. C. A. Santos and M. H. C. Dias, *Density functional theory molecular simulation of thiophene adsorption on MoS_2 including microwave effects*, *Theochem-J. Mol. Struct.*, 2007, **822**, 80-88.
55. I. Borges and A. M. Silva, *Probing topological electronic effects in catalysis: thiophene adsorption on NiMoS and CoMoS clusters*, *J. Braz. Chem. Soc.*, 2012, **23**, 1789-1799.
56. I. Borges, A. M. Silva and L. Modesto-Costa, *Microwave effects on NiMoS and CoMoS single-sheet catalysts*, *J. Mol. Model.*, 2018, **24**, 8.
57. A. Silva and I. Borges, *How to Find an Optimum Cluster Size Through Topological Site Properties: MoSx Model Clusters*, *Journal of Computational Chemistry*, 2011, **32**, 2186-2194.
58. I. Borges, R. S. S. Oliveira and M. A. S. Oliveira, in *Theoretical and Computational Chemistry*, ed. D. Mathieu, Elsevier, 2022, vol. 22, pp. 81-105.

59. T. B. Brill and K. J. James, *Thermal decomposition of energetic material. 61. perfidy in the amini-2,4,6-trinitrobenzene series of explosives*, *J. Phys. Chem.*, 1993, **97**, 8752-8758.
60. D. D. Dlott, in *Energetic Materials. Part 2. Detonation, Combustion*, eds. P. Politzer and J. S. Murray, Elsevier, Amsterdam, 2003, vol. 12.
61. D. Mathieu, *Modeling Sensitivities of Energetic Materials using the Python Language and Libraries*, *Propellants, Explosives, Pyrotechnics*, 2020, **45**, 966-973.
62. K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev and A. Walsh, *Machine learning for molecular and materials science*, *Nature*, 2018, **559**, 547-555.
63. H. Wang, Y. Ji and Y. Li, *Simulation and design of energy materials accelerated by machine learning*, *WIREs Computational Molecular Science*, 2020, **10**, e1421.
64. A. D. Casey, S. F. Son, I. Bilionis and B. C. Barnes, *Prediction of Energetic Material Properties from Electronic Structure Using 3D Convolutional Neural Networks*, *J. Chem Inf. Model.*, 2020, **60**, 4457-4473.
65. Y. Juan, Y. Dai, Y. Yang and J. Zhang, *Accelerating materials discovery using machine learning*, *J. Mater. Sci. Technol.*, 2021, **79**, 178-190.
66. Z. J. Baum, X. Yu, P. Y. Ayala, Y. Zhao, S. P. Watkins and Q. Zhou, *Artificial Intelligence in Chemistry: Current Trends and Future Directions*, *J. Chem Inf. Model.*, 2021, **61**, 3197-3212.
67. X.-l. Tian, S.-w. Song, F. Chen, X.-j. Qi, Y. Wang and Q.-h. Zhang, *Machine learning-guided property prediction of energetic materials: Recent advances, challenges, and perspectives*, *Energetic Materials Frontiers*, 2022, **3**, 177-186.
68. P. O. Dral, in *Advances in Quantum Chemistry*, eds. K. Ruud and E. J. Brändas, Academic Press, 2020, vol. 81, pp. 291-324.
69. P. O. Dral, *Quantum Chemistry in the Age of Machine Learning*, *The Journal of Physical Chemistry Letters*, 2020, **11**, 2336-2347.
70. P. O. Dral, F. Ge, B.-X. Xue, Y.-F. Hou, M. Pinheiro, J. Huang and M. Barbatti, *MLatom 2: An Integrative Platform for Atomistic Machine Learning*, *Topics in Current Chemistry*, 2021, **379**, 27.
71. J. George and G. Hautier, *Chemist versus Machine: Traditional Knowledge versus Machine Learning Techniques*, *Trends in Chemistry*, 2021, **3**, 86-95.
72. V. Gupta, K. Choudhary, F. Tavazza, C. Campbell, W.-k. Liao, A. Choudhary and A. Agrawal, *Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data*, *Nat. Commun.*, 2021, **12**, 6595.
73. Z. Yao, Y. Lum, A. Johnston, L. M. Mejia-Mendoza, X. Zhou, Y. Wen, A. Aspuru-Guzik, E. H. Sargent and Z. W. Seh, *Machine learning for a sustainable energy future*, *Nature Reviews Materials*, 2022, DOI: 10.1038/s41578-022-00490-5.
74. J. Xu, L. Zhu, D. Fang, L. Wang, S. Xiao, L. Liu and W. Xu, *QSPR studies of impact sensitivity of nitro energetic compounds using three-dimensional descriptors*, *Journal of Molecular Graphics and Modelling*, 2012, **36**, 10-19.
75. R. Todeschini and V. Consonni, *Handbook of molecular descriptors*, John Wiley & Sons, 2008.
76. 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*, *Chemometrics and Intelligent Laboratory Systems*, 2021, **215**, 104331.
77. H. Nefati, J. M. Cense and J. J. Legendre, *Prediction of the impact sensitivity by neural networks*, *Journal of Chemical Information and Computer Sciences*, 1996, **36**, 804-810.

78. S. G. Cho, K. T. No, E. M. Goh, J. K. Kim, J. H. Shin, Y. D. Joo and S. Seong, *Optimization of neural networks architecture for impact sensitivity of energetic molecules*, *Bull. Korean Chem. Soc.*, 2005, **26**, 399-408.
79. M. H. Keshavarz and M. Jaafari, *Investigation of the Various Structure Parameters for Predicting Impact Sensitivity of Energetic Molecules via Artificial Neural Network*, *Propellants, Explosives, Pyrotechnics*, 2006, **31**, 216-225.
80. L. H. Hall and L. B. Kier, *Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information*, *Journal of Chemical Information and Computer Sciences*, 1995, **35**, 1039-1045.
81. R. Wang, J. Jiang, Y. Pan, H. Cao and Y. Cui, *Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices*, *J. Hazard. Mater.*, 2009, **166**, 155-186.
82. R. Wang, J. Jiang and Y. Pan, *Prediction of Impact Sensitivity of Nonheterocyclic Nitroenergetic Compounds Using Genetic Algorithm and Artificial Neural Network*, *J. Energ. Mater.*, 2012, **30**, 135-155.
83. 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.
84. K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen and R. Barzilay, *Analyzing Learned Molecular Representations for Property Prediction*, *J. Chem Inf. Model.*, 2019, **59**, 3370-3388.
85. K. Swanson, S. Trivedi, J. Lequieu, K. Swanson and R. Kondor, *Deep learning for automated classification and characterization of amorphous materials*, *Soft Matter*, 2020, **16**, 435-446.
86. M. Y. McGrady, S. M. Colby, J. R. Nuñez, R. S. Renslow and T. O. Metz, *AI for Chemical Space Gap Filling and Novel Compound Generation*, *arXiv 2201.12398 (Pre-Print)*, 2022, DOI: 10.48550/ARXIV.2201.12398.
87. J. S. Smith, B. T. Nebgen, R. Zubatyuk, N. Lubbers, C. Devereux, K. Barros, S. Tretiak, O. Isayev and A. E. Roitberg, *Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning*, *Nat. Commun.*, 2019, **10**, 2903.
88. 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.
89. R. Roscher, B. Bohn, M. F. Duarte and J. Garcke, *Explainable Machine Learning for Scientific Insights and Discoveries*, *IEEE Access*, 2020, **8**, 42200-42216.
90. R. Todeschini and P. Gramatica, *New 3D molecular descriptors: the WHIM theory and QSAR applications*, *Perspectives in Drug Discovery and Design*, 1998, **9**, 355-380.
91. D. C. Elton, Z. Boukouvalas, M. S. Butrico, M. D. Fuge and P. W. Chung, *Applying machine learning techniques to predict the properties of energetic materials*, *Sci Rep*, 2018, **8**, 12.
92. D. M. Badgujar, M. B. Talawar, S. N. Asthana and P. P. Mahulikar, *Advances in science and technology of modern energetic materials: An overview*, *J. Hazard. Mater.*, 2008, **151**, 289-305.
93. A. Jezuita, K. Ejsmont and H. Szatylowicz, *Substituent effects of nitro group in cyclic compounds*, *Struct. Chem.*, 2021, **32**, 179-203.

94. A. Géron, *Hands-on machine learning with Scikit-learn, Keras and Tensor Flow: concepts, tools and techniques to build intelligent systems*, O'Reilly Media, Inc., Sebastopol, CA, 2017.
95. J. H. Friedman, *Greedy function approximation: A gradient boosting machine*, *The Annals of Statistics*, 2001, **29**, 1189-1232, 1144.
96. Y. Freund and R. E. Schapire, *A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting*, *Journal of Computer and System Sciences*, 1997, **55**, 119-139.
97. Y. Freund and R. E. Schapire, *A Short Introduction to Boosting*, *Journal of Japanese Society for Artificial Intelligence*, 1999, **14(5)**, 771-780.
98. P. Geurts, D. Ernst and L. Wehenkel, *Extremely randomized trees*, *Machine Learning*, 2006, **63**, 3-42.
99. L. Breiman, *Random Forests*, *Machine Learning*, 2001, **45**, 5-32.
100. 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 E. Duchesnay, *Scikit-learn: Machine Learning in Python*, *J. Mach. Learn. Res.*, 2011, **12**, 2825-2830.
101. E. Bisong, in *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, ed. E. Bisong, Apress, Berkeley, CA, 2019, DOI: 10.1007/978-1-4842-4470-8_7, pp. 59-64.
102. P. Probst, M. N. Wright and A. L. Boulesteix, *Hyperparameters and tuning strategies for random forest*, *Wiley Interdiscip. Rev.-Data Mining Knowl. Discov.*, 2019, **9**, 15.
103. P. Refaeilzadeh, L. Tang and H. Liu, in *Encyclopedia of Database Systems*, eds. L. Liu and M. T. ÖZsu, Springer US, Boston, MA, 2009, DOI: 10.1007/978-0-387-39940-9_565, pp. 532-538.
104. M. Saarela and S. Jauhiainen, *Comparison of feature importance measures as explanations for classification models*, *SN Applied Sciences*, 2021, **3**, 272.
105. S. M. Lundberg and S.-I. Lee, presented in part at the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017.
106. P. Politzer and J. S. Murray, *High Performance, Low Sensitivity: Conflicting or Compatible?*, *Propellants Explos. Pyrotech.*, 2016, **41**, 414-425.
107. C. Y. Zhang, *Investigation of the correlations between nitro group charges and some properties of nitro organic compounds*, *Propellants Explos. Pyrotech.*, 2008, **33**, 139-145.