Classifying and predicting the electron affinity of diamond nanoparticles using machine learning

First evidence of nanomaterial class-specific structure/property relationships

C. A. Feigl, B. Motevalli, A. J. Parker, B. Sun, A. S. Barnard

CSIRO Data61, Docklands, Victoria, Australia

www.data61.csiro.au

Diamond nanoparticles (nanodiamonds) are proving invaluable in bio-medical applications as a result of their bio-compatibility and high surface area, combined with 'tuneable' properties which can be tailored to meet application-specific surface chemistry requirements. This property-tuning is achieved through control over structural features; however, prediction of the relevant structure/property relationship(s) is made difficult by their multivariate nature, and the challenges of experimental control. We employed machine learning to perform the multivariate analysis, and electronic structure simulations to overcome experimental constraints, to form predictions of electronic properties including the electron affinity which is crucial to determining surface chemistry. Through the use of a suite of machine learning methods, we have found evidence of class-dependent structure/property relationships in the electron affinity, which have not been reported in nanomaterials before[1].

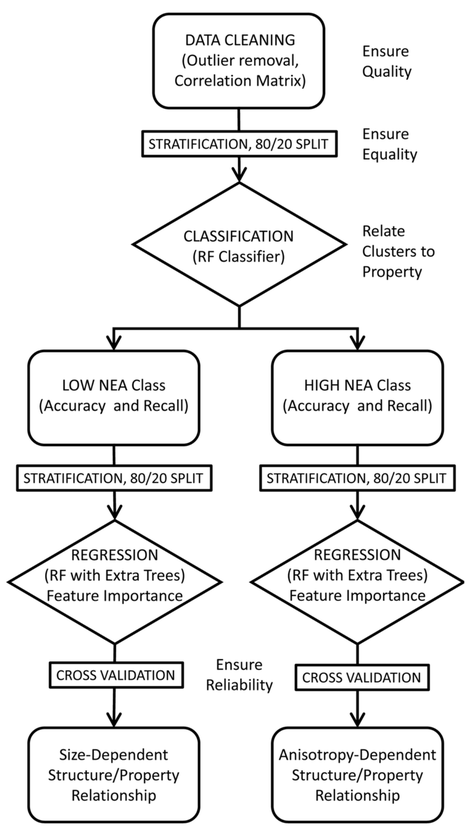

Figure 1: Conceptual workflow applied to the classification and regression analysis of the electron affinity with respect to the structural and morphological features listed in Table 1.

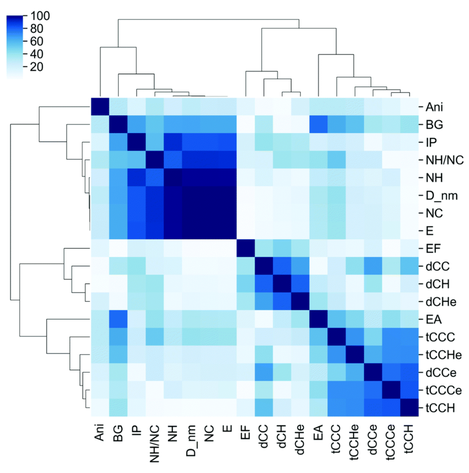

Figure 2: Correlation matrix used to identify pairs of dependent structural features. Strongly correlated features should be avoided as they overly complicate models. Accordingly, NH/NC, NH, and NC where omitted from further analysis, as they are strongly correlated with size (*D*_nm).

Abbreviation	Feature	Description
NC	Total number of carbon atoms	Global, first order
NH	Total number of hydrogen atoms	Global, first order
NH/NC	Surface-to-volume ratio	Global, first order
*D*_nm	Average particle diameter	Global, first order
Ani	Particle anisotropy	Global, first order
dCC	Average C–C bond length	Local, first order
dCCe	Uncertainty in the C–C bond length	Local, second order
dCH	Average C–H bond length	Local, first order
dCHe	Uncertainty in the C–H bond length	Local, second order
tCCC	Average C–C–C bond angle	Local, first order
tCCCe	Uncertainty in the C–C–C bond angle	Local, second order
tCCH	Average C–C–H bond angle	Local, first order
tCCHe	Uncertainty in the C–C–H bond angle	Local, second order

Table 1: Initial structural and morphological features used to described the nanodiamonds contained in the dataset.

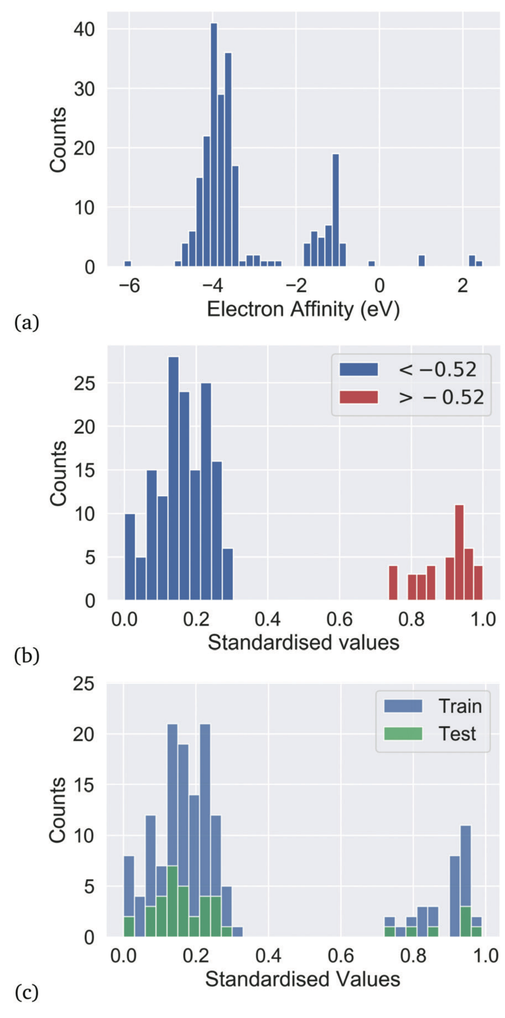

Figure 3: The histogram plots for (a) the electron affinity over the full set of 231 samples prior to removal of outliers, (b) the binary split of the dataset identified using *k*-means clustering after the removal of outliers with centroids corresponding to -3.9eV and -1.2eV, and (c) the stratified 80/20 split used for training/testing of the models.

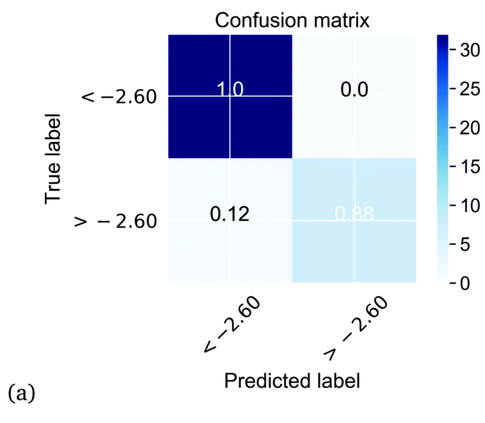

Figure 4: (a) Confusion matrix used to assess the sensitivity and specificity of the random forest classifier, and (b) the corresponding feature importances. The degree of variation in the bond angles and lengths, as well as changes in the magnitude of the bond angles and lengths, are the most important features in determining whether a nanodiamond belongs to a high or low EA class, while size is the least important.

Figure 5: Stratified split of the (a) low EA energy and (b) high EA energy classes into training and testing sets. Stratifying imbalanced datasets prior to train/test splits reduces the potential for biasing.

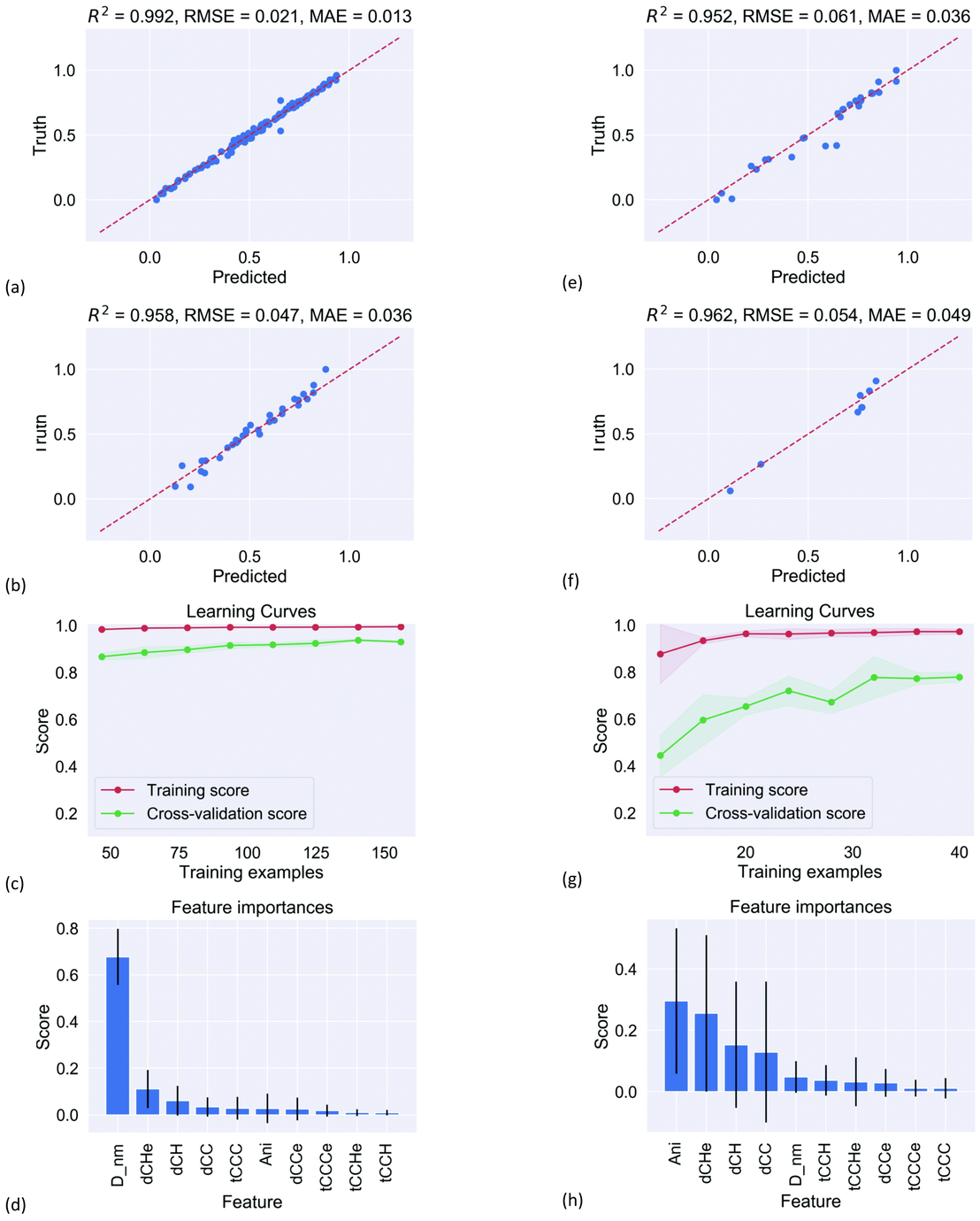

Figure 6: Results of the random forest regression model for (a),(e) training data, (b),(f) testing data, (c),(g) the learning curves, and (d),(h) the feature importances in predicting the EA values of (a)-(d) the low energy class, and (e)-(h) the high energy class. While (d) shows that the EA of the low energy class is strongly size-dependent, (h) shows that the high energy class EA depends on the overall anisotropy of the particle and the nature of hydrogen passivation at the surface.

Conclusions

In the past, attempts to predict structure/property relationships have used small and selective nanoparticle datasets with simple curve fitting algorithms, and have treated all the particles together in the same group, leading to potentially misleading results. The methods presented here provided evidence of class-specific structure/property relationships in the electron affinity of nanodiamonds, suggesting a mechanism for purification or screening of this biologically relevant material. This specificity has not been reported in nanomaterials before, and so the insights presented here provide guidance to researchers undertaking this type of analysis, as well as to those developing nanodiamond-based applications.

FOR FURTHER INFORMATION

Amanda Barnard

e amanda.barnard@data61.csiro.au

Chris Feigl

e chris.feigl@data61.csiro.au

REFERENCES

[1] C.A Feigl, B. Motevalli, A. J. Parker, B. Sun, A. S. Barnard, Nanoscale Horiz., 2019, Advance Article

ACKNOWLEDGEMENTS

Computational resources for this project were supplied by the National Computing Infrastructure national facility under grant q27.

NCI AUSTRALIA PROVIDING AUSTRALIAN RESEARCHERS WITH WORLD-CLASS HIGH-END COMPUTING SERVICES