Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Machine Learning Approach for Predicting Nucleophilicities of Organic Molecules

Contents

General Information

The nucleophilicity N index values and corresponding nucleophile and solvent information was acquired from Mayr's database.[1] Molecules were drawn on Gauss View 6 and computations were performed on Gaussian 09 D.01 suit.[2,3] The geometry optimization and frequency calculations were performed on B3LYP/6311G(d,p) for DFT and 6311G(d,p) for HF. Minima was confirmed by vibrational frequency calculations showing no imaginary frequency. NBO charges were obtained using "NPA only" keyword in the population analysis section in Gaussian. Molecular mass and steric parameters, namely, total surface area (S.A.), volume (V), radius of gyration (R_G) and diameter of molecule (R_D), were obtained using Multiwfn 3.7 software[4]. Ovality (O) was obtained from volume and surface area. Structures optimized at the DFT level of theory using B3LYP/6311G(d,p) method were used as input for calculation of these steric parameters. For machine learning all the calculations were performed on google colab using Python framework.[5] Data analysis and pre-processing was done using Pandas, NumPy and Scikit-learn libraries. Model development was performed using Scikit-learn and Keras libraries. Visualization of data was done on Seaborn and Matplotlib libraries.

Descriptors

Units

Total energy (E), energies of HOMO and LUMO, electronegativity (χ), chemical hardness (η), chemical potential (μ), electrophilicity index (ω), dipole moment (d), free energy (G), enthalpy (H) and their correlation contribution (CORR), (all quantum-mechanical descriptors in a.u., 1a.u.=27.21165 eV, 627.50956 kcal/mol, thermodynamic descriptors in Kcal/mol), dipole moment in Debye and NBO charges in Coulombs calculated with HF/6-311G (d,p) and DFT/B3LYP/6-311G(d,p) level of the theory). Molecular Mass (MM) in amu, Radius of Gyration (Rg) and Diameter in Angstrom (A), Surface Area in A^2 and Volume in A^3.[6]

Descriptor Info

Table S1. List of calculated quantum chemical descriptors used in this study.

S.No.	Notation		Definition	Source/Formula
	Nucleophile	Solvent		
1.	EHF	E*HF	Total energy-HF	Gaussian 09
2.	EDFT	E*DFT	Total energy-DFT	Gaussian 09
3.	ECORR	E*CORR	Electron correlation contribution of E	EDFT-EHF
4.	EHF(HOMO)	E*HF(HOMO)	Energy of highest occupied molecular orbital-HF	Gaussian 09
5.	EDFT(HOMO)	E*DFT(HOMO)	Energy of highest occupied molecular orbital-DFT	Gaussian 09
6.	ECORR(HOMO)	E*CORR(HOMO)	Electron correlation contribution of HOMO	EDFT(HOMO)- EHF(HOMO)
7.	EHF(LUMO)	E*HF(LUMO)	Energy of lowest unoccupied	Gaussian 09

			molecular orbital-HF	
8.	EDFT(LUMO)	E*DFT(LUMO)	Energy of lowest unoccupied molecular orbital-DFT	Gaussian 09
9.	ECORR(LUMO)	E*CORR(LUMO)	Electron correlation contribution of LUMO	EDFT(LUMO)- EHF(LUMO)
10.	χHF	χ*HF	Electronegativity - HF	-(EHF(HOMO)+EHF(LUMO))/2
11.	χDFT	χ*DFT	Electronegativity - DFT	-(EDFT(HOMO)+EDFT(LUMO))/2
12.	χCORR	χ*CORR	Electron correlation contribution of EN	χDFT- χHF
13.	ηHF	η*HF	Chemical hardness - HF	EHF(LUMO) – EHF(HOMO)
14.	ηDFT	η*DFT	Chemical hardness - DFT	EDFT(LUMO) – EDFT(HOMO)
15.	ηCORR	η*CORR	Electron correlation contribution of η	ηDFT- ηHF
16.	μHF	μ*HF	Chemical potential - HF	(EHF(HOMO) + EHF(LUMO))/2
17.	μDFT	μ*DFT	Chemical potential - DFT	(EDFT(HOMO) + EDFT(LUMO))/2
18.	μCORR	μ*CORR	Electron correlation contribution of μ	μDFT- μHF
19.	ωHF	ω*HF	Electrophilicity index - HF	$\mu HF^2/2\eta HF$
20.	ωDFT	ω*DFT	Electrophilicity index - DFT	$\mu DFT^2/2\eta DFT$
21.	ωCORR	ω*CORR	Electron correlation contribution of ω	ωDFT- ωHF
22.	dHF	d*HF	Dipole moment - HF	Gaussian 09
23.	dDFT	d*DFT	Dipole moment - DFT	Gaussian 09
24.	dCORR	d*CORR	Electron correlation contribution of d	dDFT- dHF
25.	HHF	H*HF	Enthalpy - HF	Gaussian 09
26.	HDFT	H*DFT	Enthalpy - DFT	Gaussian 09
27.	HCORR	H*CORR	Electron correlation contribution of H	HDFT- HHF
28.	GHF	G*HF	Gibbs free energy - HF	Gaussian 09
29.	GDFT	G*DFT	Gibbs free energy - DFT	Gaussian 09

30.	GCORR	G*CORR	Electron correlation contribution of G	GDFT- GHF
31.	NBOHF		Natural bond orbital charge - HF	Gaussian 09
32.	NBODFT		Natural bond orbital charge - DFT	Gaussian 09
33.	NBOCORR		Electron correlation contribution of NBO	NBODFT- NBOHF
34.	Mol. Mass		Molecular Mass	Multiwfn 3.7
35.	Surface Area		Surface Area	Multiwfn 3.7
36.	Volume		Volume	Multiwfn 3.7
37.	Ovality		Ovality	$A/4\pi(3V/4\pi)^{2/3}$ (V=Volume, A=Surface area)
38.	Diameter		Diameter of molecule	Multiwfn 3.7
39.	Rg		Radius of gyration	Multiwfn 3.7

Figures

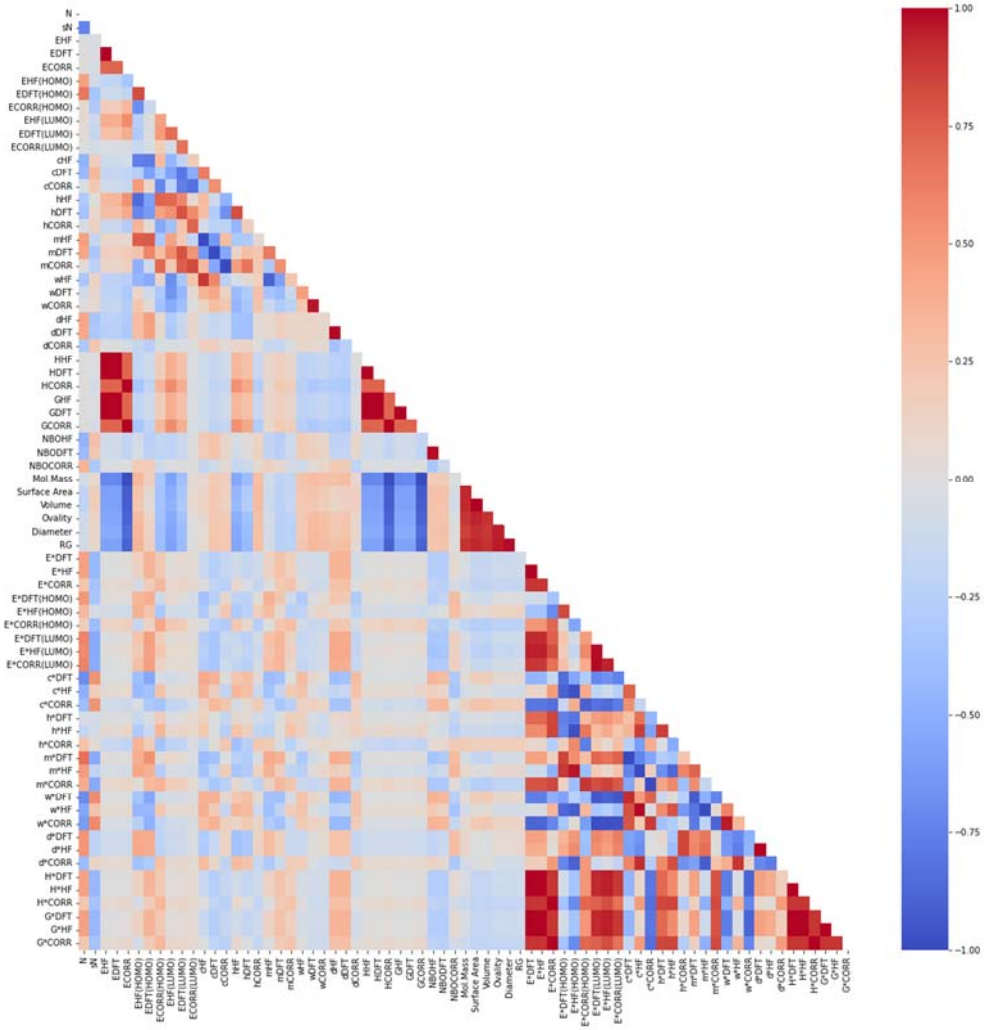

Figure S1. Correlation matrix of whole data set.

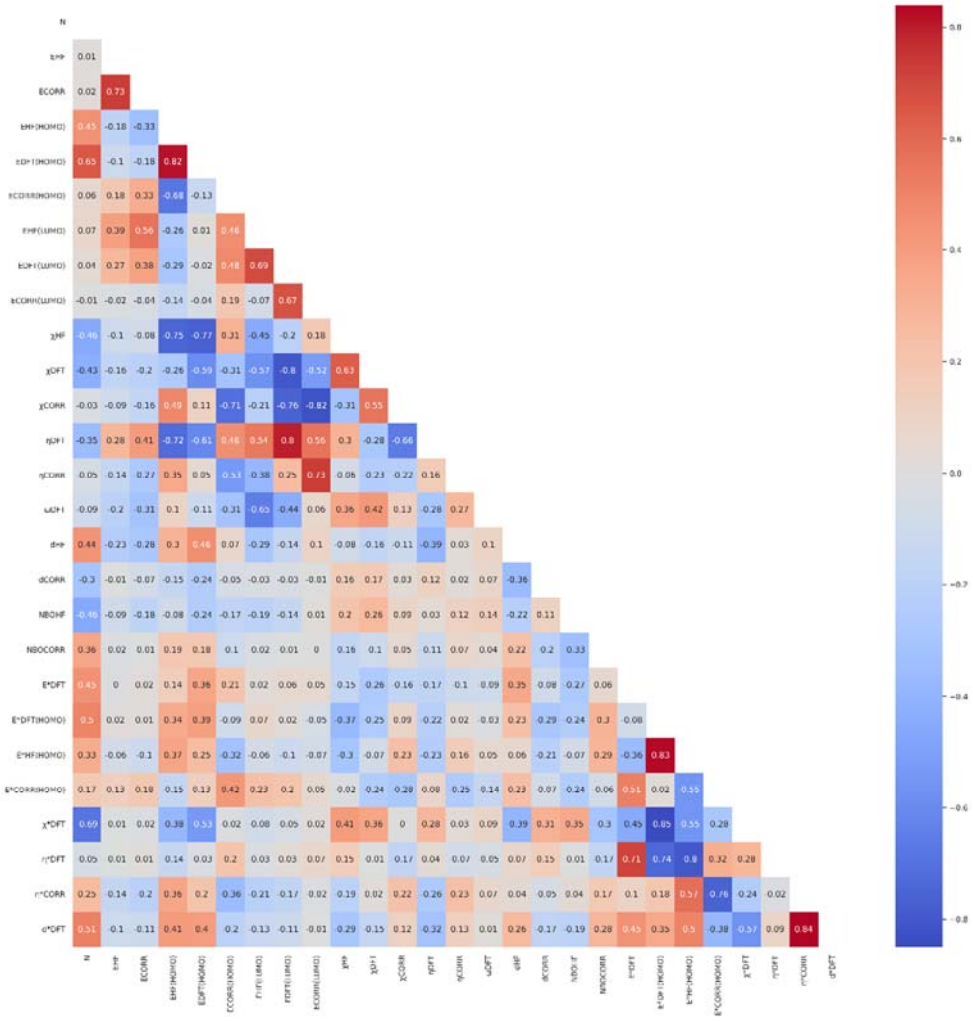

Figure S2. Correlation matrix of truncated data set (consisting of 26 descriptors)

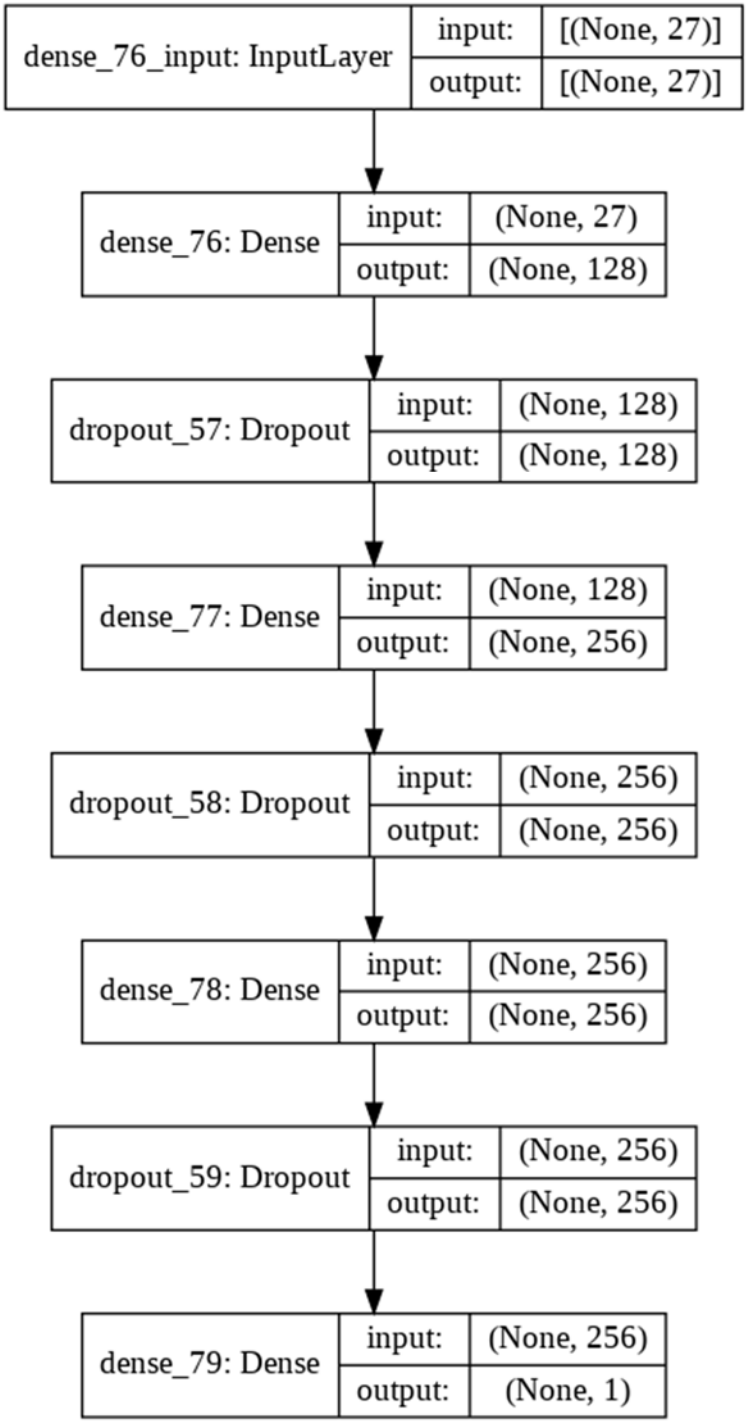

Figure S3. Summary of the hyper parameters used in Neural Network model.

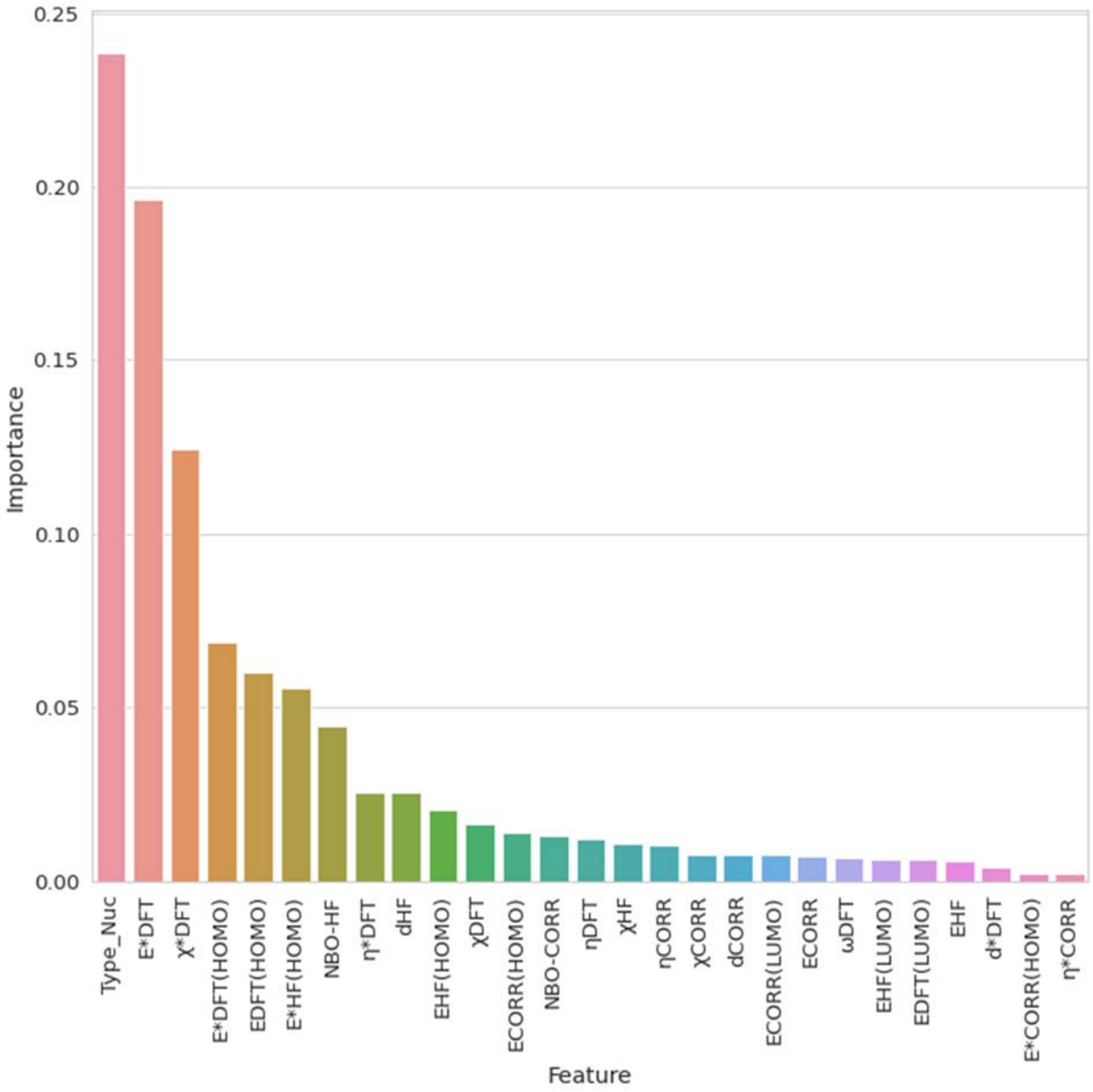

Figure S4. Relative feature importance depicting Pearson coefficient of 27 descriptors with N.

References

(1) Link for Mayr's database: https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank/

(2) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision D.01; Gaussian, Inc.: Wallingford CT, 2009.

(3) Dennington, Roy; Keith, Todd A.; Millam, John M. Semichem Inc., Shawnee Mission, KS, *GaussView,* Version 6, 2016.

(4) Lu, T.; Chen, F., Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012,** *33* (5), 580-592.

(5) Van Rossum, G., & Drake Jr, F. L. (1995). *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

(6) Karelson, M.; Lobanov, V. S.; Katritzky, A. R., Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996,** *96* (3), 1027-1044.