

Characterization of impurity contamination in solar cells with the assistance of machine learning



Oleg Olikh
Oleksii Zavhorodnii

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine



Non-destructive methods aimed at estimating the concentration of recombination-active defects are numerical:

Fourier-transform infrared spectroscopy



deep level transient spectroscopy



minority carrier lifetime measurements



electron-paramagnetic resonance

...

However, these techniques

- ✓ *are time-consuming*
- ✓ *involve the use of unconventional conditions (low temperatures, magnetic fields, etc)*

- ✓ *require special equipment*
- ✓ *require specially prepared samples.*



artificial intelligence in

...everyday life



ChatGPT

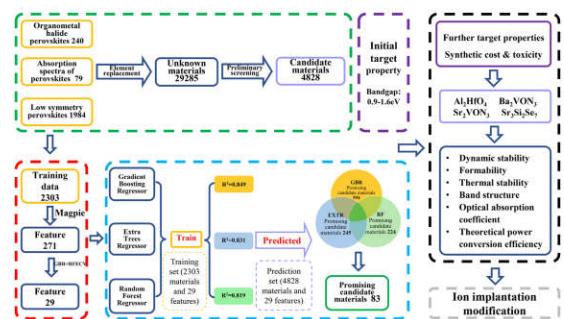
DALL-E 2



...and photovoltaics

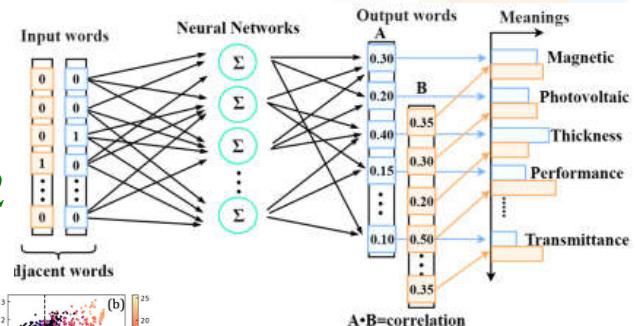
identify potentially important PV materials based on
their optical and symmetry properties

mentions of names in literature



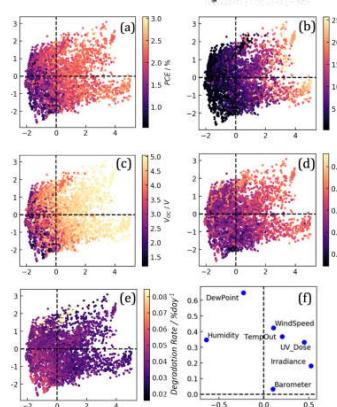
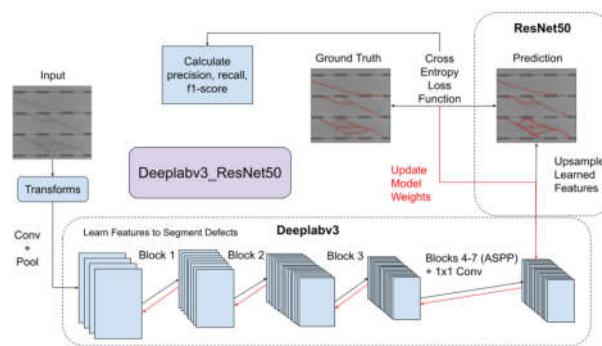
Appl. Phys. Lett. **119** 231902

J Appl. Phys. **131** 064902



defect detection
from EL images

IEEE J. Photovoltaics
12 p.53



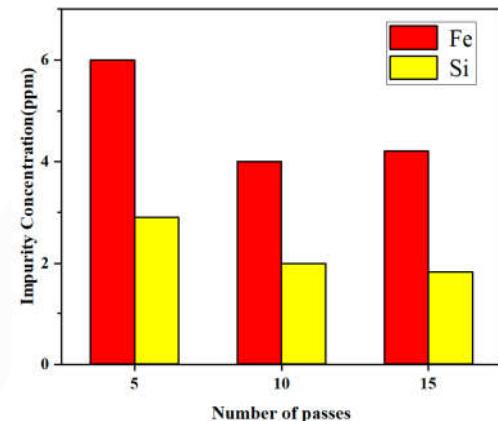
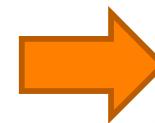
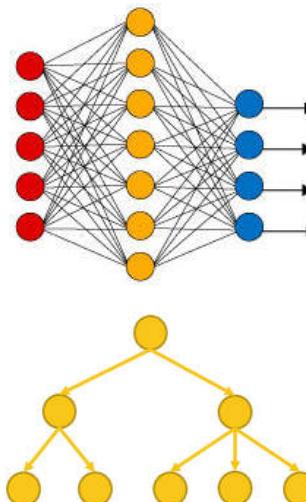
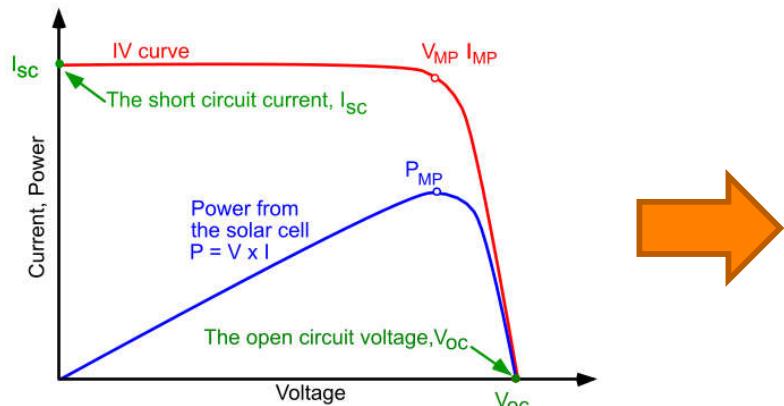
SC degradation
prediction

Prog. PV Res. Appl. **29**
p.127



Main goal:

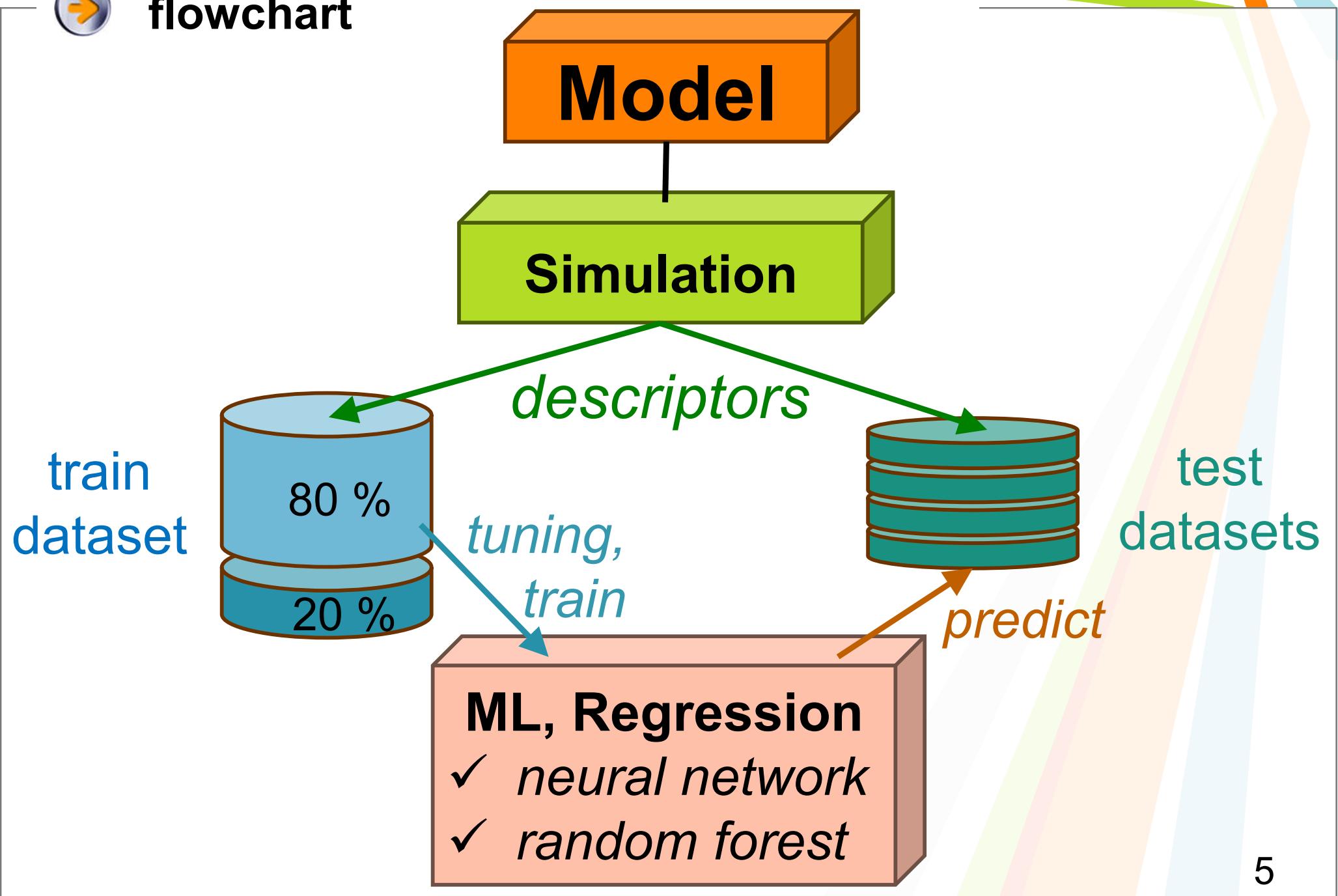
method aimed at assessing impurity concentrations in solar cells, based on applying machine learning to data obtained from current-voltage characteristic



method is **express**, **low-cost** and **does not require additional equipment**



flowchart





Model

silicon
 n^+ - p - p^+ structure

solar cell

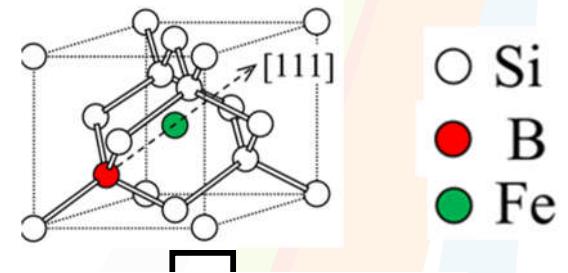
iron

impurity

iron

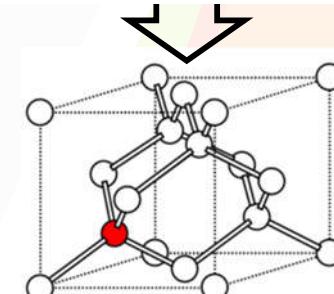
- is one of the main and most harmful impurities in silicon SC
- has two configurations with significantly different recombination properties in Si:B

$\text{Fe}_i \text{B}_s$



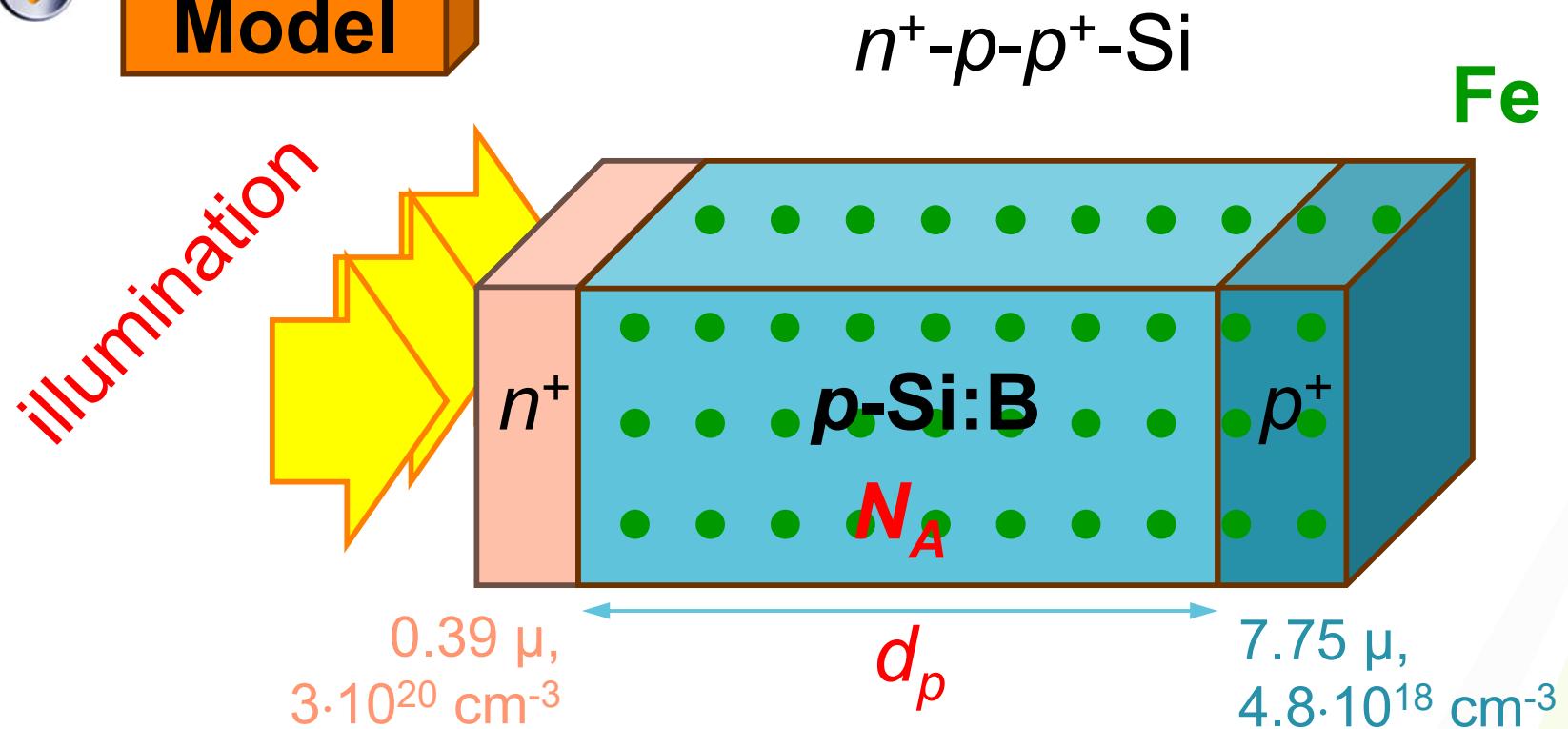
intense illumination
 high-temperature treatment

Fe_i





Model



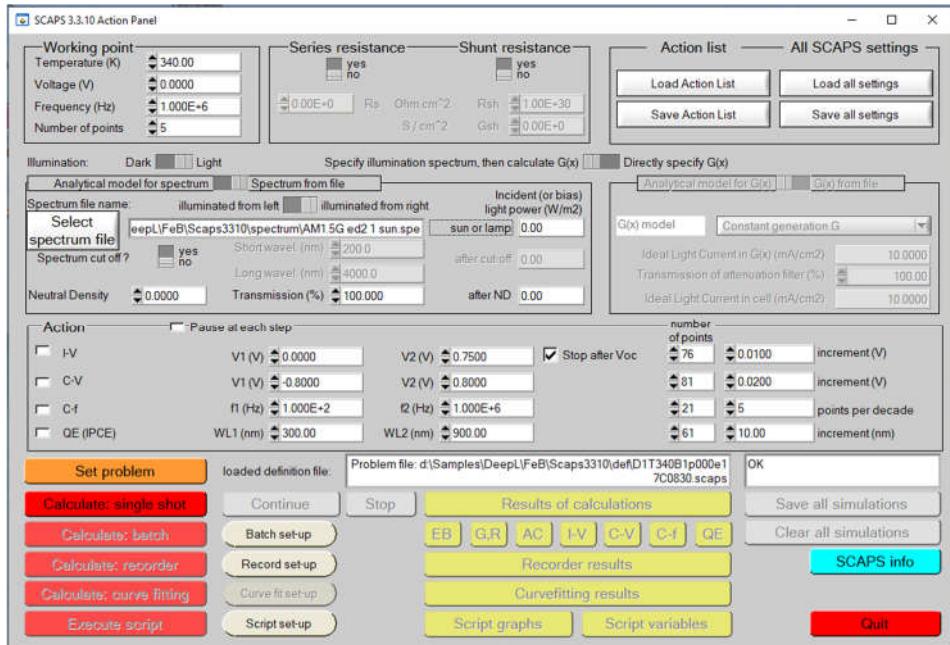
Variable parameters:

- ✓ base doping level N_A : $10^{15}\text{-}10^{17} \text{ cm}^{-3}$
- ✓ iron contamination N_{Fe} : $10^{10}\text{-}10^{14} \text{ cm}^{-3}$
- ✓ illumination: AM1.5, 1000 W/m^2
 $940 \text{ nm}, 4 \text{ W/m}^2$
- ✓ base depth d_p : $180\text{-}380 \mu$
- ✓ temperature T : $290\text{-}340 \text{ K}$



Simulation

one-dimensional code SCAPS 3.3.10

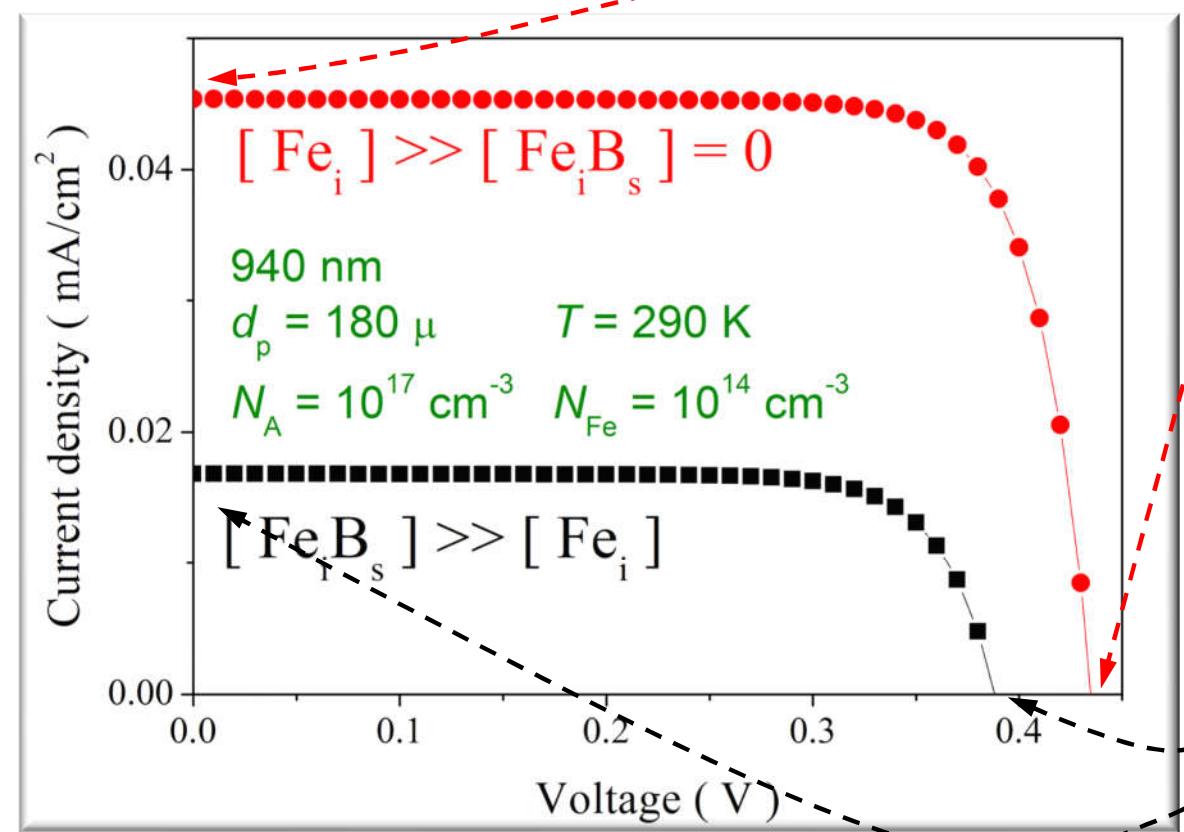


- ✓ temperature dependence of the light absorption in silicon [Photovoltaics 30 164];
- ✓ bandgap temperature dependence according to Passler equation [Phys. Rev.B 66 085201];
- ✓ doping induced bandgap narrowing [J.Appl. Phys. 116 194505];

- ✓ thermal carrier velocities [J. Appl. Phys. 67 2944];
- ✓ temperature dependence of band's state density [J. Appl. Phys. 115 093705];
- ✓ the free carrier effective masses [O'Mara Handbook of semiconductor silicon technology 1990]
- ✓ carrier mobilities according to Klaassen's theory [Solid-State Electron. 7 953];
- ✓ the temperature and doping level dependencies of band--to--band [Sol. Ener. Mat. & SC 235 111467] and Auger recombination [Sol. Ener. Mat. & SC 234 111428]
- ✓ the recombination parameters of iron-related defects [Sol. Ener. Mat. & SC 187 263]



Simulation



$$V_{OC}^{Fe}, I_{SC}^{Fe}, \eta^{Fe} = \frac{P_{MP}^{Fe}}{P_{illum}},$$

$$FF^{Fe} = \frac{P_{MP}^{Fe}}{V_{OC}^{Fe} \cdot I_{SC}^{Fe}}$$

$$V_{OC}^{FeB}, I_{SC}^{FeB}, \eta^{FeB} = \frac{P_{MP}^{FeB}}{P_{illum}},$$

$$FF^{FeB} = \frac{P_{MP}^{FeB}}{V_{OC}^{FeB} \cdot I_{SC}^{FeB}}$$

$$\varepsilon V_{OC} = \frac{V_{OC}^{FeB} - V_{OC}^{Fe}}{V_{OC}^{FeB}} \cdot 100\%,$$

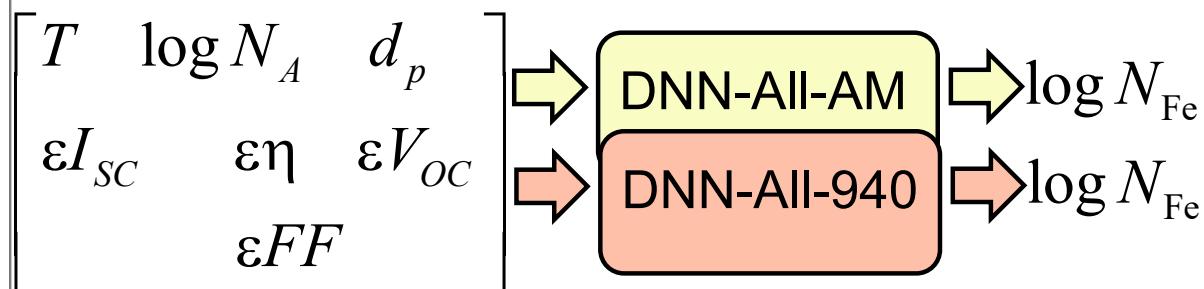
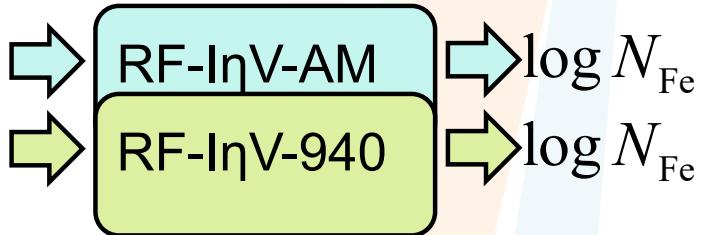
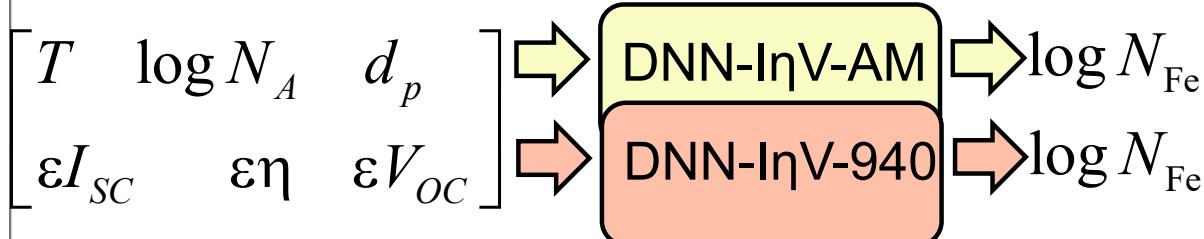
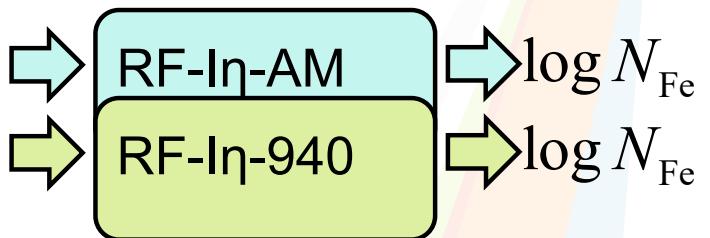
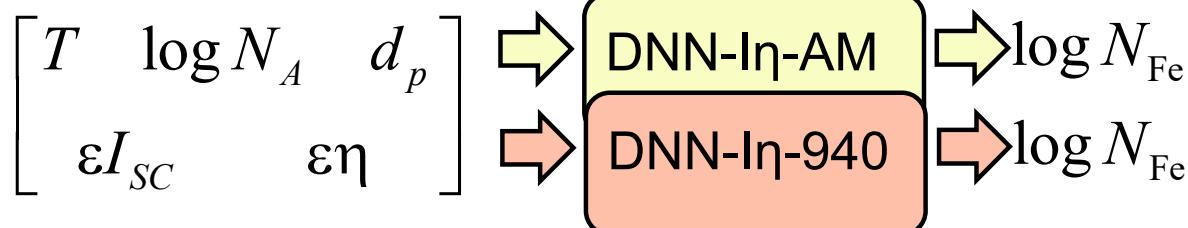
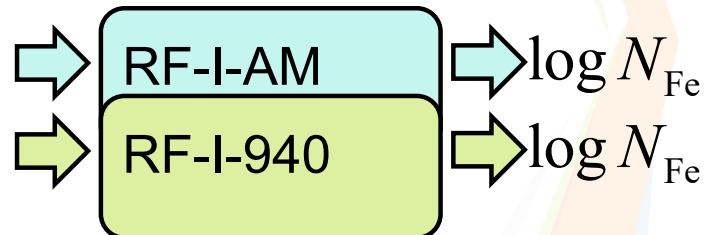
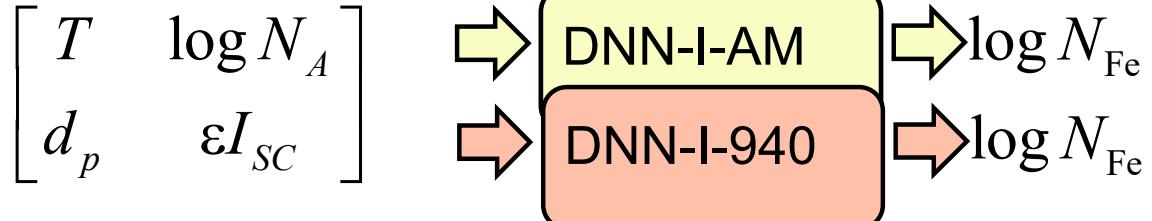
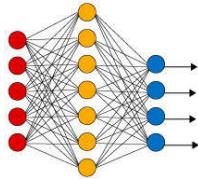
$$\varepsilon I_{SC},$$

$$\varepsilon \eta,$$

$$\varepsilon FF$$

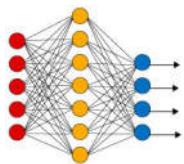


descriptors





tuning



K Keras

Weight initializer

- Xavier normal
- Xavier uniform
- He normal
- He uniform
- Random normal
- Random uniform

Optimizer

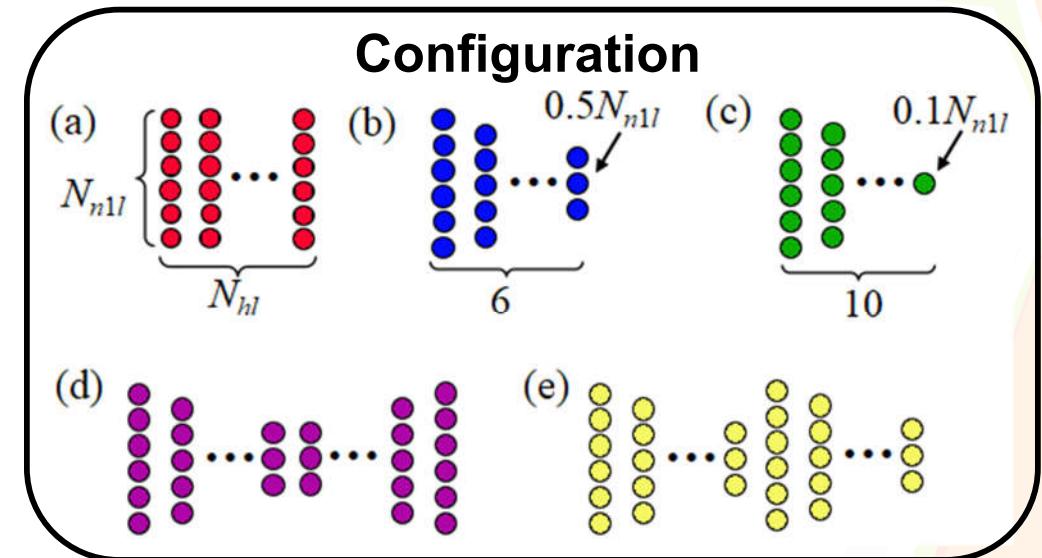
- Adam
- SGD
- RMSprop
- Adadelta
- Adagrad
- Adamax
- Nadam
- Ftrl

Activation

- relu
- sigmoid
- tanh
- selu
- elu

Batch size

- 8, 16, 32, 64, 128



N_{n1l}
5, 10, 30, 50, 75, 100, 120, 150, 200

Learning Rate

- $10^{-5} - 10^{-2}$

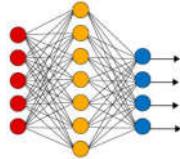
N_{hl}
4, 5, 6, 8, 10, 12, 15

Epoch

- 100 – 2000



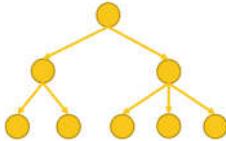
tuning result



DNN	Nodes in hidden layers	Activation	Optimizer	Initializer	Batch size	Learning rate	Epoch
I-AM	200-180-160-140-120-100	relu	Adam	Xavier Normal	16	4·10 ⁻⁴	500
I-940	200-200-200-200-200-200	relu	Adamax	Xavier Normal	64	3.7·10 ⁻³	600
Iη-AM	120-108-96-84-72-60	elu	Adamax	Xavier Uniform	128	1.8·10 ⁻³	1500
Iη-940	100-90-80-70-60-50	relu	Adamax	Xavier Normal	16	3.1·10 ⁻³	1000
IηV-AM	150-135-120-105-90-75	elu	Adamax	Xavier Uniform	16	2.5·10 ⁻³	1500
IηV-940	75-67-59-51-43-35	elu	Adamax	Xavier Normal	8	3.7·10 ⁻³	700
All-AM	120-108-96-84-72-60	tanh	Adamax	Xavier Normal	64	1.1·10 ⁻³	1000
All-940	200-180-160-140-120-100	relu	Adamax	Xavier Normal	16	1.8·10 ⁻³	1000



tuning



Min number of samples to split an internal node

2,3,4,5,6,7

Min number of samples to be in a leaf node

1, 2,3,4,5,6,7

Number of trees

100,150,200,250,300,350,400,450,500

Tree's maximum depth

10,15,20,25,30,35,40,45

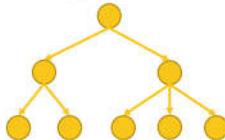
Number of features to make splits

'log2', 'sqrt', 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2

Bootstrap
True, False



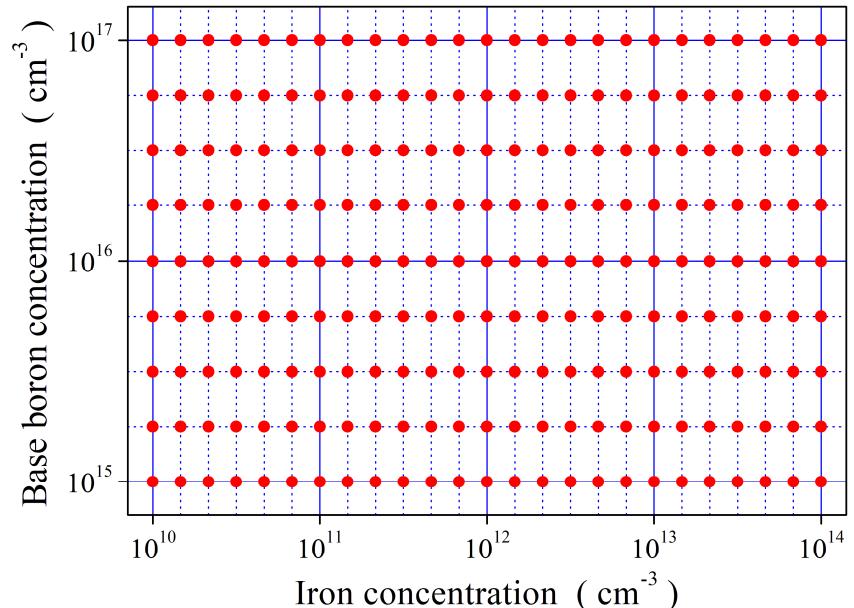
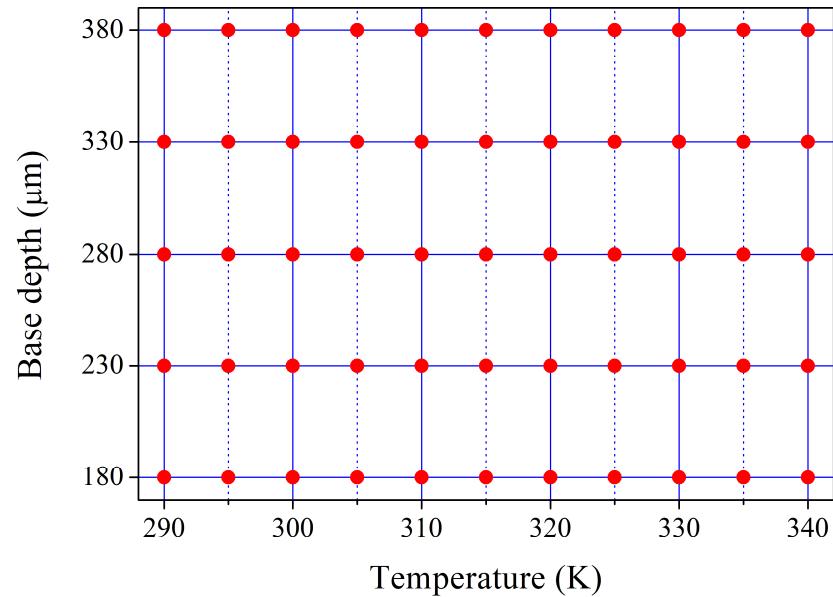
tuning result



RF	Number of trees	Depth	Min split	Min leaf	Max features	Bootstrap
I-AM	250	15	4	1	0.9	False
I-940	300	25	2	2	0.9	False
In-AM	300	35	3	1	0.9	False
In-940	250	25	3	1	0.6	False
InV-AM	400	40	3	1	0.8	False
InV-940	400	35	3	1	0.6	False
All-AM	300	30	3	1	0.6	False
All-940	300	35	2	1	0.6	False



train dataset



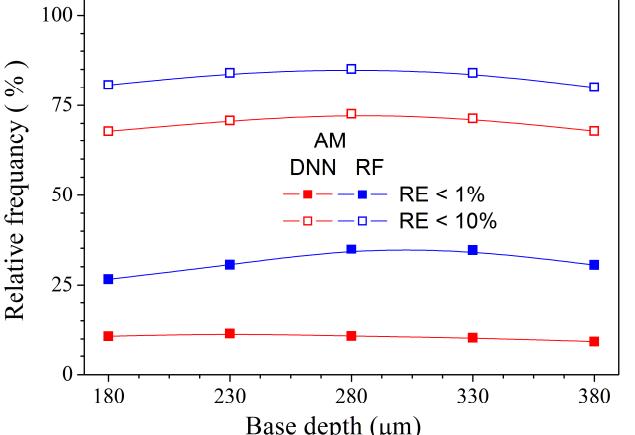
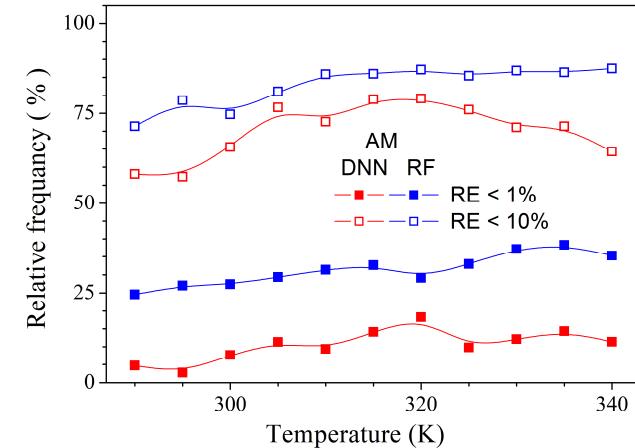
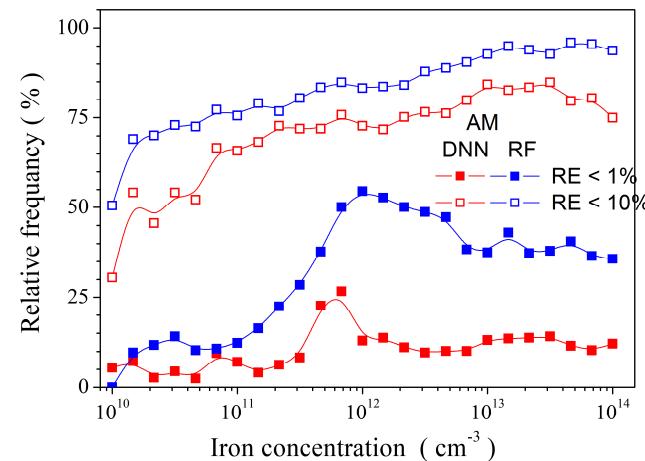
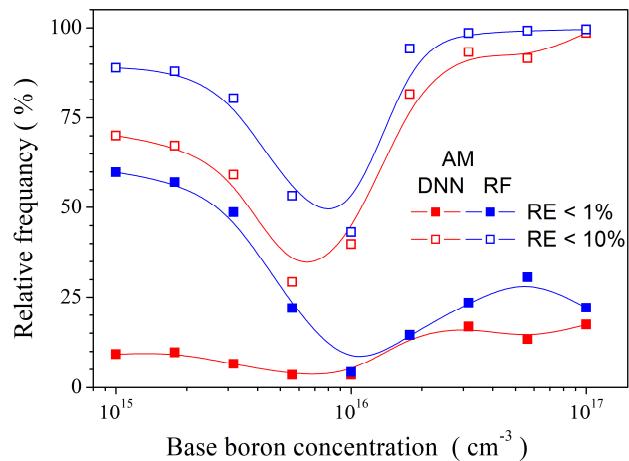
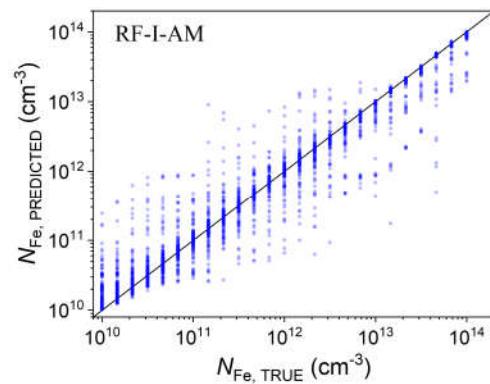
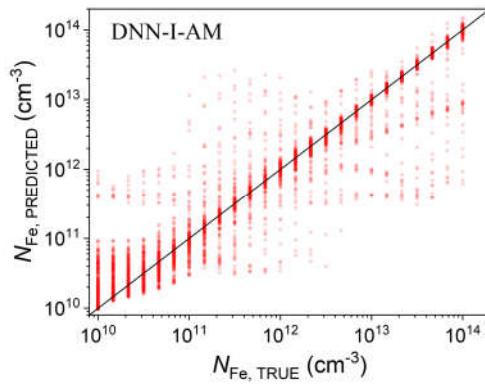
$5 \times 11 \times 25 \times 9 = 12\,375$ samples

Results of 5-fold cross-validation

		Mean squared error (MSE), 10^{-3}			
		Number of descriptors			
		4	5	6	7
DNN	AM	42±5	9±3	4±2	2±1
	940	10±5	6.1±0.4	6±2	1.5±0.7
RF	AM	33±2	11±3	5±2	4±1
	940	6±1	4.6±0.2	3.0±0.5	3.0±0.8



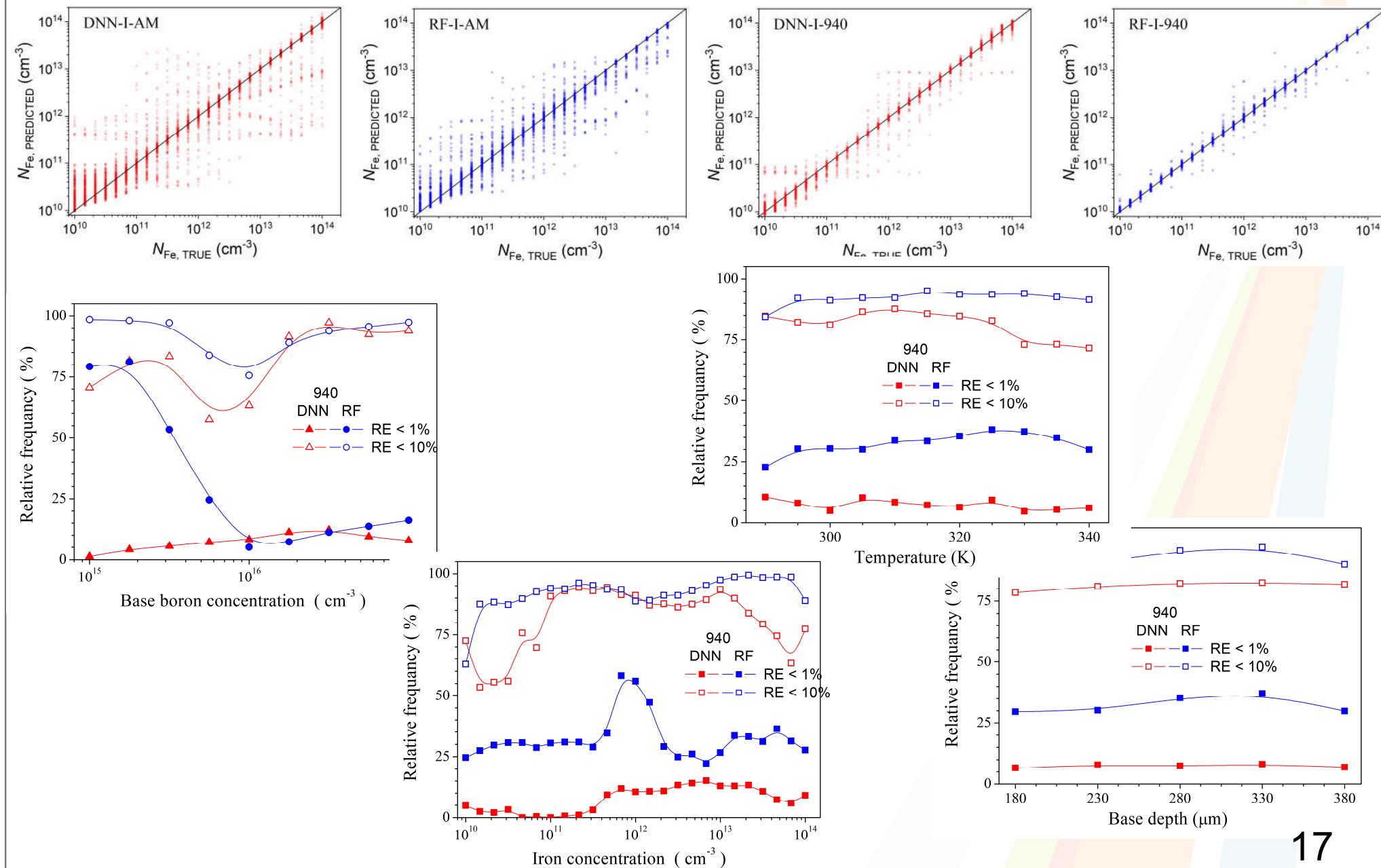
train dataset, 4 descriptors, AM (solar illumination)



$$\text{Relative Error (RE)} = \frac{|N_{Fe, TRUE} - N_{Fe, PREDICTED}|}{N_{Fe, TRUE}}$$

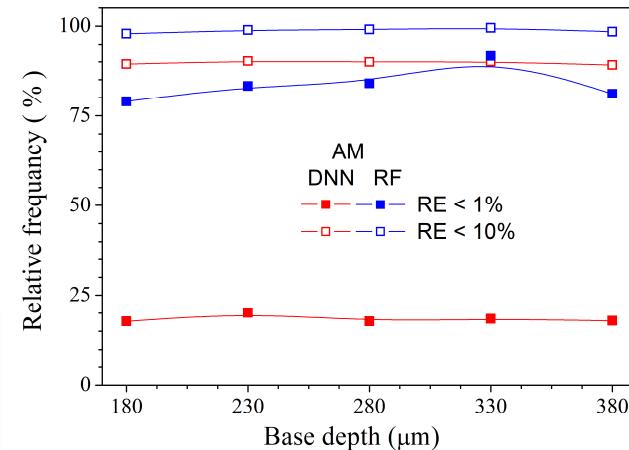
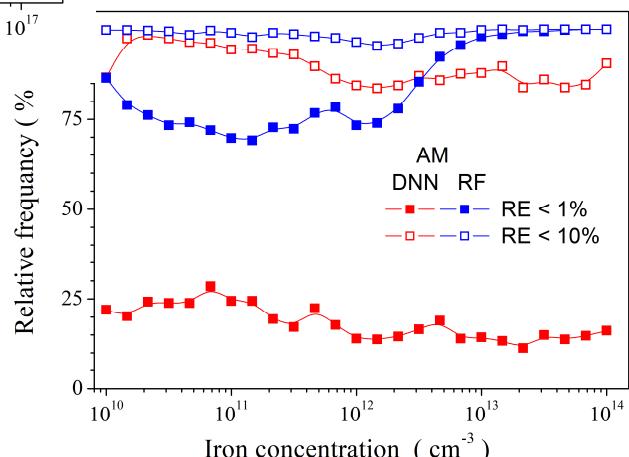
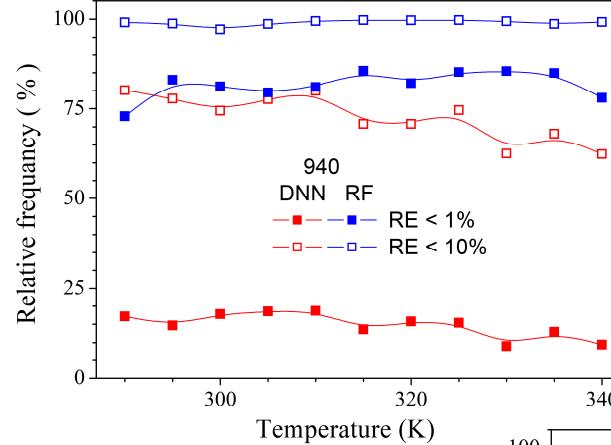
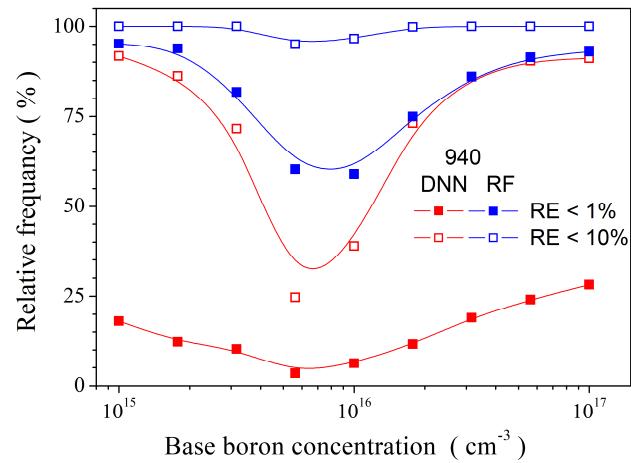
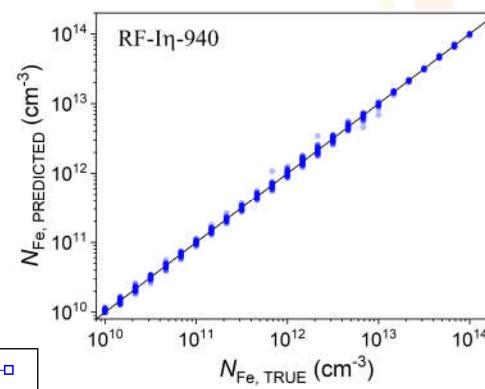
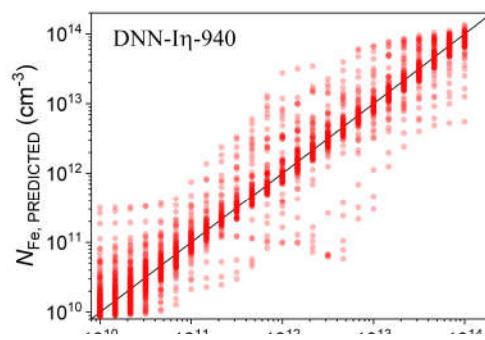
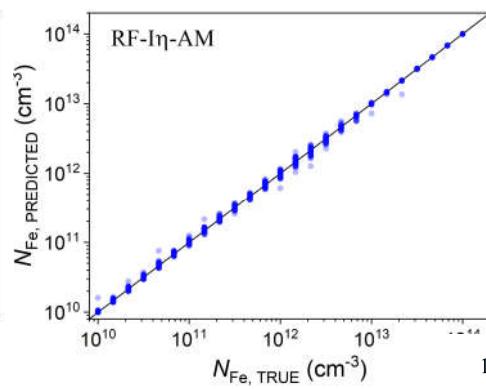
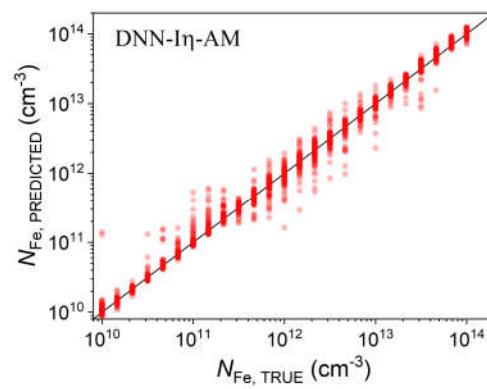


train dataset, 4 descriptors, 940 (monochromatic illumination)



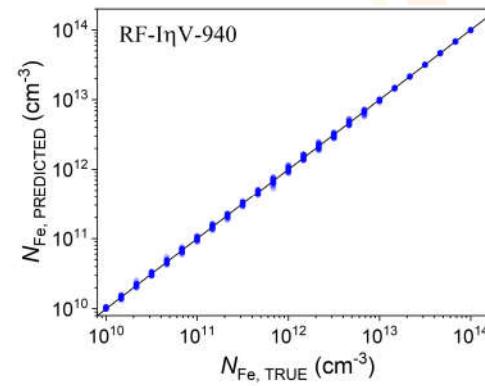
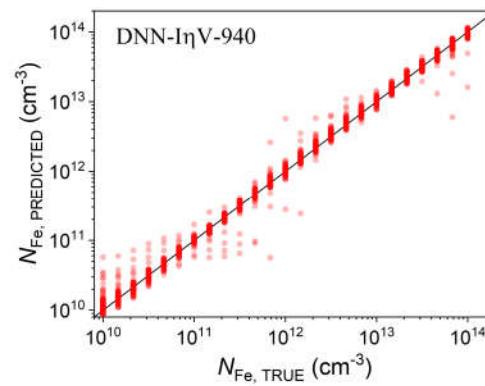
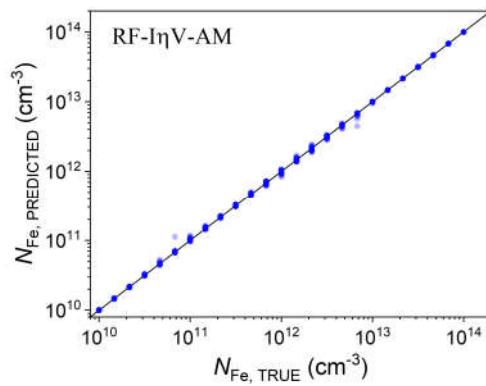
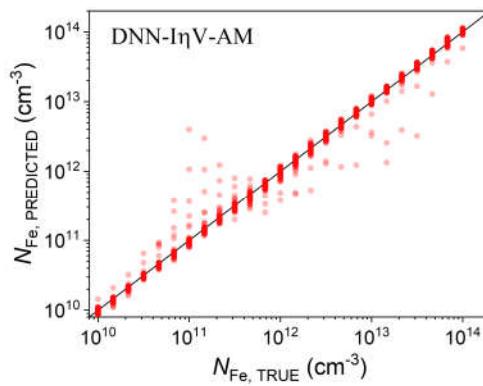


train dataset, 5 descriptors

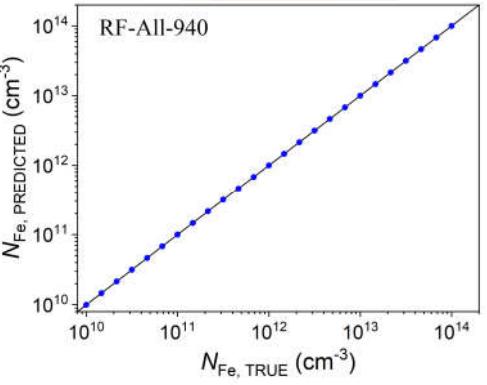
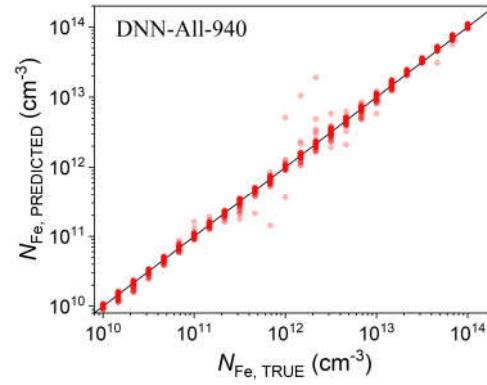
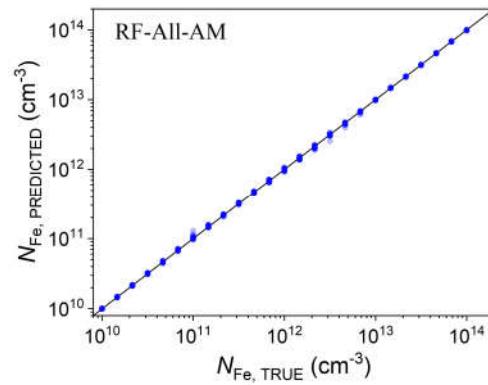
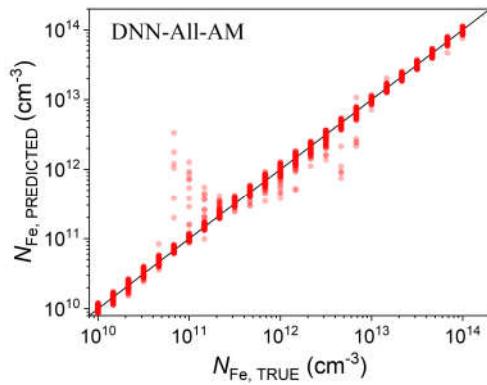




train dataset, 6 descriptors

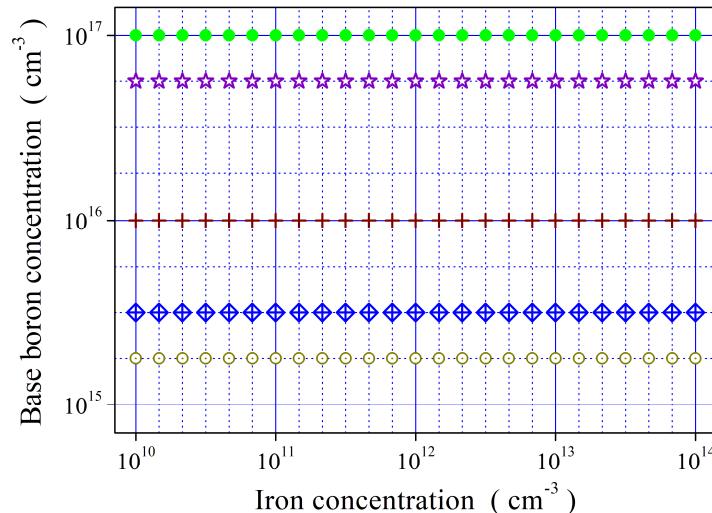
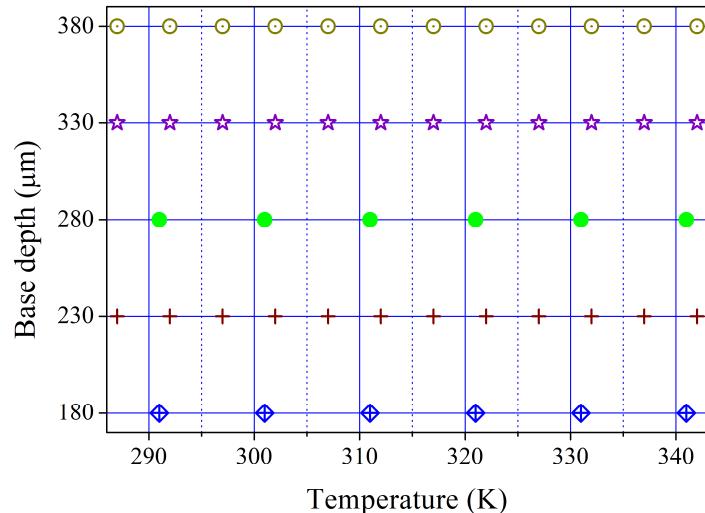


train dataset, 7 descriptors

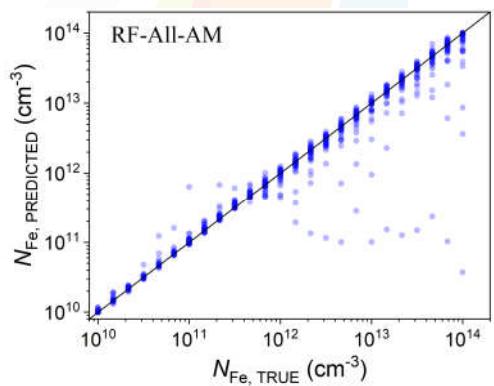
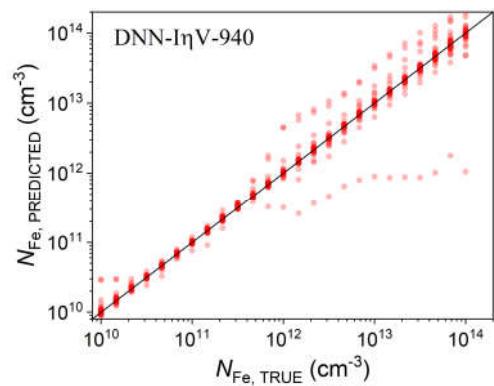
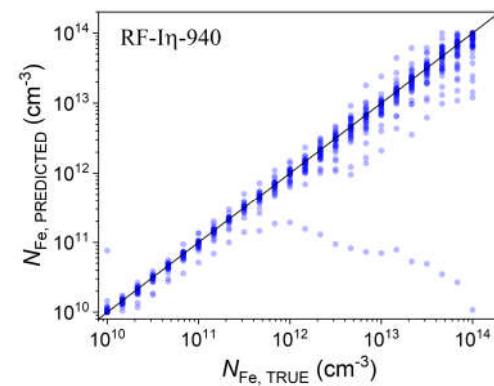
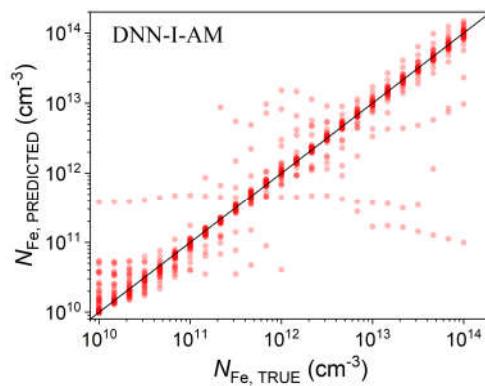




test T-varied dataset



1 200 samples

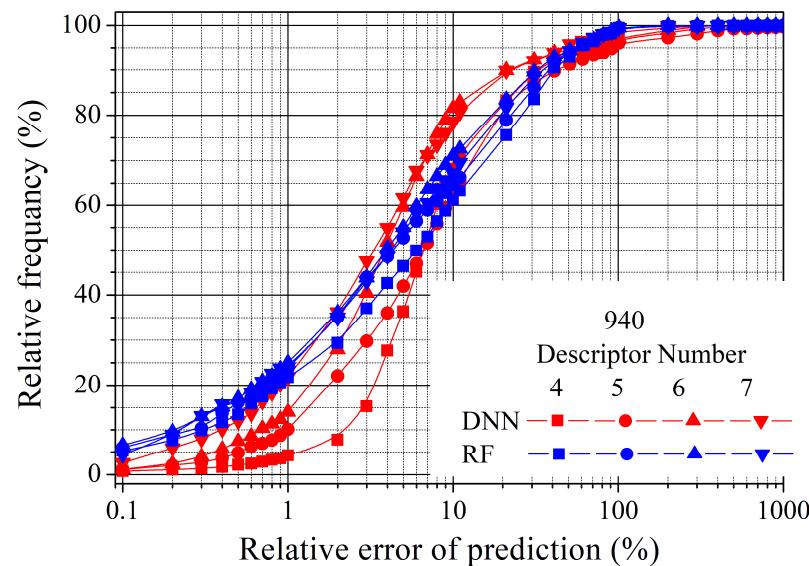
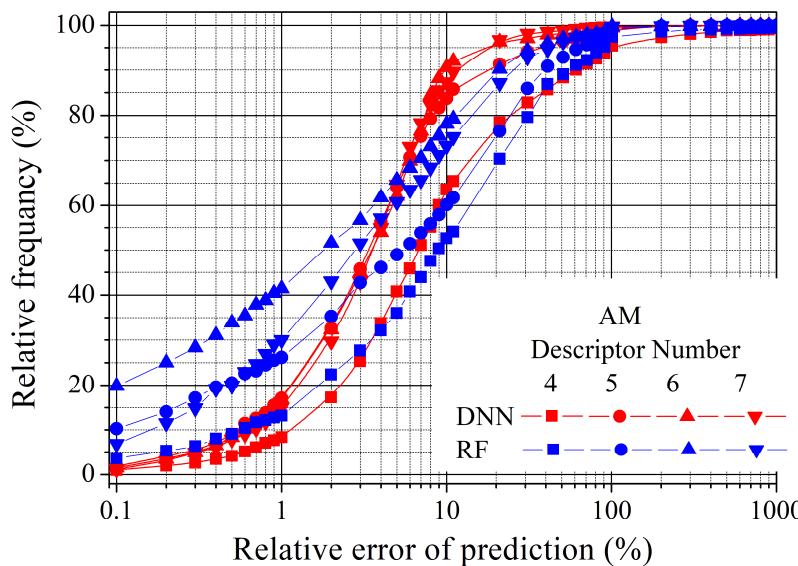




test T-varied dataset

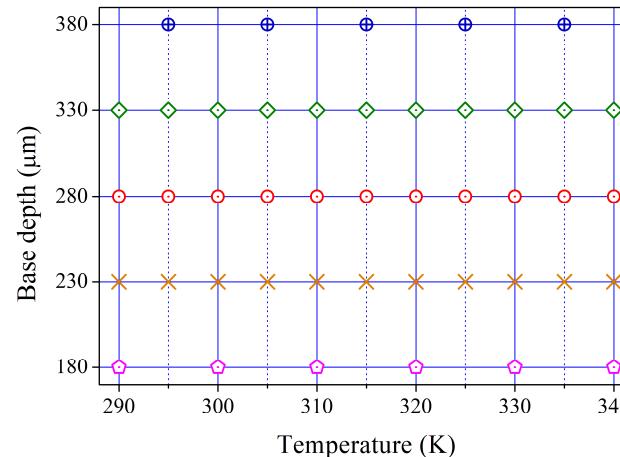
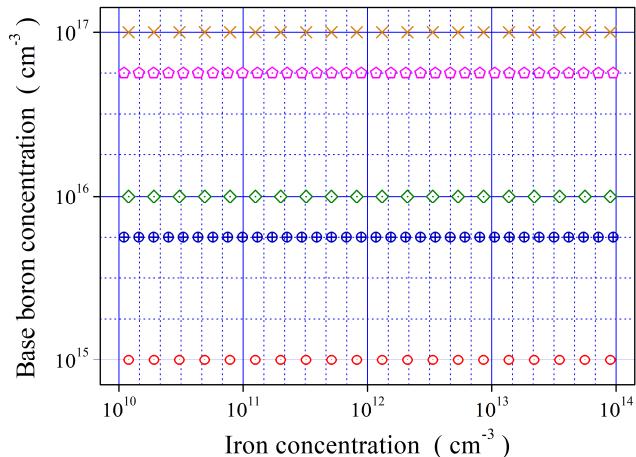


	MSE, 10^{-3}		MRE (%)		R^2	
	AM	940	AM	940	AM	940
DNN-I	64	54	40	20	0.894	0.827
DNN- $I\eta$	30	51	9	28	0.888	0.835
DNN- $I\eta V$	17	18	7	13	0.968	0.831
DNN-All	3	13	6	11	0.987	0.952
RF-I	78	71	145	27	0.876	0.839
RF- $I\eta$	53	54	15	14	0.877	0.835
RF- $I\eta V$	44	29	11	12	0.912	0.837
RF-All	39	28	10	13	0.892	0.858

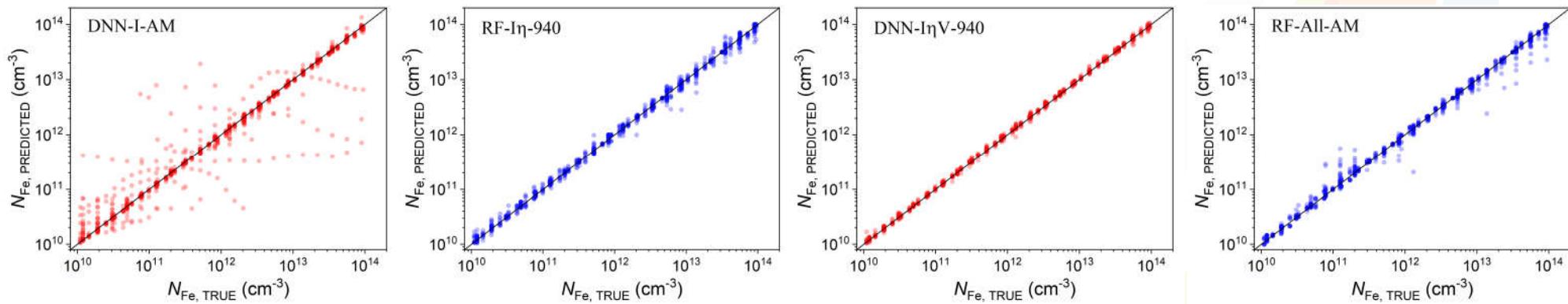




test Fe-varied dataset



1 034 samples

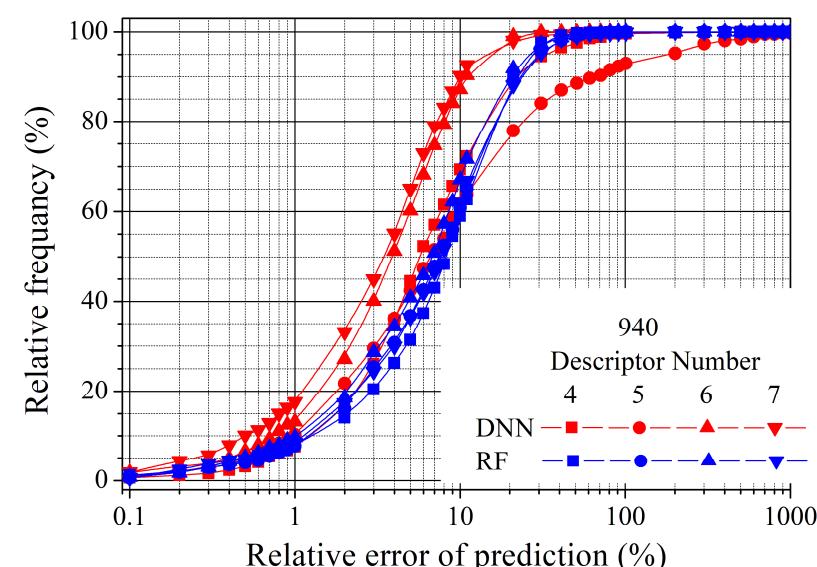
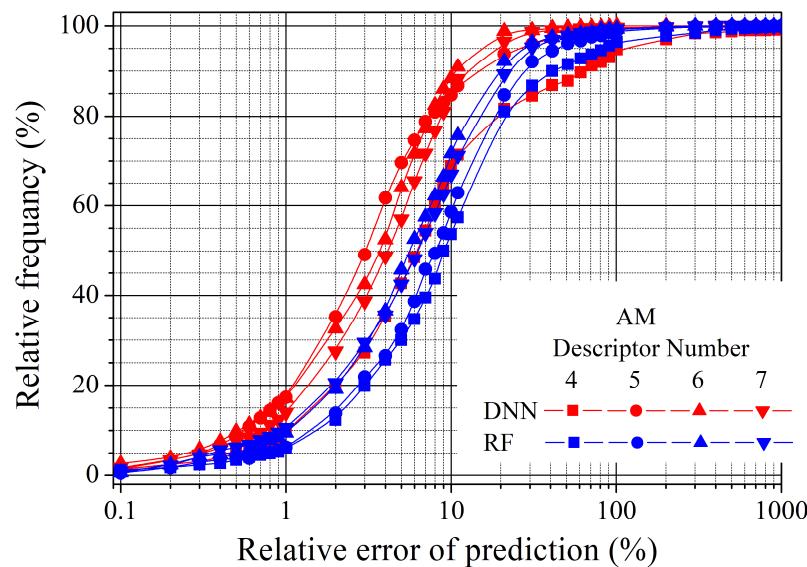


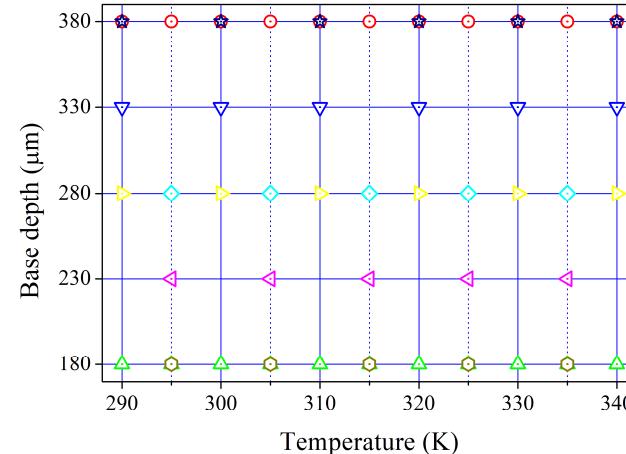
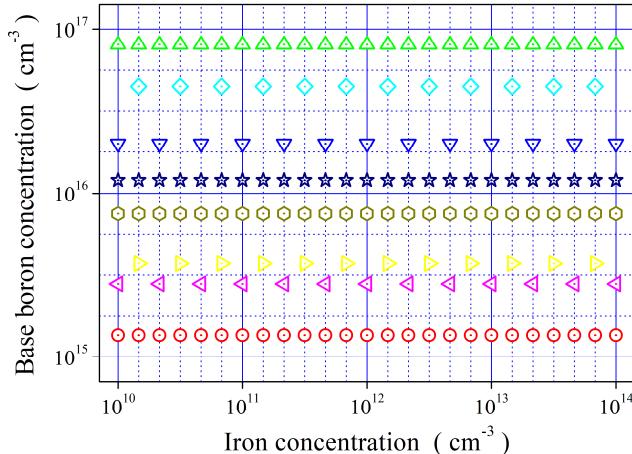
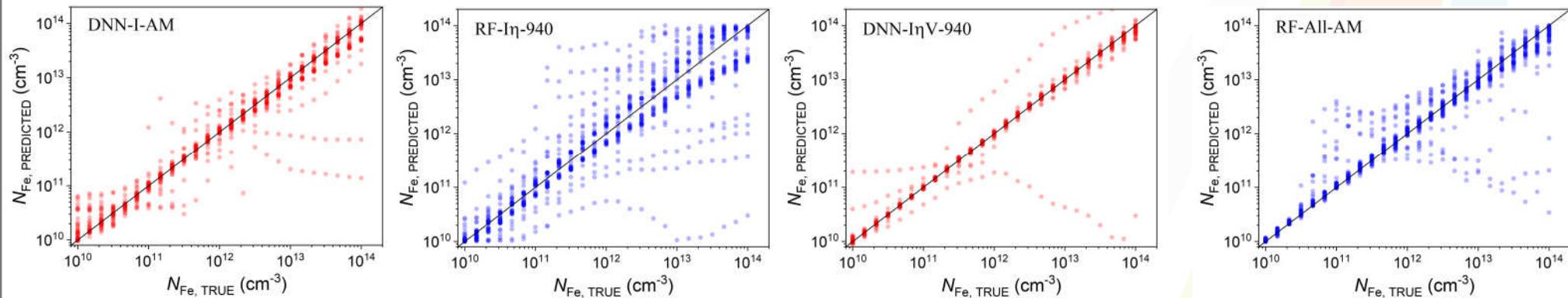


test Fe-varied dataset



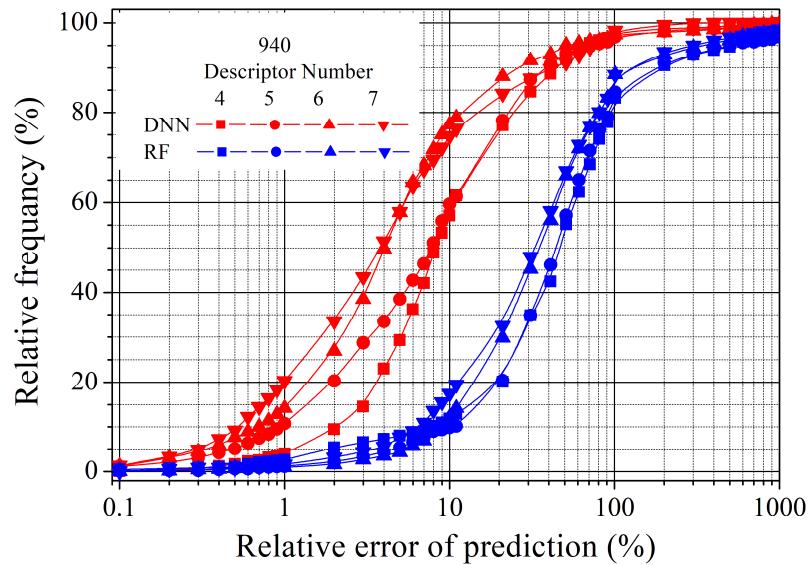
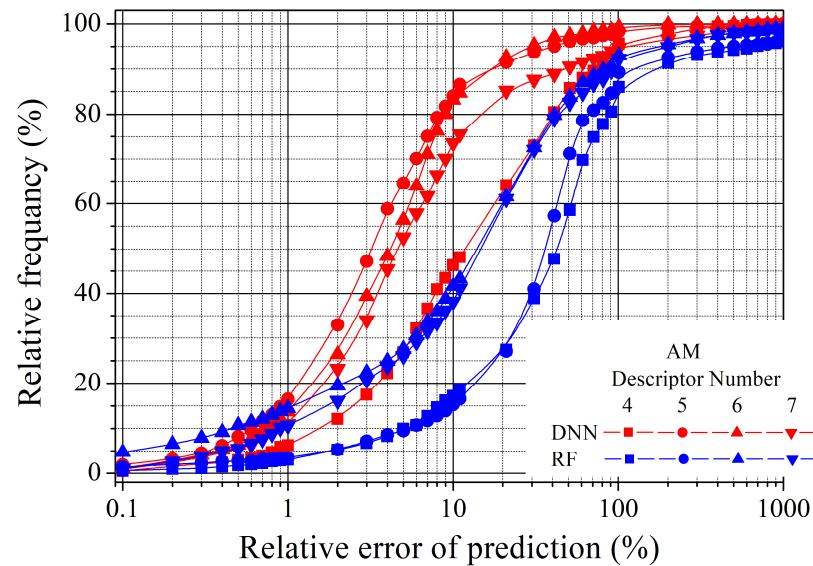
	MSE, 10^{-3}		MRE (%)		R^2	
	AM	940	AM	940	AM	940
DNN-I	58	6	53	10	0.905	0.977
DNN- η	4	33	7	36	0.988	0.881
DNN- ηV	0.9	0.6	5	5	0.992	0.993
DNN-All	5	0.8	11	5	0.990	0.988
RF-I	41	3	142	11	0.930	0.968
RF- η	10	3	15	10	0.959	0.967
RF- ηV	4	3	10	9	0.972	0.956
RF-All	5	3	11	10	0.958	0.963



**test B-varied dataset****1 100 samples**



test B-varied dataset



	MSE, 10^{-3}		MRE (%)		R^2	
	AM	940	AM	940	AM	940
DNN-I	66	63	33	23	0.788	0.743
DNN- \ln	17	70	12	32	0.919	0.649
DNN- $\ln V$	41	66	9	22	0.946	0.726
DNN-All	33	37	39	14	0.939	0.874
RF-I	221	226	204	186	0.392	0.186
RF- \ln	204	223	149	175	0.443	0.301
RF- $\ln V$	111	204	74	116	0.719	0.429
RF-All	113	174	64	93	0.755	0.486

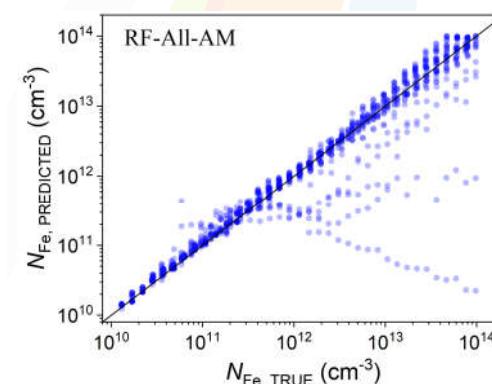
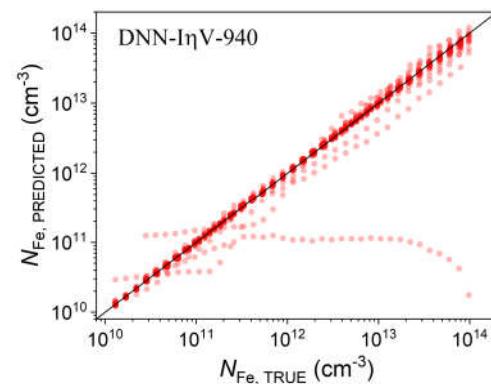
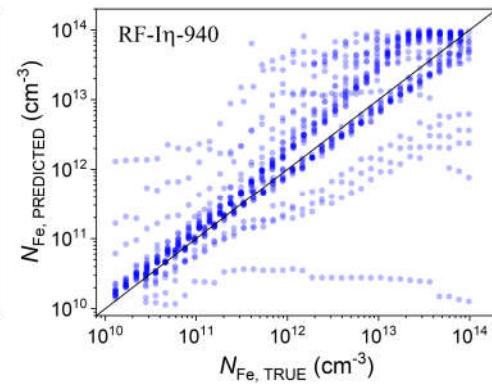
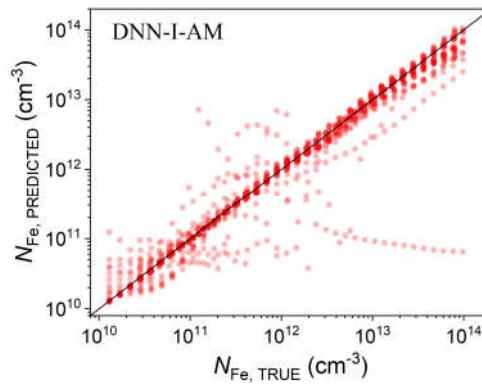
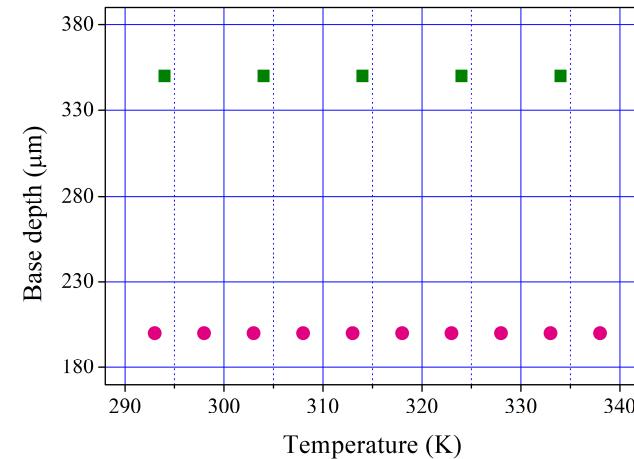
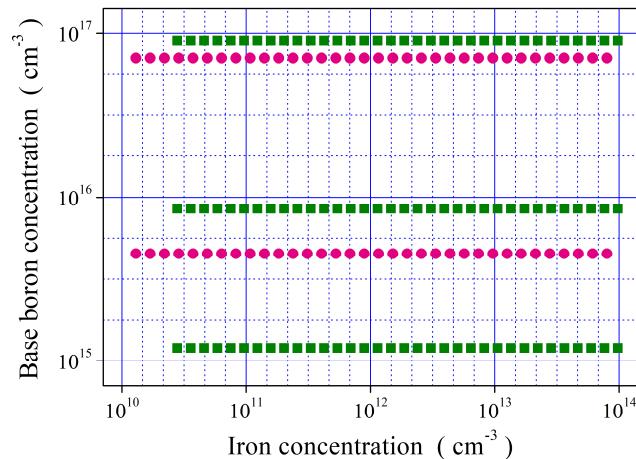




test All-varied dataset



1 190 samples

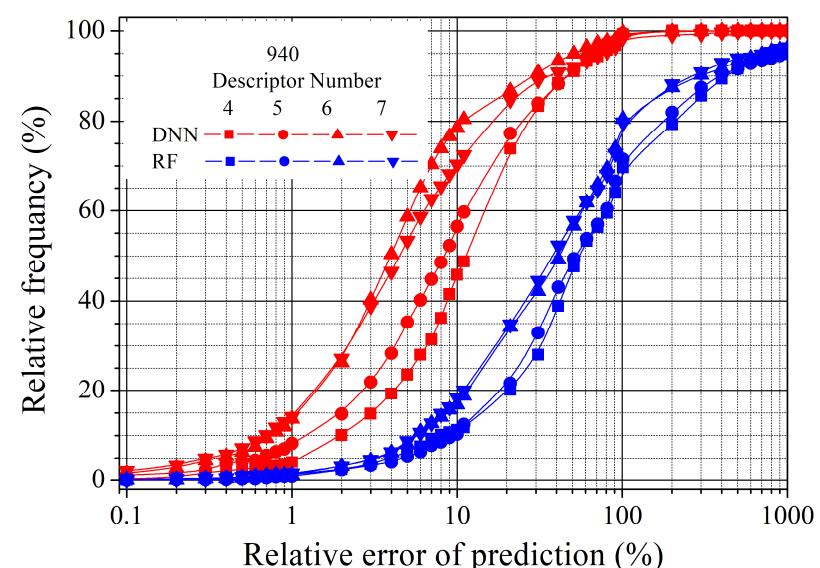
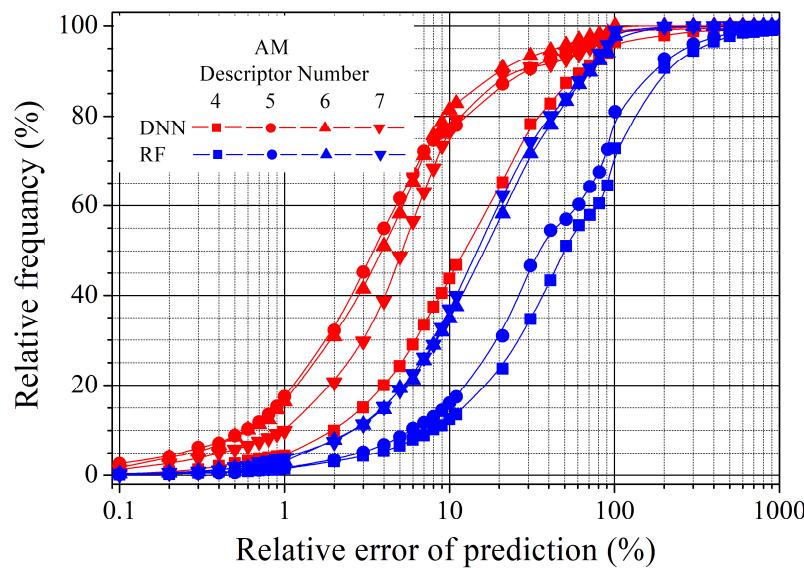




test All-varied dataset

	MSE, 10^{-3}		MRE (%)		R^2	
	AM	940	AM	940	AM	940
DNN-I	87	80	37	19	0.825	0.798
DNN-In	46	50	14	17	0.924	0.842
DNN-InV	69	61	9	11	0.929	0.907
DNN-All	62	44	17	16	0.902	0.832
RF-I	227	272	101	232	0.131	-0.304
RF-In	220	265	77	240	0.205	-0.078
RF-InV	104	232	28	187	0.724	0.288
RF-All	103	220	26	178	0.728	0.414

All-varied





CONCLUSIONS

- ✓ A new method that uses machine learning to evaluate the concentration of recombination-active impurities in solar cells has been proposed. This method is based on photoelectric parameters and has shown the ability to predict iron concentration with a mean relative error of up to 5% and an R^2 score of 0.99
- ✓ It has been found that deep neural networks outperform the random forest model for similar tasks.
- ✓ It has been shown that including open-circuit voltage in addition to short-circuit current and efficiency as descriptors has been shown to improve prediction accuracy. Additionally, it is crucial to create a training set that accurately reflects the base doping levels of real structures.



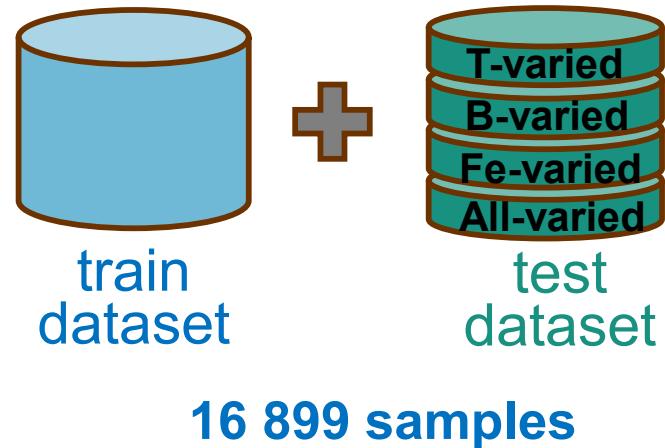
**THANK YOU
FOR YOUR
ATTENTION**



olegolikh@knu.ua



full dataset



Results of 5-fold cross-validation

		Mean squared error (MSE), 10^{-3}			
		Number of descriptors			
		4	5	6	7
DNN	AM	44 ± 7	8 ± 1	4 ± 2	3.0 ± 0.8
	940	15 ± 5	11 ± 3	10 ± 4	2.0 ± 0.7
RF	AM	39 ± 3	19 ± 1	10 ± 1	9 ± 1
	940	12 ± 2	9 ± 2	7 ± 2	8.9 ± 0.7

