IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 25, 2020, accepted June 11, 2020, date of publication June 24, 2020, date of current version July 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3004519

Prediction of the Electrical Strength and Boiling Temperature of the Substitutes for Greenhouse Gas SF_6 Using Neural Network and Random Forest

HAO SUN, LUQI LIANG, CHUNLIN WANG, YI WU, (Senior Member, IEEE), FEI YANG, (Senior Member, IEEE), AND MINGZHE RONG, (Senior Member, IEEE)
State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University, Xi'an 710049, China

Corresponding author: Yi Wu (wuyic51@xjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51707144 and Grant 51877165, in part by the Shaanxi Province Key Research and Development Program under Grant 2019ZDLGY18-05, and in part by the State Key Laboratory of Electrical Insulation and Power Equipment under Grant EIPE19302.

ABSTRACT Finding substitutes for sulfur hexafluoride (SF_6), a gas with extremely high global warming potential, has been a persistent effort for years in the field of high voltage power equipment, which focuses on the evaluation of the electrical strength and boiling temperature for the practical purpose. Following up the previous proposed linear regression models, this work introduces machine learning algorithms including artificial neural network (ANN) and random forest (RF) as the potential approaches to predict the electrical strength and boiling temperature. Based on a series of descriptors derived from the molecular structure of 74 molecules, the performance of three different methods: multiple linear regression, artificial neural network and random forest are compared and assessed in terms of the sensitivity to the sample size, prediction accuracy and stability, and the interpretability of predictors. Considering the available data are limited, random forest shows superior performance with higher robustness and efficiency. The same approaches were applied to the boiling temperature and random forest produced better results as well. Besides, the variable importance ranked by RF improves understanding of the correlation between the molecular properties and electrical strength. It provides important insights to analyze the properties of the SF_6 substitutes during the design and synthesis of the new eco-friendly gases in power equipment.

INDEX TERMS SF_6 substitutes, artificial neural network, random forest, electrical strength, boiling temperature.

I. INTRODUCTION

Sulfur hexafluoride (SF_6), a synthetic gas first commercially used in 1947, has been used on many occasions including the power system, which brought great revolution in the field of high voltage power equipment since 1967 [1] due to its high electrical strength and arc quench ability. However, SF_6 can effectively absorb the infrared radiation, especially the wavelength around 10.5 mm. Additionally, different from the natural greenhouse gas like CO_2 or CH_4, the chemical dissociation and photolysis of SF_6 are difficult and its lifetime could reach 3200 years. Therefore, its impact on the environment can be accumulated and the global warming potential (GWP) of SF_6 is evaluated to be 23500 times of CO_2. In 1995, this high GWP value of SF_6 was widely known [2] and it is listed as one the 6 green gases in Kyoto Protocol [3].

Since 1997, considerable efforts have been made to reduce SF_6 emission by improving the sealing of the device, developing the gas processing or designing the recycling system. The yield of SF_6 was decreased from 9000 tons per year in 1996 to around 1000 tons per year now, despite the increasing installed capacity of power equipment. Based on the statistics from US National Oceanic and Atmospheric Administration (NOAA) in 2009 [4], the atmospheric concentrations of SF_6 could reach 50 ppt with an increasing rate of 8.7%,

The associate editor coordinating the review of this manuscript and approving it for publication was Xue Zhou.

leading to the global warming by 0.02°, which is considerably lower than CO_2 (0.8°).

However, considering its long lifetime, reducing the usage of SF_6 is quite necessary. Much research has been done to seek for an alternative to substitute SF_6 in the field of high voltage power equipment [5]–[9]. In the case of gas insulated switchgear (GIS), gas insulated line (GIL), gas insulated transformer, high voltage circuit breaker (HVCB), etc., the crucial parameter is the electrical strength [10]. In the meantime, it is necessary to consider the boiling temperature in practical use because the power equipment could be installed in low temperature area where the risk of liquefaction must be taken into account. This problem can be solved by mixing other gases to some extent, but we still want to identify gases with lower liquefaction temperature, as well as higher electrical strength, for the applicability in a wider range of industrial occasions. Therefore, this work proposes potent approaches to predict both the electrical strength and boiling temperature [11] of the SF_6 substitutes based on machine learning algorithms including artificial neural network (ANN) and random forest (RF). It should be emphasized that the arc quenching performance is also an important parameter in some power equipment like the switch gears [12]. However, the arc quenching is quite complicated and randomly influenced by several behaviors coupling with each other, like fluid dynamics, electro-magnetic field, radiation, particle transport and interaction between the arc plasma and the material. Yet there is no certain criterion to assess the arc quenching performance. Therefore, this work just focuses on the electrical strength that is the key parameter in all kinds of power equipment.

II. MOTIVATION FOR ADOPTING ANN AND RF

For the topic of SF_6 alternatives, natural gases were first investigated, like N_2 and CO_2. N_2 molecule has very large vibrational cross sections of the electron impact, which could reduce the electron energy efficiently by the collisions between the electrons and molecules. Nevertheless, unlike SF_6, CO_2, CH_4 or O_2, N_2 cannot attach the electrons to form the negative ion, making its electrical strength only 0.38 of SF_6. Therefore, SF_6/N_2 mixture seems to be a choice due to the synergistic effect. Still, industries do not tend to use this mixture since the electrical strength would saturate (around 0.85 of pure SF_6) when the ratio of SF_6 in this mixture is higher than 50% [13]. Additionally, mixed gas leads to the difficulty of the recycling and purification process. CO_2 is another option and has been used in the real products by ABB company [14]. Despite its relatively low electrical strength (0.35 of SF_6), its dissociation products, like CO and O_2, are to the benefit of the arc quench performance. In general, although N_2 and CO_2 are potential, the inferior electrical strength to SF_6 limits their applications in many occasions, because increasing the size of the equipment to enhance the insulation is contradictory to the demand of miniaturization nowadays.

Lately, many fluorinated gases were introduced as the insulation medium with low GWP. These gases have similar or even superior electrical strength to SF_6. Since 2015, the synthetic gases announced by 3M company, known as $C_5F_{10}O$ and C_4F_7N have drawn close attention [15], [16]. It indicates that by properly designing the molecular structure, we can invent potent gases to replace SF_6.

In this case, the problem exists that what the relationship is between the molecular structure and the electrical strength or the boiling temperature. As mentioned by Rabie *et al.* [16], the breakdown experiment could be quite expensive, considering the expense for the gas manufacturing. Theoretical analysis also seems not feasible because the breakdown process includes the electron diffusion, drifting and reactions, which also need the experimental data of electron impact cross sections to obtain accurate results. A simple way to predict the properties of the gases even before it is manufactured is necessary to efficiently seek for the SF_6 replacement in the future.

The molecular properties can determine the electrical strength and boiling temperature in many ways, and establishing an empirical model is considered regardless of the detailed physical mechanism. As in [17], many previous works indicated that some of the molecular properties can influence the electrical strength and boiling temperature to some extent [18]–[22]. The potent properties could be the integrated optical absorption spectrum [21], polarizability, ionization energy [16], other properties presented in [22] and the positive surface area in our previous work [23]. Rabie *et al.*'s work [16] is quite promising by testing 67 molecules and the best correlation coefficients R^2 were 0.71 and 0.92 for polar molecules and nonpolar molecules, respectively. In their work, the fitting formula for polar molecules and nonpolar molecules are different, and for all the 67 molecules, R^2 drops down to around 0.5. In Yu *et al.*'s work [22], 43 molecules were tested and the correlation coefficient R^2 could reach 0.985. However, after we tested all the 67 molecules in [16] based on Yu *et al.*'s work, R^2 decreased to 0.67, which will be discussed in detail in later sections. Therefore, to get the acceptable R^2 in different conditions, the fitting formula should be deliberately modified.

In the past few decades, the nonlinear statistical methods based on machine learning algorithms such as artificial neural network (ANN), support vector machine (SVM) and random forest (RF) have already been developed and used in the field of electrical engineering in some cases [24]–[34]. If the machine learning algorithms can be used to learn the relationship between the electrical strength or boiling temperature and the molecular properties, it could achieve better efficiency and accuracy with good stability. Additionally, the physical mechanism of how the molecular properties influence the electrical strength has already been discussed in Rabie *et al.*'s work [16], and thus this work mainly focuses on the prediction methods and results by ANN and RF.

III. CALCULATION METHOD

A. MOLECULAR PROPERTIES AS 'PREDICTORS'

This work follows and compares the results by Rabie *et al.* [16] and Yu *et al.* [22] Several properties were investigated as 'predictors' to evaluate the electrical strength or boiling temperature of various gases. In [16], the predictors were more straightforward and included those can be directly derived based on the molecular structure, which were: polarizability α, dipole moment μ, vertical ionization energy ε_{iv}, adiabatic ionization energy ε_{ia}, vertical electron affinity ε_{av}, adiabatic electron affinity ε_{aa}, molecular mass m, electron number N_e, highest occupied molecular orbital energy *HOMO* and lowest unoccupied molecular orbital energy *LUMO*. In [22], the electrostatic potential surface $V_s(r)$, together with the molecular properties (including the surface area A_s, polarizability α and electronegativity χ defined as $(-E_{LUMO}-E_{HOMO})/2$), were used based on a series of equations: the statistical variance of the surface potential as in (1), the average deviation Π of $V_s(r)$ as in (2), the parameter ν based on (1) as in (3).

$$s_{\text{tot}}^2 = s_+^2 + s_-^2 = \frac{1}{n}\sum_{i=1}^{n}\left[V_s^+(r_i) - \overline{V}_s^+\right]^2 + \frac{1}{m}\sum_{i=1}^{m}\left[V_s^-(r_i) - \overline{V}_s^-\right]^2 \quad (1)$$

$$\Pi = \frac{1}{n+m}\sum_{i=1}^{n+m}\left|V_s(r_i) - \overline{V}_s\right| \quad (2)$$

$$\nu = \frac{s_+^2 s_-^2}{\left(s_{\text{tot}}^2\right)^2} \quad (3)$$

where the positive and negative symbols on $V_s(r)$ refer to the positive and negative potentials, respectively, n and m refer to the numbers of the potentials.

In our previous work involving 36 gases, it was found that the Pearson correlation coefficient (PCC) [35] between the positive surface area and the electrical strength could reach 0.905. Therefore, in section IV-B the final prediction model includes the positive surface area, as well as the molecular volume V_m. The specific calculation method will be introduced in the next section. The Pearson correlation coefficient (PCC) can measure a linear relationship between two variables x and y. It is formally defined as in (4).

$$r = \frac{\sum(x_i - \bar{x})\sum(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \quad (4)$$

where $\bar{x}, \bar{y}$ denotes the mean of two variables. Namely, r is calculated by covariance of x and y divided by the product of their standard deviations.

B. CALCULATION OF THE MOLECULAR PROPERTIES

To ensure the consistency of the data, we used the same data of predictors as in Rabie's work. However, Yu's work only includes 43 gases and here the data of electrostatic potential surface $V_s(r)$ not mentioned in that work were calculated based on density functional theory (DFT), in which the molecular structure was in the electronic ground state [16] and optimized by Gaussian09 program with the B3LYP function and 6-311g∗ set [36]–[38]. Then the electrostatic potential distribution plotted by [39] as in Fig. 1 can be derived. It is also necessary to define the molecular surface as in the blue line of Fig. 1, and in this work the van der Waals surface was adopted [40]. Then equation (5) rigorously defines the electrostatic potential around the molecular surface.

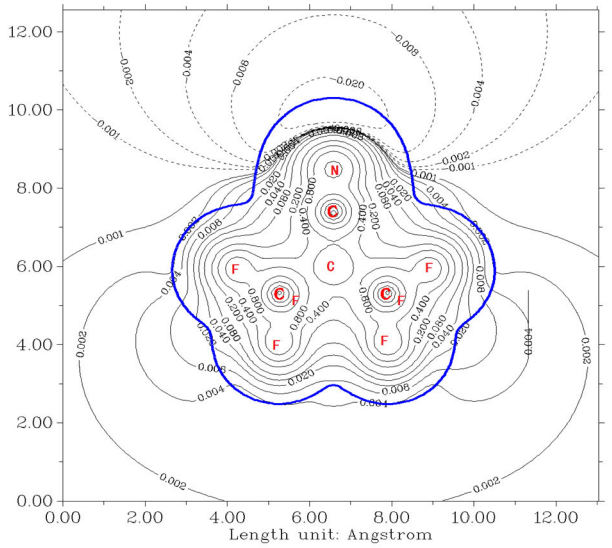

FIGURE 1. Electrostatic potential distribution and molecular surface of C_4F_7N (blue line is the molecular surface).

$$V_s(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\rho(r')\,dr'}{|r' - r|} \quad (5)$$

where Z_A is the charge of nucleus A, R_A is the location of nucleus A, and ρ is the electron density varying with the location. Additionally, the total surface area A_s and the molecular volume V_m were both derived from the molecular surface in Fig. 1.

As mentioned above, the positive surface area (*PA*) has a good correlation with the electrical strength. In this work, the positive surface area is the integration of the area around the molecular surface, where the value of electrostatic potential is positive. Correspondingly, the negative surface area can be calculated. However, the PCC between the negative surface area and the electrical strength is only 0.32 as in [23], which will not be considered here.

C. THE TOPOLOGY AND CALCULATION OF ARTIFICIAL NEURAL NETWORK

The toolbox of neural fitting (nftool) provided by Matlab was adopted, which is to solve the input-output fitting problem with a two-layer feed-forward neural network by modifying the weights and parameters to minimizing error function of validation samples. And it is easy to approach. There are

several other topologies of neural network available. Still, in this work the quantity of the sample is too small. This work is straightforward and focuses on the introduction of neural network to predict properties. Therefore, the other complicated topologies were not used. According to the introduction in Matlab, this network has sigmoid hidden neurons and linear out neurons, which can fit multi-dimensional mapping problems well and be proper to solve the current problem with several predictors as the input and the electrical strength or boiling temperature as the output.

In this work with the above toolbox, to establish the network, dozens of samples were used and they were randomly divided during the training: 80% as the training samples, 10% as the validation samples and 10% as the testing samples. It should be noted that since the samples were divided randomly, the network could differ with different training, some of which could result in low correlation. Therefore, it is necessary to choose the final network carefully considering the small sample quantity due to the lack of test results. During the fitting, the number of hidden neurons was set as 20 and two training algorithms Levenberg-Marquardt and Bayesian Regularization back propagation were tried. Although the former calculation is more efficient, the later can obtain a better result for some noisy and small sample problems. After the network was obtained, it was used to predict other samples to check its own validity.

D. THE TOPOLOGY AND CALCULATION OF RANDOM FOREST

Random forest (RF) is an ensemble learning technique derived from classification and regression trees (CART) [41]. RF shows a good performance in regression of small numbers of observations [42]. It consists of a large number of trees that are generated by the bootstrap sample of training data. At each node, the best split point of each tree is created within a randomized subset selected from all input predictors. There are two key features in RF: out-of-bag (OOB) data and variable importance. Approximately one third of the overall training data which are not included in bootstrap sample are called out-of-bag (OOB). They can be used to compute an unbiased estimation of the generalization error [41]. This measure is similar to k-fold cross-validation and can avoid overfitting. The assessment of variable importance according to the mean decrease in accuracy can also provide a more advanced understanding of the relationship between observations and response compared with linear models [42].

In this work, three parameters of RF were optimized and set as follows. The number of trees (*ntree*) was set to 1000. The number of variables to try at each split (*mtry*) and minimum size of terminal nodes (*nodesize*) were both set to 3. The randomness of RF introduced by growing trees from different bootstrap samples and by selecting *mtry* split points randomly can decrease the correlation of different trees and improve the predictive ability. The 'Random Forest' package within Matlab environment software was implemented [43]. After the random forest was constructed, other independent

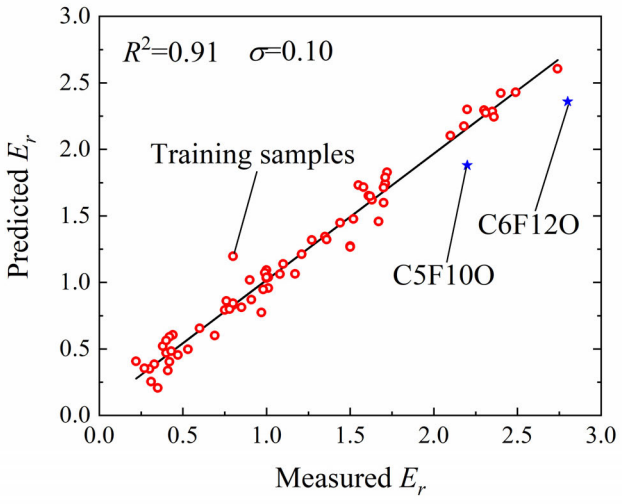

FIGURE 2. Comparison between the predicted and measured electrical strength E_r trained by 67 samples as in Rabie's work and tested by 2 samples based on ANN (Bayesian Regularization).

samples were applied to validate the generalization capability of this regression model.

IV. PREDICTION OF ELECTRICAL STRENGTH

Same as Rabie *et al.* [16], 67 molecules with test results of electrical strength relative to SF_6 (E_r) were chosen as the training samples as in Table 1. And 66 training data of boiling temperature (T_B) were completed in this work. In Table 1, the compound ID (CID) was given to avoid the confusion of isomers and can be clearly checked on the web site [44]. The test samples will be discussed in the following separated sections.

A. FOLLOW-UP OF PREVIOUS WORK

1) FOLLOW-UP OF RABIE'S WORK

In Rabie *et al.*'s work, 67 molecules were tested and divided as 18 nonpolar molecules ($\mu = 0$) and 49 polar molecules, with best R^2-values equal 0.92 and 0.71, respectively. The advantage of their work is that the fitting formula was quite simple. Still, the prediction should be done separately based on the dipole moment of molecules, which reduces the number of training samples in either group. Following their work, this work took the dipole moment μ directly as the input in the neural network, as well as the polarizability α, vertical ionization energy ε_{iv}, adiabatic ionization energy ε_{ia}, vertical electron affinity ε_{av}, adiabatic electron affinity ε_{aa}, molecular mass m, electron number N_e. The advantage of neural network is that it is not necessary to sort the gases by the dipole and we can use all the 67 gases as the samples together. As in Fig. 2, with neural network, the polar and nonpolar molecules can be trained together and the R^2-value could reach 0.91 by training the 67 samples as in Table 1.

In the nftool of Matlab, as mentioned above, during the establishment of the network, 80% of the 67 samples were chosen randomly for training and 10% were used

TABLE 1. Training set with the experimental data of E_r and T_b and all the molecular properties.

CID	Molecules	E_r(rel. SF_6)	T_b (K)	PA (Å^2)	V_m (Å^3)	$HOMO$ (eV)	$LUMO$ (eV)	α (Å^3)	μ (D)	A_s (nm^2)	σ_{tot}^2 (kcal/mol)^2	Π (eV)	ν
783	H2	0.22	20.35	20.18	18.32	-0.50	0.042	0.7	0	0.34	7.81	0.13	0.06
977	O2	0.33	90.19	27.50	33.02	-0.34	-0.133	1.37	0	0.51	33.05	0.26	0.17
947	N2	0.38	77.35	36.06	36.86	-0.51	0.017	1.65	0	0.54	10.30	0.18	0.25
948	N2 O1	0.47	184.7	35.32	47.47	-0.41	-0.002	2.76	0.12	0.67	69.05	0.43	0.15
281	C1 O1	0.4	81.68	32.20	36.97	-0.45	0.028	1.92	0.19	0.55	25.39	0.27	0.22
280	C1 O2	0.35	194.7	31.44	45.33	-0.45	-0.004	2.48	0	0.65	114.69	0.54	0.11
10039	C1 O1 S1	0.91	223.2	54.31	64.34	-0.36	-0.005	4.85	0.79	0.82	51.34	0.28	0.24
1119	O2 S1	1	263.2	36.17	57.41	-0.41	-0.137	3.68	1.58	0.76	145.71	0.77	0.18
17358	F6 S1	1	209.4	77.03	87.24	-0.54	-0.046	4.64	0	1.04	8.14	0.09	0.00
297	H4 C1	0.43	109.2	35.03	42.55	-0.46	0.004	2.47	0	0.60	7.02	0.13	0.08
6327	H3 C1 Cl1	0.31	249	42.98	64.23	-0.36	-0.003	4.2	1.9	0.82	45.44	0.55	0.25
6323	H3 C1 Br1	0.44	276.7	45.11	72.15	-0.34	-0.007	5.05	1.88	0.88	49.98	0.52	0.23
6328	H3 C1 I1	0.97	315.6	51.02	84.02	-0.31	-0.015	7.16	1.68	0.98	45.19	0.46	0.21
6345	H2 C1 F2	0.3	221.6	34.04	51.30	-0.43	0.001	2.64	1.85	0.71	86.08	0.70	0.21
6344	H2 C1 Cl2	0.6	313.2	43.15	86.06	-0.37	-0.006	6.21	1.61	1.02	111.55	0.48	0.10
6373	H1 C1 F3	0.27	191	27.66	56.17	-0.49	0.001	2.78	1.54	0.76	161.08	0.55	0.05
6212	H1 C1 Cl3	1.67	334.9	51.74	107.44	-0.38	-0.012	8.29	1.04	1.19	110.38	0.33	0.05
6372	H1 C1 F2 Cl1	0.42	313	36.89	73.53	-0.41	-0.001	4.56	1.38	0.91	154.51	0.44	0.06
6370	H1 C1 F1 Cl2	0.99	282.2	45.58	90.62	-0.39	-0.005	6.41	1.22	1.06	130.05	0.38	0.06
6393	C1 F4	0.41	144.2	40.68	61.08	-0.55	0.017	2.87	0	0.81	32.26	0.18	0.01
6392	C1 F3 Cl1	0.53	192.1	59.20	78.10	-0.43	0.011	4.62	0.31	0.96	25.20	0.17	0.03
6391	C1 F2 Cl2	1.01	243.4	61.79	94.97	-0.40	-0.007	6.48	0.36	1.10	27.60	0.18	0.08
6389	C1 F1 Cl3	1.72	297	59.00	111.67	-0.39	-0.024	8.37	0.33	1.23	28.93	0.18	0.07
6384	C1 F3 Br1	0.75	214.2	54.68	86.01	-0.39	-0.019	5.51	0.33	1.02	36.40	0.19	0.03
9625	C1 F2 Cl1 Br1	1.52	269.2	58.09	102.81	-0.38	-0.030	7.34	0.37	1.16	35.88	0.19	0.06
6342	H3 C2 N1	0.8	354.8	50.18	66.53	-0.40	-0.010	4.42	3.99	0.85	148.68	0.88	0.21
6431	C2 F6	0.78	194.2	69.48	94.12	-0.49	0.018	4.94	0	1.12	20.90	0.15	0.01
9635	C2 F3 Cl3	2.49	319	68.24	142.35	-0.39	-0.023	10.45	0.59	1.46	33.92	0.18	0.07
6428	C2 F3 Cl3	2.3	320.9	77.18	142.62	-0.40	-0.015	10.53	0.41	1.47	32.76	0.19	0.10
6594	C2 F3 Cl1	0.69	244.8	61.51	92.44	-0.33	0.006	6.26	0.42	1.10	38.64	0.25	0.11
6326	H2 C2	0.42	189.2	30.29	49.99	-0.35	0.009	3.28	0	0.67	110.79	0.54	0.16
6432	C3 F8	0.98	236.5	87.30	126.03	-0.47	0.014	7	0.13	1.38	20.42	0.14	0.01
12672	H3 C3 F3	0.8	256	55.38	95.07	-0.37	-0.013	6.14	2.43	1.12	74.25	0.49	0.18
8302	C3 F6	1.08	243.8	71.85	108.56	-0.36	-0.011	6.56	0.99	1.26	64.42	0.29	0.04
8263	C4 F8	1.27	267.1	68.92	135.86	-0.42	-0.038	8.12	0	1.47	84.27	0.25	0.00
69654	C4 F6	1.71	248.6	80.42	125.88	-0.42	-0.022	8.26	0	1.47	36.22	0.25	0.03
9640	C6 F12	2.35	323.6	145.89	194.84	-0.45	-0.022	12.15	0	1.86	11.60	0.11	0.01
5943	C1 Cl4	2.36	349.7	57.97	128.16	-0.39	-0.036	10.26	0	1.34	28.86	0.19	0.03
6365	H4 C2 Cl2	1.01	330.4	58.90	109.39	-0.37	-0.004	8.25	2.09	1.21	64.24	0.46	0.19
2775851	C4 F8	1.7	274.5	77.45	141.01	-0.39	-0.056	8.74	0	1.56	68.28	0.27	0.01
16843	C1 F3 I1	1.21	250.5	60.71	97.59	-0.35	-0.029	7.56	0.67	1.12	55.54	0.23	0.04
136213	C1 F8 S1	1.61	252.6	74.30	119.54	-0.48	-0.021	7.13	0.85	1.31	27.24	0.17	0.04
67631	C1 O2 F4 S1	1.55	251	87.98	107.30	-0.45	-0.031	6.69	1.44	1.23	55.86	0.35	0.23
24548	O1 F2 S1	1.44	229.3	35.17	64.93	-0.42	-0.029	3.94	1.54	0.84	172.97	0.61	0.14
9551	C1 N1 O2 F3	1.35	242	67.87	87.71	-0.43	-0.089	5.21	1.46	1.06	83.32	0.41	0.19
70228	C2 O2 F6	0.85	214	77.33	117.44	-0.43	-0.032	6.79	0	1.37	23.72	0.16	0.03
17607	O2 F2 S1	0.76	218	59.08	73.45	-0.46	-0.038	4.27	0.91	0.92	69.71	0.38	0.15
24558	F6 Se1	1.1	238.5	63.35	93.64	-0.55	-0.136	5.45	0	1.10	17.70	0.14	0.00
12767	C4 F6	1.7	278	80.89	118.29	-0.39	-0.020	7.65	1.51	1.36	82.12	0.39	0.05
69636	C4 F6	1.5	279	66.89	122.47	-0.31	-0.024	9.08	0	1.40	46.38	0.33	0.09
9638	C4 F10	1.36	270.5	107.79	156.99	-0.46	0.008	9.09	0	1.61	17.24	0.12	0.02
11212	C5 F8	2.2	300	89.02	148.71	-0.39	-0.035	9.77	1.16	1.59	68.48	0.31	0.03
9805	C6 F6	1	353	73.73	143.67	-0.34	-0.032	11.02	0	1.55	46.74	0.39	0.04
67738	C6 F10	2.1	326	102.45	178.91	-0.39	-0.051	11.88	0.71	1.79	51.51	0.22	0.02
6429	C2 F4 Cl2	1.71	276.6	82.03	126.71	-0.42	0.000	8.54	0.62	1.36	26.28	0.17	0.05
9775	C2 F4 Cl2	1.63	276	78.34	126.55	-0.41	-0.012	8.59	0.56	1.36	27.73	0.17	0.05
6430	C2 F5 Cl1	1.17	234.3	69.90	110.50	-0.43	0.004	6.75	0.43	1.24	22.19	0.16	0.03
6278	H3 C2 Cl3	2.31	347	61.96	130.11	-0.37	-0.006	10.27	1.84	1.35	51.94	0.36	0.16
9868	H3 C2 F3	1.62	226	46.18	79.67	-0.47	-0.002	4.54	2.31	0.98	47.25	0.56	0.20
67710	C2 N1 F3	1.58	209	78.13	81.33	-0.46	-0.008	4.8	1.35	1.01	91.03	0.33	0.24
67906	C3 N1 F5	2.18	238	107.83	113.98	-0.46	-0.020	6.91	1.39	1.31	68.57	0.26	0.25
67808	C4 N1 F7	2.74	271	132.38	145.54	-0.46	-0.026	9.03	1.51	1.56	65.21	0.23	0.25
137046	C1 N1 F5 S1	2.4	267	53.18	108.89	-0.41	-0.050	7.11	1.73	1.26	174.72	0.46	0.07
136206	C2 F6 S1	1.5	-	70.34	121.57	-0.37	-0.010	7.93	0.62	1.34	41.29	0.24	0.04
12597988	Cl3 Br1	0.9	398	82.82	128.77	-0.36	-0.100	13.69	0.04	1.43	45.50	0.27	0.03
24387	P1 Cl3	0.8	349.3	63.24	118.76	-0.35	-0.047	10.18	0.58	1.28	57.14	0.27	0.05
24807	S2 Cl2	0.4	410.1	73.87	120.52	-0.34	-0.055	11.51	1.11	1.32	37.77	0.28	0.12

for validation and test, respectively. After the network is established, it is still necessary to further test it with the samples other than 67 molecules in Table 2. In this section, the widely investigated two gases were used for test, which were $C_5F_{10}O$, commercially known as 3M Novec 5110, and $C_6F_{12}O$. The predicted E_r of the two gases were 1.88 and 2.36, whereas the measured values were 2.2 and 2.8, indicating the neural network can potentially predict the electrical strength sufficiently.

2) FOLLOW-UP OF YU'S WORK

To further test the feasibility of the ANN, we also followed up Yu *et al.*'s work [22], using the exact same training samples. Fig. 3 shows the comparison between the predicted and

TABLE 2. Test set with the experimental data of E_r and T_b and all the molecular properties.

CID	Molecules	E_r(rel. SF_6)	T_b (K)	PA (Å^2)	V_m (Å^3)	$HOMO$ (eV)	$LUMO$ (eV)	α (Å^3)	μ (D)	A_s (nm^2)	s_{tot}^2 (kcal/mol)^2	Π (eV)	ν
92944	C6 F12	2.3	-	108.32	197.36	-0.44	-0.04	10.67	0.24	1.92	48.94	0.18	0.007
10313228	C2 O1 F6	1	-	66、68	104.38	-0.49	0.01	4.85	0.04	1.21	27.55	0.16	0.02
12633773	C4 N1 F7	2.2	268	131.88	145.15	-0.45	-0.03	7.85	1.38	1.55	59.97	0.22	0.24
2782408	C6 O1 F12	2.8	322	151.30	213.82	-0.41	-0.10	12.55	0.42	2.07	45.08	0.18	0.22
12520922	C4 O1 F8	1.6	-	93.37	149.18	-0.44	0.01	7.80	0.30	1.62	54.79	0.22	0.03
14175377	C5 O1 F10	2.1	300	127.57	182.28	-0.41	-0.11	11.11	0.44	1.83	55.90	0.21	0.18
5708720	HFC 1234ze(e)	0.85	253.8	54.69	48.84	-0.31	-0.04	4.88	1.18	-	-	-	-
13129	CF3CH2F	-	247	45.17	41.68	-0.37	0.04	3.50	2.00	-	-	-	-
13550690	c-C3F4	-	253	57.02	43.67	-0.32	-0.12	4.34	1.40	-	-	-	-
136190	C2F5NF2	-	235	96.96	62.29	-0.34	-0.04	4.86	0.55	-	-	-	-
21572606	NF(CF3)2	-	236	97.37	61.74	-0.35	-0.05	4.68	0.45	-	-	-	-
24547	FOF	-	128	34.87	19.46	-0.35	-0.12	1.36	0.30	-	-	-	-
24553	NF3	-	144	47.06	25.37	-0.37	-0.05	1.87	0.19	-	-	-	-
574364	SF2NCFO	-	322	59.96	50.27	-0.33	-0.09	5.63	1.96	-	-	-	-
6368	CH3CHF2	-	248	40.17	34.50	-0.34	0.05	3.63	2.19	-	-	-	-
674	NH(CH3)2	-	281	56.46	35.85	-0.22	0.04	4.89	0.98	-	-	-	-
67716	CF3CFO	-	214	67.56	43.62	-0.36	-0.10	3.52	0.77	-	-	-	-
67925	c-C3F6O	-	235	95.28	60.91	-0.37	-0.01	4.99	0.49	-	-	-	-
9633	CF3CHF2	-	227	56.82	45.33	-0.38	0.04	3.48	1.52	-	-	-	-
9871	CF3C(CH3)O	-	295	60.20	51.39	-0.29	-0.08	5.23	2.73	-	-	-	-

measured electrical strength E_r trained by 43 samples as in Yu's work and tested by 30 other samples as in Table 1 and 2. According to Yu *et al.*, with the fitting formula as in (6), the R^2-value could reach as high as 0.985. Unfortunately, when (6) was adopted to predict the other 30 molecules, the dispersion can be clearly seen in Fig. 3(a) and the R^2-value of the 73 molecules sharply drops to around 0.67. Similar problem can be found in Fig. 3(b) based on neural network: by the 43 samples the network with the R^2-value of 0.97 can be easily achieved but the R^2-value of the 73 molecules drops to 0.68.

$$E_r = 0.36A_s^2 + 0.054\nu s_{\text{tot}}^2 - 1.97\Pi + 0.33\alpha\chi + 0.36 \quad (6)$$

As in Yu's work, when the molecules with different structures were included, for example the F-containing molecules, modifications on the fitting formula are necessary.

B. RESULTS

1) PREDICTORS

To increase the electrical strength of the insulation gas, the one important factor is the electronegativity, which requires the molecule to capture electrons during the collision. Hence, it is necessary to evaluate the possibility of the molecule to interact with the electrons. In our previous work, it was found that the positive surface area (PA) [23], indicating the total area where the value of electrostatic potential is positive as mentioned above, has a good correlation with the electrical strength. The PA has the potential to describe the possibility of collisions between molecules and electrons. Based on the 67 molecules from Table 1, its PCC could reach 0.675. PCC of molecular volume V_m is even higher as 0.754. As investigated by Rabie, the polarizability α and molecular mass m are also potential predictors, of which all the PCCs are around 0.66. Although the correlation of dipole moment μ is quite low, it is necessary to be considered to distinguish whether the molecule is polar as in Rabie's [16]. Additionally, following Yu's work, $HOMO$ and $LUMO$ are also considered in this work. In total, as in Fig. 4, 7 predictors were employed to construct the model and predict E_r. It is worth noting that the value of all 7 predictors can be calculated by density functional theory (DFT) based on the specific molecular structure for whether known molecules or new molecules designed to replace SF_6.

2) MULTIPLE LINEAR REGRESSION

Referring to the aforementioned two traditional linear regression models from Rabie *et al.* and Yu *et al.* [16], [22], in our previous work [45], we attempted to combine 7 predictors in various possible polynomial functions and finally, selected 4 predictors which had better correlation to establish a new regression model with R^2 of 0.686 as in (7)

$$E_r = 1.106V_m + 3.047\chi + 0.262PA \cdot \mu - 0.907 \quad (7)$$

where the part $PA \cdot \mu$ was introduced and only worked if molecules are nonpolar which can simplify the form of formula to avoid separating polar and nonpolar particles.

74 samples as shown in Table 1 and 2 were used to fit the formula. The results were assessed quantitatively by the root mean square error ($RMSE$) and the coefficient of determination (R^2) as shown in Fig. 5. The coefficient of determination (R^2), also called the multiple correlation coefficient, indicating the proportionate amount of the dependent variable explained by the independent variables, is defined as [46], [47]

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^{n}(P_i - M_{ave})^2}{\sum_{i=1}^{n}(P_i - M_{ave})^2 + \sum_{i=1}^{n}(P_i - M_i)^2} \quad (8)$$

where SSR is the sum of squared regression and SST is the sum of squared total. In the latter expression, P_i, M_i, M_{ave} denote predictive value, measured value and the average of measured value, respectively, and i is the number of samples.

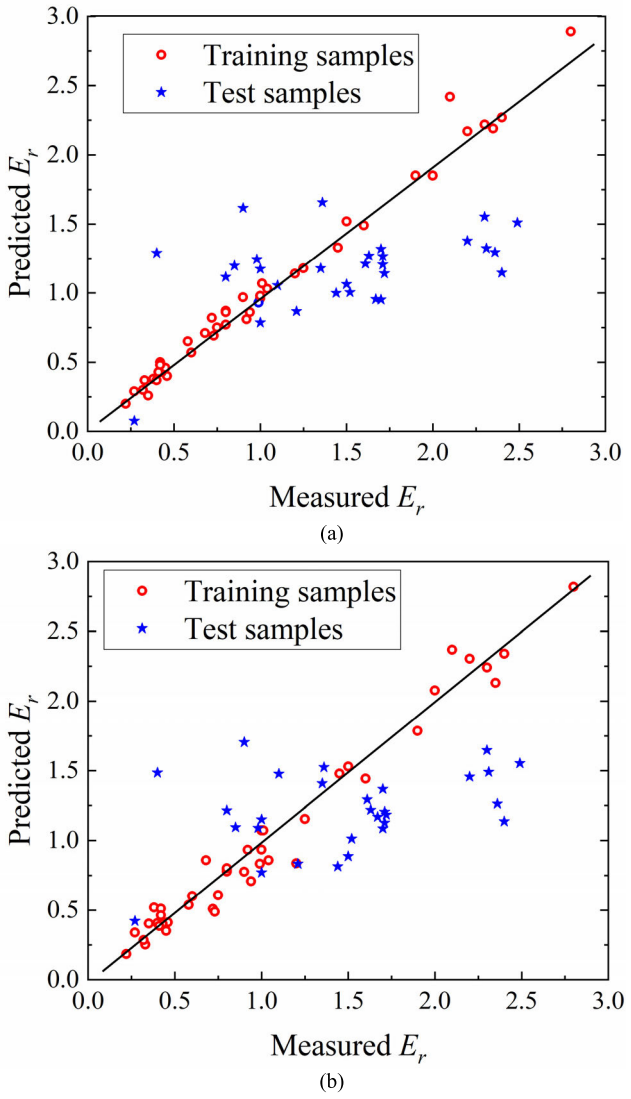

FIGURE 3. Comparison between the predicted and measured electrical strength E_r trained by 43 samples as in Yu's work and tested by 30 samples based on ANN (Bayesian Regularization).

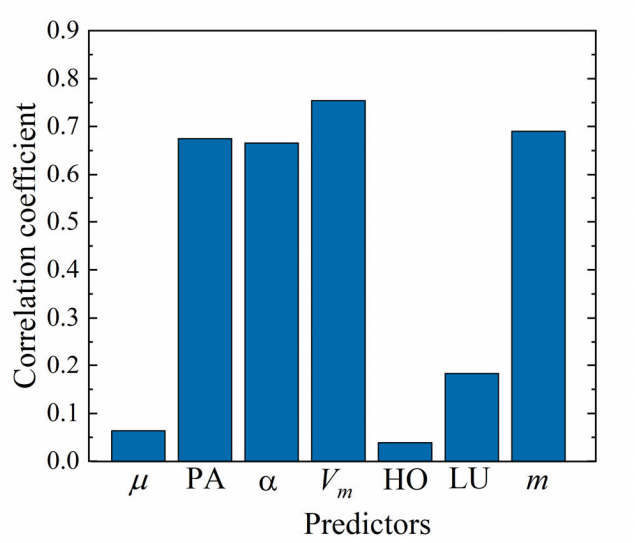

FIGURE 4. PCCs between the electrical strength and 7 predictors in this work based on the 67 molecules in Table 1.

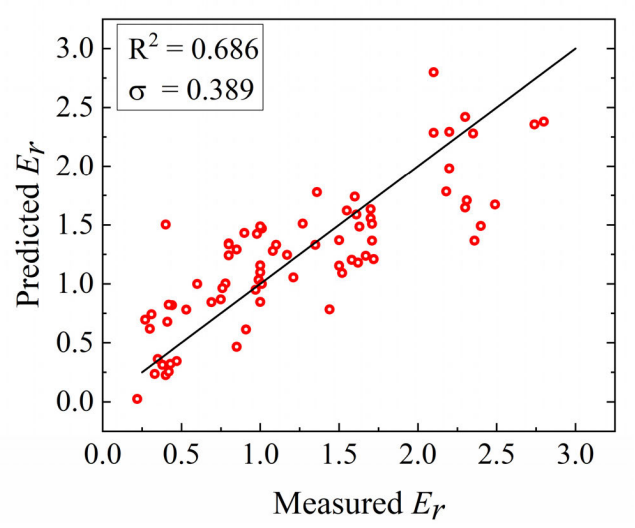

FIGURE 5. Comparison between the predicted and measured electrical strength E_r trained by 74 samples based on multiple linear regression.

3) ARTIFICIAL NEURAL NETWORKS

Two different training algorithms, Levenberg-Marquardt and Bayesian Regularization back propagation were utilized for training the network based on 7 predictors and 5 predictors without μ and *LUMO*, respectively. Reducing predictors to 5 was meant to maintain a similar procedure with the RF algorithm which will be explained in detail in later sections. For every attempt of learning, 60 samples were imported as training inputs which were randomly selected from the entire dataset including 67 samples. After running the network, 7 independent samples as in Table 2 were used as testing input to measure network generalization. Considering the error created by different initial conditions and sampling, the network was retrained for 10000 times and the average value of R^2 and *RMSE* were shown in Table 3.

It should be noted that when changing into Bayesian Regularization algorithm with 7 predictors, the R^2-value of testing samples could reach 0.803 and *RMSE* sharply drops from 0.806 to 0.376. Such results indicate that the choice of training algorithm has a great influence on the models' accuracy.

TABLE 3. Average value of R^2 and RMSE for ANN.

	R^2		RMSE	
Model	Train	Test	Train	Test
Levenberg-Marquardt				
5 predictors	0.763	0.616	0.374	0.774
7 predictors	0.779	0.586	0.366	0.806
Bayesian Regularization				
5 predictors	0.569	0.773	0.426	0.364
7 predictors	0.651	0.803	0.376	0.376

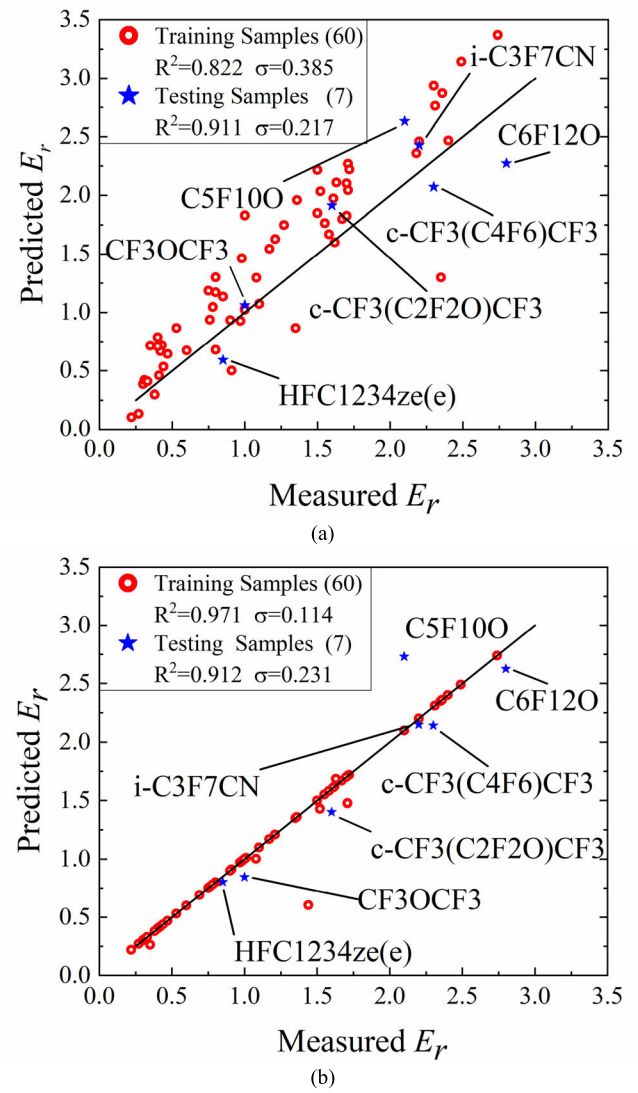

FIGURE 6. The best performance of comparison between the predicted and measured electrical strength E_r based on (a) Levenberg-Marquardt (b)Bayesian Regularization trained by 60 samples and tested by 7 samples for 7 predictors.

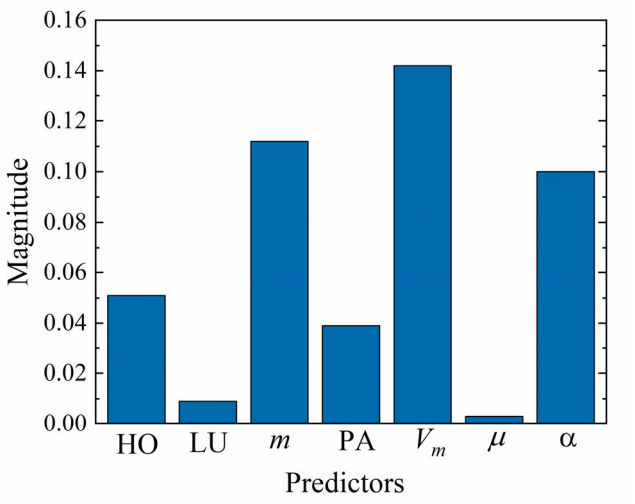

FIGURE 7. Predictor variable importance according to the mean decrease in accuracy ranked by RF between the electrical strength E_r and 7 predictors in this work based on the 67 molecules in Table 1.

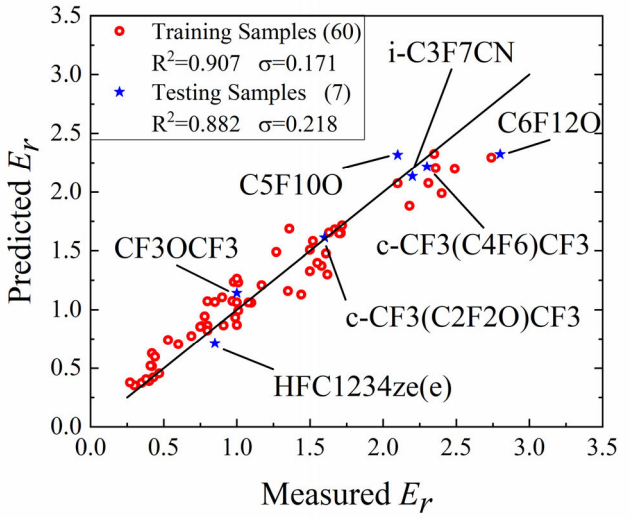

FIGURE 8. The best performance of comparison between the predicted and measured electrical strength E_r trained by 60 samples and tested by 7 samples based on RF for 5 predictors.

Nevertheless, for the same algorithm, changing the quantity of the predictor does not influence the results remarkably. In addition, Fig. 6 shows the comparison between the predicted and measured E_r of two types of networks with the best performance among 10000 networks. The predicted value is expected to be as close as possible to the measured, while the black line in the figure is y=x that represents the ideal result.

4) RANDOM FOREST

In terms of the RF, the method of calculating the average value of R^2 and *RMSE* were consistent with that of the ANN, namely executing the loop that selecting 60 samples randomly to construct the RF and testing by the same 7 independent samples for 10000 times. Furthermore, according to the variable importance ranked by RF as in Fig. 7, some predictors obtained the lower rate of the mean decrease in accuracy. This demonstrated that the correlation between them and electric strength might not be so strong. We evaluated the performance of the RF regression model by removing the weakest predictor one by one. Finally, the model eliminating two least promising variable, dipole moment μ and *LUMO*, was proved to be the most effective as shown in Fig. 8. In Table 4, averaged R^2 of testing samples with 5 predictors is as high as 0.826 whereas the value is 0.774 with 7 predictors, and the *RMSE* drops from 0.291 to 0.260.

C. DISCUSSION

A comparative analysis of the predictive performance of three different regression models, which are multiple linear regression, the ANN and RF, is discussed in this section.

The multiple linear regression and RF regression models exhibit some differences in ranking 7 predictors according to their relative contribution, and ANN is a black-box technique

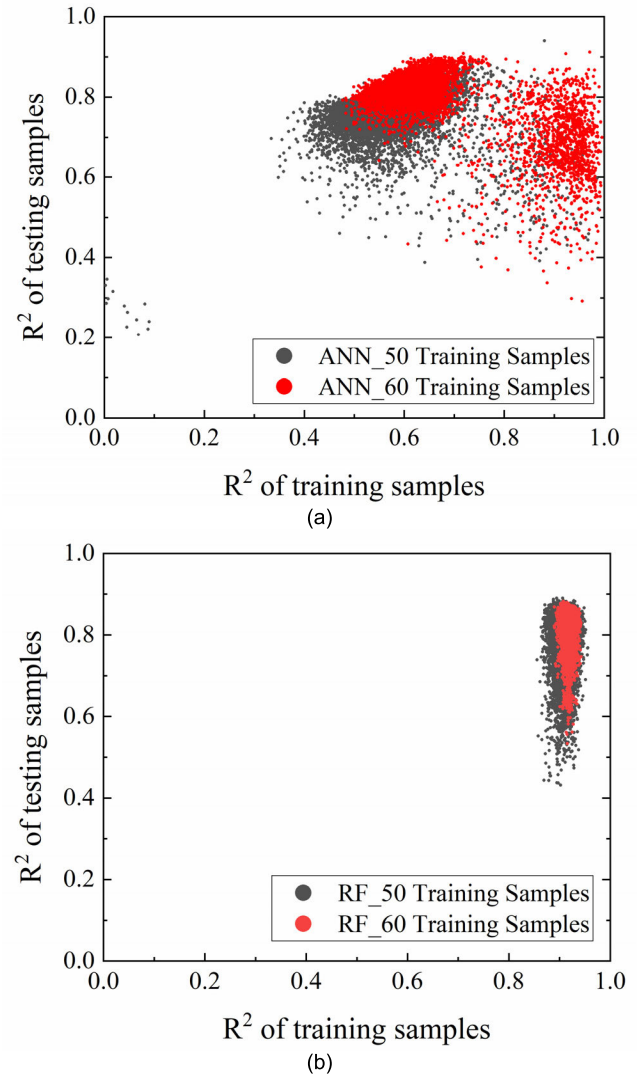

FIGURE 9. Scatter diagram (a) based on ANN (Bayesian Regularization) for 7 predictors (b) based on RF for 5 predictors compared by 50 and 60 training samples between the training and testing R^2 retrained for 10000 times.

TABLE 4. Averaged value of R^2 and RMSE for RF.

	R^2		RMSE	
Model	Train	Test	Train	Test
5 predictors	0.917	0.826	0.170	0.260
7 predictors	0.920	0.774	0.165	0.291

and does not provide information about the role of predictors in the modeling. As in Fig. 4 and Fig. 7, V_m, α and PA have good correlation and both μ and $LUMO$ perform not well. As to the one most important predictor m in the linear model, its value is lower in RF but it is still important. However, the value of $HOMO$ in the linear model is quite low but it is more significant in RF. This may suggest that a nonlinear and more complex association between $HOMO$ and electric strength exists and is better recognized by the RF model, even though the direction of the association cannot be revealed

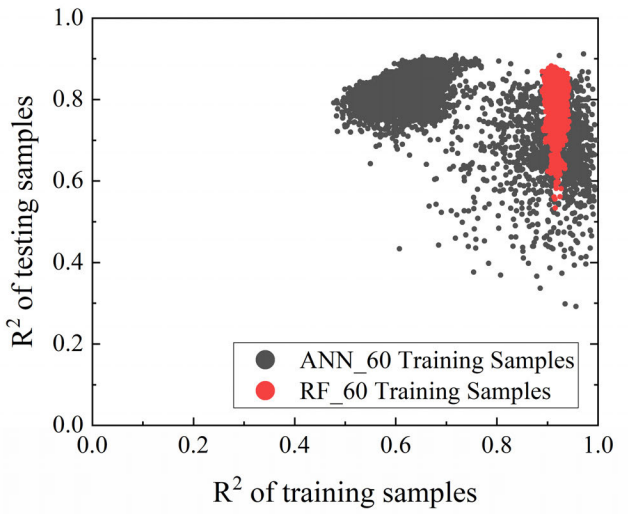

FIGURE 10. Scatter diagram compared by RF of 5 predictors and ANN (Bayesian Regularization) of 7 predictors between the training and testing R^2 retrained for 10000 times of 60 training samples.

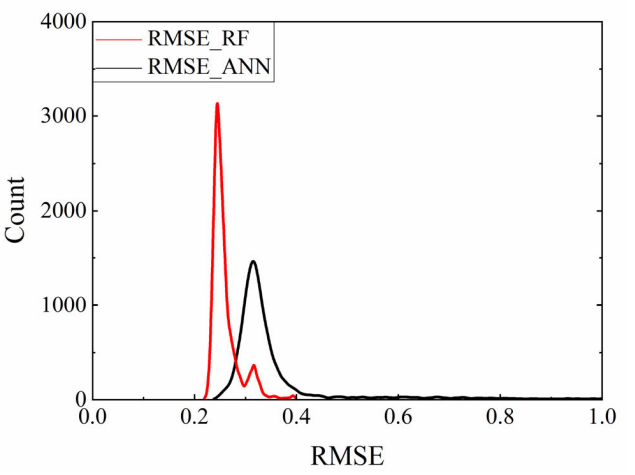

FIGURE 11. Distribution of *RMSE* compared by RF of 5 predictors and ANN (Bayesian Regularization) of 7 predictors retrained for 10000 times of 7 testing samples.

from the RF [48]. On the other hand, the robustness of the ANN responds poorly to the reduction of the number of input predictors which was described in section IV-B. On the contrary, for the RF, the smaller number of predictors offers the better predictive ability. The reason for this behavior is likely that the importance allocated to μ and $LUMO$ are quite low, and if these two weak predictors replace a related stronger predictor, it will bring the unbalance when selecting the split of trees and increase the error [42].

From the point of the sample size, Fig. 9 shows the sensitivity of ANN and RF models to the size of data. Besides the result of 60 samples, the R^2 was computed using smaller 50 training samples which were also selected randomly and the same 7 testing samples. With the decrease of sample size, the distribution of RF appears to be more scattered but that of ANN is ambiguous. It indicates that increasing sample size can significantly improve the prediction accuracy of the RF.

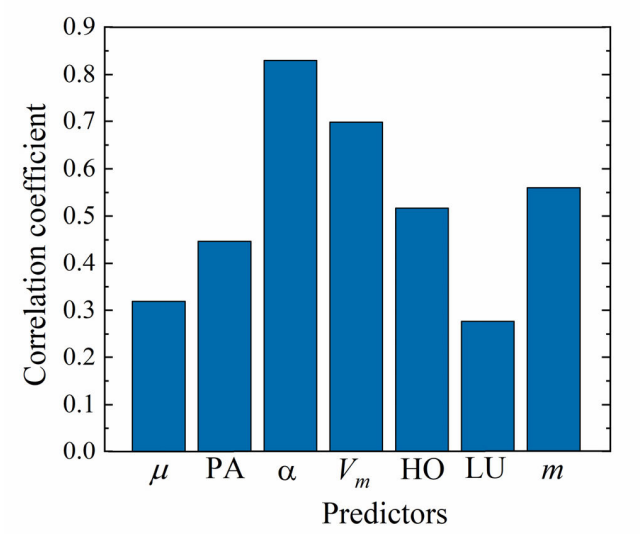

FIGURE 12. PCCs between the boiling temperature and 7 predictors in this work based on the 66 molecules in Table 1.

With regard to the assessment of generalization, on the one hand, the two nonlinear regression models produce higher accuracy than the multiple linear regression model. For further exploration, comparing Fig. 6(b) and Fig. 8, the fitting capability of two models with the best performance for ANN and RF are both excellent. However, as shown in Fig. 9, among ANN results, the R^2 of training data can reach as high as 0.95 or even higher, but it performs not as stable as RF. The average value of R^2 for testing data is 0.803 for ANN which is also lower than 0.826 of RF from Table 3 and 4. Additionally, there are a few points in the lower left corner of the figure that have worse performance, although these points only account for 0.13% of all results. This is because the neural network assigned the initial weights and biases randomly at the beginning which may cause great deviation from the measurement in ANN. Therefore, for the practical purpose, the network should be carefully selected to avoid overfitting. As for RF, the error is relatively smaller but one has to be aware that it generally tends to partly underestimate the E_r especially for larger molecules as in Fig. 8.

On the other hand, we also make statistical analysis for two methods of the R^2 distribution with the total of 10000 calculation results as in Fig. 10. Despite a part of networks show good results, the overall distribution of R^2 is scattered which demonstrates the predictive ability of ANN is not robust enough. Besides, the accuracy of training data for ANN is generally lower whereas the RF maintains a higher level both for training and testing. Similarly, a good statistical result of *RMSE* achieved by the RF is shown in Fig. 11.

To sum up, three models show distinct results. The multiple linear regression model reflects a positive but relatively weak relationship between the dependent variable and the predictors which suggests the existence of nonlinear trend. ANN could achieve high prediction accuracy, but only for a few specific cases and others are not stable enough, which is probably due to the lack of training samples since no more than 100 sets of data are available. It is promising for ANN if more observations of the electrical strength can be obtained and new predictors are found as the input. Based on the present study, the best performance of regression is implemented by RF. It is quite robust and stable with good prediction accuracy. The parameters of OOB data and variable importance make RF more effective and can also avoid overfitting.

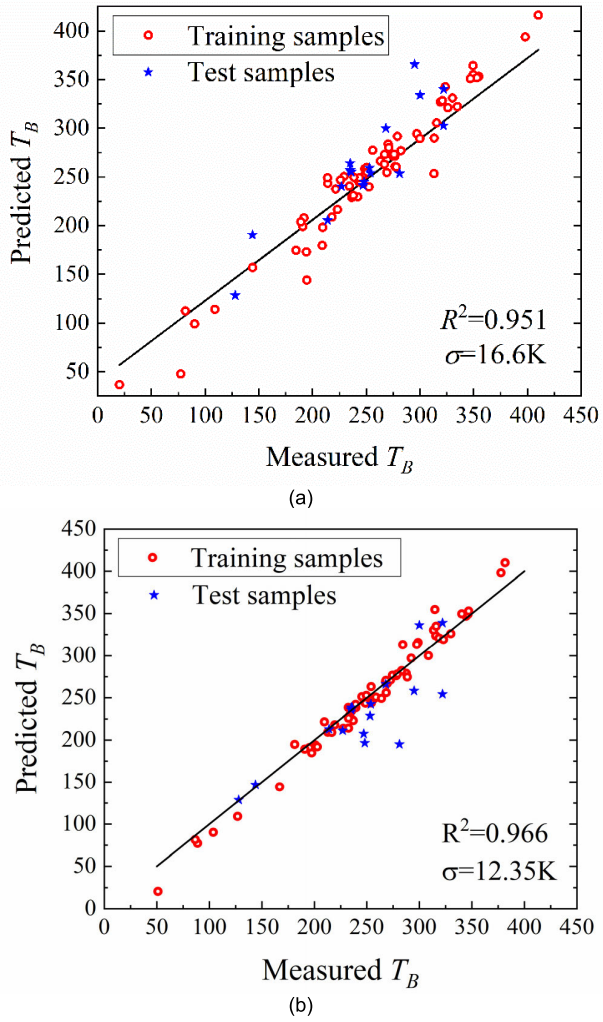

FIGURE 13. Comparison between the predicted and measured boiling temperature T_B using (a) ANN (Bayesian Regularization) (b) RF trained by 66 samples in Table 1 and tested by 17 sample in Table 2 for 7 predictors in Fig. 12.

Besides, it should be emphasized that no absolute judgment can be made to decide which method is the best for all types of problems because the performance of different methods might vary with different datasets. Considering the available data, RF is the most recommended method to predict the electrical strength.

V. PREDICTION OF BOILING TEMPERATURE

As mentioned in section I, for the industrial purpose the boiling temperature must be taken into account because in cold areas the liquidation should be avoided. This section

discusses about the boiling temperature T_B at 1 atm predicted by ANN and RF. Fig. 12 separately shows the PCCs between the boiling temperature and various predictors, based on the 66 molecules from Table 1 (the boiling temperature of C2F6S1 is not available). The PCCs are generally higher compared with Fig. 4, especially for the polarizability of which PCC could be over 0.8. The reason is that the polarizability will greatly enhance the Van der Waals' force, which will increase the boiling temperature. However, the correlation between *PA* and boiling temperature is weak. The reason is that the boiling temperature is related to the intermolecular attractive forces and *PA* does not influence the forces directly.

By combing all the predictors in Fig. 12, the R^2-value trained by the 66 molecules could easily reach 0.951 in ANN and 0.966 in RF as in Fig. 13, indicating a better prediction of the boiling temperature compared with the electrical strength. The test samples also have good agreement with the test results. Combined with *RMSE*, it can be observed RF performs better which is consistent with results of E_r.

VI. CONCLUSION

A. POTENTIAL AND CHALLENGE OF RF

As discussed above, three different models were presented to predict E_r and comparative analysis of results was carried out from different perspectives. Considering the limited data, RF outperforms the ANN and multiple linear regression methods with higher stability and prediction accuracy. It can also estimate the importance of different features which demonstrates that this method has the potentiality to be applied in designing and synthesizing the new eco-friendly gases in power equipment. RF shows good applicability to predicting the electrical strength as well as boiling temperature in this work. Still, challenges of using RF exist. The outliners underestimated by RF are probably due to the lack of training samples since no more than 100 sets of data are available. It encourages further development to obtain more test data on the electrical strength and find new predictors as the input.

B. DISCUSSION ABOUT THE IMPACT ON THE ENVIRONMENT

The goal of this work is to find eco-friendly gas to substitute SF_6 as the insulation gas in the power industry. Therefore, the impact of the gases on the environment should be checked. Table 5 lists the lifetime and GWP (at 100 years) of the 36 molecules [49]. Comparing with the data of predictors in Table 1, unfortunately it can be seen that all the predictors in this work have 30very weak correlation with the lifetime or GWP. In contrast, the GWP is strongly dependent on the radiative efficiency and the reactivity with the other particles, for example the reactions with the hydroxyl radical OH [31]. To predict the impact on the environment, future work should be done to check the feasibility of introducing the RF. Additionally, the decomposition products of the larger molecules after discharge [50], which contains CF_4 [51], should also be considered.

TABLE 5. Values of lifetime and GWP of 36 molecules.

CID	Molecules	Liftime/year	GWP (100 years)
17358	F6 S1	3200	23500
6393	F4 C1	50000	6630
6327	H3 C1 Cl1	1	12
6323	H3 C1 Br1	0.8	2
6345	H2 C1 F2	5.2	677
6344	H2 C1 Cl2	0.4	9
6373	H1 C1 F3	222	12400
6212	H1 C1 Cl3	0.4	16
6372	H1 C1 F2 Cl1	11.9	1760
6370	H1 C1 F1 Cl2	1.7	148
6392	C1 F3 Cl1	640	13900
6391	C1 F2 Cl2	100	10200
6389	C1 F1 Cl3	45	4660
6384	C1 F3 Br1	65	6290
9625	C1 F2 Cl1 Br1	16	1750
6431	C2 F6	10000	11100
6428	C2 F3 Cl3	85	6130
6432	C3 F8	2600	8900
8302	C3 F6	0.013	<1
8263	C4 F8	3200	9540
5943	C1 Cl4	26	1730
2775851	C4 F8	0.085	2
16843	C1 F3 I1	0.005	0.4
136213	C1 F8 S1	800	17400
17607	O2 F2 S1	36	4090
69636	C4 F6	0.003	<1
9638	C4 F10	2600	9200
11212	C5 F8	0.085	2
6429	C2 F4 Cl2	190	10000
6430	C2 F5 Cl1	1020	7670
6278	H3 C2 Cl3	5	160
9868	H3 C2 F3	47.1	4800
12633773'	C4 N1 F7	22	1490
2782408'	C6 O1 F12	0.014	<1
14175377'	C5 O1 F10	0.044	<1
24553'	NF3	500	16100
5708720'	HFC 1234ze(e)	0.045	<1

ACKNOWLEDGMENT

The authors would like to thank C. Franck for the suggestions.

REFERENCES

[1] M. Rabie and C. M. Franck, "Assessment of eco-friendly gases for electrical insulation to replace the most potent industrial greenhouse gas SF6," *Environ. Sci. Technol.*, vol. 52, no. 2, pp. 369–380, Jan. 2018.
[2] P. O'Connell, "SF6 in the electric industry, status 2000," *Electra*, vol. 200, pp. 16–25, 2002.
[3] K. Protocol, "United Nations framework convention on climate change," *Kyoto Protocol, Kyoto*, vol. 19, p. 497, 1997.
[4] *NOAA Earth System Research Laboratory GMD Carbon Cycle-Interactive Atmospheric Data Visualisation*, US National Oceanic and Atmospheric Administration. Accessed: Feb. 7, 2009. [Online]. Available: http://www.esrl.noaa.gov/gmd/ccgg/iadv
[5] B. Zhang, N. Uzelac, and Y. Cao, "Fluoronitrile/CO_2 mixture as an eco-friendly alternative to SF_6 for medium voltage switchgears," *IEEE Trans. Dielectrics Elect. Insul.*, vol. 25, no. 4, pp. 1340–1350, Aug. 2018.
[6] Y. Li, X. Zhang, Q. Chen, M. Fu, R. Zhuo, S. Xiao, D. Chen, and J. Tang, "Study on the dielectric properties of C_4F_7N/N_2 mixture under highly non-uniform electric field," *IEEE Access*, vol. 6, pp. 42868–42876, 2018.
[7] Yi Li, X. Zhang, S. Tian, S. Xiao, Q. Chen, D. Chen, Z. Cui, and J. Tang, "Insight into the compatibility between $C_6F_{12}O$ and metal materials: Experiment and theory," *IEEE Access*, vol. 6, pp. 58154–58160, 2018.
[8] H. Sun, M. Rong, Y. Wu, Z. Chen, F. Yang, A. B. Murphy, and H. Zhang, "Investigation on critical breakdown electric field of hot carbon dioxide for gas circuit breaker applications," *J. Phys. D, Appl. Phys.*, vol. 48, no. 5, Feb. 2015, Art. no. 055201.
[9] T. Uchii, Y. Hoshina, H. Kawano, K. Suzuki, T. Nakamoto, and M. Toyoda, "Fundamental research on SF6-free gas insulated switchgear adopting CO_2 gas and its mixtures," in *Proc. Int. Symp. Ecotopia Sci.*, 2007, pp. 516–520.
[10] Y. Wang, D. Huang, J. Liu, Y. Zhang, and L. Zeng, "Alternative environmentally friendly insulating gases for SF6," *Processes*, vol. 7, no. 4, p. 216, Apr. 2019.

[11] M. Rabie and C. Franck, "Computational screening of new high voltage insulation gases with low global warming potential," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 22, no. 1, pp. 296–302, Feb. 2015.

[12] L. G. Christophorou, J. K. Olthoff, and D. S. Green, "Gases for electrical insulation and arc interruption: Possible present and future alternatives to pure SF_6," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, NIST Tech. Note 1425, 1997.

[13] L. G. Christphorou and R. J. Brunt, "SF_6N_2 mixtures: Basic and HV insulation properties," *IEEE Trans. Dielectrics Elect. Insul.*, vol. 2, no. 5, pp. 952–1003, Oct. 1995.

[14] *Gas-Insulated Switchgear*. Accessed: Mar. 27, 2019. [Online]. Available: http://www.ABB.com/highvoltage

[15] J. D. Mantilla, N. Gariboldi, S. Grob, and M. Claessens, "Investigation of the insulation performance of a new gas mixture with extremely low GWP," in *Proc. IEEE Electr. Insul. Conf. (EIC)*, Jun. 2014, pp. 469–473.

[16] M. Rabie, D. A. Dahl, S. M. A. Donald, M. Reiher, and C. M. Franck, "Predictors for gases of high electrical strength," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 20, no. 3, pp. 856–863, Jun. 2013.

[17] C. M. Franck, D. A. Dahl, M. Rabie, P. Haefliger, and M. Koch, "An efficient procedure to identify and quantify new molecules for insulating gas mixtures," *Contrib. Plasma Phys.*, vol. 54, no. 1, pp. 3–13, Jan. 2014.

[18] A. E. D. Heylen and T. J. Lewis, "The electric strength of hydrocarbon gases," *Brit. J. Appl. Phys.*, vol. 7, no. 11, pp. 411–415, Nov. 1956.

[19] K. P. Brand, "Dielectric strength, boiling point and toxicity of gases-different aspects of the same basic molecular properties," *IEEE Trans. Elect. Insul.*, vol. EI-17, no. 5, pp. 451–456, Oct. 1982.

[20] J. Devins, "Replacement gases for SF_6," *IEEE Trans. Electr. Insul.*, vol. EI-15, no. 2, pp. 81–86, Apr. 1980.

[21] N. Meurice, E. Sandre, A. Aslanides, and D. P. Vercauteren, "Simple theoretical estimation of the dielectric strength of gases," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 11, no. 6, pp. 946–948, Dec. 2004.

[22] X. Yu, H. Hou, and B. Wang, "Prediction on dielectric strength and boiling point of gaseous molecules for replacement of SF_6," *J. Comput. Chem.*, vol. 38, no. 10, pp. 721–729, Apr. 2017.

[23] Y. Wu, C. Wang, H. Sun, M. Rong, A. B. Murphy, T. Li, J. Zhong, Z. Chen, F. Yang, and C. Niu, "Evaluation of SF_6-alternative gas C5-PFK based on arc extinguishing performance and electric strength," *J. Phys. D, Appl. Phys.*, vol. 50, no. 38, Sep. 2017, Art. no. 385202.

[24] Z. E. Aygen, S. Seker, M. Bagnyanik, F. G. Bagnyanik, and E. Ayaz, "Fault section estimation in electrical power systems using artificial neural network approach," in *Proc. IEEE Transmiss. Distrib. Conf.*, Apr. 1999, pp. 466–469.

[25] G. Cardoso, J. G. Rolim, and H. H. Zurn, "Application of neural-network modules to electric power system fault section estimation," *IEEE Trans. Power Del.*, vol. 19, no. 3, pp. 1034–1041, Jul. 2004.

[26] H.-T. Yang, W.-Y. Chang, and C.-L. Huang, "A new neural networks approach to on-line fault section estimation using information of protective relays and circuit breakers," *IEEE Trans. Power Del.*, vol. 9, no. 1, pp. 220–230, Jan. 1994.

[27] D. J. Sobajic and Y.-H. Pao, "Artificial neural-net based dynamic security assessment for electric power systems," *IEEE Trans. Power Syst.*, vol. 4, no. 1, pp. 220–228, Feb. 1989.

[28] Y.-Y. Cheng, P. P. K. Chan, and Z.-W. Qiu, "Random forest based ensemble system for short term load forecasting," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Jul. 2012, pp. 52–56.

[29] X. Li, X. Wang, A. Yang, and M. Rong, "Partial discharge source localization in GIS based on image edge detection and support vector machine," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1795–1802, Aug. 2019.

[30] G. Li, X. Wang, X. Li, A. Yang, and M. Rong, "Partial discharge recognition with a multi-resolution convolutional neural network," *Sensors*, vol. 18, no. 10, p. 3512, Oct. 2018.

[31] K.-M. Tsai and P.-J. Wang, "Predictions on surface finish in electrical discharge machining based upon neural network models," *Int. J. Mach. Tools Manuf.*, vol. 41, no. 10, pp. 1385–1403, Aug. 2001.

[32] D. S. S. Lee, B. J. Lithgow, and R. E. Morrison, "New fault diagnosis of circuit breakers," *IEEE Trans. Power Del.*, vol. 18, no. 2, pp. 454–459, Apr. 2003.

[33] X. Chen, H. Cui, and L. Luo, "Fault diagnosis of transformer based on random forest," in *Proc. 4th Int. Conf. Intell. Comput. Technol. Automat.*, Shenzhen, Guangdong, Mar. 2011, pp. 132–134.

[34] L. Zhong, "Fast prediction of electron-impact ionization cross sections of large molecules via machine learning," *J. Appl. Phys.*, vol. 125, no. 18, May 2019, Art. no. 183302.

[35] H. Zhou, Z. Deng, Y. Xia, and M. Fu, "A new sampling method in particle filter based on Pearson correlation coefficient," *Neurocomputing*, vol. 2016, pp. 208–215, Dec. 2016.

[36] Y. Fu, X. Wang, X. Li, A. Yang, G. Han, Y. Lu, Y. Wu, and M. Rong, "Theoretical study of the decomposition pathways and products of C5-perfluorinated ketone (C5 PFK)," *AIP Adv.*, vol. 6, no. 8, 2016, Art. no. 085305.

[37] F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, and K. N. Kudin, *Gaussian 09, Revision a. 02. Gaussian*. Wallingford, U.K.: Wallingford, 2009, p. 28.

[38] C. W. Bauschlicher, Jr., and H. Partridge, "A modification of the Gaussian-2 approach using density functional theory," *J. Chem. Phys.*, vol. 103, no. 5, pp. 1788–1791, 1995.

[39] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, Feb. 1996.

[40] R. F. W. Bader, A. Larouche, C. Gatti, M. T. Carroll, P. J. MacDougall, and K. B. Wiberg, "Properties of atoms in molecules: Dipole moments and transferability of properties," *J. Chem. Phys.*, vol. 87, no. 2, pp. 1142–1152, Jul. 1987.

[41] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[42] U. Grömping, "Variable importance assessment in regression: Linear regression versus random forest," *Amer. Statistician*, vol. 63, no. 4, pp. 308–319, Nov. 2009.

[43] *regRF*. Accessed: Jul. 9, 2019. [Online]. Available: https://code.google.com/p/randomforest-matlab/

[44] *Explore Chemistry*. Accessed: May 16, 2019. [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/

[45] L. Liang, Y. Wu, H. Sun, M. Rong, C. Wang, and H. Chang, "A new prediction model of SF_6 alternative gases in electrical strength," in *Proc. 5th Int. Conf. Electr. Power Equip.-Switching Technol.*, Oct. 2019, pp. 86–89.

[46] C. R. Rao, C. R. Rao, M. Statistiker, C. R. Rao, and C. R. Rao, *Linear Statistical Inference and Its Applications*, vol. 2. New York, NY, USA: Wiley, 1973.

[47] N. J. D. Nagelkerke, "A note on a general definition of the coefficient of determination," *Biometrika*, vol. 78, no. 3, pp. 691–692, 1991.

[48] S. Oliveira, F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. C. Pereira, "Modeling spatial patterns of fire occurrence in mediterranean europe using multiple regression and random forest," *Forest Ecol. Manage.*, vol. 275, pp. 117–129, Jul. 2012.

[49] Ø. Hodnebrog, M. Etminan, J. S. Fuglestvedt, G. Marston, G. Myhre, C. J. Nielsen, K. P. Shine, and T. J. Wallington, "Global warming potentials and radiative efficiencies of halocarbons and related compounds: A comprehensive review," *Rev. Geophys.*, vol. 51, no. 2, pp. 300–378, Apr. 2013.

[50] Y. Li, X. Zhang, J. Zhang, M. Fu, R. Zhuo, Y. Luo, D. Chen, and S. Xiao, "Experimental study on the partial discharge and AC breakdown properties of C_4F_7N/CO_2 mixture," *High Voltage*, vol. 4, no. 1, pp. 12–17, Mar. 2019.

[51] X. Zhang, Y. Li, S. Xiao, J. Tang, S. Tian, and Z. Deng, "Decomposition mechanism of C5F10O: An environmentally friendly insulation medium," *Environ. Sci. Technol.*, vol. 51, no. 17, pp. 10127–10136, Sep. 2017.

HAO SUN was born in Shandong, China, in August 1988. He received the B.S. degree in applied physics and the Ph.D. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2010 and 2016, respectively. He studied as a Joint Doctoral Student at Kanazawa University, Kanazawa, Japan, for one year, from 2014 to 2015.

From 2016 to 2019, he was a Lecturer with the State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University. Since 2019, he has been an Associate Professor with the State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an. His research interests include the design and development of switch gears, arc phenomenon in switchgears, simulation and diagnosis of arc plasmas, and thermal plasma applications.

Dr. Sun was a recipient of the Science and Technology Award from Shaanxi Province of China, in 2016, and the Excellent Doctoral Thesis from Xi'an Jiaotong University, in 2019. One of his publications received the IOP Publishing Top Cited Author Award (China), in 2019.

LUQI LIANG was born in Taiyuan, Shanxi, China, in 1995. She received the B.S. degree in electrical engineering and automation from Hainan University, Hainan, China, in 2018. She is currently pursuing the M.S. degree in electrical engineering with Xi'an Jiaotong University, Xi'an, China.

Her research interests include the assessment and prediction of electrical strength of the substitutes for greenhouse gas SF6 and gassing materials in switching devices.

CHUNLIN WANG was born in Shandong, China, in 1992. He received the B.S. in electrical engineering from Xian Jiaotong University, China, in 2015, where he is currently pursuing the Ph.D. degree.

His research interests are mainly thermal plasma and SF6 alternative gases at present.

YI WU (Senior Member, IEEE) was born in Jiangsu, China, in 1975. He received the B.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1998 and 2006, respectively.

He is currently with the State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University. His research interests include the application of fast current limiters, dc switchgears for the entire voltage range, and electric arc physics.

FEI YANG (Senior Member, IEEE) was born in Shanxi, China, in May 1982. He received the B.S. and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2005 and 2010, respectively.

He is currently with the State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University. His research interests include the simulation and experimentation of arc plasma in switching devices.

MINGZHE RONG (Senior Member, IEEE) was born in Shanxi, China, in 1963. He received the B.S. and Ph.D. degrees from the Department of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively.

He is currently with the State Key Laboratory of Electrical Insulation and Power Equipment, Xi'an Jiaotong University. He was involved in arc physics, electrical contact theory, intelligent electrical apparatus, and condition monitoring technique for switchgear. He has been authored more than 130 articles, since 1984.