

STATISTICAL SIGNIFICANCE APPROACH FOR DIFFERENT WORK FUNCTION MODELING OF SEMICONDUCTOR

Arpamong Changjan, Phanuchai Pramuanl, and Atirat Maksuwan*

Received: February 01, 2024; Revised: March 29, 2024; Accepted: April 10, 2024

Abstract

Significances of the work function in the field of surface science need knowledge of how electrons are transported out of the Fermi surface to the vacuum as well as ways the energy needed can be accurately estimated. Full knowledge of work functions as well as the surface energy of the metallic surface enhances understanding of the formation of grain boundaries and surface segregation. Statistical significance of differences in modeling approaches requires studies that can reliably distinguish between systematic approach effects and errors resulting from modeling approach variation. In this work, researchers introduced analysis of variance and correlation analysis to assess the statistical significance of differences in modeling approach variation. Comparisons of obtained results with estimating the work function of semiconductor variation data of support vector machine (SVM) and linear regression with natural logarithm transformation (LRNLT) were presented.

Keywords: Analysis of variance; Correlation analysis; Semiconductors; Statistical significance; work function

Introduction

The importance of the work function in the field of surface science necessitates in-depth knowledge of how electrons are carried from the Fermi surface to the vacuum, as well as methods for precisely estimating the energy required. The creation of grain boundaries, surface segregation, and growth rates, to name a few, is enhanced by a thorough grasp of the work functions and surface energy of metallic surfaces (Wang and Wang, 2014). Inadequate experimental determination of the minimal energy required to transfer electrons to the metal's surface (work function) using a scanning kelvin probe force

microscope (SKPFM) (Rohwerder and Turcu, 2007) necessitates the use of theoretical models.

Theoretical work function calculations may be traced back to 1935 (Wigner and Bardeen, 1935) when the estimate was made as the energy difference between two lattices containing the same number of ions but varied by one electron. This qualitative estimation of work function was far from the experimental values, necessitating revision. The hypothesis was further strengthened by employing pseudo-potential theory to account for the actions of ions (Lang and Kohn, 1971).

Department of Environmental Technology and Applied Science, Faculty of Science and Technology, Pathumwan Institute of Technology, Bangkok 10330, Thailand. Tel.: 0-2104-9099; E-mail: otto_sinkronity@yahoo.com; Phanuchai.p@gmail.com; atirat.maksuwan@gmail.com

* Corresponding author

DOI: <https://doi.org/10.55766/sujst-2024-03-e05901>

Suranaree J. Sci. Technol. 31(3):030198(1-11)

This model could only estimate simple metals while noble gases deviated from experimental values (Lang and Kohn, 1971). In the process of determining an appropriate technique of predicting work function, a stabilized jellium model with a constant potential in addition to the metal's effective potential was developed (Kiejna, 1999). The density functional theory method was recently used to offer another set of work functions for Body Centre Cubic (BCC) and Face Centre Cubic (FCC) metals (Wang and Wang, 2014).

Using a different methodology than the previous models, the model suggested in this study accurately calculates the work function of elemental semiconductors. The statistical significance of differences in modeling methodologies is assessed using analysis of variance and correlation analysis. The obtained results were compared to predicting the work function of semiconductor variation data using support vector machines (SVM) and linear regression with natural logarithm transformation (LRNLT). The results of Maksuwan *et al.* (2023a; 2023b) studies show that the development of the LRNLT model can be used better than the SVM model for superconductivity research. The results of this research show that the statistical significance of differences in modeling approaches requires further study. Because the statistical significance approach can reliably distinguish between systematic approach effects and errors resulting from modeling approach variation.

Materials and Methods

Brief Description of Analysis of Variance (ANOVA)

ANOVA is used to determine significance using the ratio of variance estimates from sample means and sample values (Krzywinski and Altman, 2014). Their ratio or test statistic F can be used between group mean squares (MS_B) and within group mean squares (MS_W), which are calculated from between treatment sum of squares (SS_B) and within treatment sum of squares (SS_W) weighted by degrees of freedom (d.f.). The degrees of freedom (d.f.) show the number of independent deviations in the treatment sum of squares (SS); the deviations are not all independent because deviations of a quantity from its own mean must sum to zero. The ANOVA null hypothesis is that all samples are from the same distribution and have equal means. Under this null, between group variation of sample means and within-group variation of sample values are predictably related. We introduced the idea that variance can be partitioned within group variance was interpreted as experimental error and between

group variance as structural and topological parameters variation. In one way ANOVA, the relevant quantities are within group variance and between group variance. Within a group, variance corresponds to variance in the sample after other sources of variation have been accounted for and represent experimental error. If some sources of error are not accounted for (e.g., structural and topological parameters variation), within group variance will be inflated. Between group variance is another estimate for within group variance, additionally inflated by the average squared deviation of treatment means from the grand mean. As noise within samples drops, a larger fraction of variation is allocated to between group variance, and the power of the test improves (Armstrong *et al.*, 2022). This suggests that it is beneficial to decrease within group variance. We can do this through a process to identify and isolate likely sources of sample variability.

Proposed Method

The uniqueness of this research work is that it adopts the ANOVA together with assessing the statistical significance of differences in the modeling approach. To demonstrate the outstanding efficiency of the ANOVA results. In this work, we compared obtained results with work function modeling of semiconductors between support vector machine (SVM) and adopting a simple model (linear regression) through which they developed by using a natural logarithm transformation (LRNLT) approach.

Working Principle of SVM

SVM stems from supporting vector machines which are derived from statistical learning theory and structural risk minimization principle (Cortes and Vapnik, 1995; Vapnik, 1995). It employs a mapping function called kernel function to map non-linear regression to high dimensional feature space where linear regression is conducted. SVM adopts support vector regression (SV regression) and takes some input parameters (such as hyperparameter λ , regularization factor C , and kernel option) by which the chosen kernel function generates a relation between the predictor and the target. The complete regression function is presented in equation (1) with the inclusion of the mapping function $K(x_i, x)$.

$$f(x) = \sum_{i=1}^k (\lambda_i^* - \lambda_i) K(x_i, x) + b \quad (1)$$

Where $\lambda_i^* \geq 0$, $\lambda_i^* \leq C$, $K(x_i, x)$ is the polynomial kernel function, and b kernel parameters.

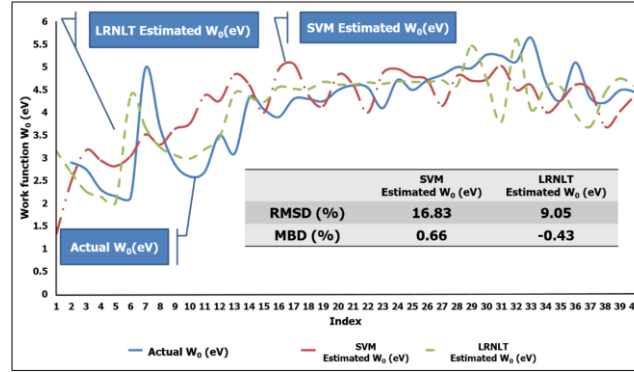


Figure 1. Work Function W_0 (eV) predictions by SVM and LRNLT

Working Principle of LRNLT

The multiple linear regression models are built on the same simple linear regression (Goldman *et al.*, 1977; Efron, 1983; Efron and Gong, 1983). However, in addition to the concepts discussed so far for simple linear regression, which remains applicable, a new concept set has to be introduced. This discussion will focus on situations where there are two predictors and one outcome variable. With all three variables, a three-dimensional figure can be used to visualize data. Models with more predictive variables follow the same principle but are difficult to visualize. We consider multiple regression and linear relation as follows,

$$Y = BX + W \quad (2)$$

in which,

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix},$$

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix}, W = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_m \end{bmatrix}$$

We develop a multiple linear regression model for predicting by using linear regression with natural logarithmic transformation (LRNLT). Start by considering the nonlinear regression problem by following Makuwan *et al.*, (2023a; 2023b).

$$y = ae^{bx}U \quad (3)$$

With the parameters a and b and the multiplication term U . If we find the natural logarithm of both sides, this will become

$$\ln(y) = \ln(a) + bx + u \quad (4)$$

in which, $u = \ln(U)$

Performance Evaluation

Performance of obtained results with work function modeling of semiconductor approach variation data is evaluated using the root mean square difference (RMSD), mean bias difference (MBD), and Coefficient Correlation is expressed by r_{XY} in Equations (5)-(7) respectively (Makuwan *et al.*, 2023a; 2023b), where N is the number of data points, T_i^{Exp} and T_i^{Model} are the i -th ($i = 1, 2, \dots, N$). $\bar{X}$ and $\bar{Y}$ are mean of a linear relationship between the X_i and Y_i variable of obtained results work function modeling of semiconductor approach variation data, respectively.

$$RMSD = \frac{\sqrt{\sum_{i=1}^N (T_i^{Model} - T_i^{Exp})^2}}{\frac{\sum_{i=1}^N T_i^{Exp}}{N}} \times 100 \quad (5)$$

$$MBD = \frac{\sum_{i=1}^N (T_i^{Model} - T_i^{Exp})}{\frac{\sum_{i=1}^N T_i^{Exp}}{N}} \times 100 \quad (6)$$

$$r_{XY} = \frac{\sum_{i=1}^N (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}} \quad (7)$$

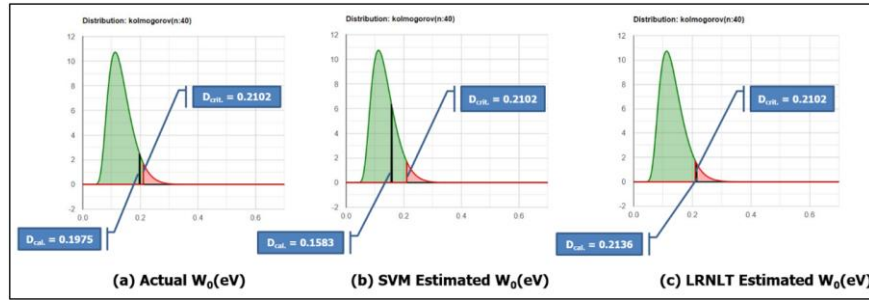


Figure 2. The Kolmogorov-Smirnov Test of Normality of (a) Actual work Function W_0 (b) SVM Estimated W_0 (eV) and (c) LRNLT Estimated W_0 (eV)

Table 1. The actual work function of semiconductors W_0 (eV) data and predictions from “SVM Estimated W_0 (eV)” and “LRNLT Estimated W_0 (eV)”

Index	Element	Number Valence electron	R (a.u.)	E_F (eV)	Actual work Function W_0 (eV)	SVM Estimated W_0 (eV)	LRNLT Estimated W_0 (eV)
1	Li	1	3.26	4.74	2.90	1.33	3.15
2	Na	1	3.99	3.24	2.75	2.47	2.68
3	K	1	4.95	2.12	2.30	3.17	2.26
4	Rb	1	5.31	1.85	2.16	2.94	2.15
5	Cs	1	5.75	1.59	2.14	2.82	2.02
6	Be	2	1.87	12.00	4.98	3.05	4.38
7	Mg	2	2.65	7.10	3.66	3.52	3.67
8	Ca	2	3.26	5.20	2.86	3.29	3.23
9	Sr	2	3.55	4.49	2.59	3.64	3.06
10	Ba	2	3.73	3.84	2.70	3.76	2.98
11	Se	2	3.72	7.54	3.50	4.38	3.19
12	Y	2	2.99	6.88	3.10	4.26	3.46
13	Ti	3	2.24	8.84	4.33	4.84	4.43
14	Zr	3	2.33	8.50	4.07	4.61	4.34
15	Hf	3	2.43	8.69	3.90	4.00	4.24
16	V	3	2.09	11.09	4.30	4.96	4.56
17	Nb	3	2.13	10.82	4.30	5.07	4.52
18	Ta	3	2.13	10.19	4.25	4.38	4.52
19	Cr	3	1.98	11.95	4.50	4.15	4.68
20	Mo	3	2.03	11.36	4.60	4.84	4.62
21	W	3	2.04	11.47	4.55	4.61	4.61
22	Mn	3	2.01	12.20	4.10	4.00	4.66
23	Re	3	2.01	10.90	4.72	4.86	4.63
24	Fe	3	1.96	11.81	4.50	4.96	4.69
25	Ru	3	1.96	10.82	4.71	4.79	4.68
26	Os	3	1.97	11.06	4.83	4.72	4.67
27	Co	3	1.93	10.52	5.00	4.15	4.71
28	Rh	3	2.07	9.26	4.98	4.81	4.59
29	Ir	4	1.87	10.26	5.27	4.70	5.47
30	Ni	3	1.92	9.66	5.25	4.73	4.74
31	Pd	3	2.92	7.69	5.12	5.03	3.80
32	Pt	4	1.91	8.79	5.65	4.50	5.61
33	Cu	2	2.12	9.03	4.65	4.61	4.07
34	Ag	2	1.39	7.84	4.26	4.01	4.58
35	Au	3	2.22	7.25	5.10	4.26	4.57
36	Zn	2	2.30	8.65	4.33	4.61	3.95
37	Cd	2	2.59	6.45	4.22	4.50	3.69
38	Hg	3	2.51	5.66	4.49	3.69	4.45
39	B	2	1.84	15.40	4.45	4.03	4.75
40	Al	3	2.07	11.30	4.28	4.38	4.59
Maximum		4	5.75	15.40	5.65	5.07	5.61
Minimum		1	1.39	1.59	2.14	1.33	2.02
Mode		3	1.87	10.82	4.30	4.61	4.52
Meadian		3	2.13	8.82	4.32	4.38	4.49
Mean		2.5	2.60	8.45	4.11	4.14	4.09
Standard Deviation		0.78	0.996	3.23	0.93	0.81	0.86
Variance		0.62	0.992	10.44	0.87	0.65	0.74
Correlation with Actual work Function W_0 (eV)		0.724	-0.767	0.607	-	-	-
Correlation with SVM Estimated W_0 (eV)		0.682	-0.441	0.532	-	-	-
Correlation with LRNLT Estimated W_0 (eV)		0.787	-0.924	0.808	-	-	-
Root Mean Square Difference (RMSD)		-	-	-	-	16.83	9.05
Mean Bias Difference (MBD)		-	-	-	-	0.66	-0.43

Note: “Number Valence electron”, “electron density parameter R (a.u.)” “Femi energy E_F (eV)” are the predictors work functions of elemental semiconductors. “Reference” shows sources of all work function of semiconductors W_0 (eV) data in (Olatunji *et al.*, 2011; 2014).

Results and Discussion

The actual data (obtained experimentally) employed in this research work was drawn from the literature (Michaelson, 1977; Wang and Wang, 2014). A total of forty raw data points were used for both training and testing, as presented in Table 1 (Olatunji *et al.*, 2011). Before the analysis of variance, we used the Kolmogorov-Smirnov test to check the normality of the set of raw data, which comes from each modeling approach in Figure 2.

The graph in Figure 2 shows the test maximum statistic D of (a) Actual work Function W_0 (b) SVM Estimated W_0 (eV) and (c) LRNLT Estimated W_0 (eV) equals 0.1975, 0.1583, and 0.2136, respectively, which had a trend lower than the critical test maximum statistic D, which equals 0.2102. In those results, it was assumed that the data distribution was not normal. In other words, the difference between the raw data sample and the normal distribution was not big enough to be statistically significant. For this reason, the raw data sample conflicted with the ANOVA under elements of the parametric statistical model (Siegel, 1956).

The procedure uses a nonparametric approach to the one-way ANOVA to compare three or more groups on a dependent variable that is measured on at least an ordinal level. It is sometimes called the Kruskal-Wallis test. The graph in Figure 3 shows the nonparametric approach to the one-way ANOVA to compare three groups: Actual work function, SVM from Equation (1), and LRNLT estimated W_0 (eV) from Equation (4). The test statistic H under the chi-square distribution equals 0.084, which is in the 95% region of null hypothesis acceptance: [0, 5.991]. This result means that the mean ranks of all groups are assumed to be equal. In other words, the difference between the mean ranks of all groups is not big enough to be statistically significant.

Result from the one-way ANOVA to compare three groups in Figure 3. Let us know just that the difference between the mean ranks of all groups was nonsignificant. However, we still don't know how the predictions of each model compare to the experimental values. Therefore, we have to make pairwise comparisons using the measure of the central tendencies of the two groups under the nonparametric approach, so-called the two-sample Mann-Whitney U test. The result of the two-sample Mann-Whitney U test is presented in Figure 4.

The graph in Figure 4 shows the test statistic Z of (a) Actual W_0 vs. SVM Estimated W_0 (eV), (b) Actual W_0 vs. LRNLT Estimated W_0 (eV), and (c) SVM Estimated W_0 (eV) vs. LRNLT Estimated W_0 (eV) equals 0.0385, 0.04812, and 0.4283,

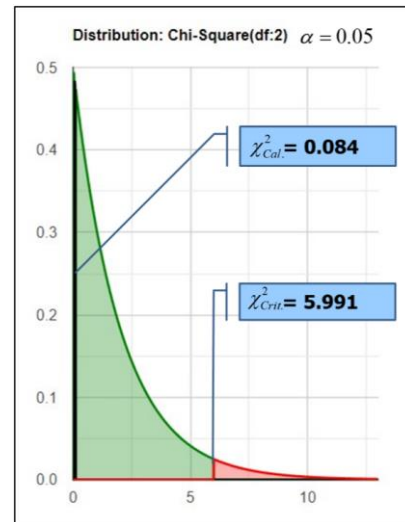


Figure 3. The test statistic H under the chi-square distribution was obtained by using a nonparametric approach to the one-way ANOVA to compare three groups: actual, SVM from Equation (1), and LRNLT from Equation (4) estimated work function

respectively, which is in the 95% region of null hypothesis acceptance as [-1.96: 1.96]. Paired comparisons for all 3 pairs show results trending in the same direction, namely that the randomly selected value of the population from all 3 pairs is assumed to be equal. In other words, the difference between the randomly selected values of the population from all 3 pairs is not big enough to be statistically significant.

Results from the two-sample Mann-Whitney U test in Figures 4(a), 4(b), and 4(c) let us know that the difference between the measures of the central tendencies of the two groups was nonsignificant. In other words, although the performance of the obtained results with %RMSD and %MBD of LRNLT was lower than the SVM modeling approach, the difference between the measures of the central tendencies from both modeling was nonsignificant with the actual work function of the semiconductor. For this reason, a comparison of the performance of the LRNLT and SVM modeling approaches with only %RMSD and %MBD is not enough to confirm that the performance of LRNLT is higher than that of the SVM modeling approach.

Therefore, we must assess the performance of the LRNLT and SVM modeling approaches to determine how well the relationship between two variables can be described using the analysis of variance of correlation under a nonparametric viewpoint. Such assessments of this performance use the so-called Spearman correlation coefficient

as presented in Tables 2, 3, and 4 for correlation with actual work function, SVM estimated W_0 (eV), and LRNLTL estimated W_0 (eV), respectively. Those tables indicate the direction of association between the predictors, namely Fermi energy (E_F), electron

density parameter (R), and number of valence electrons, and the work function of the semiconductor.

Results from the Spearman correlation coefficient as presented in Tables 2, 3, and 4 showed

Table 2. The actual work function of semiconductors W_0 (eV) data and predictions from “SVM Estimated W_0 (eV)” and “LRNLTL Estimated W_0 (eV)”

			Number Valence electron	Electron density parameter R (a.u.)	Femi energy EF (eV)	Actual work Function W0 (eV)
Spearman's rho	Number Valence electron	Correlation Coefficient	1.000	-.609**	.610**	.724**
		Sig. (2-tailed)	.	.000	.000	.000
		N	40	40	40	40
	Electron density parameter R (a.u.)	Correlation Coefficient	-.609**	1.000	-.824**	-.767**
		Sig. (2-tailed)	.000	.	.000	.000
		N	40	40	40	40
	Femi energy E_F (eV)	Correlation Coefficient	.610**	-.824**	1.000	.607**
		Sig. (2-tailed)	.000	.000	.	.000
		N	40	40	40	40
	Actual work Function W_0 (eV)	Correlation Coefficient	.724**	-.767**	.607**	1.000
		Sig. (2-tailed)	.000	.000	.000	.
		N	40	40	40	40

**. Correlation is significant at the 0.01 level (2-tailed).

Table 3. The correlation between the data set contains the predictors and the SVM Estimated W_0 (eV)

			Number Valence electron	Electron density parameter R (a.u.)	Femi energy EF (eV)	Actual work Function W0 (eV)
Spearman's rho	Number Valence electron	Correlation Coefficient	1.000	-.609**	.610**	.682**
		Sig. (2-tailed)	.	.000	.000	.000
		N	40	40	40	40
	Electron density parameter R (a.u.)	Correlation Coefficient	-.609**	1.000	-.824**	-.441**
		Sig. (2-tailed)	.000	.	.000	.000
		N	40	40	40	40
	Femi energy E_F (eV)	Correlation Coefficient	.610**	-.824**	1.000	.532**
		Sig. (2-tailed)	.000	.000	.	.000
		N	40	40	40	40
	Actual work Function W_0 (eV)	Correlation Coefficient	.682**	-.441**	.532**	1.000
		Sig. (2-tailed)	.000	.000	.000	.
		N	40	40	40	40

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4. The correlation between the data set contains the predictors and the LRNLTL Estimated W_0 (eV)

			Number Valence electron	Electron density parameter R (a.u.)	Femi energy EF (eV)	Actual work Function W0 (eV)
Spearman's rho	Number Valence electron	Correlation Coefficient	1.000	-.609**	.610**	.787**
		Sig. (2-tailed)	.	.000	.000	.000
		N	40	40	40	40
	Electron density parameter R (a.u.)	Correlation Coefficient	-.609**	1.000	-.824**	-.924**
		Sig. (2-tailed)	.000	.	.000	.000
		N	40	40	40	40
	Femi energy E_F (eV)	Correlation Coefficient	.610**	-.824**	1.000	.808**
		Sig. (2-tailed)	.000	.000	.	.000
		N	40	40	40	40
	Actual work Function W_0 (eV)	Correlation Coefficient	.787**	-.924**	.808**	1.000
		Sig. (2-tailed)	.000	.000	.000	.
		N	40	40	40	40

**. Correlation is significant at the 0.01 level (2-tailed).

the difference in the magnitude-changing trend of correlation between the predictors and work function W_0 (eV) as presented in Figure 5.

The column chart in Figure 5 lets us know that the magnitude-changing trend of correlation between the predictors and work function W_0 (eV) as presented by LRNLT Estimated W_0 (eV) (green dotted line) had the magnitude-changing trend that was likewise with work function W_0 (eV) as presented by Actual Work Function W_0 (eV) (blue dotted line). Conversely, the SVM-presented predictors and work function W_0 (eV) Estimated W_0 (eV) (red dotted line) showed a magnitude-changing pattern that was *paradoxical* with work function W_0 (eV) as provided by Actual Work Function W_0 (eV) (blue dotted line). Because of this, it was sufficient

to verify that the LRNLT modeling approach performed better than the SVM modeling approach by comparing the performance of the two approaches using %RMSD and %MBD to gather the relationship between two variables using the analysis of variance viewpoint of the correlation analysis.

Because of the effects of ions, the qualitative estimation of work function would be affected by theoretical calculation (Wigner and Bardeen, 1935). Therefore, we tested the statistical difference of work function W_0 (eV) values between the actual, SVM, and LRNLT modeling approaches. By using the number of valence electrons from raw data in Table 1 as a condition for dividing test groups, which can be divided into four groups, the numbers

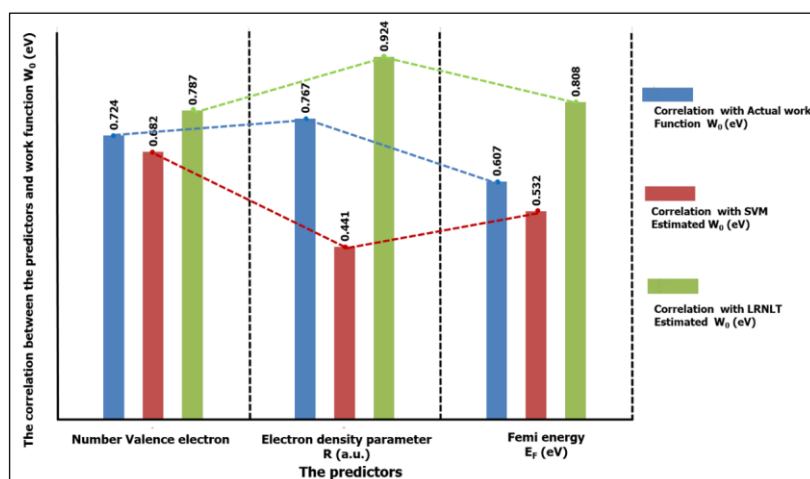


Figure 5. The magnitude-changing trend of correlation between the data set contains the predictors and work function W_0 (eV)

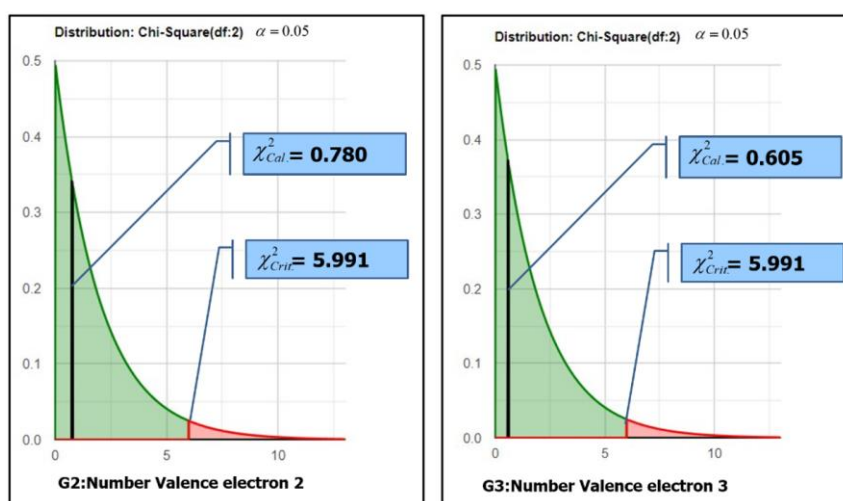


Figure 6. Kruskal-Wallis-test, using Chi-Square (df:2) distribution of G2: Number Valence electron 2 and G3: Number Valence electron 3

of valence electrons 1, 2, 3, and 4 were G1, G2, G3, and G4, respectively. In this process, we analyze the statistical differences in work function W_0 (eV) values between the actual, SVM, and LRNLT modeling approaches when n_1 , n_2 , and n_3 were the sample sizes of the actual, SVM, and LRNLT modeling approaches, respectively. By doing the one-way ANOVA by ranking the Kruskal-Wallis test as follows:

The first, G1, had sample sizes of $n_1 = 5$, $n_2 = 5$, and $n_3 = 5$. Probabilities associated with values of $H = 0.98$ equal 0.102 ($p = 0.102$). From the statistical test (Table O in Siegel, 1956), the value of probabilities (p) is less than 0.05. The second, in Figure 6, shows that the test statistic H under the chi-square distribution of G2 sample size $n_1 = 12$, $n_2 = 12$, $n_3 = 12$, and G3 sample size $n_1 = 21$, $n_2 = 21$, $n_3 = 21$, equals 0.780 and 0.605, respectively, which is in the 95% region of null hypothesis acceptance: [0, 5.991]. Lastly, sample sizes for G4 were $n_1 = 2$, $n_2 = 2$, and $n_3 = 2$. Probabilities associated with values of $H = 3.429$ are equal to 0.200 ($p = 0.200$). Based on the statistical test presented in Siegel (1956) Table O, the probability (p) value is less than 0.05. The mean ranks of all groups are taken to be equal based on the outcomes of G1, G2, G3, and G4. As stated otherwise, there is not a significant enough difference between the mean ranks of all the groups to warrant statistical significance.

Findings from a one-way ANOVA were used to compare the G1, G2, G3, and G4 groups. Tell us only that there was no statistically significant difference in the mean ranks of all the groups. However, we are still unsure of the paired comparison model's difference. Therefore, we have to make pairwise comparisons using the measure of the central tendencies of the two groups. Depending on sample size, we used the two-sample Mann-Whitney U test for G1, G2, and G3, together with the median test for two independent samples in G4. The result of paired comparisons into G1, G2, G3, and G4 is as follows:

The first is the critical probability $p = 0.05$ of G1, as follows in Table J of Siegel (1956) associated with (1) Actual vs. SVM. Estimated W_0 (eV) comparison with sample sizes $n_1 = 5$ and $n_2 = 5$, as the observed value of the test statistic $U = 8$ equals 0.210 ($p = 0.210$), therefore $p = 0.210 > p = 0.05$. (2) Sample sizes $n_1 = 12$ and $n_2 = 12$ are linked to the probability of Actual vs. LRNLT Estimated W_0 (eV), since the observed value of the test statistic $U = 11$ equals 0.421 ($p = 0.421$), and $p = 0.421 > p = 0.05$, and (3) SVM vs. LRNLT estimated W_0 (eV) comparison with sample sizes $n_1 = 5$ and $n_2 = 5$, as the observed value of the test statistic $U = 9$

equals 0.274 ($p = 0.274$), therefore $p = 0.274 > p = 0.05$.

Results from the two-sample Mann-Whitney U test in G1 let us know that the difference between the measures of the central tendencies of the all pared was not significant. In other words, the difference between the measures of the central tendencies in G1 as the numbers of valence electrons equal to 1 from all pared was not significant with the actual work function of the semiconductor.

The second is the critical value of U (U_{Crit}) equal to 42 of G2, as follows in Table K of Siegel (1956) associated with (1) Actual vs. SVM. Estimated W_0 (eV) comparison with sample sizes $n_1 = 12$ and $n_2 = 12$ for a one-tailed test at $\alpha = 0.05$. The observed value of two samples Mann-Whitney U ($U_{\text{Obs.}}$) as 62.5, so that $U_{\text{Crit.}} = 42 < U_{\text{Obs.}} = 62.5$. (2) Actual vs LRNLT Estimated W_0 (eV) comparison for a one-tailed test at $\alpha = 0.05$ with sample sizes $n_1 = 12$ and $n_2 = 12$. Since the two sample Mann-Whitney U has an observed value ($U_{\text{Obs.}}$) of 72, $U_{\text{Crit.}} = 42 < U_{\text{Obs.}} = 72$, and (3) SVM vs. LRNLT Estimated W_0 (eV) comparison with sample sizes $n_1 = 12$ and $n_2 = 12$ for a one-tailed test at $\alpha = 0.05$. The observed value of two sample Mann-Whitney U ($U_{\text{Obs.}}$) as 55.5, so that $U_{\text{Crit.}} = 42 < U_{\text{Obs.}} = 55.5$.

The findings of the two-sample Mann-Whitney U test in G2 demonstrated a significant difference between the evaluations of each pared's central tendencies. Stated differently, the findings of the two-sample Mann-Whitney U test show that the measures of the two group's central inclinations differed statistically significantly. In other words, the real work function of the semiconductor differed significantly from the measures of the central tendencies in G2, as the numbers of valence electrons were equal to 2 from all pared.

The third is the test statistic Z of G3: (a) Actual W_0 vs. SVM Estimated W_0 (eV), (b) Actual W_0 vs. LRNLT Estimated W_0 (eV), and (c) SVM Estimated W_0 (eV) vs. LRNLT Estimated W_0 (eV), which is displayed on the graph of the two-sample Mann-Whitney U test in Figure 7. This corresponds to -0.2139, -0.2139, and 1.0822, respectively, and falls within the 95% region of null hypothesis acceptance, which is [-1.96: 1.96]. In G3, all three pair's paired comparisons trend in the same direction, indicating that the population's randomly chosen value is taken to be equal in each pair. In other words, there is not a significant enough difference between the population's randomly picked values from all three pairs to warrant statistical significance.

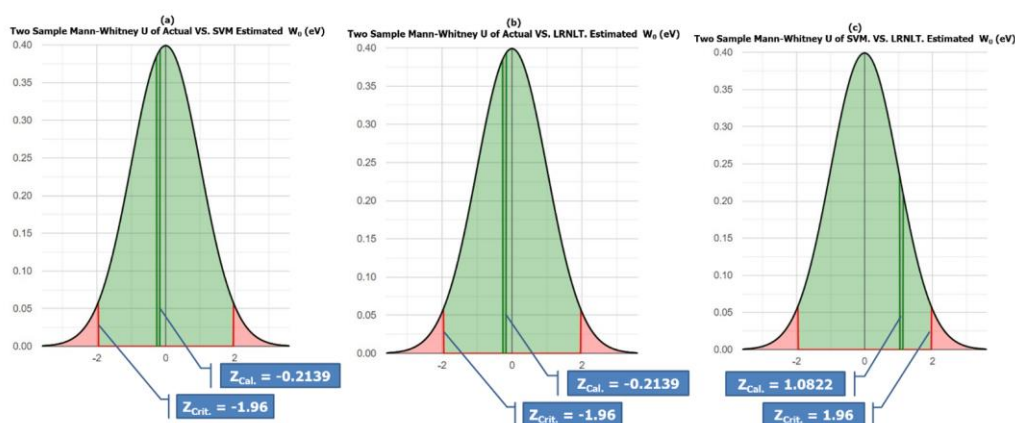


Figure 7. Two Sample Mann-Whitney U, using Normal distribution (two-tailed) of G3: Number Valence electron as 3 groups (a) Actual W_0 VS. SVM Estimated W_0 (eV) (b) Actual W_0 VS. LRNLT Estimated W_0 (eV) and (c) SVM Estimated W_0 (eV) VS. LRNLT Estimated W_0 (eV)

Figures 7(a), 7(b), and 7(c) show the results of G3 by using the two-sample Mann-Whitney U test, which indicates that there was no statistically significant difference between the measures of the central inclinations of the two groups. The outcome indicates that there was no statistically significant discrepancy between the semiconductor's actual work function and the measures of the central tendency in G3, where the number of valence electrons equals three from all three pair's paired comparisons.

Finally, the result of the median test for two independent samples and the calculation of the difference significance by the Fisher test of G4, namely (1) Actual vs. SVM. Estimated W_0 (eV) comparison with the probabilities associated with median = 4.985, $n_1 = 2$, $n_2 = 2$, and $n_1 + n_2 < 20$ as observed value by using the Fisher test equals 0.167 ($p = 0.167$), so that $p = 0.167 > p = 0.05$. (2) Actual vs. LRNLT Estimated W_0 (eV) comparison with the probabilities associated with median = 5.541, $n_1 = 2$, $n_2 = 2$ and $n_1 + n_2 < 20$ as observed value by using Fisher test equals 0.667 ($p = 0.667$) so that, $p = 0.667 > p = 0.05$. (3) SVM. vs. LRNLT estimated W_0 (eV), the probability is related to median = 5.084, $n_1 = 2$, $n_2 = 2$, and $n_1 + n_2 < 20$ as the observed value. Calculating the estimated W_0 (eV) comparison using the Fisher test yields a result of 0.167 ($p = 0.167$), meaning that $p = 0.167 > p = 0.05$.

The difference in the measures of the central tendency of the entire sample was not significant, according to the results of the median test for two independent samples in G4. In other words, there was no statistically significant difference between the measures of the central tendency in G4 as the

number of valence electrons equals 4 from all paired and the semiconductor's actual work function.

The results of paired comparisons into G1, G2, G3, and G4 showed the effects of ions; the qualitative estimation of work function would be affected according to the theoretical calculation by (Wigner and Bardeen, 1935). The results of paired comparisons into G1, G3, and G4 showed the difference between the measures of the central tendencies from the LRNLT estimated W_0 (eV) was not significant with the actual and SVM estimated W_0 (eV) of the semiconductor. Performance of LRNLT estimated the W_0 (eV) of G1, such as Li, Na, Cs, and G3, namely Al, corresponding to estimated by using Lang and Kohn's (1971) model under the concept of pseudo-potential theory, which includes the effects of ions. Such an impact will affect the minimum energy required to move electrons to the surface of the metal (work function).

Due to the fact that germanium (Ge) and silicon (Si) are far superior electrical conductors to diamond, the performance of LRNLT estimated the W_0 (eV) is likely to be compatible with the predictions of G4, namely Si and Ge. The difference between the measures of the central tendencies from the LRNLT estimated W_0 (eV) was not significant compared with the actual and SVM estimated W_0 (eV) of Si and Ge work function predictions from research by Owolabi *et al.* (2014). Also, in such research, the work functions estimated using SVM were compared with the experimental results as well as the results from other theoretical models, namely the excellent agreement with the experimental work function in comparison with the metallic plasma model (MPM) of Olatunji *et al.* (2011).

The explanation virtually explains the properties of metals using Band theory (Singleton, 2001). The set of forbidden energies does not correspond to any allowed combinations of atomic orbitals. They extend farther from the nucleus, the valence orbitals of adjacent atoms, and interact much more strongly with one another than do the filled core levels; as a result, the valence bands have a larger bandwidth. So the result is overlapping bands. Molecular orbitals are derived from two or more different kinds of valence electrons that have similar energies. These have molecular orbitals derived from two or more valence orbitals with similar energies. The valence band is filled with one, two, or three electrons per atom for Na (G1), Mg (G2), and Al (G3), respectively.

The combined band that arises from the overlap of the 3s and 3p bands is also filling up; it has a total capacity of eight electrons per atom (two electrons for each 3s orbital and six electrons for each set of 3p orbitals). With Na, therefore, which has one valence electron (G1), the combined valence band is one-eighth filled; with Mg (two valence electrons; G2), it is one-fourth filled; and with Al (G3), it is three-eighths filled, as indicated. The effect of the combined band that arises from the overlap leads to different unoccupied energy levels at an infinitesimally small energy level above the highest occupied level.

Although the performance of the obtained results with G1, G3, and G4 of LRNLT was not significantly different from the SVM estimated W_0 (eV) of the semiconductor. But both modeling approaches were significantly different from the actual work function W_0 (eV) with G2. Due to this fact, the prediction of work function from both LRNLT and SVM modeling approaches covers the variance due to the effect of the combined band that arises from the overlap of G1, G3, and G4, but it is not covered by G2. Therefore, the LRNLT modeling approach must be developed. It is proposed to include factors related to the effect of the combined band that arises from the overlap in the next model as well.

Conclusions

The main goal of this research is to compare the obtained results with estimating the work function of semiconductor variation data from SVM and LRNLT. The findings of predicting the work function from semiconductor variation data reveal that the LRNLT outperforms the SVM modeling strategy. The statistical significance technique revealed a comparison of the performance of the

LRNLT and SVM modeling approaches with only %RMSD and %MBD, which is insufficient to confirm that the LRNLT performs better than the SVM modeling approach. As a result, we must evaluate the effectiveness of the LRNLT and SVM modeling approaches in order to establish how effectively the relationship between two variables can be described using the analysis of variance of correlation from a nonparametric perspective. In addition, the statistical significance approach shows the imperfection of LRNLT modeling that is caused by factors related to the effect of the combined band that arises from the overlap, which leads to the development of LRNLT modeling being perfect next time.

Acknowledgement

The financial support of Thailand Science Research and Innovation (TSRI) under contract No. 2610293, Pathumwan Institute of Technology is acknowledged.

References

- Armstrong, R.A., Eperjesi, F., and Gilmartin, B. (2022). The application of analysis of variance (ANOVA) to different experimental designs in optometry. *Ophthalmic and Physiological Optics*, 22(3):248-256. <https://doi.org/10.1046/j.1475-1313.2002.00020.x>
- Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20:273-297. <https://doi.org/10.1007/BF00994018>
- Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American Magazine*, 248(5):116-130. <https://doi.org/10.1038/scientificamerican0583-116>
- Efron, B. and Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36-48. <https://doi.org/10.1080/00031305.1983.10483087>
- Goldman, L., Caldera, D.L., Nussbaum, S.R., Southwick, F.S., Krogstad, D., Murray, B., Burke, D.S., O'Malley, T.A., Goroll, A.H., Caplan, C.H., Nolan, J., Carabello, B., and Slater, E.E. (1977). Multifactorial index of cardiac risk in noncardiac surgical patients, *N Engl J Med.*, 297(16):845-850. <https://doi.org/10.1056/NEJM197710202971601>
- Kiejna, A. (1999). Stabilized jellium-simple model for simple-metal surfaces. *Progress in Surface Science*, 61(5-6):85-125. [https://doi.org/10.1016/S0079-6816\(99\)00011-8](https://doi.org/10.1016/S0079-6816(99)00011-8)
- Krzywinski, M. and Altman, N. (2014). Analysis of variance and blocking, *Nature Methods*, 11(7):699-700. <https://doi.org/10.1038/nmeth.3005>
- Lang, N.D. and Kohn, W. (1971). Theory of metal surfaces: work function. *Physical Review B*, 3(4):1215-1223. <https://doi.org/10.1103/PhysRevB.3.1215>
- Maksuwan, A., Changjan, A. and Pramuanl, P. (2023b). The application of analysis of variance to different iron-based superconductor critical temperature modeling. *Suranaree Journal of Science and Technology*, 30(4):030128(1-9). <https://doi.org/10.55766/sujst-2023-04-e02609>

- Maksuwan, A., Changjan, A., and Pramuanl, P. (2023a). The application of analysis of variance to different yttrium barium copper oxide superconducting transition temperature modeling. *Suranaree Journal of Science and Technology*, 30(4):030124(1-8). <https://doi.org/10.55766/sujst-2023-05-e02267>
- Michaelson, H.B. (1977). The work function of the elements and its periodicity. *Journal of Applied Physics*, 48(11):4729-4733. <https://doi.org/10.1063/1.323539>
- Olatunji, S.O., Selamat, A., and Abdulraheem, A. (2014). A hybrid model through the fusion of type-2 fuzzy logic systems and extreme learning machines for modelling permeability prediction. *Information Fusion*, 16:29-45. <https://doi.org/10.1016/j.inffus.2012.06.001>
- Olatunji, S.O., Selamat, A., and Raheem, A.A.A. (2011). Predicting correlation properties of crude oil systems using type-2 fuzzy logic systems. *Expert Systems with Applications*, 38(9):10911-10922. <https://doi.org/10.1016/j.eswa.2011.02.132>
- Owolabi, T., Akande K.O., and Olatunji, S.O. (2014). Prediction of superconducting transition temperatures for Fe-based superconductors using support vector machine. *Advances in Physics Theories and Applications*, 35:12-26.
- Rohwerder, M. and Turcu, F. (2007). High-resolution Kelvin probe microscopy in corrosion science: Scanning Kelvin probe force microscopy (SKPFM) versus classical scanning Kelvin probe (SKP). *Electrochimica Acta*, 53(2):290-299. <https://doi.org/10.1016/j.electacta.2007.03.016>
- Siegel, S. (1956). *Nonparametric Statistics for The Behavioral Sciences*. McGraw-Hill Book Company, INC, New York Toronto London.
- Singleton, J. (2001). *Band Theory and Electronic Properties of Solids*. Oxford Master Series in Physics. Oxford University Press. <https://doi.org/10.1093/oso/9780198506454.001.0001>
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. 1st ed. Springer New York, NY, 188p. <https://doi.org/10.1007/978-1-4757-2440-0>
- Wang, J. and Wang, S.-Q. (2014). Surface energy and work function of fcc and bcc crystals: Density functional study. *Surface Science*, 630:216-224. <https://doi.org/10.1016/j.susc.2014.08.017>
- Wigner, E. and Bardeen, J. (1935). Theory of the work functions of monovalent metals. *Physical Review*, 48(1):84-87. <https://doi.org/10.1103/PhysRev.48.84>