



مدل سازی و مطالعه ارتباط کمی ساختار- خاصیت (QSPR)، جهت پیش بینی مقادیر آنتالپی تبخیر دسته وسیعی از ترکیبات آلی با استفاده از توصیف کننده های ساختاری

مهدی نکویی*، مجید محمدحسینی

دانشگاه آزاد اسلامی، واحد شاهرود، دانشکده علوم پایه، گروه شیمی، شاهرود، ایران

تاریخ ثبت اولیه: ۱۳۹۲/۸/۱، تاریخ دریافت نسخه اصلاح شده: ۱۳۹۲/۹/۵، تاریخ پذیرش قطعی: ۱۳۹۲/۱۰/۲۵

چکیده

مطالعه ارتباط کمی ساختار- خاصیت (QSPR) جهت پیش بینی مقادیر آنتالپی تبخیر دسته وسیعی از ترکیبات آلی با استفاده از توصیف کننده های ساختاری انجام شد. با توجه به اهمیت موضوع، یک مدل ساده، قوی، توصیفی و قابل تفسیر بر اساس ارتباط کمی ساختار- خاصیت بوسیله روش رگرسیون خطی چند گانه (MLR) برای پیش بینی آنتالپی تبخیر ساخته شد. با استفاده از ساختار مولکول و محاسبه توصیف کننده ها، نه توصیف کننده مهم تر که با مقادیر آنتالپی تبخیر ارتباط بیشتری دارند انتخاب و با استفاده از آنها، مدل ساخته شد. روش پیشنهادی توسط چند تکنیک از جمله ارزیابی تقاطعی و تقسیم داده ها به دو گروه آموزش و پیش بینی مورد ارزیابی قرار گرفت. نتایج نشان می دهد که تکنیک های خطی مانند رگرسیون خطی چند گانه که با یک روش انتخاب متغیر مناسب کوپل شده باشد، قادر است مدل های مناسبی برای پیش بینی آنتالپی تبخیر ترکیبات ارائه نماید. یک مدل با خطای پیش بینی کم و ضریب همبستگی بالا بدست آمد. این مدل می تواند برای پیش بینی آنتالپی تبخیر ترکیبات مشابه که در مدل بکار گرفته نشده اند بکار رود.

واژه های کلیدی: مطالعه ارتباط کمی ساختار، خاصیت- رگرسیون خطی چند گانه- آنتالپی تبخیر.

۱. مقدمه

مقدار کل گرمای لازم برای تبخیر یک مول از مایع در دمای معین، آنتالپی تبخیر مولی آن مایع نامیده می شود. در مراجع علمی، اندازه گیری آنتالپی تبخیر ترکیبات از اهمیت بالایی برخوردار است. روش های آزمایشگاهی جهت اندازه گیری این پارامتر، وقت گیر، هزینه بر و نسبتاً مشکل است [۴-۱]. بنابراین اندازه گیری آن با روش های محاسباتی و کمومتریکس از اهمیت بالایی برخوردار است. یکی از این روش ها که امروزه از آن جهت پیش بینی خواص ترکیبات استفاده می شود روش ارتباط کمی ساختار- خاصیت (QSPR) می باشد. با استفاده از این روش ها می توان ارتباطی منطقی بین آنتالپی تبخیر با تعدادی از توصیف کننده های مولکولی که برگرفته شده از ساختار مولکولی می باشند برقرار و از این مدل ها جهت محاسبه آنتالپی تبخیر سایر ترکیبات مجهول استفاده نمود. توصیف کننده ها کمیت هایی هستند که به ساختار

*عهده دار مکاتبات: مهدی نکویی

نشانی: شاهرود - دانشگاه آزاد اسلامی - دانشکده علوم - گروه شیمی

تلفن: ۰۲۳۳۲۳۹۴۲۸۹ پست الکترونیک: E-Mail: m_nekoei1356@yahoo.com

مولکول ارتباط دارند و برای هر مولکول مقادیر خاصی را به خود اختصاص می دهند. یافتن رابطه ریاضی بین یک خاصیت ترکیب و توصیف کننده های مولکولی با استفاده از تکنیک های آماری مختلف امکان پذیر است. این ارتباط معمولاً توسط تکنیک های چند متغیره از جمله رگرسیون خطی چندگانه (MLR)، آنالیز اجزای اصلی (PCA) و حداقل مربعات جزئی (PLS)، شبکه های عصبی مصنوعی و ... ایجاد می شود. روشی که معمولاً بیشتر مورد استفاده قرار می گیرد، روش MLR است. برای نیل به این هدف مراحل زیر جهت مدل سازی آنتالپی تبخیر ترکیبات مختلف انجام شد.

۲. انتخاب سری داده ها

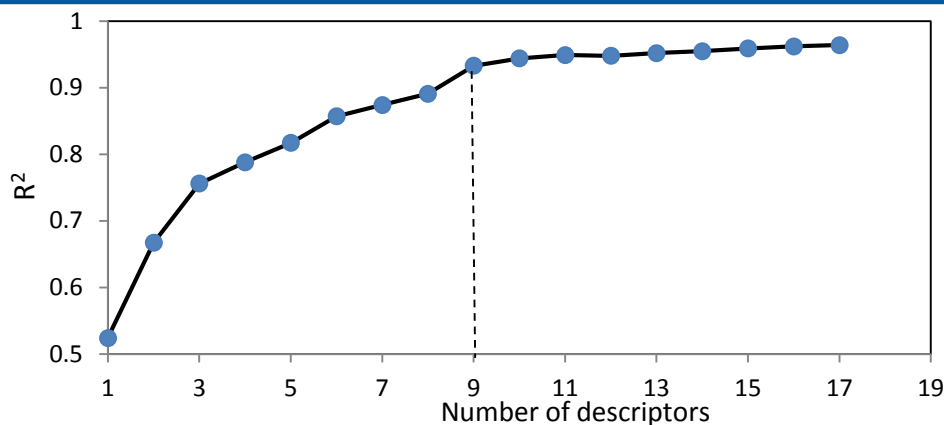
سری داده ها که برای مطالعه ارتباط کمی ساختار-خاصیت انتخاب گردید، مربوط به آنتالپی تبخیر ۲۹۹ ترکیب آلی می باشد [۵]. در این کار این ترکیبات به صورت تصادفی به دو گروه سری آموزش و سری تست تقسیم شدند (جدول ۱)، سری آموزش شامل ۲۳۸ مولکول و سری تست شامل ۶۱ مولکول می باشند. مقادیر آنتالپی تبخیر به عنوان متغیر وابسته و توصیف کننده ها به عنوان متغیر مستقل انتخاب شدند. سری آموزش جهت ایجاد یک مدل مناسب و سری تست جهت ارزیابی مدل مورد استفاده قرار گرفت.

۳. محاسبه توصیف کننده ها

در ابتدا برای محاسبه توصیف کننده ها، ساختار ترکیبات به کمک نرم افزار Hyperchem رسم شدند. سپس ساختارهای مولکولی رسم شده، به وسیله الگوریتم AM1 بهینه شدند. با استفاده از این نرم افزار می توان اطلاعات فراوانی نظیر زوایای پیوندی، طول پیوندها، زوایای پیچش، بار اتم ها، انرژی تشکیل مولکول و... را بدست آورد. ساختارهای بهینه شده به نرم افزار دراگون منتقل و توصیف کننده ها به تعداد ۱۴۹۷ مورد به کمک این نرم افزار محاسبه شدند. این نرم افزار برای محاسبه هیجده دسته از توصیف کننده های مولکولی مورد استفاده قرار می گیرد [۶-۷].

۴. مدل سازی با روش رگرسیون خطی چندگانه

برای ساختن مدلی که بیان گر ارتباط ساختاری ترکیبات مورد بررسی با آنتالپی تبخیر آنها باشد، از روش رگرسیون خطی چندگانه (MLR) استفاده شد. طبیعی است که تعداد زیاد توصیف کننده ها باعث پیچیدگی محاسبات شده و همچنین احتمال وجود فاکتورهای دارای برهم کنش با هم را افزایش می دهد. لذا تعدادی از این توصیف کننده ها که دارای ۹۰٪ مقادیر یکسان بودند حذف شدند و همچنین از بین توصیف کننده هایی که دارای همبستگی بیش از ۰/۹ بودند توصیف کننده ای که دارای همبستگی کمتری با متغیر وابسته بود، از سری داده ها حذف شد [۸-۹]. سپس مقادیر توصیف کننده های محاسبه شده توسط نرم افزار Dragon پس از کاهش تعداد متغیرها بعنوان متغیرهای مستقل و مقادیر آنتالپی تبخیر مولکول های مورد نظر بعنوان متغیرهای وابسته بعنوان ورودی به نرم افزار SPSS وارد شد. در نهایت با استفاده از منوی آنالیز، گزینه ی رگرسیون خطی و روش مرحله ای انتخاب و نهایتاً چندین مدل مختلف به طور جداگانه به دست آمد، که با توجه به خصوصیات آماری آنها از جمله ضریب همبستگی (R)، آماره F و خطای استاندارد و پس از رسم مقادیر R و R² و SE بر حسب تعداد توصیف کننده ها بهترین مدل که دارای بیشترین مقدار R و F و کمترین مقدار خطای استاندارد و شامل توصیف کننده های تا حد امکان قابل توجه باشد، به عنوان مدل نهایی برای ارتباط آنتالپی تبخیر مولکول ها با ساختار آنها انتخاب شد. با این روش مدل نهم با تعداد ۹ توصیف کننده به عنوان مناسب ترین آنها انتخاب شد و توسط روش MLR مدل سازی و مورد ارزیابی قرار گرفت. شکل ۱ تأثیر تعداد توصیف کننده ها را بر مقدار R² نشان می دهد. همانطور که ملاحظه می شود، تغییرات R² بعد از ۹ توصیف کننده خیلی کم می باشد. بنابراین ۹ توصیف کننده جهت مدل سازی انتخاب گردید.



شکل ۱. تأثیر تعداد توصیف کننده ها بر مقدار R^2 .

فهرست توصیف کننده‌های انتخاب شده توسط نرم افزار SPSS به همراه توصیف مختصری از آنها در جدول ۲ آورده شده است.

جدول ۲. توصیف کننده‌های انتخاب شده با SPSS و توصیف آنها.

توصیف کننده	نوع توصیف کننده	علامت	ضریب
Narumi simple topological index (log).	topological	SNar	1.203
Maximal electrotopological positive variation	geometrical	MAXDP	0.167
O= Atom-centred fragments	Atom-centred fragments	O-058	-0.569
Moran autocorrelation – lag 1 / weighted by atomic polarizabilities	2D autocorrelations	MATS1p	4.362
Rotatable bond fraction	Constitutional	RBF	10.598
Complementary information content (neighborhood symmetry of 0-order).	topological	CIC0	-2.031
Radial Distribution Function – 1.0 / weighted by atomic Sanderson electronegativities	RDF	RDF010e	17.525
Radial Distribution Function – 1.0 / unweighted	RDF	RDF010u	-19.091
Absolute eigenvalue sum on geometry matrix	geometrical	SEig	+0.269
Constant			76.978

۴-۱. ایجاد مدل با استفاده از SW-MLR

پس از انتخاب مناسب‌ترین توصیف کننده‌ها توسط روش مرحله‌ای با استفاده از SPSS، مرحله بعدی ایجاد مدل، میان توصیف کننده‌های انتخاب شده و آنتالپی تبخیر ترکیبات می‌باشد. از نرم افزار SPSS برای این منظور استفاده گردید. بین توصیف کننده‌ها و آنتالپی تبخیر ترکیبات سری آموزش با استفاده از روش MLR را بطه‌ی زیر بدست آمد:

$$\Delta H_{\text{vap}} = 76.978 + 1.203 \text{ SNar} + 0.167 \text{ MAXDP} - 0.569 \text{ O-058} + 4.362 \text{ MATS1p} + 10.598 \text{ RBF} - 2.031 \text{ CIC0} + 17.525 \text{ RDF010e} - 19.091 \text{ RDF010u} + 0.269 \text{ SEig} \quad (1)$$

سپس از این معادله برای پیش‌بینی آنتالپی تبخیر سری پیش‌بینی استفاده گردید.

مقادیر واقعی و پیش‌بینی شده آنتالپی تبخیر و همچنین خطاهای پیش‌بینی برای کلیه ترکیبات مجموعه آموزش و پیش‌بینی در جدول ۱ آورده شده است. شکل ۲ نمودار مقادیر آنتالپی تبخیر محاسبه شده برای سری آموزش و پیش‌بینی بر حسب مقادیر تجربی را نشان می‌دهد. در این شکل، میزان نزدیکی داده‌ها به خط راست قدرت پیشگویی مدل را نشان می‌دهد. شکل ۳، مقادیر باقیمانده‌ی خطاها (اختلاف مقدار پیش‌بینی شده و مقدار واقعی آنتالپی تبخیر) را نسبت به مقادیر تجربی نشان می‌دهد. میزان پراکندگی خطاها در اطراف محور نشان دهنده این است که خطای سیستماتیک در مدل وجود ندارد.

جدول ۱. مقادیر تجربی و محاسبه شده آنتالپی تبخیر برای ترکیبات مختلف برای مجموعه های آموزشی و پیش بینی در مدل SW-MLR همراه با مقادیر خطای نسبی

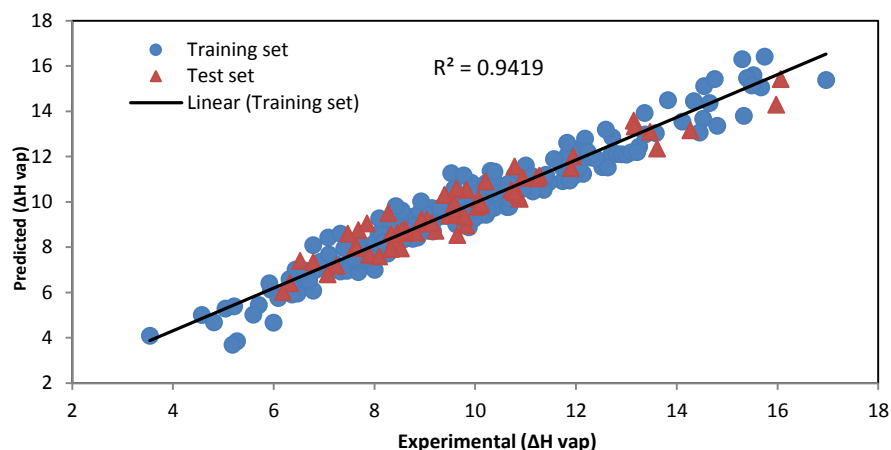
Compound	ΔH_{vap} (Exp.)	SW- MLR	%E	Compound	ΔH_{vap} (Exp.)	SW- MLR	%E
Training set				Training set			
Propane	3.54	4.07	15.23	2,4,4- Tri methylhexane	9.54	9.89	3.71
n-Butane	5.04	5.29	5.03	3,3-Diethylpentane	10.13	10.48	3.48
2-Methyl Propane	4.57	5.00	9.56	3-Ethyl-2,2-dimethylpentane	9.89	9.86	-0.28
2-Methyl butane	5.94	6.11	2.97	3-Ethyl-2,3-dimethylpentane	10.18	10.12	-0.53
2,2-Dimethylpropane	5.21	5.39	3.47	3-Ethyl-2,4-dimethylpentane	9.99	10.37	3.81
n-Hexane	7.54	7.44	-1.25	2,2,3,4- Tetramethylpentane	9.75	9.56	-1.85
2-Methylpentane	7.14	7.15	0.26	2,2,4,4- Tetramethylpentane	9.15	9.32	1.90
2,2-Dimethyl butane	6.62	6.48	-2.06	2,3,3,4- Tetramethylpentane	10.21	9.43	-7.60
2,3 -Dimethyl butane	6.96	7.06	1.52	n-Decane	12.28	11.95	-2.64
n-Heptane	8.74	8.52	-2.50	n-Dodecane	14.65	14.36	-1.91
2-Methyl hexane	8.32	8.24	-0.91	pentaneCyclo	6.82	6.97	2.31
3- Methyl hexane	8.42	8.40	-0.21	Cyclohexane	7.90	7.42	-6.04
2,2-Dimethylpentane	7.75	7.55	-2.47	Methyl cyclopentane	7.55	8.08	7.14
2,3-Dimethylpentane	8.19	8.22	0.36	Ethyl Cyclopentane	8.72	9.33	7.03
2,4-Dimethylpentane	7.86	8.03	2.17	Cis-1,2-Dimethyl Cyclopentane	8.55	8.93	4.54
2,2,3-Trimethylbutane	7.66	7.46	-2.52	Cis-1,3-Dimethyl Cyclopentane	8.20	8.76	6.91
n-Octane	9.92	10.83	9.24	1,1- Dimethyl Cyclopentane	8.08	8.57	6.13
2-Methylheptane	9.48	9.35	-1.30	Cis-1,2-Dimethyl Cyclohexane	9.49	9.78	3.07
3- Methylheptane	9.52	9.39	-1.31	Cis-1,3-Dimethyl Cyclohexane	9.14	9.72	6.34
3-Ethylhexane	9.48	9.35	0.39	Cis-1,4-Dimethyl Cyclohexane	9.33	9.63	3.28
2,2-Dimethylhexane	8.91	8.64	-3.01	Ethylcyclohexane	9.67	9.92	2.68
2,3- Dimethylhexane	9.27	9.33	0.70	n-Propyl cyclohexane	10.78	11.17	3.63
2,4- Dimethylhexane	9.03	9.20	1.95	n-Butyl cyclopentane	11	11.60	5.48
3,3- Dimethylhexane	8.97	8.78	-2.04	n-Buthylcyclohexane	11.96	12.34	3.17
3,4- Dimethylhexane	9.14	9.40	2.85	n-Pentyl Cyclopentane	12.18	12.78	5.00
2-Methyl-3- ethylpentane	9.21	9.47	2.90	n-Hexyl Cyclopentane	13.36	13.92	4.20
3- Methyl-3-ethylpentane	9.08	8.96	-1.27	n-Heptyl Cyclopentane	14.54	15.11	3.94
2,2,4- Trimethylpentane	8.40	8.64	2.88	1-Butene	4.81	4.67	-2.85
2,3,3- Trimethylpentane	8.90	8.66	-2.62	1-Pentene	6.09	5.75	-5.46
2,3,4- Trimethylpentane	9.01	8.93	-0.88	2-Methyl -2-butene	6.47	5.95	-7.96
n-Nonane	11.10	10.83	-2.37	3-Methyl -1-butene	5.7	5.43	-4.60
3- Methyl octane	10.73	10.56	-1.56	1-Hexene	7.32	6.93	-5.30
4- Methyl octane	10.68	10.55	-1.12	3,3-Dimethyl-1-butene	6.36	5.91	-6.98
3- Ethylheptane	10.64	10.63	-0.02	4,4-Dimethyl-1-pentene	7.45	6.95	-6.60
4- Ethylheptane	10.56	10.68	1.14	3,3- Dimethyl-1-pentene	8.00	7.00	-12.46
2,3-Dimethylheptane	10.42	10.47	0.48	2,2,3-Trimethyl-1-butene	7.68	6.89	-10.19
2,4- Dimethylheptane	10.18	10.31	1.34	1-Octene	9.62	9.01	-6.32
2,5- Dimethylheptane	10.33	10.23	-0.94	1-Decene	12.05	11.33	-5.91
2,6- Dimethylheptane	10.30	10.21	-0.81	1-Undecene	13.25	12.44	-6.11
3,4- Dimethylheptane	10.42	10.53	1.10	1-Dodecene	14.52	13.67	-5.84
3,5- Dimethylheptane	10.25	10.35	1.00	1,3-Butadiene	5.27	3.84	-27.03
4,4- Dimethylheptane	9.92	9.93	0.16	Methylbenzene	9.09	8.88	-2.22
3-Ethyl-2-methylhexane	10.30	10.56	2.60	1,4-Dimethylbenzene	10.23	9.98	-2.35
3-Ethyl-4-methylhexane	10.37	10.58	2.05	1,2-Dimethylbenzene	10.38	9.74	-6.14
4-Ethyl-2-methylhexane	10.13	10.48	3.53	1,3-Dimethylbenzene	10.20	10.01	-1.86
2,2,3- Tri methylhexane	9.89	9.75	-1.32	1,2,3-Trimethylbenzene	11.73	10.91	-6.96
2,2,4- Tri methylhexane	9.51	9.66	1.58	1,2,4-Trimethylbenzene	11.46	10.86	-5.18
2,3,3- Tri methylhexane	9.94	9.79	-1.41	1,3,5-Trimethylbenzene	11.35	11.10	-2.18
2,3,4- Tri methylhexane	10.25	10.26	0.09	1-Methyl-2-ethylbenzene	11.40	11.19	-1.81
2,3,5- Tri methylhexane	9.92	10.22	3.04	1-Methyl-4-ethylbenzene	11.14	11.07	-0.57
n-Propylbenzene	11.05	11.05	0.07	Isopropyl acetate	8.89	9.50	2.59
Isopropylbenzene	10.79	10.88	0.84	Methyliso butanoate	8.92	10.02	6.94
1-Methyl-2-isopropylbenzene	12.10	12.23	1.14	Ethyl propionate	9.37	9.77	12.37

Compound	ΔH_{vap} (Exp.)	SW- MLR	%E	Compound	ΔH_{vap} (Exp.)	SW- MLR	%E
1-Methyl-4-isopropylbenzene	12.02	12.07	0.45	Butyl acetate	10.21	10.83	4.35
n- Butylbenzene	11.98	12.18	1.70	Propyl propionate	10.38	10.87	6.09
Sec- Butylbenzene	11.83	12.07	2.04	Ethyl isobutanoate	9.25	11.26	4.77
Tert- Butylbenzene	11.73	11.50	-1.92	Methyl petanoate	10.3	11.36	18.30
n-Hexylbenzene	14.34	14.44	0.70	Methyl amine	5.59	5.01	10.38
n-Heptylbenzene	15.52	15.58	0.44	Dimethyl amine	5.99	4.66	-10.24
Butanone	8.26	7.72	-6.52	Isopropyl amine	6.78	8.08	-22.19
2-Pentanone	9.16	8.69	-5.07	Trimethyl amine	5.18	3.68	19.24
3-Methyl-2-butanone	8.85	8.42	-4.76	Butyl amine	8.54	9.62	-28.76
2-Hexanone	10.26	9.63	-6.09	Isobutyl amine	8.09	9.28	12.64
3-Hexanone	10.12	9.68	-4.28	Cyclopentylamine	9.77	11.7	14.71
3,3-Dimethyl-2-butanone	9.15	8.72	-4.63	Butylethylamine	9.6	10.38	18.95
2-Heptanone	11.13	10.65	-4.29	Isopropyl propyl amine	8.42	9.8	8.30
2-Methyl-3-hexanone	10.44	10.41	-0.24	Tri ethylamine	8.33	8.97	10.36
2,4-Dimethyl-3- pentanone	9.73	10.20	4.90	Dipropyl amine	9.57	10.37	14.38
2-Methyl-3- heptanone	11.87	11.39	-3.99	Cyclohexylamine	10.38	11.33	8.12
5-Nonanone	12.72	12.84	0.99	Di butyl amine	11.82	12.60	16.46
2,6-Dimethyl-4- heptanone	12.23	12.21	-0.08	Benzenamine	13.21	12.20	7.69
6-Undecanone	15.67	15.06	-3.88	2-Methylbenzenamine	12.59	13.19	8.38
2-Undecanone	15.33	13.79	-9.98	N- Methylbenzenamine	11.56	11.88	9.22
Butanal	7.99	7.81	-2.23	N-Ethylbenzenamine	13.37	12.98	6.65
Ethanol	10.11	9.42	-0.79	N,N-Dimethylbenzenamine	12.63	11.52	-7.59
1-Propanol	11.36	10.52	-6.79	N,N-Diethylbenzenamine	14.11	13.54	4.79
1-Butanol	12.52	11.54	-7.32	2-Methyl pyridine	10.12	9.98	2.80
2-Butanol	11.87	11.26	-7.82	3- Methyl pyridine	10.63	9.76	-2.88
2-Methyl-1-propanal	12.14	11.23	-5.09	4- Methyl pyridine	10.67	9.78	-8.77
2-Methyl-2- propanal	11.14	10.46	-7.46	2,3- Dimethyl pyridine	11.33	11.01	-4.03
2-Pentanol	12.70	12.08	-6.03	2,5- Dimethyl pyridine	11.13	11.13	-1.36
3-Pentanol	12.80	12.11	-4.81	2,6- Dimethyl pyridine	10.84	11.03	-8.11
2-Methyl-1-butanol	12.90	12.09	-5.32	3,4- Dimethyl pyridine	11.96	11.15	-8.31
3-Methyl-1-butanol	13.00	12.09	-6.23	3,5- Dimethyl pyridine	11.87	10.94	-2.74
3-Methyl-2-butanol	12.40	11.90	-6.96	Butanenitrile	9.40	9.51	0.05
1-Hexanol	14.80	13.35	-4.02	2-Methylpropanenitrile	8.87	9.04	1.75
4-Methyl-1-pentanol	14.45	13.06	-9.74	Pentanenitrile	10.42	10.38	-6.74
2-Methyl-2- pentanol	13.09	12.18	-9.57	3-Methylbutanenitrile	9.95	10.06	-7.81
1-Octanol	16.96	15.38	-6.94	Hexanenitrile	11.40	11.14	1.23
Phenol	13.82	14.48	-9.30	Octanenitrile	13.58	13.03	1.92
o-Cresol	15.39	15.46	4.81	Decanenitrile	15.49	15.14	-0.32
m-Cresol	14.75	15.42	0.50	Chloroethane	5.91	6.40	1.11
2,4-Xylenol	15.74	16.41	4.56	2-Chloropropane	6.43	7.01	-2.27
3-Ethylphenol	15.30	16.30	4.31	1-Chlorobutane	8.01	8.19	-4.03
Diethyl ether	6.51	6.84	6.55	2-Chlorobutane	7.54	7.86	-2.25
Isopropylmethyl ether	6.31	6.59	5.21	2-Chloro -2- methyl propane	6.93	7.34	8.36
Butyl methyl ether	7.74	7.95	4.49	2- Chloropentane	8.61	8.74	9.05
Methyl pentyl ether	8.81	9.11	2.77	1-Chloro-3-methylbutane	8.66	8.85	2.31
Butylethyl ether	8.63	9.13	3.48	1,1-Dichloroethane	7.32	8.59	4.35
Dipropylether	8.53	9.13	5.20	1,2-Dichloroethane	8.40	8.93	6.00
Methyl formiate	6.78	6.08	7.07	Bromoethane	6.70	6.52	1.535
Methyl acctate	7.77	7.45	-10.27	1-Bromopropane	7.65	7.61	2.22
Ethyl formiate	7.63	7.37	-4.10	2-Bromopropane	7.21	7.28	17.42
Ethyl acetate	8.53	8.72	-3.35	1-Bromobutane	8.76	8.37	6.37
Methyl propionate	8.66	8.68	2.32	2-Bromo-2-methylpropane	7.60	7.50	-2.56
Propyl acctate	9.49	9.73	0.32	1-Bromopentane	9.87	8.87	-0.45
1,2-Dibromoethane	9.97	9.25	0.98	1-Hexanthiol	10.95	10.58	-0.32
1,2-Dibromopropane	9.96	9.60	-4.40	2,3-Dimethyl-2- butanethiol	9.41	9.59	1.38
1-Propanethiol	7.66	8.06	-1.21	Cyclohexanethiol	10.69	10.11	-0.85
2-Propanethiol	7.09	7.67	-10.09	1-Heptanethiol	12.05	11.51	-2.24

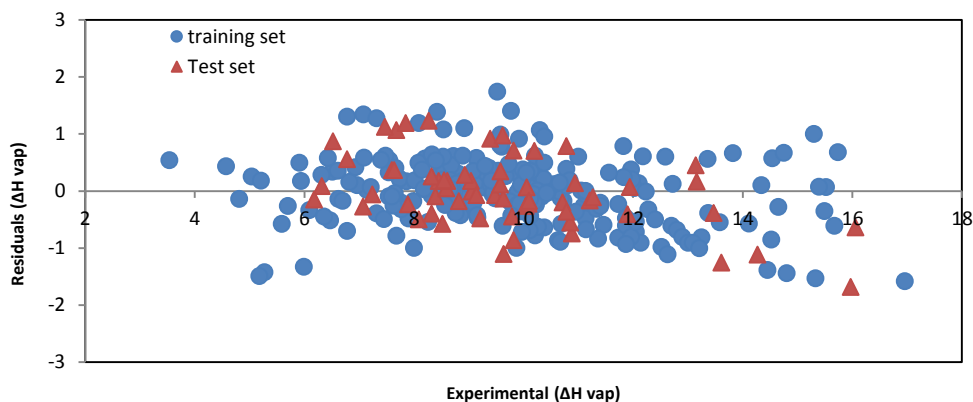
Compound	ΔH_{vap} (Exp.)	SW- MLR	%E	Compound	ΔH_{vap} (Exp.)	SW- MLR	%E
1-Butanethiol	8.78	8.91	-7.16	Methylthio methane	6.61	6.96	-3.30
2-Butanethiol	8.15	8.52	-3.56	1-Methylthio propane	8.66	8.70	1.92
2-Methyl-2-propanethiol	7.40	7.94	5.32	2-Methylthio propane	8.18	8.22	-5.36
1-Pentanethiol	9.89	9.64	8.23	1-Methylthiobutane	9.78	9.35	-4.44
2-Methyl-1-butanethiol	9.43	9.37	1.52	2-Methyl-2-methylthio propane	8.57	8.32	5.38
3-Methyl-1-butanethiol	9.45	9.41	4.57	2-Ethylthio propane	9.08	9.42	0.48
3-Methyl-2-butanethiol	8.98	9.10	7.38	1-Ethylthio butane	10.66	10.80	0.50
2,2-Dimethyl-1-propanethiol	8.72	8.64	-2.44	2-Ethylthio butane	10.05	10.37	-4.36
Cyclopentanethiol	9.93	9.70	-0.55	2-Ethylthio-2-methylpropane	9.3	9.74	-2.84
Test set				Test set			
n-Pentane	6.32	6.40	1.35	2-Propanol	10.85	10.29	-5.08
3-Methylpentane	7.24	7.17	-0.85	1-Pentanol	13.61	12.35	-9.21
3-Methylhexane	8.39	8.29	-1.07	2-Methyl-2-butanol	11.9	11.49	-3.39
3,3-Dimethylpentane	7.89	7.65	-2.93	1-Heptanol	15.97	14.28	-10.52
4-Methylheptane	9.48	9.41	-0.73	p-Cresol	16.06	15.42	-3.98
2,5-Dimethylhexane	9.05	9.03	-0.11	Ethyl n-propyl ether	7.6	7.96	4.81
2,2,3-Trimethylpentane	8.82	8.64	-2.02	Diisopropyl ether	7.68	8.74	13.91
2-Methyloctane	10.71	10.51	-1.85	Propylformiate	8.59	8.63	0.54
2,2-Dimethylheptane	10.04	9.78	-2.57	Methylbutanoate	9.39	10.30	9.77
3,3-Dimethylheptane	10.09	9.91	-1.68	Ethylbutanoate	10.20	10.90	6.92
3-Ethyl-3-methylhexane	10.06	10.12	0.61	Propylamine	7.47	8.59	15.06
3,2,5-Trimethylhexane	9.63	9.49	-1.37	Sec-Butylamine	7.85	9.04	15.21
3,3,4-Trimethylhexane	10.09	9.86	-2.18	Hexylamine	10.78	11.56	7.30
2,2,3,3-Tetramethylpentane	9.82	8.95	-8.76	Diisopropylamine	8.27	9.50	14.89
n-Undecane	13.47	13.08	-2.83	3-Methylbenzenamine	14.27	13.15	-7.80
Methylcyclohexane	8.45	8.62	2.07	Pyridine	9.64	8.53	-11.46
1,1-Dimethylcyclohexane	9.04	9.21	1.93	2,4-Dimethylpyrhdine	11.27	11.16	-0.97
n-Propylcyclopentane	9.82	10.52	7.21	Propanenitrile	8.61	8.78	2.05
n-Pentylcyclohexane	13.14	13.59	3.44	2,2-Dimethylpropanenitrile	8.93	9.21	3.19
2-Methyl-1-butene	6.18	6.02	-2.44	1-Chloropropane	6.78	7.33	8.22
1-Heptene	8.52	7.94	-6.73	1-Chloropentane	9.14	9.07	-0.76
1-Nonene	10.88	10.13	-6.84	1,1,2-Trichloroethane	9.62	10.59	10.09
Benzene	8.09	7.59	-6.14	1-Bromo-2-methylpropane	8.32	7.92	-4.76
Ethylbenzene	10.1	9.93	-1.68	Ethanethiol	6.52	7.39	13.38
1-Methyl-3-ethylbenzene	11.21	9.93	-1.49	2-Methyl-propanethiol	8.32	8.57	3.10
1-Methyl-3-isopropylbenzene	11.94	12.00	0.57	2-Methyl-2-butanethiol	8.55	8.73	2.18
n-Pentylbenzene	13.16	13.32	1.26	2-Methyl-2-pentanethiol	9.57	9.68	1.187
3-Pentanone	9.21	8.72	-5.21	Methylthioethane	7.64	8.00	4.79
4-Methyl-2-pentanone	9.79	9.34	-4.57	1-Ethylthio propane	9.58	9.29	3.62
2,5-Dimethyl-3-hexanone	10.94	11.07	1.23	Methylthio cyclopentane	10.78	10.41	-3.34
Propanal	7.07	6.79	-3.82				

۴-۲. ارزیابی اعتبار مدل های انتخاب شده

یکی از مواردی که می توان با استناد به آن از معتبر بودن مدل انتخابی اطمینان پیدا کرد، به دست آوردن ضریب تعیین (R^2) می باشد که هر چه این مقدار به یک نزدیک تر باشد مدل، معتبر تر خواهد بود. برای نیل به این امر مقادیر پیش بینی شده آنتالپی تبخیر را بر حسب مقادیر تجربی آن رسم می کنیم که شکل ۲ مؤید این مطلب می باشد. در نمودار فوق مقدار ضریب تعیین، برابر ۰/۹۴۱ است که حاکی از اعتبار مدل می باشد. مورد دیگری که می توان برای اعتبار مدل به آن استناد کرد رسم مقادیر باقی مانده حاصل از اختلاف مقادیر پیش بینی شده آنتالپی تبخیر بر حسب مقادیر تجربی آن می باشد که شکل ۳ آن را به خوبی نشان می دهد. در نمودار فوق می توان پراکندگی نسبتاً یکسان نقاط را حول مقادیر صفر، ناشی از عدم وجود خطای معین در روش و اعتبار مدل دانست.



شکل ۲. نمودار مقادیر آنتالپی تبخیر محاسبه شده با کمک مدل SW-MLR برای مجموعه‌های آموزشی و پیش‌بینی بر حسب مقادیر تجربی.



شکل ۳. نمودار تغییرات خطا برای مقادیر آنتالپی تبخیر محاسبه شده با کمک مدل SW-MLR برای مجموعه‌های آموزشی و پیش‌بینی.

درصد خطای نسبی برای سری پیش‌بینی^۱ (REP) نیز جهت بررسی اعتبار مدل مورد استفاده قرار می‌گیرد. هر چه مقدار REP کوچک‌تر باشد مدل، مدل مناسب‌تری خواهد بود. مقدار این آماره برابر $۶/۲۰۶\%$ می‌باشد که مقدار نسبتاً کوچک آن اعتبار مدل را نشان می‌دهد. یکی دیگر از روش‌ها که قبلاً به آن پرداخته شد اعتبارسنجی تقاطعی است که یکی از رایج‌ترین تکنیک‌ها جهت اعتبارسنجی مدل محسوب می‌شود. در این روش هر بار یک^۲ (LOO) یا تعداد معینی^۳ (LGO) از مولکول‌ها کنار گذاشته می‌شوند و مدل با مولکول‌های باقی مانده ساخته می‌شود، سپس کمیت مدل‌سازی برای مولکول‌های کنار گذاشته شده، توسط مدل حاصله پیش‌بینی می‌شود. این عمل به صورت چرخه‌ای برای تمام مولکول‌ها تکرار می‌شود به طوری که همه مولکول‌ها یک بار در سری پیش‌بینی قرار می‌گیرند. پارامتر آماری Q^2 برای ارزیابی مدل حاصل از ارزیابی تقاطعی بکار می‌رود. مقدار Q^2_{LOO} برابر $۰/۹۳۷$ و مقدار Q^2_{LGO} برابر $۰/۹۲۵$ نشان دهنده اعتبار مدل می‌باشد.

و در نهایت برای اینکه نشان دهیم مدل ارایه شده یک مدل شانس نیست از روش Y-randomization استفاده گردید. در این روش تمام آنتالپی‌های تبخیر، بطور تصادفی بهم ریخته و سپس مدل‌سازی و پیش‌بینی انجام می‌شود. مسلماً باید ضریب هم‌بستگی بسیار پایین بدست آید که اگر

^۱ - Relative error of prediction (REP)

^۲ - Leave-one-out (LOO)

^۳ - Leave-group-out (LGO)

اینطور شد نشان دهنده این است که مدل اولیه شانس نبوده و دارای اعتبار می‌باشد. نتایج ۱۰ بار تکرار در جدول ۳ آورده شده است. همانطور که مشاهده می‌شود در همه تکرارها ضریب هم‌بستگی بسیار پایین می‌باشد که نشان دهنده اعتبار مدل ارائه شده و شانس نبودن مدل می‌باشد.

جدول ۳. مقادیر R^2 و Q^2 برای ۱۰ بار Y-randomization.

Iteration	R^2	Q^2
1	0.072	0.025
2	0.201	0.085
3	0.084	0.141
4	0.072	0.213
5	0.142	0.135
6	0.305	0.044
7	0.008	0.052
8	0.142	0.023
9	0.301	0.085
10	0.125	0.149

۳-۴. پارامترهای آماری

جدول ۴ پارامترهای آماری مختلف را برای مدل SW-MLR نشان می‌دهد. به عنوان مثال مقادیر R^2 برای سری آموزش و پیش‌بینی به ترتیب عبارتند از: ۰/۹۴۱ و ۰/۹۱۹ و همچنین درصد خطای نسبی پیشگویی برای آنتالپی تبخیر محاسبه شده برای مجموعه پیش‌بینی با کمک این مدل برابر ۶/۲۰۶٪ می‌باشند. با توجه به پارامترهای آماری مناسب از جمله خطای کم آن، می‌توان از این مدل، برای محاسبه‌ی آنتالپی تبخیر سایر ترکیبات مجهول و مشابه استفاده نمود.

جدول ۴. پارامترهای مختلف آماری برای مدل SW-MLR.

Model	R^2_{training}	$R^2_{\text{prediction}}$	REP(%)	RMSEP	Q^2_{LOO}	Q^2_{LGO}	F_{train}	F_{test}
SW-MLR	0.941	0.919	6.206	0.602	0.937	0.925	442.27	57.39

۵. نتیجه‌گیری

در کار ارائه شده مطالعه ارتباط کمی ساختار-خاصیت جهت پیش‌بینی آنتالپی تبخیر دسته وسیعی از ترکیبات آلی انجام شد. در ابتدا توصیف کننده‌های ساختاری ایجاد و مناسب‌ترین آنها که بیشترین ارتباط را با آنتالپی تبخیر داشتند به کمک روش مرحله‌ای انتخاب گردیدند و به روش رگرسیون خطی چندگانه مدل مناسب ایجاد شد. مدل ایجاد شده دارای ۹ توصیف‌کننده می‌باشد که این توصیف‌کننده‌ها از دسته‌های گوناگون از جمله توپولوژیکی، هندسی، ساختاری و ... می‌باشند. با استفاده از این توصیف‌کننده‌ها مدلی ایجاد شد که قدرت پیش‌بینی آنتالپی تبخیر ترکیبات مشابه را داشت. این مدل با استفاده از روشهای مختلف از جمله ارزیابی تقاطعی مورد ارزیابی قرار گرفت. پارامترهای آماری مختلف نشان می‌دهد که مدل ارائه شده می‌تواند جهت پیش‌بینی آنتالپی تبخیر سایر ترکیبات قرار گیرد در ضمن از این روش می‌توان جهت مدل‌سازی و پیش‌بینی سایر خواص فیزیکوشیمیایی استفاده نمود.

قدردانی

نویسندگان از حوزه معاونت محترم پژوهشی دانشگاه آزاد اسلامی واحد شاهرود جهت حمایت از این کار پژوهشی تشکر و قدردانی می‌کنند.

مراجع

- [1] J. Garai, *Fluid Phase Equilibria* 283(2009) 89.
- [2] X. Ge, X. Wang, *Ind. Eng. Chem. Res.* 48(2009)2229.
- [3] X. Ge, X. Wang, *J. Sol. Chem.* 38(2009)1097.
- [4] S. Zemansky, *University Physics*, Addison-Wesley Publishing Company, Sixth ed., (1982).
- [5] D. Dalmazzone, A. Salmon, S. Guella, *Fluid Phase Equilibria*, 242 (2006) 29.
- [6] J. Hunger, G. Huttner, *J. Comput. Chem.*, 20 (1999) 455.
- [7] S. Ahmad, M.M. Gromiha, *J. Comput. Chem.*, 24 (2003) 1313.
- [8] C.L. Waller, M.P. Bradley, *J. Chem. Inf. Comput. Sci.*, 39 (1999) 345.
- [9] J. Aires-de-Sousa, M.C. Hemmer, J. Casteiger, *Anal. Chem.*, 74 (2002) 80.