Gmail DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>

FW: Paper ID CE20029 /Requesting paper review for JESTEC, First round of Review Process/

13 messages

Jestec <Jestec@taylors.edu.my> Sat, Jan 30, 2021 at 9:09 PM

Dear Dr.

Greetings from the Editorial Board of JESTEC.

The following attached manuscript titled

THE PREDICTION OF KOVATS RETENTION INDICES OF ESSENTIAL OILS AT GAS CHROMATOGRAPHY USING GENETIC ALGORITHM

has been submitted to JESTEC for consideration for publication, and I am writing to request that you, as an expert in its topic area, review it and make a recommendation regarding its acceptability.

I hope that you will agree to review this manuscript, in which case, I appreciate receiving your review within 2-3 weeks from the date of this email.

I highly appreciate your commitment as the timely reviews are of utmost importance to authors and to the journal editor.

If you would like to have more than 4 weeks to complete the review, could you please indicate the time frame within which you expect to return the review report.

I appreciate your contribution in maintaining the quality and value of JESTEC, and look forward to your response.

Best regards

Some quick guidelines to our respected reviewers

We would appreciate if you take note that whenever appropriate, papers are evaluated on the basis of the following seven criteria. Please try not to focus on the editorial issues/mistakes as too many of them may lead to the author's frustration. We want the authors, when we revise their paper, to focus on our comments/concern related to the seven above-mentioned criteria.

1. **Research question**: why the authors do this research and what is its importance and application.

2. **Novelty**: a paper gives new ideas, derivations, applications that has been not studied before or little- or not in depth-studied.
3. **Literature review**: to identify the research gap with recent references from 2010 onwards.
4. **Research methodology**: analytical, numerical or experimental or mixed. What is the contribution of the authors, assumptions and/or approximations used, description of apparatus and its limitations, steps of experiments, etc.
5. **Quality of results**: and the depth and logic of the discussion.
6. **Insight** conveyed and recommendations that might be used by others for future work.
7. **English:** used effectively to communicate the ideas and easy to understand with least or no grammatical error or typos.

Assoc. Prof. Dr. Abdulkareem Sh. Mahdi Al-Obaidi, CEng MIMechE

Executive Editor, Journal of Engineering Science & Technology

http://jestec.taylors.edu.my

2 attachments

Review Report - 2017.docx
62K

CE20029.docx
407K

DIDING SUHANDY <diding.sughandy@fp.unila.ac.id> Sat, Jan 30, 2021 at 9:12 PM
To: Jestec <Jestec@taylors.edu.my>

Dear Jestec,
Thank you for the email. I confirmed to accept this review invitation.

Thank you,

Best regards,

DS
[Quoted text hidden]

Jestec <Jestec@taylors.edu.my> Sat, Jan 30, 2021 at 9:34 PM
To: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>

Dear Dr.

Thank you in advance for the support and accepting the review invitation.

Best regards

Abdulkareem

[Quoted text hidden]

DIDING SUHANDY <diding.sughandy@fp.unila.ac.id> Wed, Feb 17, 2021 at 5:54 PM
To: Jestec <Jestec@taylors.edu.my>

Dear Jestec Editorial Office,

I am very happy to send the review result of this paper. I have corrected some technical and language typos directly in this original submitted paper. I also put my comments directly in the paper. I believe this paper could be publishable with some major modification.

Thank you,

Best regards,

DS

On Sat, Jan 30, 2021 at 9:09 PM Jestec <Jestec@taylors.edu.my> wrote:
[Quoted text hidden]

2 attachments

Review Report - Paper ID CE20029.docx
43K

CE20029 (Review Result) Paper ID CE20029.docx
421K

Jestec <Jestec@taylors.edu.my> Wed, Feb 17, 2021 at 6:06 PM
To: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>

Dear Dr.

Thank you for your time in reviewing the said paper.

We highly appreciate your support and commitment.

Best regards

JESTEC Editor

http://jestec.taylors.edu.my

From: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>
Sent: Wednesday, February 17, 2021 6:54 PM
To: Jestec <Jestec@taylors.edu.my>
Subject: Re: FW: Paper ID CE20029 /Requesting paper review for JESTEC, First round of Review Process/

Dear Jestec Editorial Office,

[Quoted text hidden]
[Quoted text hidden]

Jestec <Jestec@taylors.edu.my> Sun, May 9, 2021 at 10:13 AM
To: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>

Dear Dr.

I would like to express on behalf of the Review panel our sincere thanks for your effort shown in reviewing this paper. We highly appreciate this effort and support and hope that we may call upon you again to review future manuscripts.

Kindly accept the attached appreciation letter.

Best regards

Assoc. Prof. Dr. Abdulkareem Sh. Mahdi Al-Obaidi, CEng MIMechE

Executive Editor, Journal of Engineering Science & Technology

http://jestec.taylors.edu.my

From: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>
Sent: Wednesday, February 17, 2021 6:54 PM
To: Jestec <Jestec@taylors.edu.my>
Subject: Re: FW: Paper ID CE20029 /Requesting paper review for JESTEC, First round of Review Process/

Dear Jestec Editorial Office,

[Quoted text hidden]
[Quoted text hidden]

AL_Diding Suhandy_1.pdf
55K

Jestec <Jestec@taylors.edu.my> Sun, May 9, 2021 at 10:13 AM
To: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>

Dear Dr.

Your Reviewer number is: 3

The paper you earlier reviewed has been revised according to your comments/concern.

Could you kindly have a look at the revised paper and check whether the author(s) addressed all your comments/concern.

We appreciate receiving your feedback before or latest by **21/5/2021**

Attached for your reference, please find

- the original paper
- your review report
- the revised paper and
- the outlining how the author(s) addressed your and other reviewers' comments.

Thank you

Assoc. Prof. Dr. Abdulkareem Sh. Mahdi Al-Obaidi, CEng MIMechE

Executive Editor, Journal of Engineering Science & Technology

http://jestec.taylors.edu.my

From: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>
Sent: Wednesday, February 17, 2021 6:54 PM
To: Jestec <Jestec@taylors.edu.my>
Subject: Re: FW: Paper ID CE20029 /Requesting paper review for JESTEC, First round of Review Process/

Dear Jestec Editorial Office,

[Quoted text hidden]
[Quoted text hidden]

4 attachments

CE20029 R1.docx
360K

Outlining of Review Report.docx
54K

Review Report - 3 corrections.docx
421K

Review Report - 3.docx
43K

Jestec <Jestec@taylors.edu.my> Sun, May 23, 2021 at 7:06 AM
To: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>

Dear Dr.

On 9/5/2021 I sent the attached documents for your kind review.

Please be reminded that this is the second round.

I hope I receive the feedback from you before or latest by 28/5/2021.

Thank you for your commitment.

Regards

Associate Professor Dr. Abdulkareem Sh. Mahdi Al-Obaidi, CEng MIMechE

Executive Editor, Journal of Engineering Science & Technology

http://jestec.taylors.edu.my

[Quoted text hidden]

4 attachments

Outlining of Review Report.docx
54K

Review Report - 3 corrections.docx
421K

Review Report - 3.docx
43K

CE20029 R1.docx
360K

DIDING SUHANDY <diding.sughandy@fp.unila.ac.id> Sun, May 23, 2021 at 9:36 AM
To: Jestec <Jestec@taylors.edu.my>

Dear Jestec Editorial Office,

Thank you for sending me the remainder. I apologize for the delay. I forgot to ask for a little bit of an extension due to my health situation. However, today I have finished it. All my comments have been well addressed by the authors. I am satisfied with the revised paper and it can be published in this current form.

Thank you

Best regards,

Diding Suhandy
University of Lampung

[Quoted text hidden]

Jestec <Jestec@taylors.edu.my> Sun, May 23, 2021 at 9:07 PM
To: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>

Dear Dr.

Thank you for your kind reply.

I did not know about your health issue.

I wish you speedy recovery.

Best regards

JESTEC Editor

[Quoted text hidden]

DIDING SUHANDY <diding.sughandy@fp.unila.ac.id> Mon, May 24, 2021 at 7:14 PM
To: Jestec <Jestec@taylors.edu.my>

Dear Jestec Editorial Office,

Thank you for your email. I am in better condition now.
Please if you provide a certificate of reviewer, could you please send it to me?

Thank you,

Best regards,

Diding Suhandy
University of Lampung

[Quoted text hidden]

Jestec <Jestec@taylors.edu.my> Mon, May 24, 2021 at 11:03 PM
To: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>

here you go

Best regards

JESTEC Editor

http://jestec.taylors.edu.my

From: DIDING SUHANDY <diding.sughandy@fp.unila.ac.id>
Sent: Sunday, May 23, 2021 10:37 AM

[Quoted text hidden]

[Quoted text hidden]

AL_Diding Suhandy_1.pdf
55K

DIDING SUHANDY <diding.sughandy@fp.unila.ac.id> Sat, May 29, 2021 at 9:44 AM
To: Jestec <Jestec@taylors.edu.my>

Well received with thanks.
[Quoted text hidden]

REVIEW FORM

Title of paper: THE PREDICTION OF KOVATS RETENTION INDICES OF ESSENTIAL OILS AT GAS CHROMATOGRAPHY USING GENETIC ALGORITHM

For sections A & B, please tick a number from 0 to 5, where 0 = strongly disagree and 5 = strongly agree.

A. Technical aspects

1. The paper is within the scope of the Journal.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5
2. The paper is original.	☐ 0	☐ 1	☐ 2	☐ 3	☑ 4	☐ 5
3. The paper is free of technical errors.	☐ 0	☐ 1	☑ 2	☐ 3	☐ 4	☐ 5

B. Communications aspects

1. The paper is clearly readable.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5
2. The figures are clear & do clearly convey the intended message.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5
3. The length of the paper is appropriate.	☐ 0	☐ 1	☐ 2	☑ 3	☐ 4	☐ 5

C. Comments to the authors (You may use another sheet of paper.)

Please see the comments and corrections in the original submitted paper.

D. Recommendation (Tick one)

1. Accepted without modifications.	☐
2. Accepted with minor corrections.	☐
3. Accepted with major modification.	☑
4. Rejected.	☐

E. Comments to the editors (These comments will not be sent to the authors)

Dear Editor,
I am very happy to send the review result of this paper. I have corrected some technical and language typos directly in this original submitted paper. I also put my comments directly in the paper. I believe this paper could be publishable with some major modification.

Thank you,

Best regards,

DS

Journal of Engineering Science and Technology
Vol. XX, No. Y (Year) PPP - QQQ

THE PREDICTION OF KOVATS RETENTION INDICES OF ESSENTIAL OILS AT GAS CHROMATOGRAPHY USING GENETIC ALGORITHM

Commented [DS1]: The title is a little bit misleading. The model prediction of KRI (Kovats retention index) was performed by MLR and SVR. Genetic algorithms were a feature selection used to select the input variables for KRI prediction. Please consider to change the title.

Abstract

The Kovats retention indices (KRI) of 340 essential oil compounds had been successfully predicted using molecular descriptor data through a genetic algorithm (GA) approach. The ~~genetic algorithm~~GA was used to select the optimal molecular descriptors. The selected molecular descriptors were then used to predict the K~~k~~ovats retention indices using multiple linear regression (MLR) and support vector regression (SVR) methods. The molecular descriptors of the essential oil compounds were calculated using Online Chemical Database (OCHEM) software, where it generated 184 molecular descriptors for 340 compounds in the essential oils. The GA was run for 10 times, with 1000 iteration each, suggesting 5 most optimal molecular descriptors to be used in constructing Kovats retention index prediction model. As the results, MLR had R^2 training = 0.970, R^2 testing = 0.970, RMSE training = 56.55, and RMSE testing = 56.99. Meanwhile SVR model produced R^2 training = 0.981, R^2 testing = 0.973, RMSE training = 44.62 and RMSE testing = 53.60. In Kovats retention index prediction test, the MLR model generated the average difference of the predicted values as high as 3.8% for the training set and 3.4% for the testing set. Meanwhile, SVR yielded 2.4% difference for the training set and 3.4% for the testing set. The number of compounds with the difference of predicted value above 10% from MLR model was 13 compounds for the training set and 1 compound for the testing set, while in the SVR model, the number reached 7 compounds for the training set and 1 compound for the testing set. In comparison with MLR, the SVR model successfully gave higher R^2, lower RMSE, the lower average difference of the predicted value, and a smaller number of compounds with the difference above 10%. It suggests SVR model ability to predict the ~~k~~Kovats retention indices in higher accuracy, where the predicted ~~k~~Kovats retention indices are closed to the observed ones. In the conclusion of this research, the SVR is better than MLR method in predicting Kovats retention indices of the essential oil compounds.

Keywords: Genetic Algorithm, Molecular Descriptors, Multiple Linear Regression, Support Vector Regression, Kovats Retention Index, Essential Oils

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Commented [DS2]: Why the 10% difference used as a threshold for evaluating the performance of prediction model? It should be explained.

Formatted: Superscript

1. Introduction

Essential oils are known as natural products extracted from plants, attracting many interests due to their various biological properties [1]. The composition of essential oils has been extensively investigated owing to the increase in commercial demand [2]. Essential oils have been utilized in a wide range of fields, such as in the production of antifungal [3], antibiofilm [4], algae controlling agent for water ecosystem [5], and fungal growth inhibitor in food products [6].

There are many types of essential oils contained within plants. Moreover, the essential oil content between one plant to another is very diverse [7]. A hyphenated method, Chromatography-Mass Spectroscopy, is the most popular method used in the identification of organic components within plants [8,9], including the essential oil. Plant extract components are separated through a chromatography based on their respective polarity [10], followed by the identification of the separated chemical components using Mass Spectrometry [11].

In the component analysis using gas chromatography (GC), retention time (a chromatography parameter) does not represent any information of the psycho-chemistry or thermodynamic of the component. It is ascribed to the inaccuracy of the retention time against the changes in analytical conditions. In other words, the retention time of a component can be changed throughout the changes in the analytical condition [12]. To obtained more useful information, the retention time is converted to a retention index. The retention index ~~has a correlation~~correlates with the carbon chain of an organic compound, thus, it can be used to predict the relative carbon number and polarity of the analyzed component [13,14].

Commented [DS3]: In this part, please introduce the Kovats retention index (KRI) in detail along with its definition.

In general, a retention index is calculated by determining the dead time at the same analytical condition used to analyze the retention time. The dead time value can be determined using the inert gas or homologous series method [15]. Furthermore, the retention time and dead time values can be employed in calculating the retention index with the help of an algorithm. In this method, the accuracy of determining the dead time is very crucial [16,17], because it determines the accuracy in a retention index calculation.

The correlation between a carbon number and a retention index indicates the correlation between the chemical structure and the retention index of a compound. Multivariate analysis also statistically reveals the pattern similarity among the retention indices of the compounds with similar chemical structures [18,19]. On that basis, we can calculate retention indices without firstly acquiring the dead time value, by utilizing the quantitative information of the chemical structure of a compound along with a genetic algorithm (GA).

GA is initially proposed and developed by John Holland in 1960-1970s [20]. It is a part of the stochastic methods used to solve the optimization problem defined by fitness criteria. It applies Darwin's evolution hypothesis and several genetic functions such as mutation and crossover [21]. Unlike the conventional optimization technique, which only relies on a single point-based searching; GA does the searching through a population of a solution. Therefore, it allows GA to have the probability of reaching the global optimum and helps to avoid the local stationary point [22]. Furthermore, the advantage of GA is an approach to solve phenomena by adopting the uncertainty principle as a stochastic concept.

GA has been used in many applications, including in solving the optimization of the problem variation in engineering and sciences fields [23]. It also has been

used in other applications, such as arranging the research consultation schedule [24], predicting a dengue outbreak [25], classifying diabetic diseases [26], and classifying big data [27]. The previous studies had conducted the prediction of Kovats retention indices on flavor and fragrance compounds using GA and multiple linear regression, yielding a pretty good prediction [28]. Parveen et al. also had conducted a comparative study on MLR, support vector regression (SVR), and artificial neural network (ANN) models to predict the heavy metal sorption, where the SVR model appeared to be more superior than MLR and ANN [29].

Nowadays, the chemical structure information of a compound has been widely described and quantified in several parameters, known as molecular descriptors. Several descriptors, such as BCUTc(-11) (eigenvalue), AMR (atom addition logP and molar refractivity), MOMI(-R) (moments of inertia and ratios of the principal moments), etc., have been extensively used in computational chemistry. Molecular descriptor calculation can be conducted with easy-to-operate and cost-free Online Chemical Database software [30].

In this research, molecular descriptors of the essential oil compounds are calculated with Online Chemical Database. Afterward, the descriptor selection is conducted through GA to obtain the optimal molecular descriptors. The selected molecular descriptors will further be used to develop a prediction model for Kovats retention indices using MLR and SVR methods. The results obtained from the MLR and SVR will be compared to determine which model yields the best results.

Commented [DS4]: The authors must describe in detail the reason why MLR ad SVR was selected? The comparison of MLR and SVR is linear vs nonlinear approach?

2. Materials and Method

2.1. Integrated Development Environment (IDE)

The source-code editor used for this research was Visual Studio Code. It was used to build the GA by using the Perl programming language. Descriptor calculation was conducted by using Online Chemical Database (OCHEM). RStudio as IDE for R was also used to build MLR and SVR models.

2.2. Dataset Collection

The dataset used for this research was that of essential oil compounds obtained from Babushok et al. [31]. There are 340 compounds within the dataset, where 90% of the dataset were used as the training set and the other 10% were used as the testing set. It was conducted as such to ensure the fitness of the built model. The data distribution of the compounds within the training set and testing set can be seen in Fig. 1.

Commented [DS5]: Please explain in detail how the authors divided the samples into two sets?

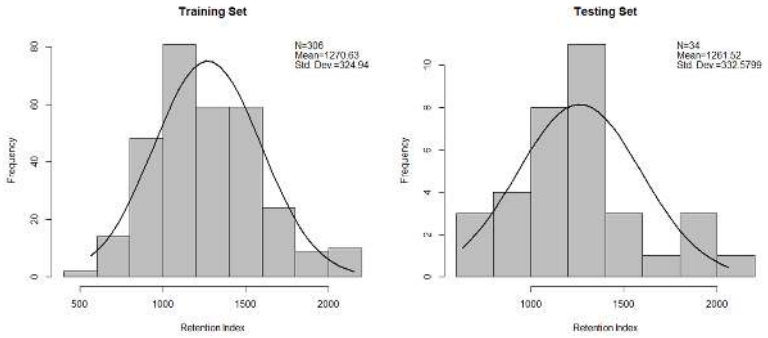


Fig. 1. The distribution of the compounds within the training set and testing set.

2.3. Descriptor Calculation

Molecular descriptors of the essential oil compounds were calculated with calculating descriptors feature using OCHEM. Several examples of the calculated descriptor class are topological, geometrical, constitutional, and hybrid descriptors. From the calculation, as many as 184 molecular descriptors were obtained, which was employed later.

2.4. Descriptor Selection

In constructing the model, molecular descriptors of the essential oil compounds were used as an independent variable. In practice, the obtained number of molecular descriptors is overwhelmingly high, therefore, a method to select optimal molecular descriptors is required. This step was meant to ease the interpretation, prediction ability, and speed up the model construction [30].

GA method was used for the selection of the most optimal molecular descriptor. There are five major steps in GA, including the initialization of the initial population, calculation of the fitness value for each individual within the population, selection of the individual as parent candidates, crossover to produce offspring, and mutation. Fig. 2 presents the pseudocode of GA.

1. initial population $P \leftarrow$ GenerateIndividuals()
2. generation $G \leftarrow 0$
3. fitness $f(S) \leftarrow$ EvaluateIndividual($S \in P$)
4. best individual $S^* \leftarrow$ BestIndividual($S \in P$) and $f(S^*) \leftarrow f$(BestIndividual($S \in P$))
5. **while** termination criteria is not TRUE **do**
6. parents $P \leftarrow$ Selection(P)
7. offsprings $Q \leftarrow$ Recombination(P)
8. $Q \leftarrow$ Mutation(Q)
9. EvaluatePopulation(Q)
10. $P \leftarrow$ Selection($Q \cup P$)
11. **if** $f(S^*) \geq f$(BestIndividual($S \in Q$)) **then** %update best individual %
    $S^* \leftarrow$ BestIndividual($S \in Q$)
    $f(S^*) \leftarrow f$(BestIndividual($S \in Q$))
    endif
    $K \leftarrow K + 1$
    end while
12. **Return** S^*

Fig.2. Pseudocode of GA [23].

2.5. Constructing Prediction Model

The results of molecular descriptor selection using GA were then used to construct the prediction model of the Kovats retention indices for the essential oils. The regression model was build using MLR and SVR methods.

The linear equation used to calculate the linear correlation between a Kovats retention index and a molecular descriptor in the MLR method is expressed below:

$$RI_{mlr} = c_o + \sum_{i=1}^{n} c_i D_i \quad (1)$$

Where c_o represents intercept, c_i represents the regression coefficient of the molecular descriptor (D_i), and n represents the number of the selected molecular descriptor [32]. For the SVR method, the regression equation is presented as follow:

$$f(x) = w \bullet \phi(x) + b \quad (2)$$

Where w and b respectively represent slope and offset of the regression line, x is high dimensional input space, and ϕ is the kernel function that can map the input space x to higher-dimensional space. The function of f(x) can be calculated by minimizing the following equation:

$$\frac{1}{2} w^T w + \frac{1}{n} \sum_{i=1}^{n} c(f(x_i), y_i) \quad (3)$$

Where 1/2 w^T w is a term characterizing the model complexity, c(f(x_i),y_i) is a loss function, y is the target and n is the number of samples [33,34].

2.6. Result Analysis

Commented [DS6]: Rewrite in correct form.

Commented [DS7]: Rewrite in correct form.

Commented [DS8]: Rewrite in correct form.

Commented [DS9]: What kind of kernel function used in this study? More description is needed.

Commented [DS10]: Rewrite in correct form.

Commented [DS11]: Rewrite in correct form.

Commented [DS12]: Rewrite in correct form.

The coefficient of determination (R^2) and RMSE obtained was used to judge the result of the prediction model of the Kovats retention indices. R^2 indicates the value of the independent variable combination collectively affecting the value of the dependent variable, meanwhile, RMSE is a method used to evaluate the accuracy of the results yielded by a prediction model. The better model can be judged based on the higher R^2 and lower RMSE. R^2 and RMSE are defined in the following equations:

Commented [DS13]: It is better to describe values for acceptable R^2 or RMSE.

$$R^2 = \frac{\sum_{i=1}^{n}(y_i^{fit} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)}{n}} \quad (5)$$

Where y_i^{fit} is the fitted value, $\bar{y}$ is the average of observed values, y_i is the observed value, $\hat{y}_i$ is the predicted value and n is the number of data [34].

3. Results and Discussions

In this research, GA was built to select the five most optimal molecular descriptors used as independent variables in constructing the prediction model of the Kovats retention indices. GA was run 10 times with 1000 iterations, respectively. GA was constructed based on the flowchart in Fig. 3.

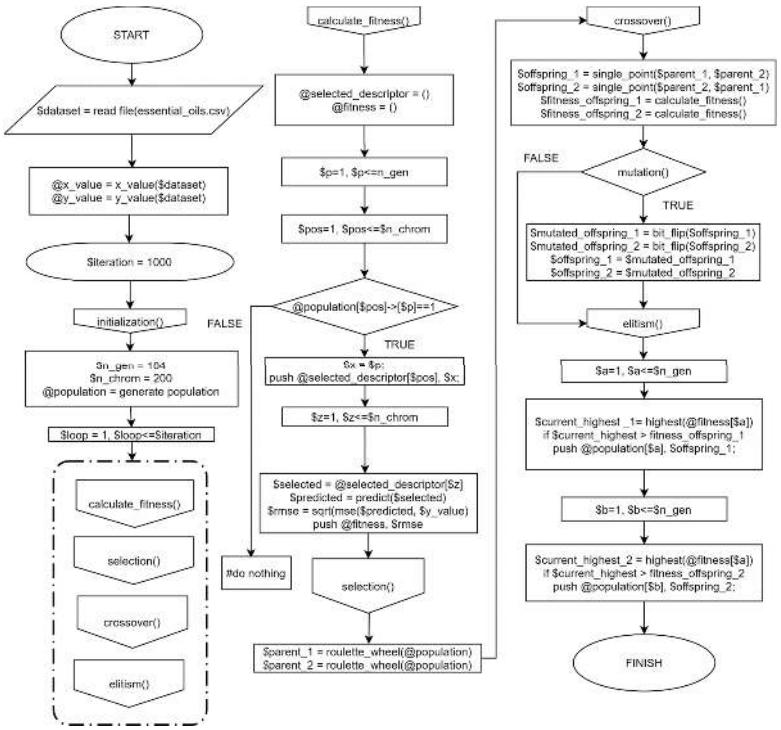


Fig. 3. GA flowchart to predict the Kovats retention indices.

The first step taken was reading the dataset of the essential oils and determining the values of x and y variables. The next step included the determination of the iteration number. Afterwards, GA started the initialization() process which aimed to call the initial population with the gene number as many as $n_gen and chromosome number as many as $n_chrom.

The process was then followed by the calculate_fitness(), selection(), crossover(), and mutation() as many as $iteration. In calculate_fitness(), the program checked the value of each gene in each chromosome within the initial population. Molecular A molecular descriptor would be used if the gene had the value of 1, and would not be used if the value was 0. The RMSE values of each chromosome within the population were calculated and stored in @fitness.

Next, the parent selection was conducted using the roulette wheel selection method. The selected parents were used in the crossover(), by crossing both parents using the single-point crossover method, resulting in offsprings. RMSE of the offsprings were was then calculated. There are possibilities where the offsprings experienced a mutation, in such cases, the bit flip mutation method was used. Then, the elitism() process was run to replace the chromosome of the initial population with higher RMSE through a comparison with the RMSE of the offsprings. The program would stop if the iteration number reached $iteration—the obtained result after running GA as many as 10 times can be seen in Table 1.

Table 1. The results of GA.

Test	Selected Descriptors	R^2	RMSE
1	ATSc2, VCH-6, SP-2, TPSA, WTPT-1	0.968	57.59
2	ATSc2, VCH-5, SP-1, WNSA-3,TopoPSA	0.967	58.93
3	ATSc1,VPC-6,SP-1,tpsaEfficiency,nAtomLC	0.966	59.67
4	ATSc2,VPC-5,SP-1,RPSA,WTPT-1	0.964	61.58
5	ATSc1,VC-4,PPSA-1,WNSA-3,MW	0.958	66.46
6	ATSp1,C3SP3,SP-1,WNSA-3,MDEC-11	0.960	64.78
7	ATSc2,C2SP3,SP-1,WNSA-3,Weta1.unity	0.966	59.63
8	ATSc2,SPC-5,VP-0,RNCS,MW	0.966	60.04
9	ATSc1,VCH-7,SP-1,Kier1,MLogP	0.970	56.55
10	ATSc1,SCH-6,SP-1,Kier1,MLogP	0.969	57.10

Table 1. presents the molecular descriptor selected using GA with 10 times testing. The best result was obtained at the ninth test, where the RMSE was 56.55 and R^2 was 0.970. Molecular descriptors selected at the ninth test are ATSc1, VCH-7, SP-1, Kier1, and MLogP. Generally, the distribution portrayal for the R^2 and RMSE can be observed in the boxplot. The boxplot generated from the test results of the GA can be observed in Fig. 4.

Formatted: Superscript

Formatted: Superscript

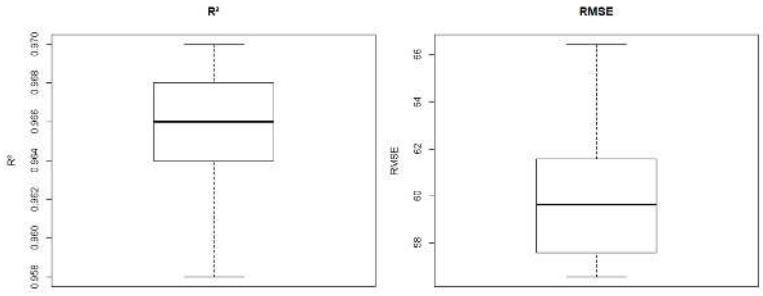


Fig. 4. The boxplot of the GA test results.

The higher the correlation coefficient of the prediction model, the smaller the errors in the built prediction model. The explanation of the selected molecular descriptors along with their correlation can be seen in Table 2. In this table, the correlation refers to Pearson's correlation, where SP-1 shows a high correlation against the retention indices (0.971), followed by MW (0.961), WTPT-1 (0.956), and VP-0 (0.924).

Table 2. ~~Explanation~~ The explanation for the selected molecular descriptors along with their correlations.

No.	Name	Definition	Correlation
1	SP-1	Evaluates the Kier & Hall Chi path indices of orders 0,1,2,3,4,5,6 and 7	0.971
2	MW	Descriptor based on the weight of atoms of a certain element type. If no element is specified, the returned value is the Molecular Weight	0.961
3	WTPT-1	The weighted path (molecular ID) descriptors described by Randic. They characterize molecular branching.	0.956
4	VP-0	Evaluates the Kier & Hall Chi path indices of orders 0,1,2,3,4,5,6 and 7	0.924
5	Kier1	Descriptor that calculates Kier and Hall kappa molecular shape indices.	0.896
6	MLogP	Moriguchi octanol-water partition coefficient	0.856
7	SP-2	Evaluates the Kier & Hall Chi path indices of orders 0,1,2,3,4,5,6 and 7	0.815
8	ATSp1	The Moreau-Broto autocorrelation descriptors using polarizability	0.758
9	PPSA-1	A variety of descriptors combining surface area and partial charge information	0.646
10	C2SP3	Characterizes the carbon connectivity in terms of hybridization	0.643
11	nAtomLC	Returns the number of atoms in the largest chain	0.468

No.	Name	Definition	Correlation
12	Weta1.unity	Holistic descriptors described by Todeschini et al .	0.258
13	SPC-5	Evaluates the Kier & Hall Chi path cluster indices of orders 4,5 and 6	0.197
14	MDEC-11	Evaluate molecular distance edge descriptors for C, N, and O	0.190
15	VPC-6	Evaluates the Kier & Hall Chi path cluster indices of orders 4,5 and 6	0.184
16	VPC-5	Evaluates the Kier & Hall Chi path cluster indices of orders 4,5 and 6	0.175
17	C3SP3	Characterizes the carbon connectivity in terms of hybridization	0.145
18	VC-4	Evaluates the Kier & Hall Chi cluster indices of orders 3,4,5,6 and 7	0.107
19	ATSc1	The Moreau-Broto autocorrelation descriptors using partial charges	0.082
20	VCH-7	Evaluates the Kier & Hall Chi chain indices of orders 3,4,5 and 6	0.030
21	VCH-6	Evaluates the Kier & Hall Chi chain indices of orders 3,4,5 and 6	0.018
22	TopoPSA	Calculation of topological polar surface area based on fragment contributions .	0.016
23	SCH-6	Evaluates the Kier & Hall Chi chain indices of orders 3,4,5 and 6	0.006
24	VCH-5	Evaluates the Kier & Hall Chi chain indices of orders 3,4,5 and 6	0.003
25	TPSA	Calculation of topological polar surface area based on fragment contributions	-0.071
26	ATSc2	The Moreau-Broto autocorrelation descriptors using partial charges	-0.147
27	WNSA-3	A variety of descriptors combining surface area and partial charge information	-0.256
28	RPSA	A variety of descriptors combining surface area and partial charge information	-0.267
29	tpsaEfficiency	Topological polar surface area efficiency	-0.422
30	RNCS	A variety of descriptors combining surface area and partial charge information	-0.481

The molecular descriptor, selected at the ninth test, was used to construct the prediction model of the Kovats retention indices using MLR and SVR. The obtain parameters from MLR are compared with the SVR model to identify which model gives the best result.

Table 3. The test result of the MLR and SVR models.

Model	Training set		Testing set	
	R^2	RMSE	R^2	RMSE
MLR	0.970	56.55	0.970	56.99
SVR	0.981	44.62	0.973	53.60

In Tabel 3., the comparison between MLR and SVR models can be observed. MLR model generates R^2 training = 0.970, R^2 testing = 0.970, RMSE training = 56.55 and RMSE testing = 56.99. Meanwhile SVR yields R^2 training = 0.981, R^2 testing = 0.973, RMSE training = 44.62 and RMSE testing = 53.60. The results suggest that SVR has higher R^2 and lower RMSE in comparison with the MLR model. The prediction plots of Kovats retention indices for MLR and SVR models are presented in Fig. 5.

Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript
Formatted: Superscript

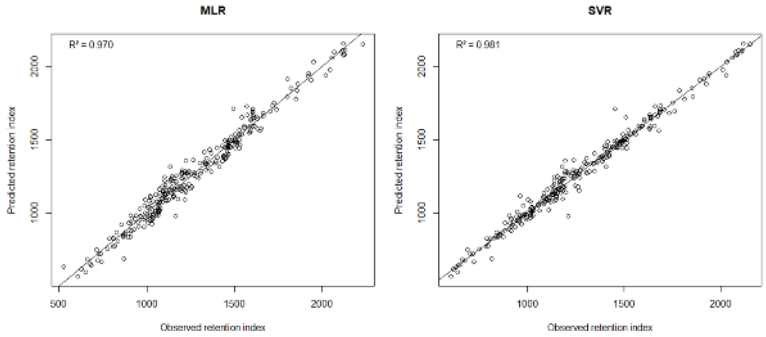
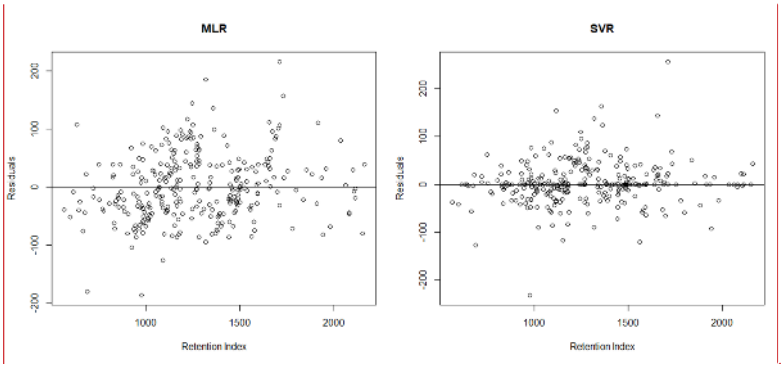


Fig. 5. Prediction plots of Kovats retention indices for MLR and SVR models.

In Fig. 5, it can be observed the Kovats retention index prediction of the essential oil compounds using MLR and SVR models. In the plot using MLR, the R^2 obtained is 0.970; while higher R^2 is given by SVR (0.982). In the next figure, the residual (the difference between the predicted value and the observed value of the Kovats retention indices obtained from the dataset) is presented (Fig. 6).

Formatted: Superscript
Formatted: Superscript



Commented [DS14]: Is Figure 6 and Table 4 generated from same data (the difference between the predicted value and the observed value of the Kovats retention indices obtained from the dataset)? If so, why you need to show both similar data in different way?

Fig. 6. Residual plots of Kovats retention indices from MLR and SVR models.

The comparison between the observed and predicted Kovats retention indices obtained from the MLR and SVR models on the training set could be in Table 4.

Table 4. The comparison of predicted and observed values of the MLR and SVR models for the training set.

No.	Compounds	Observed	Predicted		Difference(%)	
			MLR	SVR	MLR	SVR
1	Abietatriene	2033	1953	2033	4.1	0.0
2	Acetic acid	633	526	633	20.4	0.0
3	Acetoin	684	662	665	3.3	2.9
4	Acetophenone	1042	1015	1024	2.7	1.8
5	2-Acetylfuran	884	907	884	2.5	0.0
6	Alloaromadendrene	1459	1460	1402	0.1	4.1
7	allo-Ocimene	1116	1020	962	9.4	16.0
8	Anethole, (E)-	1265	1201	1265	5.4	0.0
9	p-Anisyl alcohol	1250	1143	1155	9.4	8.3
10	Ar-Curcumene	1471	1515	1480	2.9	0.6
11	Aromadendrene	1439	1460	1402	1.4	2.7
12	Artemisia alcohol	1072	1074	1072	0.2	0.0
13	Artemisia ketone	1048	1027	1058	2.0	1.0
14	Benzaldehyde	937	949	928	1.2	1.0
15	Benzeneacetaldehyde	1016	1045	1063	2.8	4.4
16	Benzyl benzoate	1734	1719	1691	0.9	2.6
17	Benzyl salicylate	1837	1859	1786	1.2	2.8
18	Bicycloelemene	1336	1367	1342	2.3	0.4
19	Bicyclogermacrene	1490	1473	1450	1.2	2.8
20	*β*-Bisabolene	1500	1505	1488	0.4	0.8
21	*α*-Bisabolol	1668	1605	1634	3.9	2.1
22	*β*-Bisabolol	1659	1610	1638	3.1	1.3
23	Borneol	1153	1090	1148	5.7	0.4
24	Bornyl acetate	1270	1275	1270	0.4	0.0
25	*β*-Bourbonene	1382	1444	1382	4.3	0.0
26	*α*-Bulnesene	1501	1527	1513	1.7	0.8
27	Bulnesol	1653	1617	1635	2.2	1.1
28	Butan-1-ol, 2-methyl-	722	738	718	2.2	0.6
29	Butanal, 2-methyl-	643	683	643	5.8	0.0
30	2,3-Butanedione	566	605	604	6.5	6.2
31	Butanoic acid, 2-methyl-	828	808	828	2.5	0.0
32	1-Butanol	652	677	655	3.8	0.4
33	2-Buten-1-ol, 3-methyl-	751	712	689	5.5	9.1
34	Cadalene	1655	1543	1512	7.2	9.5
35	*α*-Cadinene	1527	1529	1515	0.1	0.8
36	*α*-Cadinol	1640	1629	1640	0.7	0.0
37	Camphene	947	1013	955	6.6	0.8
38	Camphene hydrate	1136	1104	1136	2.9	0.0
39	*α*-Campholenal	1107	1098	1115	0.9	0.8

No.	Compounds	Observed	Predicted		Difference(%)	
			MLR	SVR	MLR	SVR
40	Camphor	1125	1050	1127	7.2	0.2
41	3-Carene	1007	1018	948	1.1	6.2
42	Carotol	1593	1604	1593	0.7	0.0
43	Carvacrol	1283	1218	1197	5.3	7.2
44	Carvacrol acetate	1354	1394	1360	2.8	0.4
45	Carvone	1218	1134	1179	7.4	3.3
46	Caryophyllenyl alcohol	1560	1591	1561	2.0	0.1
47	α-Cedrene	1411	1458	1411	3.2	0.0
48	Cedrol	1597	1558	1576	2.5	1.3
49	Chamazulene	1710	1494	1453	14.5	17.7
50	β-Chamigrene	1470	1487	1472	1.1	0.1
51	Chavicol	1237	1198	1186	3.2	4.3
52	Chrysanthenone	1104	1087	1134	1.5	2.6
53	1,8-Cineole	1022	1073	1112	4.8	8.1
54	Cinnamaldehyde, cis-	1178	1145	1194	2.9	1.3
55	Cinnamaldehyde, trans-	1239	1145	1194	8.2	3.8
56	Citronellal	1134	1111	1145	2.1	0.9
57	Citronellol	1212	1168	1167	3.8	3.8
58	Citronellyl acetate	1336	1338	1306	0.1	2.3
59	α-Copaene	1376	1450	1394	5.1	1.3
60	p-Cresol	1052	989	977	6.4	7.6
61	p-Cresol, 2-methoxy-	1163	1127	1152	3.2	0.9
62	α-Cubebene	1352	1433	1387	5.7	2.5
63	β-Cubebene	1384	1435	1384	3.6	0.0
64	Cubebol	1505	1499	1505	0.4	0.0
65	β-Curcumene	1503	1503	1493	0.0	0.7
66	β-Cyclocitral	1196	1105	1140	8.2	4.9
67	p-Cymen-7-ol	1270	1226	1259	3.6	0.9
68	p-Cymen-8-ol	1165	1154	1161	1.0	0.4
69	m-Cymene	1012	1070	1025	5.4	1.3
70	o-Cymene	1032	1077	1032	4.2	0.0
71	p-Cymene	1015	1070	1025	5.1	1.0
72	Cyperene	1399	1467	1424	4.6	1.7
73	2,4-Decadienal, (2E,4E)-	1291	1204	1238	7.3	4.3
74	2,4-Decadienal, (2E,4Z)-	1273	1204	1238	5.8	2.8
75	Decanal	1186	1212	1244	2.1	4.7
76	1-Decanol	1259	1268	1260	0.7	0.0
77	2-Decenal, (E)-	1239	1205	1239	2.9	0.0
78	Decyl acetate	1392	1437	1392	3.2	0.0
79	Dendrolasin	1561	1647	1682	5.2	7.2
80	Dihydrocarveol	1182	1185	1182	0.3	0.0
81	Dill apiole	1596	1589	1596	0.4	0.0
82	1,4-Dimethoxybenzene	1138	1073	1084	6.0	5.0
83	2,5-Dimethoxy-p-cymene	1407	1369	1346	2.7	4.5
84	Dimethyl trisulfide	948	1034	948	8.3	0.0

No.	Compounds	Observed	Predicted		Difference(%)	
			MLR	SVR	MLR	SVR
85	Dodecanoic acid	1564	1529	1523	2.3	2.7
86	1-Dodecanol	1460	1465	1460	0.4	0.0
87	2-Dodecenal, (E)-	1444	1402	1443	3.0	0.0
88	*β*-Elemene	1388	1437	1422	3.4	2.4
89	Elemicin	1521	1479	1521	2.8	0.0
90	Elemol	1536	1527	1560	0.6	1.6
91	4,5-Epoxy-2-decenal, (E)-	1362	1263	1238	7.8	10.0
92	Ethyl acetate	598	650	639	7.9	6.5
93	Ethyl benzoate	1151	1199	1182	4.0	2.6
94	Ethyl decanoate	1380	1456	1409	5.2	2.0
95	Ethyl dodecanoate	1578	1653	1632	4.5	3.3
96	Ethyl hexadecanoate	1978	2047	2012	3.4	1.7
97	Ethyl hexanoate	983	1062	1014	7.4	3.1
98	Ethyl isovalerate	836	907	876	7.8	4.6
99	Ethyl linoleate	2151	2231	2151	3.6	0.0
100	Ethyl octanoate	1181	1259	1185	6.2	0.3
101	Ethyl tetradecanoate	1778	1850	1812	3.9	1.9
102	2-Ethylfuran	689	869	815	20.7	15.5
103	*α*-Eudesmol	1641	1593	1606	3.0	2.2
104	*β*-Eudesmol	1634	1595	1604	2.4	1.9
105	Eugenol	1340	1333	1329	0.5	0.9
106	Eugenol acetate	1485	1516	1486	2.1	0.1
107	*β*-Farnesene, (E)-	1449	1464	1466	1.0	1.1
108	*α*-Farnesene, (E,E)-	1496	1466	1463	2.1	2.3
109	*α*-Farnesene, (Z,E)-	1481	1466	1463	1.0	1.2
110	*β*-Farnesene, cis-	1444	1464	1466	1.3	1.5
111	Farnesol, (2Z,6E)-	1705	1604	1687	6.3	1.1
112	Farnesol, (2E,6E)	1710	1604	1687	6.6	1.4
113	Farnesol, (2Z,6Z)-	1687	1604	1687	5.2	0.0
114	Farnesol, (2E,6Z)-	1691	1604	1687	5.4	0.2
115	Farnesyl acetone, (5E,9E)	1914	1804	1914	6.1	0.0
116	Fenchone	1073	1061	1121	1.1	4.2
117	Furaneol	1030	960	1030	7.2	0.0
118	Furfural	807	836	798	3.4	1.1
119	Furfural, 5-methyl-	933	898	883	3.9	5.6
120	Furfuryl alcohol	832	895	838	7.0	0.7
121	Geranial	1247	1103	1138	13.1	9.6
122	Geraniol	1239	1156	1167	7.2	6.2
123	Geranyl butanoate	1537	1543	1536	0.4	0.1
124	Geranyl formate	1283	1243	1205	3.2	6.5
125	Geranyl isobutanoate	1491	1496	1500	0.4	0.6
126	Geranyl isovalerate	1588	1585	1588	0.2	0.0
127	Geranyl propanoate	1449	1444	1429	0.3	1.4
128	Geranylacetone	1431	1360	1431	5.2	0.0
129	Germacrene A	1491	1511	1489	1.3	0.1
130	Germacrene B	1535	1505	1496	2.0	2.6

No.	Compounds	Observed	Predicted		Difference(%)	
			MLR	SVR	MLR	SVR
131	Germacrene D	1476	1518	1480	2.7	0.3
132	Gleenol	1574	1634	1632	3.7	3.6
133	o-Guaiacol	1064	1068	1085	0.4	1.9
134	Guaiacol, 4-ethyl-	1254	1237	1254	1.4	0.0
135	Guaiacol, p-vinyl-	1284	1239	1256	3.7	2.2
136	α-Guaiene	1442	1528	1514	5.6	4.8
137	Guaiol	1589	1618	1637	1.8	2.9
138	2-Heptadecanone	1883	1861	1927	1.2	2.3
139	2,4-Heptadienal, (2E,4E)-	983	908	907	8.3	8.4
140	Heptanal	881	916	915	3.8	3.7
141	Heptanoic acid	1077	1037	1070	3.9	0.7
142	2-Heptanol	886	929	905	4.6	2.1
143	1-Heptanol	956	973	956	1.7	0.0
144	2-Heptanone	868	876	868	1.0	0.0
145	5-Hepten-2-ol, 6-methyl-	975	968	944	0.8	3.3
146	2-Heptenal, (E)-	931	909	908	2.4	2.5
147	Hexadec-9-enoic acid, (Z)-	1935	1920	1935	0.8	0.0
148	Hexadecanal	1797	1803	1857	0.3	3.2
149	Hexadecanoic acid	1955	1923	1940	1.7	0.8
150	Hexanal	777	818	796	5.0	2.4
151	Hexanoic acid	985	938	973	5.0	1.3
152	1-Hexanol	855	874	856	2.2	0.1
153	1-Hexanol, 2-ethyl-	1015	1049	1031	3.2	1.5
154	2-Hexen-1-ol, (E)-	850	865	847	1.8	0.3
155	3-Hexen-1-ol, (E)-	837	866	848	3.4	1.3
156	3-Hexen-1-ol, (Z)-	842	866	848	2.8	0.7
157	2-Hexen-1-ol, acetate, (E)-	993	1037	986	4.2	0.7
158	2-Hexenal, (E)-	827	811	788	2.0	4.9
159	3-Hexenal, (Z)-	770	808	785	4.7	2.0
160	3-Hexenyl acetate, (Z)-	986	1037	986	4.9	0.0
161	3-Hexenyl benzoate, (Z)-	1550	1586	1550	2.2	0.0
162	3-Hexenyl butanoate, (Z)-	1166	1252	1177	6.9	0.9
163	Hexyl benzoate	1554	1593	1555	2.4	0.1
164	Hexyl butanoate	1177	1259	1185	6.5	0.7
165	α-Himachalene	1445	1505	1493	4.0	3.2
166	β-Himachalene	1501	1500	1501	0.0	0.0
167	Himachalol	1648	1605	1638	2.7	0.6
168	α-Humulene	1449	1472	1462	1.6	0.9
169	β-Humulene	1448	1476	1458	1.9	0.7
170	Indole	1273	1199	1206	6.1	5.6
171	Isoborneol	1148	1090	1148	5.3	0.0
172	Isobornyl acetate	1271	1275	1270	0.3	0.1

No.	Compounds	Observed	Predicted		Difference(%)	
			MLR	SVR	MLR	SVR
173	Isobutanol	616	625	616	1.4	0.0
174	Isoitalicene	1378	1403	1372	1.8	0.4
175	Isopentyl acetate	858	892	860	3.9	0.2
176	Isophorone	1092	990	1005	10.3	8.7
177	Isophytol	1939	2021	2031	4.1	4.5
178	Isovaleric acid	826	787	805	4.9	2.6
179	Lavandulol	1155	1127	1134	2.5	1.9
180	Ledol	1583	1557	1583	1.7	0.0
181	Limonen-4-ol	1158	1159	1162	0.0	0.4
182	Limonene	1024	1060	1012	3.4	1.2
183	Linalool	1086	1112	1100	2.3	1.3
184	Linalool acetate	1242	1296	1272	4.2	2.3
185	Linalool propanoate	1318	1413	1408	6.7	6.4
186	Linoleic acid	2105	2112	2108	0.3	0.1
187	Longifolene	1404	1466	1413	4.2	0.6
188	p-Menth-2-en-1-ol, cis-	1115	1156	1150	3.5	3.0
189	p-Menth-2-en-1-ol, trans-	1114	1156	1150	3.6	3.1
190	p-Mentha-1,5-dien-8-ol	1145	1149	1156	0.4	1.0
191	p-Mentha-2,8-dien-1-ol, cis-	1117	1151	1151	3.0	3.0
192	p-Mentha-2,8-dien-1-ol, trans-	1107	1151	1151	3.8	3.8
193	Menthofuran	1153	1219	1270	5.4	9.2
194	Menthol	1163	1189	1180	2.2	1.4
195	Menthone	1137	1140	1175	0.2	3.2
196	Menthyl acetate	1281	1367	1323	6.3	3.1
197	Methional	866	827	840	4.7	3.0
198	Methyl 3-phenylpropionate	1247	1285	1247	2.9	0.0
199	p-Methyl anisole	1002	991	1002	1.1	0.0
200	Methyl benzoate	1074	1091	1091	1.5	1.6
201	Methyl chavicol	1178	1198	1261	1.7	6.6
202	Methyl decanoate	1309	1347	1280	2.8	2.3
203	Methyl eugenol	1376	1340	1306	2.7	5.3
204	Methyl hexadecanoate	1909	1938	1892	1.5	0.9
205	Methyl hexanoate	907	953	920	4.8	1.4
206	Methyl linoleate	2079	2123	2079	2.1	0.0
207	Methyl octadecanoate	2112	2131	2091	0.9	1.0
208	Methyl octanoate	1110	1150	1081	3.5	2.7
209	Methyl oleate	2081	2128	2086	2.2	0.2
210	Methyl salicylate	1173	1228	1190	4.5	1.4
211	Methyl tetradecanoate	1709	1741	1709	1.8	0.0
212	3-Methyl-1-butanol	721	723	705	0.3	2.3
213	6-Methyl-5-hepten-2-one	964	914	917	5.4	5.2
214	p-Methylacetophenone	1161	1072	1104	8.3	5.1
215	2-Methylpropyl 3-	989	1052	995	6.0	0.6

No.	Compounds	Observed	Predicted		Difference(%)	
			MLR	SVR	MLR	SVR
	methylbutanoate					
216	Myrcene	983	1020	962	3.6	2.1
217	Myrcenol	1097	1178	1182	6.9	7.2
218	Myristicin	1494	1445	1494	3.4	0.0
219	Myrtenol	1182	1171	1182	1.0	0.0
220	Naphthalene	1165	1197	1165	2.7	0.0
221	Neral	1220	1103	1138	10.6	7.2
222	Nerol	1216	1156	1167	5.2	4.2
223	Nerol oxide	1140	1224	1200	6.9	5.0
224	Neryl acetate	1344	1327	1297	1.2	3.6
225	2,4-Nonadienal, (2E,4E)-	1187	1105	1132	7.4	4.8
226	2,6-Nonadienal, (2E,6Z)-	1126	1102	1129	2.2	0.3
227	Nonanal	1084	1113	1139	2.6	4.8
228	Nonanoic acid	1269	1234	1228	2.9	3.3
229	2-Nonanol	1090	1126	1101	3.2	1.0
230	1-Nonanol	1157	1170	1157	1.1	0.0
231	2-Nonanone	1073	1073	1091	0.0	1.6
232	2-Nonenal, (Z)-	1125	1106	1133	1.7	0.7
233	Nonyl acetate	1294	1339	1274	3.4	1.5
234	Nopinone	1107	1023	1107	8.2	0.0
235	Octadecanoic acid	2159	2120	2117	1.8	2.0
236	1-Octadecanol	2060	2056	2060	0.2	0.0
237	Octanal	982	1015	1029	3.2	4.6
238	1-Octanol	1057	1071	1055	1.3	0.2
239	3-Octanol	984	1043	1016	5.7	3.2
240	Octanol acetate	1194	1240	1167	3.7	2.3
241	3-Octanone	966	991	997	2.5	3.1
242	2-Octen-1-ol, (E)-	1054	1062	1053	0.8	0.1
243	1-Octen-3-ol	966	1037	1015	6.9	4.9
244	1-Octen-3-one	956	989	996	3.3	4.0
245	1-Octen-3-yl acetate	1091	1218	1148	10.4	5.0
246	2-Octenal (E)-	1036	1008	1023	2.8	1.2
247	Oleic acid	2113	2117	2113	0.2	0.0
248	Pentadecanal	1696	1704	1761	0.5	3.7
249	Pentadecanoic acid	1854	1825	1852	1.6	0.1
250	2-Pentadecanone	1681	1664	1734	1.0	3.1
251	Pentanal	675	719	678	6.1	0.4
252	1-Pentanol	754	776	753	2.8	0.2
253	1-Penten-3-ol	666	742	723	10.2	7.9
254	2-Pentylfuran	979	1166	1211	16.0	19.1
255	Perilla alcohol	1282	1219	1249	5.2	2.6
256	Perilla aldehyde	1252	1167	1222	7.3	2.4
257	α-Phellandrene	999	1066	1020	6.3	2.0
258	β-Phellandrene	1021	1069	1019	4.4	0.2
259	Phenol	957	930	917	2.9	4.3
260	Phenylacetonitrile	1098	1083	1098	1.4	0.0

No.	Compounds	Observed	Predicted		Difference(%)	
			MLR	SVR	MLR	SVR
261	Phenylethyl 3-methylbutanoate	1465	1532	1502	4.4	2.4
262	2-Phenylethyl alcohol	1088	1104	1112	1.4	2.2
263	Phytol	2099	2069	2077	1.4	1.1
264	α-Pinene	935	1008	961	7.2	2.7
265	β-Pinene	973	1008	963	3.5	1.0
266	α-Pinene oxide	1085	1013	1085	7.1	0.0
267	Pinocarvone	1140	1090	1131	4.6	0.8
268	Piperitenone	1317	1132	1180	16.3	11.6
269	Piperitone	1233	1137	1180	8.4	4.5
270	Pulegone	1223	1136	1177	7.6	3.9
271	Sabinene	968	1034	965	6.4	0.3
272	Safrole	1271	1305	1309	2.6	2.9
273	α-Santalene	1416	1327	1416	6.7	0.0
274	β-Santalene	1453	1477	1440	1.6	0.9
275	Santolina triene	903	983	937	8.1	3.6
276	α-Selinene	1490	1504	1485	0.9	0.3
277	β-Selinene	1481	1506	1481	1.6	0.0
278	α-Sinensal	1728	1572	1660	10.0	4.1
279	β-Sinensal	1670	1574	1663	6.1	0.4
280	Styrene	979	964	979	1.6	0.0
281	Terpinen-4-ol	1165	1162	1161	0.3	0.3
282	α-Terpinene	1011	1065	1019	5.1	0.8
283	Terpinolene	1079	1056	1017	2.2	6.1
284	α-Terpinyl acetate	1333	1337	1302	0.3	2.4
285	Tetradecanal	1595	1606	1659	0.7	3.9
286	Tetradecanoic acid	1753	1726	1753	1.6	0.0
287	1-Tetradecanol	1663	1662	1673	0.0	0.6
288	α-Thujene	926	1030	968	10.1	4.3
289	Thymol	1272	1218	1197	4.4	6.3
290	Thymol acetate	1343	1394	1360	3.6	1.3
291	Tricyclene	922	855	922	7.8	0.0
292	Tridecanoic acid	1659	1628	1641	1.9	1.1
293	2-Tridecanone	1479	1467	1518	0.8	2.6
294	Umbellulone	1152	1108	1148	3.9	0.3
295	Undecanal	1286	1310	1347	1.8	4.5
296	Undecanoic acid	1458	1431	1411	1.9	3.4
297	2-Undecanone	1276	1270	1306	0.4	2.3
298	2-Undecenal, (E)-	1341	1303	1341	2.9	0.0
299	Valencene	1483	1509	1491	1.7	0.5
300	Vanillin	1358	1223	1195	11.1	13.7
301	Veratrole	1113	1075	1089	3.5	2.2
302	Verbenene	946	1023	946	7.5	0.0
303	Verbenone	1184	1086	1129	9.0	4.8
304	Viridiflorene	1489	1460	1415	2.0	5.2
305	α-Ylangene	1370	1450	1394	5.5	1.7
306	α-Zingiberene	1483	1512	1483	1.9	0.0

Formatted Table

For the training set test using MLR, the difference was obtained being 3.8%, where there were 13 compounds having more than 10% difference. On the contrary, SVR gave an average difference of 2.4%, where only 7 compounds have more than 10% difference. The comparison of the observed and predicted Kovats retention indices obtained using MLR and SVR models on the testing set could be observed in Table 5.

Table 5. Comparison between the predicted and observed value of the MLR and SVR for the testing set.

No	Compounds	Observed	Predicted		Difference (%)	
			MLR	SVR	MLR	SVR
1	Abietadiene	2062	1949	2034	1.4	1.4
2	p-Anisaldehyde	1223	1088	1093	11.9	11.9
3	Benzyl acetate	1141	1175	1159	1.5	1.6
4	Benzyl alcohol	1015	1003	991	2.4	2.4
5	Butanoic acid	807	741	747	8.0	7.9
6	Carvotanacetone	1221	1138	1181	3.5	3.4
7	Citronellyl formate	1260	1252	1207	4.2	4.4
8	Cuparene	1505	1473	1438	4.5	4.6
9	Decanoic acid	1364	1332	1312	3.8	3.9
10	Dodecanal	1389	1409	1449	4.2	4.1
11	Ethyl butanoate	785	865	843	6.6	6.8
12	Ethyl pentanoate	883	963	934	5.3	5.4
13	Farnesyl acetate, (2E,6E)-	1818	1773	1769	2.8	2.8
14	Geranyl acetate	1361	1327	1297	4.9	5.0
15	1-Hexadecanol	1862	1859	1888	1.4	1.4
16	Hexahydrofarnesylacetone	1833	1825	1933	5.5	5.2
17	Hexyl 2-methyl butanoate	1224	1326	1265	3.0	3.2
18	Hexyl acetate	996	1043	995	0.1	0.1
19	Isopentyl isovalerate	1088	1150	1089	0.1	0.1
20	α-Longipinene	1351	1429	1387	2.6	2.6
21	Methyl cinnamate, trans-	1362	1283	1245	9.2	9.5
22	3-Methylbutanal	633	667	627	0.8	0.8
23	Myrtenal	1171	1121	1161	0.9	0.8
24	2-Nonenal, (E)-	1136	1106	1133	0.2	0.2
25	Octanoic acid	1175	1135	1151	2.1	2.0
26	Patchouli alcohol	1653	1600	1590	4.0	4.0
27	2-Pentenol, (Z)-	747	766	740	0.9	1.0
28	2-Phenylethyl acetate	1230	1274	1239	0.7	0.7
29	Salicylaldehyde	1020	1089	1116	8.9	8.7
30	Spathulenol	1566	1554	1587	1.4	1.3
31	α-Terpineol	1176	1150	1149	2.4	2.4
32	Tridecanal	1491	1507	1554	4.1	4.0
33	1-Undecanol	1358	1366	1360	0.2	0.2
34	Yomogi alcohol	988	1037	1025	3.6	3.6

For the testing set, both MLR and SVR models gave ~~the~~ an average difference of 3.4%, and only 1 compound was obtained with the difference being more than 10%. As can be seen from the testing results of the training set and testing set, in

Commented [DS15]: The average?

Commented [DS16]: Why the 10% difference used as a threshold for evaluating the performance of prediction model? It should be explained.

Commented [DS17]: Please carefully check the calculated difference values.

Formatted Table

Commented [DS18]: 5.8%?

Commented [DS19]: 12.4%?

Commented [DS20]: Why the 10% difference used as a threshold for evaluating the performance of prediction model? It should be explained.

comparison with MLR, the SVR model gives lower differences and a smaller number of compounds that have more than 10% difference.

Commented [DS21]: Why the 10% difference used as a threshold for evaluating the performance of prediction model? It should be explained.

4. Conclusion

This research has successfully constructed a GA ~~which~~ that can be used to select molecular descriptors as the independent variables for the Kovats retention index prediction model of the essential oil compounds. The descriptors, ATSc1, VCH-7, SP-1, Kier1, and MLogP, were selected as the optimal molecular descriptors. These molecular descriptors were later used for constructing MLR and SVR models.

From the test conducted, MLR model gave R^2 training = 0.970, R^2 testing = 0.970, RMSE training = 56.55, and RMSE testing = 56.99. Meanwhile SVR model produced R^2 training = 0.981, R^2 testing = 0.973, RMSE training = 44.62 and RMSE = 53.60. In Kovats retention index prediction, the MLR model generated average predicted values difference as high as 3.8% for the training set and 3.4% for the testing set. Meanwhile, SVR yielded 2.4% difference for the training set and 3.4% for the testing set. The number of compounds with the predicted value difference above 10% obtained from the MLR model was 13 compounds for the training set and 1 compound for the testing set. Meanwhile from the SVR model, the number reached 7 compounds for the training set and 1 compound for the testing set.

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

In comparison with the MLR model, the SVR model successfully had higher R^2, lower RMSE, lower average predicted value differences, and fewer compounds with the predicted value difference above 10%. It suggests the SVR model has higher accuracy in predicting the Kovats retention indices, where the predicted values are close to the results obtained from the observation. In conclusion, the SVR method is better compared to the MLR method in predicting Kovats retention indices for essential oils.

Commented [DS22]: Please don't repeat this result in conclusion. Rewrite conclusion. For example, SVR is superior than MLR. However, please consider that SVR is nonlinear based mathematical models which is often found difficult to understand and deploy in a practical setting. How did the authors consider about this?

Formatted: Superscript

References

Commented [DS23]: Please add a table for nomenclature and abbreviations used in this study.

[1] Baser, K.H.C.; and Buchbauer, G. (2015) *Handbook of essential oils: science, technology, and applications*. CRC press, .

[2] Noorizadeh, H.; Farmany, A.; and Noorizadeh, M. (2011) Application of GA-PLS and GA-KPLS calculations for the prediction of the retention indices of essential oils. *Quimica Nova*. 34 (8), 1398–1404.

[3] Jahani, M.; Pira, M.; and Aminifard, M.H. (2020) Antifungal effects of essential oils against Aspergillus niger in vitro and in vivo on pomegranate (Punica granatum) fruits. *Scientia Horticulturae*. 264 109188.

[4] Pratiwi, S.U.T.; Lagendijk, E.L.; Weert, S. de; Idroes, R.; Hertiani, T.; and Hondel, C. Van den (2015) Effect of Cinnamomum burmannii Nees ex Bl. and Massoia aromatica Becc. Essential Oils on Planktonic Growth and Biofilm formation of Pseudomonas aeruginosa and Staphylococcus aureus In Vitro. *, International Journal of Applied Research in Natural Products*. 8 (2), 1–13.

[5] Zerrifi, S.E.A.; Kasrati, A.; Redouane, E.M.; Tazart, Z.; Khalloufi, F. El,

Abbad, A.; et al. (2020) Essential oils from Moroccan plants as promising ecofriendly tools to control toxic cyanobacteria blooms. *Industrial Crops and Products*. 143 111922.

[6] Granata, G.; Stracquadanio, S.; Leonardi, M., Napoli, E.; Consoli, G.M.L.; Cafiso, V.; et al. (2018) Essential oils encapsulated in polymer-based nanocapsules as potential candidates for application in food preservation. *Food Chemistry*. 269 286–292.

[7] Zahi, M.R.; Liang, H.; Khan, A.; and Yuan, Q. (2014) Identification of Essential Oil Components in Chinese Endemic Plant Achnatherum inebrians. *Asian Journal of Research in Chemistry*. 7 (6), 576–579.

[8] Earlia, N., Rahmad, R., Amin, M., Prakoeswa, C., Khairan, K., and Idroes, R. (2019) The Potential Effect of Fatty Acids from Pliek U on Epidermal Fatty Acid Binding Protein: Chromatography and Bioinformatic Studies. *Sains Malaysiana*. 48 (5), 1019–1024.

[9] Helwani, Z.; Ramli, M., Saputra, E.; Bahruddin, B.; Yolanda, D.; Fatra, W.; et al. (2020) Impregnation of CaO from Eggshell Waste with Magnetite as a Solid Catalyst (Fe3O4/CaO) for Transesterification of Palm Oil Off-Grade. *Catalysts*. 10 (2), 164.

[10] Estevam, E.C.; Griffin, S.; Nasim, M.J.; Zieliński, D.; Aszyk, J., Osowicka, M.; et al. (2015) Inspired by Nature: The Use of Plant-derived Substrate/Enzyme Combinations to Generate Antimicrobial Activity in situ. *Natural Product Communications*. 10 (10), 1733–8.

[11] Earlia, N.; Suhendra, R.; Amin, M.; Prakoeswa, C.R.S.; and Idroes, R. (2019) GC/MS Analysis of Fatty Acids on Pliek U Oil and Its Pharmacological Study by Molecular Docking to Filaggrin as a Drug Candidate in Atopic Dermatitis Treatment. *The Scientific World Journal*. 2019.

[12] Hao, Z.; Xiao, B.; and Weng, N. (2008) Impact of column temperature and mobile phase components on selectivity of hydrophilic interaction chromatography (HILIC). *Journal of Separation Science*. 31 (9), 1449–1464.

[13] Idroes, R. (2005) Determination of Absolute Retention Index System in High Performance Liquid Chromatography (RP-HPLC). *Malaysian J. Anal. Sci*. 9 224–32.

[14] Idroes, R. (2009) Evaluasi Waktu Mati Kromatografi untuk Penentuan Indeks Retensi pada RP-HPLC Menggunakan Beberapa Deret Homolog. *Indonesian J. Pharmacy*. 20 133–40.

[15] Idroes, R.; Muslem; Saiful; Mahmudi; Idroes, G.M.; Suhendra, R.; et al. (2019) Dead Time Determination of 2-Alkanone Homologues Series using Methanol/Water Eluent in High Performance Liquid Chromatography

System by Indirect Method. *IOP; Science Material and Engineering,*.

[16] Idroes, R.; Husna, I.; Muslem; Mahmudi; Rusyana, A.; Helwani, Z.; et al. (2019) Analysis of temperature and column variation in gas chromatography to dead time of inert gas and n-alkane homologous series using randomized block design. in: IOP Conf. Ser. Earth Environ. Sci., IOP Publishing, p. 12020.

[17] Idroes, R.; Muslem; Mahmudi; Saiful; Idroes, G.M.; Suhendra, R.; et al. (2020) The effect of column and temperature variation on the determination of the dead time in gas chromatographic systems using indirect methods. *Heliyon*. 6 (2), e03302–e03302.

[18] Idroes, R.; Japnur, A.F.; Suhendra, R.; and Rusyana, A. (2019) Kovats Retention Index analysis of flavor and fragrance compound using Biplot Statistical method in gas chromatography systems. *IOP Conference Series: Materials Science and Engineering*. 523 012007.

[19] Husna, I.; Rusyana, A.; Muslem; Idroes, G.M.; Suhendra, R.; and Idroes, R. (2020) Grouping of Retention Index on Gas Chromatography using Cluster Analysis. *IOP Conference Series: Materials Science and Engineering*. 796 012064.

[20] Mitchell, M. (1998) *An introduction to genetic algorithms*. MIT press, .

[21] Riahi, S.; Ganjali, M.R.; Pourbasheer, E.; and Norouzi, P. (2008) QSRR Study of GC Retention Indices of Essential-Oil Compounds by Multiple Linear Regression with a Genetic Algorithm. *Chromatographia*. 67 (11–12), 917–922.

[22] Sivanandam, S.N.; and Deepa, S.N. (2008) *Introduction to Genetic Algorithms*. Springer Berlin Heidelberg, Berlin, Heidelberg.

[23] Shelokar, P.; Kulkarni, A.; Jayaraman, V.K.; and Siarry, P. (2014) Metaheuristics in Process Engineering: A Historical Perspective. in: Appl. Metaheuristics Process Eng., Springer International Publishing, Champp. 1–38.

[24] Idroes, R.; Maulana, A.; Noviandy, T.R.; Suhendra, R.; Sasmita, N.R.; Lala, A.; et al. (2020) A Genetic Algorithm to Determine Research Consultation Schedules in Campus Environment. *IOP Conference Series: Materials Science and Engineering*. 796 012033.

[25] Husin, N.A.; Mustapha, N.; and Sulaiman, M.N. (2017) Hybridization of Genetic Algorithm and Neural Network on Predicting Dengue Outbreak. *International Review on Computers and Software (IRECOS)*. 12 (5), 219.

[26] Vaishali, R.; Sasikala, R.; Ramasubbareddy, S.; Remya, S.; and Nalluri, S. (2017) Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians Diabetes dataset. *Proceedings of*

the IEEE International Conference on Computing, Networking and Informatics, ICCNI 2017. 2017-Janua 1–5.

[27] Shi, H.; and Xu, M. (2018) A Data Classification Method Using Genetic Algorithm and K-Means Algorithm with Optimizing Initial Cluster Center. *2018 IEEE International Conference on Computer and Communication Engineering Technology, CCET 2018*. 224–228.

[28] Idroes, R.; Noviandy, T.R.; Maulana, A.; Suhendra, R.; Sasmita, N.R.; Muslem, M.; et al. (2019) Retention Index Prediction of Flavor and Fragrance by Multiple Linear Regression and the Genetic Algorithm. *International Review on Modelling and Simulations (IREMOS)*. 12 (6), 373.

[29] Parveen, N.; Zaidi, S.; and Danish, M. (2017) Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr(VI). *Process Safety and Environmental Protection*. 107 428–437.

[30] Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A.K.; Rupp, M.; Teetz, W.; et al. (2011) Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *Journal of Computer-Aided Molecular Design*. 25 (6), 533–554.

[31] Babushok, V.I.; Linstrom, P.J.; and Zenkevich, I.G. (2011) Retention Indices for Frequently Reported Compounds of Plant Essential Oils. *Journal of Physical and Chemical Reference Data*. 40 (4), 043101.

[32] Mihaleva, V. V.; Verhoeven, H.A.; de Vos, R.C.H.; Hall, R.D.; and van Ham, R.C.H.J. (2009) Automated procedure for candidate compound selection in GC-MS metabolomics based on prediction of Kovats retention index. *Bioinformatics*. 25 (6), 787–794.

[33] Üstün, B.; Melssen, W.J.; Oudenhuijzen, M.; and Buydens, L.M.C. (2005) Determination of optimal support vector regression parameters by genetic algorithms and simplex optimization. *Analytica Chimica Acta*. 544 (1–2), 292–305.

[34] Zhang, J.; Zheng, C.H.; Xia, Y.; Wang, B.; and Chen, P. (2017) Optimization enhanced genetic algorithm-support vector regression for the prediction of compound retention indices in gas chromatography. *Neurocomputing*. 240 183–190.