Article

Hybrid Machine Learning Model of Support Vector Machine and Fruit Fly Optimization Algorithm for Prediction of Remaining Service Life of Flexible Pavement

Nader Karballaeezadeh [1], Adrienn Dineva [2], Amir Mosavi [3], Narjes Nabipour [4,*], Shahaboddin Shamshirband [5,6,*], Danial Mohammadzadeh [7,8]

[1] Department of Civil Engineering, Shahrood University of Technology, Shahrood, Iran: N.karballaeezadeh@shahroodut.ac.ir
[2] Kalman Kando Faculty of Electrical Engineering, Obuda University, 1034 Budapest, Hungary
[3] Queensland University of Technology (QUT), 130 Victoria Park Road, Queensland 4059, Australia
[4] Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam;
[5] Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[6] Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Vietnam
[7] Department of Civil Engineering, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran
[8] Department of Elite Relations with Industries, Khorasan Construction Engineering Organization, Mashhad 9185816744, Iran
* Correspondence: Shahaboddin.shamshirband@tdtu.edu.vn

Abstract: Remaining service life (RSL) of pavement, as a sign of future pavement performance, has always received growing attention from pavement engineers. The RSL describes the time from the moment of pavement inspection until such a time when a major repair or reconstruction is required. The conventional approach to determining RSL involves using non-destructive tests. These tests, in addition to being costly, interfere with traffic flow and compromise users' safety. In this paper, surface distresses of pavement have been used to estimate the pavement's RSL in order to eliminate the aforementioned problems and challenges. To implement the proposed theory, 105 flexible pavement segments were taken from Shahrood-Damghan Highway (Highway 44) in Iran. For each pavement segment, the type, severity, and extent of surface damage and pavement condition index (PCI) were determined. The pavement RSL was then estimated using non-destructive tests include Falling Weight Deflectometer (FWD) and Ground Penetrating Radar (GPR). After completing the dataset, the modeling was conducted to predict RSL using three techniques include Support Vector Regression (SVR), Support Vector Regression Optimized by Fruit Fly Optimization Algorithm (SVR-FOA), and Gene Expression Programming (GEP). All three techniques estimated the RSL of the pavement by selecting the PCI as input. The Correlation Coefficient (CC), Nash-Sutcliffe efficiency (NSE), Scattered Index (SI), and Willmott's Index of agreement (WI) criteria were used to examine the performance of the three techniques adopted in this study. In the end, it was found that GEP with values of 0.874, 0.598, 0.601, and 0.807 for CC, SI, NSE, and WI criteria, respectively, had the highest accuracy in predicting the RSL of pavement.

Keywords: hybrid machine learning model; transportation infrastructure; flexible pavement; remaining service life prediction; pavement condition index; support vector regression; fruit fly optimization algorithm (FOA); gene expression programming (GEP); SVR-FOA

Peer-reviewed version available at *Mathematics* **2019**, *7*, 1198; doi:10.3390/math7121198

1. Introduction

Predicting future pavement conditions and estimating its service life is one of the fundamental tasks of pavement engineers in pavement management systems (PMSs), as the future network conditions act as a prerequisite for the planning, prioritization, and allocation of resource [1]. In general, pavement management activities are split into two categories [2]: network-level management and project-level management. Project-level management specifies roads that need to be repaired, repair process, and repair timetable. Therefore, predicting future conditions of pavement is essential for network-level management [2]. Forecasting pavement future conditions will require ongoing pavement assessment and inspection that will improve the operational quality of maintenance operations [3,4]. On the other hand, network-level management focuses on determining the budget required to preserve the pavement network at the standard level. Hence, it is indispensable to determine the remaining service life (RSL) of payment at this level of management [2]. Various factors such as traffic, characteristics of pavement materials, subgrade properties, climatic conditions, and maintenance quality have a destructive impact on the road pavement. As a result, pavement service life is surrounded by uncertainties that complicate the prediction of RSL [5].

In recent years, several indices have been developed for pavement evaluation. One of these indicators is the pavement condition index (PCI), which represents the general conditions of the pavement surface and ranges from zero for a practically unusable pavement to 100 for a flawless pavement. PCI is determined based on pavement inspection results in terms of type, severity, and extent of distresses [6,7]. The PCI estimation requires an experienced inspector to determine PCI after a thorough inspection of pavement surface distresses. Conversely, RSL of pavement, which is one of the pillars of network-level pavement management, is determined by applying Falling Weight Deflectometer (FWD) non-destructive test. In this test, a traffic lane is first blocked. Then an impulsive loading is applied to the pavement to induce pavement surface deflections. By analyzing pavement surface deflections, the RSL is calculated. A key point in calculating the RSL using the above method is knowing the pavement layers thickness for the analysis of deflections. The thickness of the pavement layers is determined by the Ground Penetrating Radar (GPR) non-destructive test. As a result, two non-destructive tests are required to determine the RSL of pavement. Also, traffic interference during FWD and GPR tests should not be forgotten[8].

In light of the above, the current method of determining RSL is not only costly but also compromises the safety of inspectors due to traffic interference during testing. Given the limited budget resources of transportation agencies, determining the RSL of pavement to manage a pavement network represents one of the ongoing concerns of such companies[4]. One way to overcome these problems is to employ artificial intelligence models. In this paper, three methods of Support Vector Regression (SVR), SVR-FOA (Support Vector Regression Optimized by Fruit Fly Optimization Algorithm), and Gene Expression Programming (GEP) have been applied to predict the RSL of road pavements. For the ease of application and implementation, PCI has been used as the only input of these methods.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at *Mathematics* **2019**, *7*, 1198; doi:10.3390/math7121198

PCI is an index adopted in the project-level pavement management process. Hence, this index is specified before entering network-level pavement management. As such, using PCI for network-level management operations, such as estimation of RSL, will contribute to the overlapping of activities and saving time. The methods presented in this paper, by excluding non-destructive tests from the process of determining the RSL of pavement, drastically reduce costs associated with the pavement network management. On the other hand, by eliminating non-destructive tests, traffic interference during testing and the potential safety hazards to inspectors are also eliminated. The data required for developing the methods proposed in this paper was collected from Shahrood-Damghan highway in Semnan province, Iran. The pavement type of this highway is flexible. In this highway, a 100-m long pavement segment was taken from the beginning of each kilometer, and after assessing the surface distresses of each segment, its PCI was calculated. In the next step, FWD and GPR tests were applied to the selected segments to determine the RSL. Finally, using three SVR, SVR-FOA and GEP methods, the RSL modeling based on PCI was implemented.

This paper is organized as follows: Section 2 offers a review of studies on the prediction of RSL. Section 3 introduces the method employed in this study. This section is made of four sub-sections titled PCI, RSL, artificial intelligence techniques, and case study. In Section 4, the results of the analyses are presented and discussed. Finally, the conclusions are drawn in Section 5.

2. Literature Review

PMS is an assessment management system (AMS) used by road network administrators to maintain the entire network at the desired level. Predicting pavement performance is a key factor in PMSs [9]. The models of predicting pavement's RSL fall in the category of the pavement performance prediction models. In this section, studies carried out by other researchers on predicting RSL are reviewed. Table 1 lists the results of studies that have strived to estimate RSL to date.

Table 1. Models for the prediction of RSL[10-17]

Category	Model inputs	Equation	Author
1. Based on pavement responses	ε_t = Tensile strain at the bottom of the asphalt layer, E_1 = Elastic modulus of asphalt.	$RSL_{fatigue} = f_1(\varepsilon_t)^{-f_2}(E_1)^{-f_3}$	Huang (1993)
	ε_c = Compressive strain at the top of the subgrade.	$RSL_{rutting} = f_4(\varepsilon_c)^{-f_5}$	Huang (1993)
	ε_t = Tensile strain at the asphalt layer bottom,	$RSL_{fatigue} = 0.1001(\varepsilon_t) - 3.565(M_R)^{-1.4747}$	Das & Pandey (1999)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at *Mathematics* **2019**, *7*, 1198; doi:10.3390/math7121198

	M_R = Resilient modulus.		
	ε_r = Horizontal tensile strain at the bottom of the asphalt layer, E_{AC} = Modulus of asphalt.	$\ln(RSL_{fatigue}) = a - b\ln(\varepsilon_r) - c\ln(E_{AC})$	Hossain & Wu (2002)
	ε_t = Tensile strain at the asphalt layer bottom.	$RSL_{fatigue} = K(\varepsilon_t)^{-C}$	Park & Kim (2003)
2. Based on pavement quality indices	IRI = International roughness index.	$RSL = \frac{\ln(\frac{IRI_{terminal}}{a})}{b} - (\text{Current age})$	Al-suleiman & Shiyab (2003)
	PCI = Pavement Condition Index.	$RSL = 4.1872\ln(PCI) - 14.728$	Setyawan et al (2015)
3. Based on the result of the non-destructive test	δ = Pavement surface curvature, $\delta = D_0 - D_{20}$	$RSL_{fatigue} = \alpha\left(\frac{1}{0.0023\delta + 0.00002}\right)^{\beta}$	Saleh (2016)
	AUPP = Area under pavement profile, $AUPP = \frac{5D_0 - 2D_{30} - 2D_{60} - D_{90}}{2}$	$RSL_{fatigue} = \alpha(\frac{1}{0.0000023AUPP^{0.912}})^{\beta}$	Saleh (2016)

In Table 1, the models of determining RSL of pavement are divided into three categories based on the model inputs:

• First category: Models that predict RSL based on the response (stress and strain) of pavement to the applied loads.

• Second category: Models that predict RSL based on pavement quality indices.

• Third Category: Models that predict RSL based on the results of pavement non-destructive tests.

Among the above categories, models that predict remaining pavement service based on qualitative indices appear to be more appropriate. It is because such models neither call for the analysis of pavement behavior and response, as in the first category nor require non-destructive tests, like the third category. Instead, they estimate the RSL of the pavement by assessing pavement and calculating a qualitative index in the simplest possible way. In light of the above points, in this paper, PCI has been adopted as a qualitative index for predicting RSL. Setyawan et al. conducted a similar study the results of which are displayed in Table 1. The model presented in their study was based on data collected from only 5 pavement segments in the Microsoft Excel software, which cast doubt on the reliability of the model. In contrast, the methods proposed in this paper are based on the data gathered from more than 100 pavement segments using artificial intelligence methods.

In general, PCI offers a valid index accepted by all transportation agencies around the world, and it is widely used in their evaluations. Compared to the current method of estimating the RSL of pavements (using two non-destructive FWD and GPR tests), the proposed method provides a far simpler, safer and less costly way of estimating RSL.

3. Methodology

3.1. Pavement Condition Index (PCI)

PCI was developed by U.S. Army Crops of Engineers as a performance benchmark for PMSs[7]. Extensively used in roads, parking lots, and airports, PCI is recognized as a standard practice by many organizations around the world, including the Federal Aviation Administration, the American Public Works Association, and the U.S. Air Force[18]. PCI is a numerical index that expresses the rate of pavement surface distresses. PCI exhibits structural integrity and Surface operational condition but is not able to measure structural capacity[19].

To determine PCI in a pavement segment, the pavement surface of the segment is inspected, and their surface distresses are recorded in the assessment form. The PCI of a pavement sample unit depends on the type, extent, and severity of its surface damages. In the PCI calculation process, a perfect and flawless pavement receives a maximum score of 100. For defective pavements, the score is deducted from 100 incommensurate with the type, extent, and severity of the damage[7,18,19]. In general, PCI can be computed using Eq. 1[4]:

$$\mathrm{PCI} = 100 - \sum_{i=1}^{n}(\text{Distress Score}) \qquad (1)$$

where: PCI = Pavement Condition Index, distress Score = Score based on type, extent, and severity of distresses, n = Number of distresses. The instructions for calculating distress scores are fully described in ASTM D6433-07. The classification of a pavement segment based on PCI is according to Table 2.

Table 2. PCI rating scale[19]

Rating Scale	0-10	10-25	25-40	40-55	55-70	70-85	85-100
Description	Failed	Serious	Very Poor	Poor	Fair	Satisfactory	Good

3.2. Remaining Service Life (RSL)

Peer-reviewed version available at *Mathematics* **2019**, *7*, 1198; doi:10.3390/math7121198

The RSL of pavement under operation is a key factor in implementing PMSs. It is because learning about the future conditions of the pavement network is essential for decision making, life cycle cost analysis, planning and budget allocation[8,20]. In general, the definitions of RSL by different agencies and departments of transportation can be split into two general categories[21]:

- *The remaining time to reach a level of distress when the pavement needs to be rehabilitated or reconstructed*. For example, the Minnesota Department of Transportation (MnDOT) defines the RSL as the time until the next major rehabilitation.
- *The time until pavement conditions reach a specific condition index limit*. For example, the Michigan Department of Transportation (MDOT) defines the RSL based on the Michigan Ride Quality Index, assuming an RSL of zero when the said index is 50.

The RSL of pavement segments in this paper has been determined using Heavy Falling Weight Deflectometer (HWD). The HWD is an FWD, the application of which is not constricted to the road and could be used for airport pavement assessment. Figure 1 shows the HWD device employed in this study.

Fig. 1. HWD used in this study for determining RSL

The HWD applies a tension similar to the standard axle load (8.2 tons) to the pavement surface over 10 to 35 milliseconds. A number of geophones are placed on the pavement surface at specified distances from the loading center. The task of the geophones is to record the pavement deflections induced by the load applied with the HWD device. Standard axle load simulation in HWD is generated by a series of weight drops on a loading plate placed on the pavement surface[8]. Table 3 reveals the details of the HWD test undertaken in this study.

Peer-reviewed version available at Mathematics 2019, 7, 1198; doi:10.3390/math7121198

Table 3. HWD test details[22]

Parameter	Value
Tension (Kpa)	600 – 900
Number of geophones	9
Geophone distance from center of laoding plate (cm)	0, 20, 30, 45, 60, 90, 120, 150, and 180
Number of weights falling	Four times
Loading plate radius (mm)	150

The deflections recorded by geophones are transferred to the central computer in HWD. In this computer, the analysis of pavement surface deflections is performed using ELMOD6 software. One output of this analysis is the estimation of pavement remaining service life under test.

A prerequisite of estimating RSL in accordance with the process described in this section is knowing the thickness of pavement layers. For determining the pavement layer's thickness, ground-penetrating radar (GPR) non-destructive testing was carried out for all pavement segments. GPR is capable of calculating pavement thickness as a continuous profile by transmitting electromagnetic waves through a transmit antenna and receiving recursive signals[23]. Table 4 shows the complete details of the GPR experiment carried out in this study.

Table 4. GPR test details[24,25]

Parameter	Value
Antenna type	1000 MHz
Record speed (km/hr)	10
Number of scanning (per meter)	10
Maximum pulse penetrating depth (ns)	32

3.3. Artificial intelligent techniques

In this paper, the main objective is to present a new approach for predicting the RSL of flexible pavements based on PCI using artificial intelligence techniques. To do so, a dataset of 105 pavement segments was collected from Shahrood-Damghan highway in Iran. The data set consisted of the RSL and PCI of all segments under study, which were analyzed using the GEP, SVR and SVR-FOA methods. GEP is an evolutionary algorithm that investigates the relationship between input and output variables by developing computer programs[26]. Different from both the Genetic Algorithm (GA) and Genetic Programming (GP), GEP is a combination of both introduced by Ferreira in 2001[27]. SVR is a supervised machine learning technique employed to solve regression problems. SVR is especially popular due to its desirable management and performance in handling nonlinear issues[28]. The success of an SVR in problem-solving depends on the values of its basic parameters. The SVR basic parameters include C, ε, and kernel function parameters[8]. Improper values of the basic parameters in the SVR can lead to under-fitting or over-fitting. Thus, optimum values must be selected for these parameters during training[29]. FOA is an intelligent swarm algorithm introduced by Pan in 2012, which utilizes the food searching strategy of a fruit fly to find the optimal values for SVR basic parameters[30]. These methods are introduced in the following subsections.

3.3.1. Gene Expression Programming (GEP)

GEP is a developed GP method that solves a problem by creating expression trees (ETs). In fact, GEP is an evolutionary algorithm for creating computer programs. The designed computer programs have sophisticated tree structures that are trained similar to a living organism by changing size, shape, and composition, and adapted to the conditions. Like living organisms, GEP programs are coded as simple fixed-length linear chromosomes. Hence, GEP is a genotype-phenotype system that employs a simple genome to store and transmit genetic information and adopts a complex phenotype to explore and adapt to the environment. The genome consists of a chromosome or a fixed-length string that combines one or more genes of the same size. In fact, each chromosome contains one or more genes known as Sub-ETs. In GEP, all Sub-TEs are linked through the root with connection functions. The connection functions in GEP include division, multiplication, subtraction, and addition[31]. Figure 2 shows an instance of the genotype-phenotype structure in GEP.

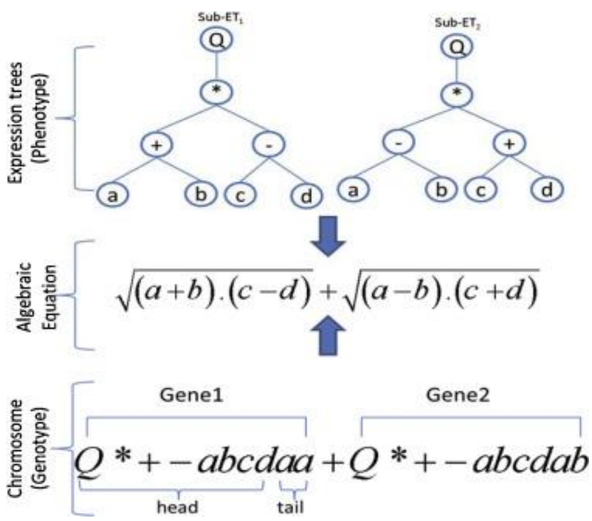


Fig. 2. A sample of the genotype-phenotype structure in GEP[32]

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at Mathematics 2019, 7, 1198; doi:10.3390/math7121198

These genes, despite their fixed length, are coded for ETs of varying size and shape. It implies that the size of the coding region varies from one gene to another to allow for progressive adaptation and evolution. Each gene has a coding area called Open Reading Frame (ORF), which after being coded as an expression tree, provides a solution to the problem[33]. Figure 3 demonstrates the coding region (ORF), non-coding region, and the expression tree for a gene.

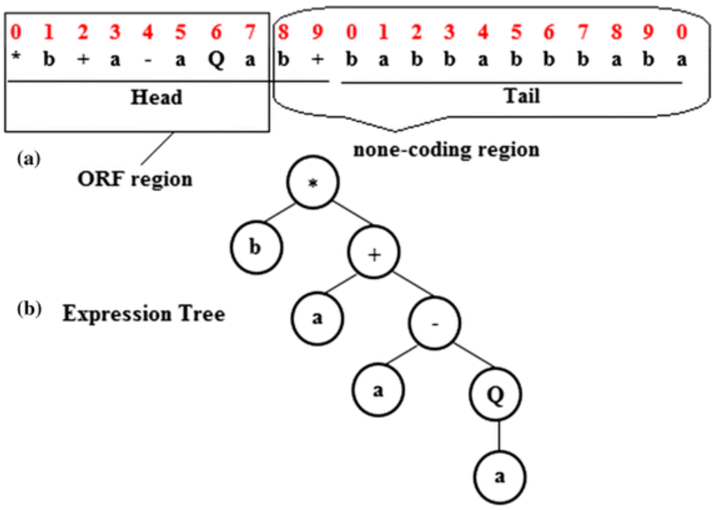


Fig. 3. The coding region (ORF), no-coding region, and expression tree in a gene [34]

Like other evolutionary approaches, GEP begins by randomly generating initial population chromosomes. In the first population, each chromosome is assessed based on the fitness function and receives a fitness value. Various fitness functions have been used in GEP, including Root Relative Squared Error (RRSE), Relative Square Error (RSE), Root Mean Square Error (RMSE), and Mean Square Error (MSE)[26]. The proper chromosomes are more likely to be picked in the next generation. After being selected, chromosomes are amended by genetic operators (including transposition, inversion, mutation, recombination, and gene crossover) and then reconstructed. This process is sustained until a suitable solution or the maximum number of generations is reached[26,35].

3.3.2. Support Vector Regression (SVR)

The SVR is actually a support vector machine (SVM) used for regression problems. Supervisor machine learning techniques, including SVR, utilize Structural Risk Minimization (SRM), while conventional neural networks use Empirical Risk Minimization (ERM). ERM minimizes the error of training samples, but SRM is able to minimize a higher level of error. As a result, SVM is capable of overcoming the deficiencies of conventional neural networks[8,28]. The main idea in SVR is to map nonlinear information into a higher dimensional space and then solve a linear regression problem in the new space[28,36,37]. In the new space, a simple linear kernel function is adopted to solve the problem. However, in complex problems, a simple linear kernel function will be inadequate. The kernel function k(x,z) for all $x, z \in X$ is defined as follows[28]:

$$k(x,z) =< \varphi(x).\varphi(z) > \qquad (2)$$

Each kernel function must have two features[28]:

1. Symmetricity

$$k(x,z) =< \varphi(x).\varphi(z) >=< \varphi(z).\varphi(x) >= k(z,x) \qquad (3)$$

2. Compliance with the Cauchy-Schwartz criterion

$$k(x,z)^2 =< \varphi(x).\varphi(z) >^2 \leq \|\varphi(x)\|^2\|\varphi(z)\|^2 \qquad (4)$$

These two conditions guarantee that the new space is definable by the kernel function. The most famed kernel functions are Polynomial kernel, Radial Basis Functional kernel, Linear kernel, and Sigmoid kernel [37]. Given the above explanation, it is clear that SVR requires an appropriate function to explain the nonlinear relationship between input (x_i) and output (y_i)[37]:

$$f(x_i) = w.\varphi(x_i) + b \qquad (5)$$

where: $\varphi(x_i)$ = Transformation Function, w = Weight, b = Bias. w and b are obtained by minimizing the following function, which is known as the regularized risk function [37]:

$$R(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N} L_\varepsilon(y_i, f(x_i)) \qquad (6)$$

where:

$\frac{1}{2}\|w\|^2$ = Regularization term,

C = Penalty coefficient,

$L_\varepsilon(y_i, f(x_i))$ = ε-insensitive loss function:

$$L_\varepsilon\big(y_i, f(x_i)\big) = \max\{0, |y_i, f(x_i)| - \varepsilon\} \qquad (7)$$

where: ε = Permitted error threshold.

To solve the optimization boundaries, two factors of ξ and ξ^* are defined[37]:

$$\min f(w,\xi,\xi^*) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi,\xi^*) \qquad (8)$$

Subject to:

$$\begin{cases} y_i - [w.\varphi(x_i)] - b \leq \varepsilon + \xi & , \quad \xi \geq 0 \\ [w.\varphi(x_i)] + b - y_i \leq \varepsilon + \xi^* & , \quad \xi^* \geq 0 \end{cases} \qquad (9)$$

We now need to define a Lagrange function based on the objective function and boundary conditions[37]:

$$\max H(\partial_i^-, \partial_i^+) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\partial_i^- - \partial_i^+)(\partial_j^- - \partial_j^+)K(x_i, x_j) + \sum_{i=1}^{N} y_i(\partial_i^- - \partial_i^+) - \varepsilon\sum_{i=1}^{N} y_i(\partial_i^- + \partial_i^+) \qquad (10)$$

Subjected to:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at Mathematics 2019, 7, 1198; doi:10.3390/math7121198

$$\sum_{i=1}^{N} y_i(\partial_i^- - \partial_i^+) = 0\,, \quad \partial_i^-, \partial_i^+ \in [0\,, C] \qquad (11)$$

Therefore, the regression function can be showed as follows[37]:

$$f(x) = \sum_{i=1}^{N} (\partial_i^- - \partial_i^+) K(x_i\,, x_j) + b \qquad (12)$$

Where $K(x_i\,, x_j)$ = Kernel function.

Figure 4 summarizes the overall structure of the SVR.

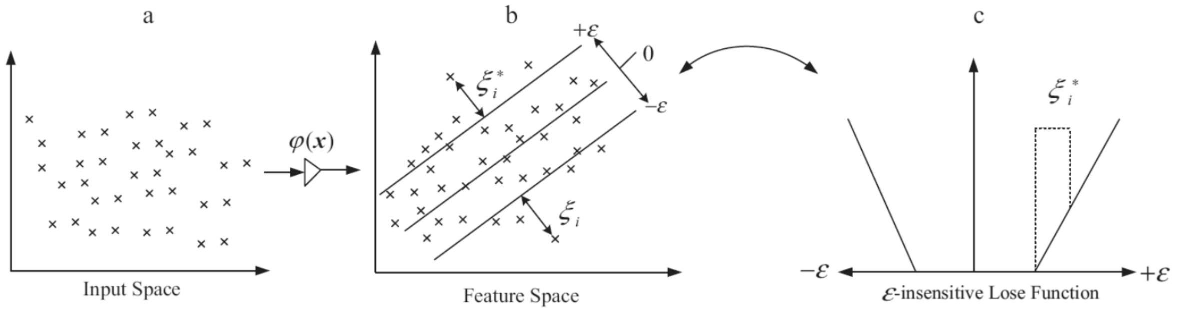


Fig. 4. The transformation process in SVR[38]

In general, SVR performance depends on its parameters. SVR parameters are[39,40]:

- ε: this parameter supervises the width of the ε-insensitive zone, used to fit the training data. The value ε can affects the number of support vectors used to build the regression function. For the bigger ε, estimates are more 'flat', and the fewer support vectors are chosen.
- C: this parameter specifies the trade-off between the complexity of the model and the grade to which deviations larger than ε are bearable in optimization formulation.
- γ: this parameter determines the relation between error minimization and smoothness of the estimated function.

These parameters are chosen by the user based on prior knowledge of SVR, so this method is not suitable for non-professional users. Various algorithms have been developed to optimize the amounts of SVR parameters. In the following subsection, one of the optimization algorithms used in this paper is introduced.

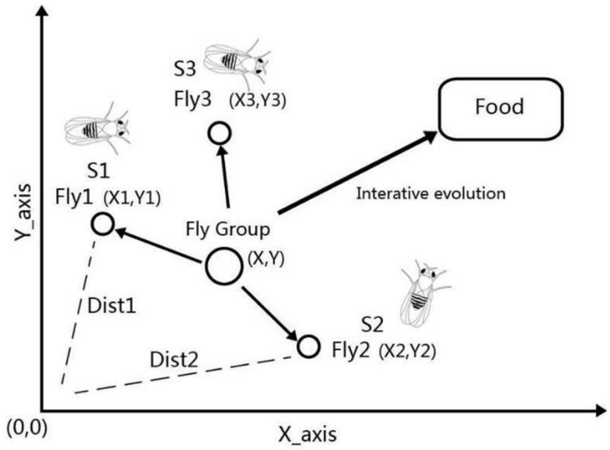
3.3.3. Fruit Fly Optimization Algorithm (FOA)

FOA is an optimization algorithm developed based on the food search behavior of the Drosophila insect[30]. The fruit fly has superior smell and vision senses, which discriminates it from other insects. Fruit flies track the smell of the food sources dispersed through the air and heads towards it. This insect is even capable of smell tracking from a distance of 40 km. When approaching the food source, the fruit fly employs a sense of vision to locate food and other fruit flies. The best information is

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at Mathematics 2019, 7, 1198; doi:10.3390/math7121198

296 shared among the fruit flies and finally, the route leading to the food source is identified[41]. Figure
297 5 illustrates the process of food search by a fruit fly.



298

299 **Fig. 5.** Food finding the iterative process of a fruit fly swarm[42].

300

301 In this paper, this optimization algorithm has been selected to find the optimal values of SVR
302 parameters, which is known as SVR-FOA. The general steps of SVR-FOA can be summarized as
303 follows[43,44]:

304 **Step 1.** SVR parameters (ε , C) initialization and kernel function determination

305 **Step 2.** Parameter initialization

306 Including the maximum number of iterations, location of initial population (X-axis, Y-axis),
307 population size, and random flight distance domain:

$$\text{X\_axis} = \text{rands}(1,2) \tag{13}$$

$$\text{Y\_axis} = \text{rands}(1,2) \tag{14}$$

308 **Step 3.** Population initialization

309 A random location (Xi, Yi) and food founding distance are assigned to each fruit fly:

$$X_i = \text{X\_axis} + \text{Random value} \tag{15}$$

$$Y_i = \text{Y\_axis} + +\text{Random value} \tag{16}$$

310 where: i = Population size.

311 **Step 4.** Population evaluation

312 The distance from origin to the food source (D) and the smell concentration parameter (S) are
313 calculated:

$$D_i = \sqrt{{X_i}^2 + {Y_i}^2} \tag{17}$$

$$S_i = \frac{1}{D_i} \tag{18}$$

314 **Step 5.** Replacement

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at Mathematics 2019, 7, 1198; doi:10.3390/math7121198

S value is substituted with the fitness function or smell concentration judgment function so that the smell concentration for each fruit fly location can be attained:

$$\mathrm{Smell}_i = \mathrm{Function}\ (S_i) \qquad (19)$$

Step 6. Detect the maximal smell concentration

At this point, the fruit fly with the highest S_i is identified and located within the population.

$$[\mathrm{bestSmellbestIndex}] = \max\ (\mathrm{Smell}) \qquad (20)$$

Step 7. Keep smell concentration

The coordinates of the maximum smell concentration are set, and the fruit fly swarm flows in that direction.

$$\mathrm{X\_axis} = \mathrm{X(bestIndex)} \qquad (21)$$

$$\mathrm{Y\_axis} = \mathrm{Y(bestIndex)} \qquad (22)$$

Step 8. Iterative optimization

Steps 3 to 6 are repeated until the smell concentration does not show any improvement compared to the previous one or the maximum number of repetitions in Step 2 is reached.

Step 9. Output the optimum parameter of SVR

3.4. Case study

To implement the theory proposed in this paper, a stretch of 105 km from Shahrood-Damghan highway in Iran was selected and inspected. Given that this highway is part of the route between Tehran (the capital of Iran) and Mashhad (the second most important city of Iran), it constitutes one of the major roads. The highway consists of two lanes in each direction and uses the flexible pavement. A prerequisite of implementing the proposed theory is to select a number of sample segments from his highway. To do so, segments 100 m in length and 7 m in width were selected from the beginning of each kilometer. By inspecting the selected segments, the surface distress data (type, severity, and extent) of each segment was recorded in the assessment forms. The PCI of each segment was calculated as described in Section 3.1. With PCI known, the RSL of the pavement segments need to be known. Hence, HWD and GPR tests were performed on all segments and the average RSL of each segment was determined. Figure 6 shows the position of the Shahrood-Damghan Highway as well as the starting and ending points of the segments understudy on this Google map.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at Mathematics 2019, 7, 1198; doi:10.3390/math7121198

Fig. 6. Highway No. 44 (Shahrood-Damghan) map used in this study.

4. Results and discussion

The results of the analysis are presented in this subsection. First, the statistical specifications of the input and output modeling variables are listed in Table 5. The data was extracted from IBM SPSS 23 software.

Table 5. Statistical characteristics of the utilized data.

Variable	mean	minimum	maximum	standard deviation	kurtosis	skewness	sig. in Kolmogorov-Smirnov test	correlation with RSL
PCI	59.97	19.00	100.00	21.51	-1.093	0.03	0.068	0.572
RSL	17.77	0.00	40.00	15.04	-1.393	0.50	0.000	1

As depicted in Table 5, the mean PCI of all pavement segments is 59.97, which according to Table 2, is indicative of the fair state of all segments. Also, the mean RSL of pavement is 17.77 years. For interpreting the mean RSL, it is worth noting that ELMOD6 software does not suggest the application of an overlay layer for this RSL. Hence, the average RSL of the segments is fairly desirable. Before calculating the correlation coefficient of the modeling input and output, the normality of the data must be determined. This is determined by kurtosis and skewness coefficients, as well as the results of the Kolmogorov-Smirnov test. According to Table 5, since PCI variables have normal distribution but RSL distribution is abnormal, the Spearman correlation test must be used. The correlation between PCI and RSL is 57.2%, which represents an average value.

doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at *Mathematics* **2019**, *7*, 1198; doi:10.3390/math7121198

As noted in subsection 3.3.2, SVR consists of three basic parameters (C, ε and γ), with the quality of SVR performance depending on the values selected for these three parameters. Table 6 shows the values of these basic parameters for the SVR as well as the optimized values of these parameters by the FOA algorithm.

Table 6. Parameters of the SVR and SVR-FOA models.

	Model		
	SVR		SVR-FOA
SVR parameter	C	1.0000	1.0022
	ε	0.0100	0.2561
	γ	0.0010	0.0760

In Table 7, the characteristics of the GEP model used in this study, including model parameters and genetic operators, are shown.

Table 7. Characteristics of the GEP model.

Parameter	Quantity
Head size	8
Number of Genes	3
Chromosomes	30
Linking Function	Addition (+)
One-Point Recombination	0.3
Two-Point Recombination	0.3
Inversion Rate	0.1
Gene Recombination Rate	0.1
Mutation Rate	0.044
Gene Transposition Rate	0.1
Used functions	+, -, ×, ÷, power

Eq. 23 shows the proposed formula of the GEP method for estimating pavement remaining service life in terms of PCI.

$$\mathrm{RSL} = -6.59964 + \frac{322.568}{\mathrm{PCI} - 8.96265} - 2.61881\sqrt{\mathrm{PCI}} + \mathrm{PCI} - 1.18921\sqrt{\mathrm{PCI}^{\frac{3}{2}}} + \frac{0.35035(3.83258\mathrm{PCI} - 9.83936)}{9.83936 + \mathrm{PCI}} \tag{23}$$

The results of scientific research are generally assessed with indicators that exhibit the accuracy and error of the analysis. In this study, four criteria entitled Correlation Coefficient (CC), Scattered Index (SI), Nash-Sutcliffe efficiency (NSE) and Willmott's Index of agreement (WI) were used to determine the quality of the outputs[45,46]:

Peer-reviewed version available at *Mathematics* **2019**, *7*, 1198; doi:10.3390/math7121198

$$\mathrm{CC} = \frac{\sum_{i=1}^{n} \mathrm{RSL_{Oi}RSL_{pi}} - \frac{1}{n}\sum_{i=1}^{n} \mathrm{RSL_{Oi}} \sum_{i=1}^{n} \mathrm{RSL_{pi}}}{\left(\sum_{i=1}^{n} \mathrm{RSL_{Oi}}^{2} - \frac{1}{n}(\sum_{i=1}^{n} \mathrm{RSL_{Oi}})^{2}\right)\left(\sum_{i=1}^{n} \mathrm{RSL_{pi}}^{2} - \frac{1}{n}\left(\sum_{i=1}^{n} \mathrm{RSL_{pi}}\right)^{2}\right)} \tag{24}$$

$$\mathrm{SI} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{RSL_{pi}} - \mathrm{RSL_{Oi}})^{2}}}{\overline{\mathrm{RSL_{Oi}}}} \tag{25}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i}^{n}\left(\mathrm{RSL_{Oi}} - \mathrm{RSL_{pi}}\right)^{2}}{\sum_{i}^{n}(\mathrm{RSL_{Oi}} - \overline{\mathrm{RSL_{Oi}}})^{2}} \tag{26}$$

$$\mathrm{WI} = 1 - \left[\frac{\sum_{i=1}^{n}\left(\mathrm{RSL_{pi}} - \mathrm{RSL_{Oi}}\right)^{2}}{\sum_{i=1}^{n}\left(\left|\mathrm{RSL_{pi}} - \overline{\mathrm{RSL_{Oi}}}\right| + \left|\mathrm{RSL_{Oi}} - \overline{\mathrm{RSL_{Oi}}}\right|\right)^{2}}\right] \tag{27}$$

where: RSL_{Oi} = Observed RSL i^{th} value, RSL_{Pi} = Predicted RSL i^{th} value, and $\overline{RSL_{Oi}}$ = Average of RSL_{Oi}. CC is a number in the range of [-1, + 1], with values of +1, -1 indicating a complete correlation between model inputs and outputs. Positive values represent direct correlation and negative values demonstrate an inverse correlation. As the absolute value of CC approaches zero, the strength of correlation decreases. SI represents an error, and smaller values indicate lower errors in modeling. The highest NSE value is one with values close to one indicating greater modeling accuracy so that NSE = 1 represents the best modeling quality. WI is an index between 0 and 1 with values close to one suggesting the higher the modeling accuracy.

Table 8 reveals the four criteria introduced for all three methods employed in this study. To shed further light on the results of Table 8, a three-dimensional bar histogram of the criteria is presented in Figure 7. Based on the description in the preceding paragraph and the values in Table 8, it can be concluded that FOA has improved SVR results. Comparing the SVR-FOA and GEP modeling results, it can be contended that the GEP results are partially superior in the modeling proposed in this paper.

Table 8. Performance evaluation indices for GEP, SVR and SVR-FOA models.

parameter	GEP	SVR	SVR-FOA
CC	0.874	0.865	0.879
SI	0.598	0.894	0.616
NSE	0.601	0.110	0.577
WI	0.807	0.369	0.786

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at *Mathematics* **2019**, *7*, 1198; doi:10.3390/math7121198

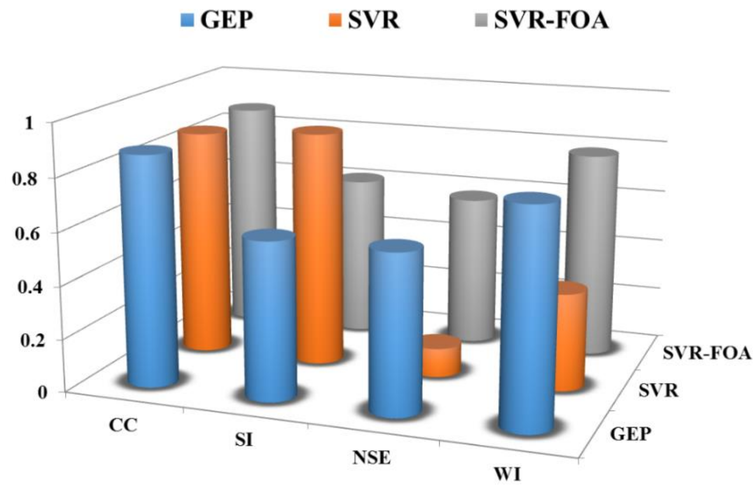


Fig. 7. Three-dimensional bar graphs of the statistical parameters.

In this study, a dataset containing information about 105 pavement segments was used. Approximately 70% of the data (75 segments) were used for training and the remaining 30 segments were utilized for testing. Figure 8 shows the RSL predicted by the three SVR, SVR-FOA and GEP methods as well as the RSL measured in the HWD test for the segments selected as the test.

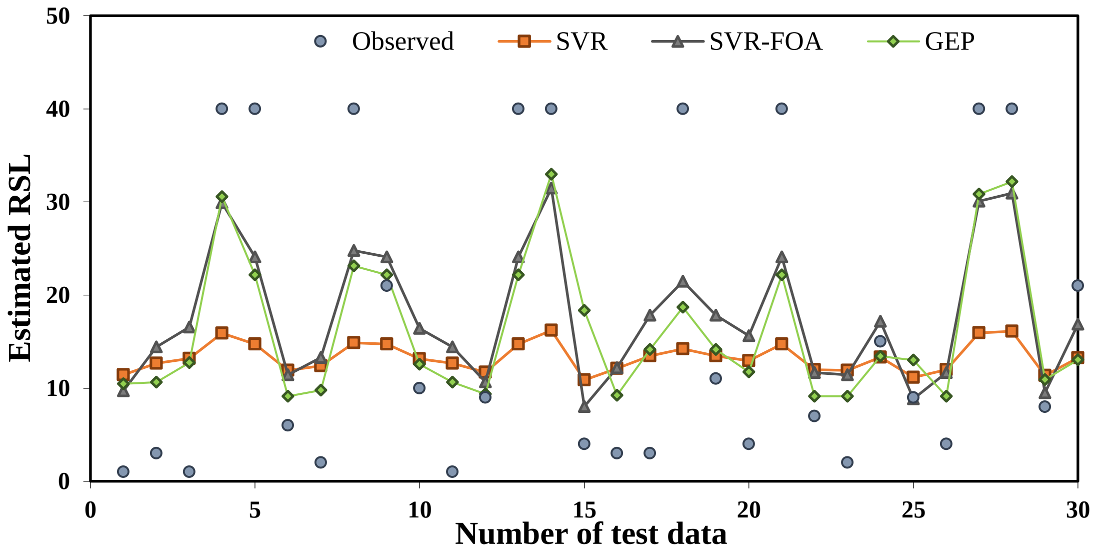


Fig. 8. Observed and estimated values of RSL with SVR, SVR-FOA and GEP models for test data.

Figure 9 displays the predicted RSL values versus the RSL values calculated by the HWD test for all three machine learning techniques adopted in this paper. In this regard, the method with the best prediction accuracy is the one that has a fit line equation of y = x, meaning that the line slope is equal to one and its intercept is equal to zero. However, since in most cases it is not possible to reach this state, it is generally stated that the highest prediction accuracy for a drawn fit line is obtained when the slope is 1 and the intercept is 0. Figure 11 is plotted for the test dataset.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1
Peer-reviewed version available at Mathematics 2019, 7, 1198; doi:10.3390/math7121198

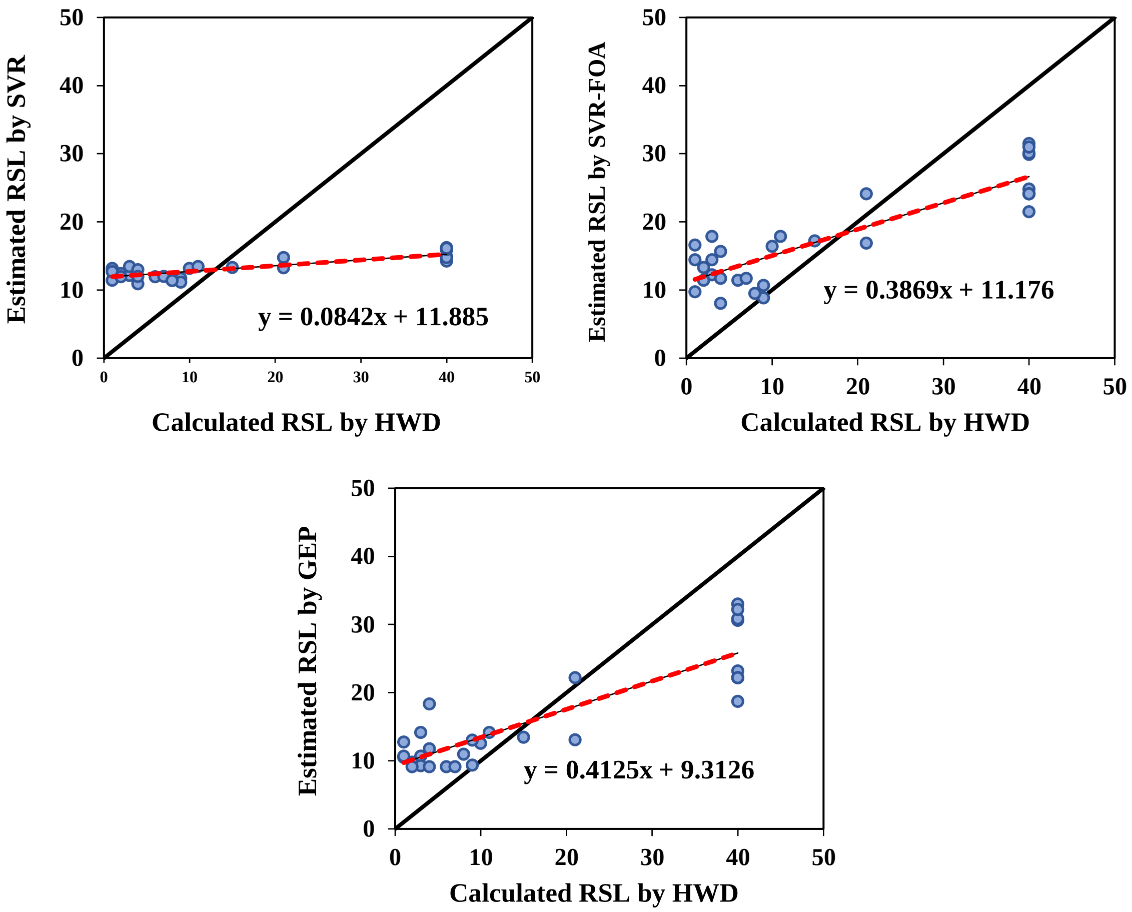


Fig. 9. The scatter plots of calculated RSL by HWD and estimated RSL by SVR, SVR-FOA and GEP models for test data

Figure 10 shows the Taylor diagram of this article. Introduced by Taylor in 2001, Taylor diagram is a mathematical diagram that graphically allows a comparison of several models of a system. In this diagram, there are three categories of contours[47]:

- **Blue contours**

It shows the Pearson correlation coefficient.

- **Orange contours**

It indicates the RMS error that is proportional to the distance from a green spot on the horizontal axis called observed.

- **Black contours**

It indicates the standard deviation proportional to the radial distance from the center.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at Mathematics 2019, 7, 1198; doi:10.3390/math7121198

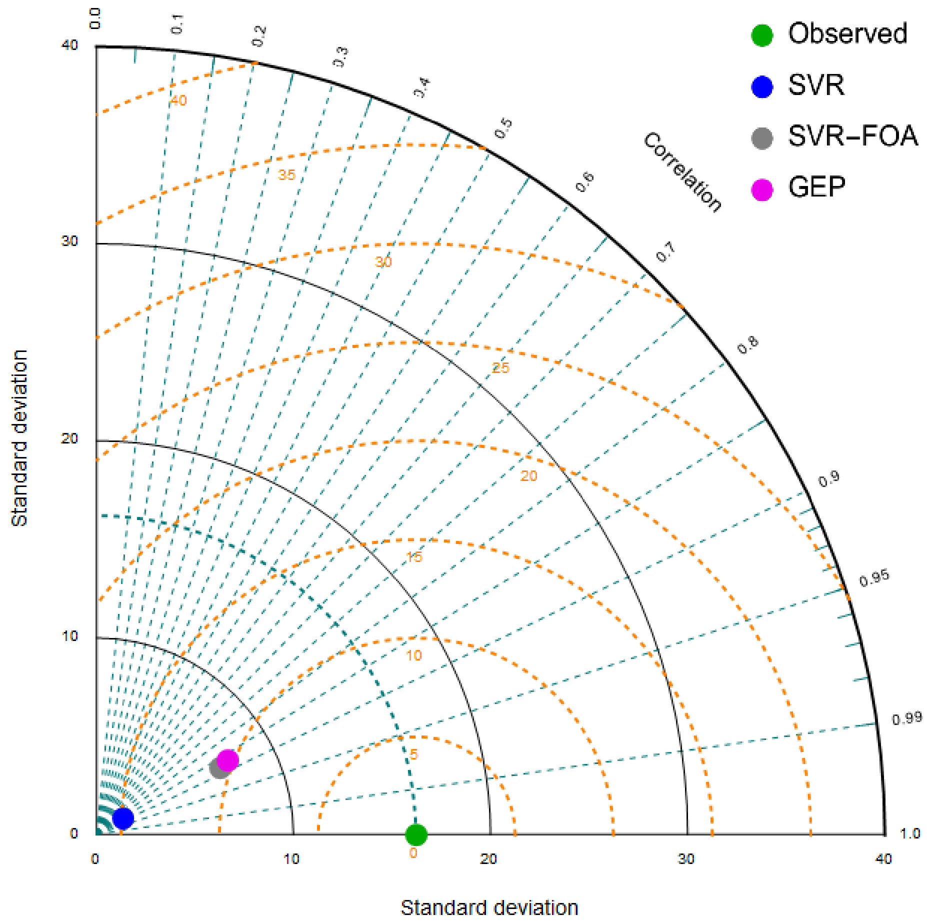


Fig. 10. Taylor diagrams of estimated RSL for all models.

By examining Figures 7 to 10, it can be concluded that the SVR technique offers an average accuracy for the purpose of this article. Using the FOA algorithm to select the basic parameters of this technique significantly enhanced the accuracy of this method. On the other hand, the GEP method provides a formula for RSL prediction. By re-examining Figures 7 to 10, it turned out that both SVR-FOA and GEP methods yielded desirable accuracy for RSL prediction. However, the accuracy of the GEP method was slightly higher than that of the SVR-FOA method.

5. Conclusion

Pavement management at both project and network levels are always associated with substantial costs. Due to the budget constraints inflicted on organizations in charge of PMS, optimizing pavement management costs is one of the priorities of any organization. RSL is a crucial factor for pavement management at the network level. The current procedure for determining RSL involves using FWD and GPR tests. These devices are not only costly but also interfere with the traffic flow and compromise the safety of pavement inspectors. The aim subject of the study was to present a new approach for predicting the RSL of flexible pavement, which eliminated the drawbacks of current methods. After a review of previous studies on estimating the pavement RSL, we decided to use pavement surface distresses as a criterion of predicting RSL. Therefore, PCI pavement was employed as input variable in modeling pavement RSL. PCI is an index that assigns a score of 0 to 100 based on the type, severity, and extent of pavement surface distress, with zero indicating the worst situation and 100 representing the highest quality. The dataset utilized for modeling was selected from Shahrood-Damghan highway in Iran. After selecting 105 pavement segments from the highway, PCI and RSL of all segments were determined. Modeling was conducted using GEP and SVR techniques after completing the dataset. The results of modeling with these techniques were evaluated based on four criteria include CC, SI, NSE, and WI to determine the most appropriate technique for estimating pavement RSL. After exploring all four criteria, it was found that the GEP outcomes were far more accurate than the SVR. Then, to improve the accuracy of the SVR method, the FOA optimization algorithm was employed to add a third technique (SVR-FOA) to the methods applied in this paper. Again, the four criteria CC, SI, NSE, and WI revealed a significant improvement in the accuracy of the SVR-FOA method compared to the SVR method but the GEP method still had the highest prediction precision. In sum, the findings of this paper suggested that the GEP method (with values of 0.874, 0.598, 0.601 and 0.807 for the four criteria CC, SI, NSE, and WI, respectively) offered an alternative to current methods of predicting pavement RSL.

Author Contributions: conceptualization, N.K.; methodology, N.K.; software, A.D., A.M., N.N., and S.S.; validation, A.D., A.M., N.N., and S.S; formal analysis, A.K.; investigation, N.K., A.D., A.M., N.N., S.S, and D.M; resources, X.X.; data curation, N.K., A.D., A.M., N.N., S.S, and D.M; writing—original draft preparation, N.K.; writing—review and editing, A.M.; visualization, N.K; supervision, S.S.; project administration, A.D.; funding acquisition, A.D.

Acknowledgments: The EU regional fund is acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Butt, A.A.; Shahin, M.Y.; Feighan, K.J.; Carpenter, S.H. *Pavement performance prediction model using the Markov process*; 1987.
2. Soncim, S.P.; de Oliveira, I.C.S.; Santos, F.B. Development of fuzzy models for asphalt pavement performance. *Acta Scientiarum. Technology* **2019**, *41*, e35626.
3. Arhin, S.A.; Williams, L.N.; Ribbiso, A.; Anderson, M.F. Predicting pavement condition index using international roughness index in a dense urban area. *Journal of Civil Engineering Research* **2015**, *5*, 10-17.
4. Mfinanga, D.A. Sampling procedure for pavement condition evaluation of local collectors and access roads. *Tanzania Journal of Engineering and Technology* **2007**, *1*, 99-109.
5. Wahyudi, W.; Sandra, P.A.; Mulyono, A.T. Analysis of Pavement Condition Index (PCI) and Solution Alternative of Pavement Damage Handling Due to Freight Transportation Overloading (Case Study: National Road Section West Sumatra Border–Jambi City). In Proceedings of Proceedings of Eastern Asia Society for Transportation Studies.
6. Marcelino, P.; Lurdes Antunes, M.d.; Fortunato, E. Comprehensive performance indicators for road pavement condition assessment. *Structure and Infrastructure Engineering* **2018**, *14*, 1433-1445.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 October 2019 doi:10.20944/preprints201910.0238.v1

Peer-reviewed version available at *Mathematics* **2019**, *7*, 1198; doi:10.3390/math7121198

7. Bryce, J.; Boadi, R.; Groeger, J. Relating Pavement Condition Index and Present Serviceability Rating for Asphalt-Surfaced Pavements. *Transportation Research Record* **2019**, 0361198119833671.
8. Karballaeezadeh, N.; Mohammadzadeh S, D.; Shamshirband, S.; Hajikhodaverdikhan, P.; Mosavi, A.; Chau, K.-w. Prediction of remaining service life of pavement using an optimized support vector machine (case study of Semnan–Firuzkuh road). *Engineering Applications of Computational Fluid Mechanics* **2019**, *13*, 188-198.
9. Luo, Z.; Chou, E.Y. Pavement condition prediction using clusterwise regression. *Transportation research record* **2006**, *1974*, 70-77.
10. Huang, Y.H. *Pavement analysis and design*; Prentice-Hall: Upper Saddle River, New Jersey, 2004.
11. Das, A.; Pandey, B. Mechanistic-empirical design of bituminous roads: an Indian perspective. *Journal of transportation engineering* **1999**, *125*, 463-471.
12. Hossain, M.; Wu, Z. *Estimation of asphalt pavement life*; 2002.
13. Park, H.M.; Kim, Y.R. Prediction of remaining life of asphalt pavement using fwd multiload level deflections. In Proceedings of Transportation Research Board, 2003 Annual Meeting Proceedings, Transportation Research Board, Washington, DC.
14. Smith, R.E. Structuring a microcomputer based pavement management system for local agencies. *Dissertation Abstracts International Part B: Science and Engineering[DISS. ABST. INT. PT. B- SCI. & ENG.]* **1987**, *47*.
15. Al-Suleiman, T.I.; Shiyab, A.M. Prediction of pavement remaining service life using roughness data—case study in Dubai. *International Journal of Pavement Engineering* **2003**, *4*, 121-129.
16. Setyawan, A.; Nainggolan, J.; Budiarto, A. Predicting the remaining service life of road using pavement condition index. *Procedia Engineering* **2015**, *125*, 417-423.
17. Saleh, M. A mechanistic empirical approach for the evaluation of the structural capacity and remaining service life of flexible pavements at the network level. *Canadian Journal of Civil Engineering* **2016**, *43*, 749-758.
18. Shahin, M.Y. *Pavement management for airports, roads, and parking lots*; Springer New York: 2005; Vol. 501.
19. ASTM, D. 6433-07 Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys. *ASTM International, West Conshohocken, Pennsylvania* **2009**.
20. Abdel-Khalek, A.; Elseifi, M.A.; Codjoe, J.; Fillastre, C. Estimating Service Life of In Situ Flexible Pavements in Louisiana Using Pavement Management System Data. In Proceedings of Construction Research Congress 2018American Society of Civil Engineers.
21. Kumar, R.; de Oliveira, M.; Lorrany, J.; Schultz, A.; Marasteanu, M. Remaining Service Life Asset Measure, Phase 1. **2018**.
22. ASTM. Standard test method for deflections with a falling-weight-type impulse load device. 2009.
23. ASTM. Standard guide for using the surface ground penetrating radar method for subsurface investigation. **2011**.
24. Cao, Y.; Dai, S.; Labuz, J.; Pantelis, J. Implementation of ground penetrating radar. **2007**.
25. Designation, A.S. D4748-87 S tandard Test Method for Determining the Thickness of Bound Pavement Layers Using Short-Pulse Radar. *American Society for Testing and Materials, Philadelphia, Pennsylvania* **1987**.

26. Ferreira, C. Gene expression programming in problem solving. In *Soft computing and industry*, Springer: 2002; pp. 635-653.
27. Ferreira, C. Gene expression programming: a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027* **2001**.
28. Suykens, J.A.; Vandewalle, J. Least squares support vector machine classifiers. *Neural processing letters* **1999**, *9*, 293-300.
29. Cherkassky, V.; Ma, Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural networks* **2004**, *17*, 113-126.
30. Pan, W.-T. A new fruit fly optimization algorithm: taking the financial distress model as an example. *Knowledge-Based Systems* **2012**, *26*, 69-74.
31. Ebtehaj, I.; Bonakdari, H.; Zaji, A.H.; Azimi, H.; Sharifi, A. Gene expression programming to predict the discharge coefficient in rectangular side weirs. *Applied Soft Computing* **2015**, *35*, 618-628.
32. Yassin, M.A.; Alazba, A.; Mattar, M.A. Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate. *Agricultural Water Management* **2016**, *163*, 110-124.
33. Lopes, H.S.; Weinert, W.R. EGIPSYS: an enhanced gene expression programming approach for symbolic regression problems. *International Journal of Applied Mathematics and Computer Science* **2004**, *14*, 375-384.
34. Faradonbeh, R.S.; Salimi, A.; Monjezi, M.; Ebrahimabadi, A.; Moormann, C. Roadheader performance prediction using genetic programming (GP) and gene expression programming (GEP) techniques. *Environmental earth sciences* **2017**, *76*, 584.
35. Ferreira, C. *Gene expression programming: mathematical modeling by an artificial intelligence*; Springer: 2006; Vol. 21.
36. Dibike, Y.B.; Velickov, S.; Solomatine, D. Support vector machines: Review and applications in civil engineering. In Proceedings of Proceedings of the 2nd Joint Workshop on Application of AI in Civil Engineering; pp. 215-218.
37. Cortes, C.; Vapnik, V. Support-vector networks. *Machine learning* **1995**, *20*, 273-297.
38. Cheng, K.; Lu, Z.; Wei, Y.; Shi, Y.; Zhou, Y. Mixed kernel function support vector regression for global sensitivity analysis. *Mechanical Systems and Signal Processing* **2017**, *96*, 201-214.
39. Cherkassky, V.; Mulier, F.M. *Learning from data: concepts, theory, and methods*; John Wiley & Sons: 2007.
40. Vapnik, V. *The nature of statistical learning theory*; Springer science & business media: 2013.
41. Xiao, C.; Hao, K.; Ding, Y. An improved fruit fly optimization algorithm inspired from cell communication mechanism. *Mathematical Problems in Engineering* **2015**, *2015*.
42. Chu, D.; He, Q.; Mao, X. Rolling bearing fault diagnosis by a novel fruit fly optimization algorithm optimized support vector machine. *Journal of Vibroengineering* **2016**, *18*.
43. Yu, Y.; Li, Y.; Li, J.; Gu, X. Self-adaptive step fruit fly algorithm optimized support vector regression model for dynamic response prediction of magnetorheological elastomer base isolator. *Neurocomputing* **2016**, *211*, 41-52.
44. Cong, Y.; Wang, J.; Li, X. Traffic flow forecasting by a least squares support vector machine with a fruit fly optimization algorithm. *Procedia Engineering* **2016**, *137*, 59-68.

45. Mehr, A.D. An improved gene expression programming model for streamflow forecasting in intermittent streams. *Journal of hydrology* **2018**, *563*, 669-678.
46. Willmott, C.J.; Robeson, S.M.; Matsuura, K. A refined index of model performance. *International Journal of Climatology* **2012**, *32*, 2088-2094.
47. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres* **2001**, *106*, 7183-7192.