A Revisited Convex Hull-Based Fuzzy Linear Regression Model for Dynamic Fuzzy Healthcare Data

Azizul Azhar Ramli[1], Shahreen Kasim[1], Mohd. Farhan Md. Fudzee[1], Nazri Mohd. Nawi[1], Hairulnizam Mahdin[1] and Junzo Watada[2]

[1] Faculty of Computer Science and Information Technology
Universiti Tun Husseion Onn Malaysia
Parit Raja, 86400 Batu Pahat, Johor Darul Takzim, Malaysia
[2]Universiti Teknologi PETRONAS,
Department of Computer and Information Sciences
32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia
azizulr@uthm.edu.my

***Abstract*— Healthcare data analysis is widely used in cancer classification and disease prediction. Hence, fuzzy linear regression integration into dynamic analysis can provide better decision making process in healthcare industry especially dealing with dynamic data analysis. Healthcare officers and related researchers require efficient regression tools to produce precise inference as an aid to save human life. However, the key problem in this circumstance is related to computational complexity and processing time. Both of these parameters drastically increase beside an increment of data size in dynamic databases. With regard to the aforementioned problem, the main objective is to improve the implementation of fuzzy linear regression for fuzzy data by addressing and mitigating some limitations of the existing methods through a convex hull approach. More specifically, we look at the realization of an incremental algorithm called Beneath-Beyond algorithm. This algorithm provides a useful vehicle to reduce the computing time and computational complexity as well. Furthermore, a real fuzzy healthcare data set which derive from healthcare industry will be selected as the main source of data sets. Additionally, there are two major procedures or components namely a formulation of a product of fuzzy number optimization and the use of the convex hull technique to the obtained locus points in hyper-rectangles polygon and each of them have their own distinctive activities. As a research output, the combination of this mathematical geometry algorithm and fuzzy linear regression analysis will produce an optimized algorithm called convex hull-based fuzzy linear regression model deliberately for dynamic fuzzy healthcare data. The proposed algorithm may help to produce a rapid decision making especially for critical area such as healthcare industry.**

***Index Terms*—Fuzzy Regression; Convex Hull; Fuzzy Data; Healthcare Data.**

I. INTRODUCTION

In real-world optimization problems, which we routinely encounter in engineering, management, economy, medicine, psychology, biotechnology and other disciplines, it is quite common to handle large amounts of various types of data ([1], [2], [3], [4], [5]). In particular, dynamic data analysis becomes more important given the growing demand to support efficient managerial practices, which call for on-line (dynamic) timely results of data analysis [6]. The main intent in this setting is to reduce computing overhead in supplying the results in dynamic.

Soft computing techniques such as fuzzy logic have become an important alternative to realize effective data analysis, especially for decision making process ([4], [7], [8]). On the other hand, regression analysis is a generic statistical tool to explore and describe dependencies among variables. When forming a synergy between these two essential modeling methodologies, we arrive at fuzzy regression, which takes full advantage of the strengths of the contributing technologies; cf. ([9]). Linear programming (LP) was used to determine the location of the centers and the spreads of the fuzzy coefficients (fuzzy numbers) of the fuzzy regression hyperplane by minimizing an objective function which takes into consideration the total spread of the outputs of the model treated as fuzzy numbers ([2], [10], [11], [12], [13]).

Let recall that convex hull is a fundamental concept present in many applications encountered in pattern recognition, image processing, and statistics. Convex hull is defined as the smallest convex polygon located in a multidimensional data space which contains all point set (vertices) [14]. In other words, convex hull polygon is corresponds to the intuitive conception of a “boundary” of a set of points and as such can be used to approximate a shape of any available objects of complex geometry.

In real-world problems such as those present in economics, bio-computing or engineering, we are concerned with the massive data sets of high dimensionality to analyze in a limited time or even in real-time or dynamic situation [15]. In addition to experimental evidence of numeric nature, some data can be described in a linguistic term, which immediately invokes the concept of fuzzy sets ([16], [17]).

Related to the methods used for dynamic application where statistical regression has been shown to be highly relevant, we can also observe some shortcomings, which arise when dealing with several characteristics of the data or making some simplifying yet not necessarily fully legitimate assumptions, cf. [18]; Difficulties with a thorough verification of assumptions about data distributions; Vagueness present in the relationships between input and output variables; Ambiguity of events or non-Boolean degrees to which they occur, and Inaccuracy and distortion introduced by linearization.

In all these scenarios, the use of the "standard" regression might raise some hesitation. Here the use of fuzzy regression arises as a viable alternative. There are two general ways supporting the development process of fuzzy regression ([18]) including, models where the relationships among the variables is inherently fuzzy, and models where the input (independent) variables themselves are fuzzy. There are some arguments that are worth highlighting with regard to fuzzy regression. As an example, there's no proper interpretation of fuzzy regression interval [9]. Besides that expert could still provide an interval of possible values but also indicate the probability of occurrence of each one of them. Obviously, this approach would require more information ([20], [21]).

Given the explanation presented above, a convex hull approach can help implement dynamic fuzzy regression model by serving as an alternative optimization vehicle. The main objective of this research is to enhance the implementation procedure of dynamic fuzzy regression model with the utilization of the designated convex hull approach. In addition, the adaptation of selected algorithm of this approach, specifically the Beneath-Beyond algorithm, helps address the limitations of the generic implementation of fuzzy regression when applied to the analysis of dynamic data.

The paper is organized as follows. Section 2 serves as a related previous works, which includes brief review of the fundamentals of fuzzy regression. Next, Section 3 presents basic Beneath Beyond algorithm and related terms. Hybrid optimization for real-time fuzzy linear regression model realized with the use of the convex hull approach clearly explain in Section 4. Section 4 is devoted to empirical experiments. Finally, Section 5 presents concluding remarks.

II. FUZZY REGRESSION MODEL

Essentially, dynamic data analysis refers to studies where data revisions (updates, successive data accumulation) or data release timing is important to a significant degree. The most important properties for dynamic data analysis are dynamic analysis and reporting, based on data entered into a system in a short interval before the actual time of the usage of the results [22]. The following sub sections were highlight some theories and techniques related this study.

Regression models are statistical constructs, which describe relations among variables. They explain a dependent variable by making use of some independent variables [10]. The variables used to explain the target variable(s), are called explanatory attributes ([22], [12], [23]).

Since Zadeh proposed fuzzy sets, the concept of fuzziness has received more attention and fuzzy data analysis has become increasingly important. In order to consider the fuzziness in regression analysis, the earlier proposal related to the study of fuzzy linear regression model which considering two factors, namely the degree of the fitting and the vagueness of the model. In addition, fuzzy linear regression is suitable for problems in which human experts use subjective judgment or some experience [24]. Furthermore, trying to build a latent structure model from such data as well as possible, any developed regression model should be based on availability of an adequate amount of raw material or samples of data which represent the actual scenario, in order to minimize the erroneous problem of the obtained result [25]. When independent variables are given, then the estimate of the output is expressed as follows:

$$Y_i = h(x_i) + \varepsilon_i, \qquad i = 1,...,n \tag{1}$$

where $h(\cdot)$ represents some function and $\varepsilon_i s$ an independent random variable with zero mean and some variance, $i-1,\dots,n$. A major difference between fuzzy linear regression and statistical regression is that the deviations (differences) between the collected data and the estimated values (coming from the model) are assumed to be associated with the vagueness (which is captured through fuzzy linear regression) rather than with the randomness as it is done in conventional regression techniques ([26], [27], [14]). In other words, fuzzy linear regression seems more appealing for estimating the relationship between the dependent variable and independent variables [28].

Furthermore, a fuzzy set-based approach can be used to gain insights into a complex system for which analytical model may not exist [18] or when dealing with human perception processes [29]. Moreover, studies dealing with fuzzy linear regression can be broadly classified into two approaches which are includes LP based approach and fuzzy least squared approach ([17], [9]). Therefore, with the incorporation of fuzzy sets, an enhancement of regression models comes in the form of a fuzzy linear regression or possibilistic regression was introduced which to reflect the non-numeric nature of relationship between the dependent and independent variables. The upper and lower regression boundaries are used in the possibilistic regression to reflect the possibilistic distribution of the output values. On the other hand, since the measure of best fit by residuals under fuzzy consideration is not presented in Tanaka's approach, the fuzzy least-squares approach been proposed; which is a fuzzy extension of the ordinary least squares based on a new defined distance on the space of fuzzy numbers [17].

Based on the several previous studies which deal with a generic form of triangular numbers, fuzzy linear regression quantifies the deviations existing between the data and the linear model [9]. Computationally, the estimation of the fuzzy parameters of the regression gives rise to in a certain problem of LP.

Let us recall that the fuzzy linear regression comes in the form

$$\tilde{Y} = \tilde{A}_0 X_0 + \tilde{A}_1 X_1 + ... + \tilde{A}_K X_K = \tilde{\mathbf{A}}\mathbf{X}^t \tag{2}$$

where $\mathbf{X} = [X_0, X_1,..., X_K]$ represents a vector of independent variables while $\tilde{\mathbf{A}} = [\tilde{A}_0, \tilde{A}_1,...\tilde{A}_K]$ is a vector of fuzzy coefficients described in the form of symmetric triangular fuzzy numbers and denoted by $\tilde{A}_j = (\alpha_j, c_j)$ The corresponding triangular membership function is described as follows

$$u_{\tilde{A}_j}(a_j) = \begin{cases} 1 - \dfrac{|\alpha_j - a_j|}{c_j}, & : c_j \neq 0, \alpha_j - c_j \leq a_j \leq \alpha_j \\ 1, & : c_j = 0, \alpha_j = a_j \\ 0, & otherwise \end{cases} \tag{3}$$

$$j = 0, \dots, K$$

where α_j and c_j are the central (modal) and spread values of the triangular fuzzy number, respectively.

Given the notation used above, (3) can be rewritten as

$$\tilde{Y}_i = (\alpha_0, c_0) + (\alpha_1, c_1)X_1 + (\alpha_2, c_2)X_2 + \ldots + (\alpha_K, c_K)X_K = (\boldsymbol{\alpha}, \mathbf{c})\mathbf{X}^t \qquad (4)$$

where $\mathbf{c} = (c_0, c_1, \ldots, c_K), \boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_K)$.

In the above model, we assume that the numeric input and output data are available, while the relation between the input and output data is defined by some fuzzy function whose parameters are governed by the corresponding membership functions ([17], [30]).

Making use of the extension principle, we derive the membership function of fuzzy number $\tilde{Y}_i$ to be in the following form:

$$u(Y_i) = \begin{cases} 1 - \dfrac{\left|Y_i - \boldsymbol{\alpha}\mathbf{X}_i^t\right|}{\mathbf{c}\left|\mathbf{X}_i^t\right|}, & : \mathbf{c}\left|\mathbf{X}_i^t\right| \neq 0, \boldsymbol{\alpha}\mathbf{X}_i^t - \mathbf{c}\left|\mathbf{X}_i^t\right| \leq Y_i \leq \boldsymbol{\alpha}\mathbf{X}_i^t \\ 1, & : \mathbf{c}\left|\mathbf{X}_i^t\right| = 0, Y_i = \boldsymbol{\alpha}\mathbf{X}_i^t \\ 0, & : otherwise \end{cases} \qquad (5)$$

where $\mathbf{c}\,|\mathbf{X}_i^t| = \sum_{j=0}^{K} c_j\,|x_{ij}|$, $x_{i0} = 1 : (\forall i = 1, \ldots, n)$.

In order to develop the fuzzy linear regression with the minimal level of fuzziness of the output, the following objective function is proposed to minimize the total spread of the resulting fuzzy number $\tilde{Y}_i$

$$\underset{\boldsymbol{\alpha},\mathbf{c}}{\text{MIN}} \sum_{i=1}^{n} \mathbf{c}\,|\mathbf{X}_i^t| = \underset{\boldsymbol{\alpha},\mathbf{c}}{\text{MIN}} \sum_{i=1}^{n}\sum_{j=0}^{K} c_j\,|x_{ij}| \qquad (6)$$

along with the constraints that each observation Y_i has at least h degree of membership in $\tilde{Y}_i$, that is, $u(y_i) \geq h \; (i = 1, \ldots, n)$. This requirement can be expressed in the following equivalent form

$$1 - \frac{|Y_i - \boldsymbol{\alpha}\mathbf{X}_i^t|}{\mathbf{c}\,|\mathbf{X}_i^t|} \geq h, \qquad \forall i = 1, \ldots, n \qquad (7)$$

The above formulation of the membership functions leads us to the following task of LP:

$$\underset{\boldsymbol{\alpha},\mathbf{c}}{\text{MIN}} \sum_{i=1}^{n}\sum_{j=0}^{K} c_j\,|x_{ij}| \qquad (8)$$

$$\text{subject to } \sum_{j=0}^{K} \alpha_j x_{ij} + (1-h)\sum_{j=0}^{K} c_j\,|x_{ij}| \geq Y_i,$$
$$\sum_{j=0}^{K} \alpha_j x_{ij} - (1-h)\sum_{j=0}^{K} c_j\,|x_{ij}| \leq Y_i, \qquad (9)$$
$$c_j \geq 0, \quad \alpha_j \in \Re, \quad x_{i0} = 1 \quad (0 \leq h < 1; \quad \forall i = 1, \ldots, n).$$

In other words, the fuzzy linear regression model is built in such a way so that it includes all potential samples of data points under the analysis process; refer to cf. [31]. In addition, this problem can be re-formulated in the language of LP [32].

When fuzzy output data $(\tilde{y}_i, \mathbf{x}_i), \tilde{y}_i = (y_i, d_i)$, are given with numeric input data such $\mathbf{x}_i = [x_{i0}, x_{i1}, x_{i2}, \ldots, x_{iK}]$ $(i = 1, \ldots, n)$ with general convex membership function $L(\cdot)$ instead of a triangular membership function of fuzzy coefficients $(\boldsymbol{\alpha}, \mathbf{c})$, the development of the fuzzy linear regression model gives rise to the following LP problem:

$$\underset{\boldsymbol{\alpha},\mathbf{c}}{\text{MIN}} \sum_{i=1}^{n} \mathbf{c}\,|\mathbf{x}_i^t| \qquad (10)$$

$$\text{subject to } \; y_i + L^{-1}(h)d_i \leq \boldsymbol{\alpha}\mathbf{x}_i^t + L^{-1}(h)\mathbf{c}\,|\mathbf{x}_i^t|$$
$$y_i - L^{-1}(h)d_i \geq \boldsymbol{\alpha}\mathbf{x}_i^t - L^{-1}(h)\mathbf{c}\,|\mathbf{x}_i^t| \qquad (11)$$
$$\mathbf{c} \geq 0 \quad (i = 1, \ldots, n)$$

where, $L(u)$ is a shape of the membership function of the parameters of the fuzzy linear regression model and h is a grade expressing to which extent a given datum is "captured" by the model.

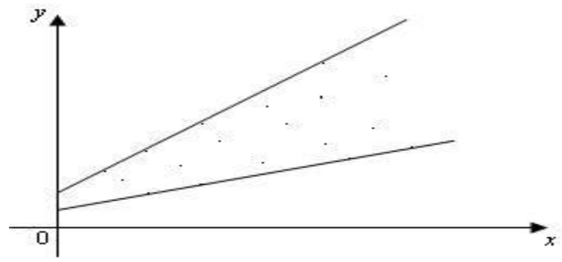

Figure 1: Fuzzy linear regression model (case of $d_j = 0$)

In essence, the LP problem formulated in this manner produces a solution for obtaining the fuzzy linear regression model to represent relationship of analyzed data ([33], [15], [20]). Referring to (4), Figure 1 illustrates the relation between the output of the model and the data. In a nutshell, as noted earlier, the model of the fuzzy linear regression should "contain" or "cover" all samples. Through the membership grades, we quantify the degrees of inclusion of the data. Furthermore in such formulation, it is possible of the fuzzy linear regression model to treat non-fuzzy data with zero width by setting the value of d to 0 as described in the above formula; refer to ([34], [35]) for details.

III. Beneath-Beyond Algorithm - A Brief Review

This algorithm incrementally builds up the convex hull by keeping track of the current convex hull P_i using an incidence graph. In order to add a new point P to the convex hull, the incremental algorithm identifies the facets below the point. These are the facets visible from the point. The boundary of the visible facets builds the set of horizontal ridges for the point. If there are no visible facets from point P, the point inside the convex hull can be discarded. Otherwise, the algorithm constructs new facets of the convex hull from horizon ridges and the processed point P and does not explicitly build the convex hulls of lower dimensional faces.

A new facet of the convex hull is a facet with point P as its apex and a horizon ridge as its base. The cone of point P is the set of all new facets. Any point chosen to be added to the current convex hull must not be in the same affine space as any of the facets of the current convex hull. For instance, if the current convex hull is a tetrahedron, a new point to be added will not be coplanar with any of the faces of the tetrahedron [39].

The Beneath-Beyond algorithm comprises the following steps:

Step 1 : Select and sort outsider points along one direction, say x_1. Let $s = \{p_0, p_1, ..., p_{n-1}\}$ be input points after sorting. Process the points in an increasing order.

Step 2 : Take the first n points, which define a facet as the initial hull.

Step 3 : Let P_i be the point to be added to the hull at the *i*th stage. Let $p_i = conv(p_0, p_1, ..., p_{i-1})$ be the convex hull polytope built so far. This step includes two possible hull updates:

a. A pyramidal update is done when $p_i \notin aff(p_0, p_1, ..., p_{i-1})$ - when P_i is not on the hyperplane defined by the current hull. A pyramidal update consists of adding a new node representing P_i to the incidence graph and connecting this node to all existing hull vertices by new edges.

A non-pyramidal update is done when the above condition is not met, i.e., P_i is in the affine subspace defined by the current convex hull. In this case, facets that are visible from P_i are removed and new facets are created.

In addition, processing a point through Quickhull (a variation of Clarkson and Shor's algorithm) and the randomized incremental algorithm comes as an implementation of the following Beneath-Beyond theorem (Grunbaum's Beneath-Beyond theorem) [39].

Figure 2 shows an example of an implementation of the Beneath-Beyond algorithm.

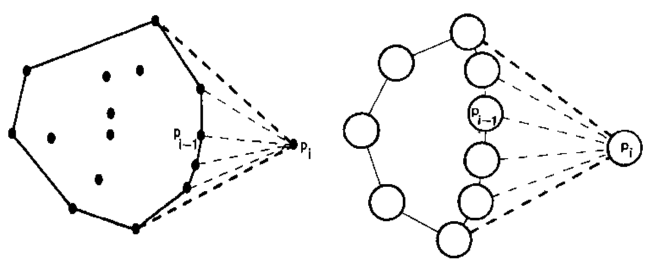

Figure 2: Illustration of the implementation of the Beneath-Beyond algorithm

The rationale behind the first condition is straightforward. The second condition describes a face of the cone that is to be created if p is at least above one facet. The ridge with one incident facet below and the other one above p is the equivalent of the edge between a visible and an invisible face for the incremental algorithm discussed above.

The efficient determination of visible facets for a given point is crucial to the runtime behavior of any incremental algorithm. As visible facets are neighbors; once one visible facet has been found, the others can be easily detected. The main idea behind Quickhull is to maintain a set for each facet in which points are stored that are outside the respective facet. A point is outside a facet *iff* the signed distance between the facet and the point is positive. Each unprocessed point or newly inserted point that appears in the particular field belongs to exactly one outside set. It can be shown that if a point is on the outside of multiple facets, it does not matter to which of the corresponding outside sets the point belongs. These outside sets represent a partitioning of the set of unprocessed points.

IV. Hybrid Optimization with Fuzzy Linear Regression

In order to elucidate the procedure of the proposed approach, we refer to the overall flow of processing as visualized in Figure 3.

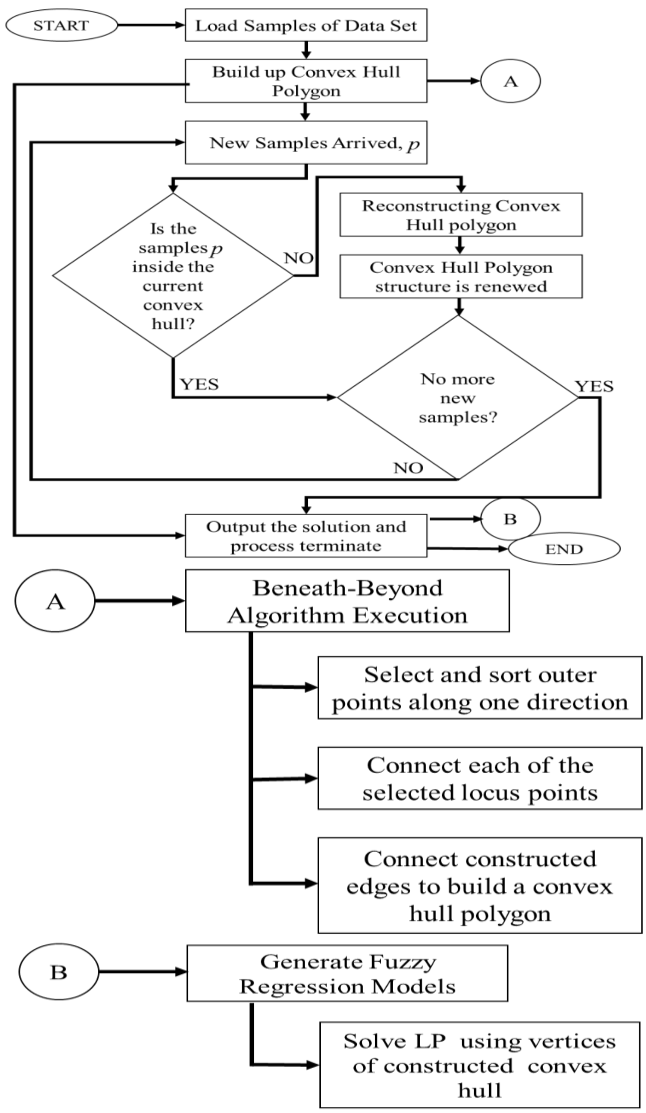

Figure 3: An overall flow of processing realized in the proposed approach and related processes for Beneath-Beyond implementation and Fuzzy Regression Models generation

Given the discussion presented so far, let us outline a flow of computing of the dynamic fuzzy regression model. Two major processes should be considered, see following below.

Process_1	For given points of sample $p_1, p_2, p_3 ..., p_n$, create a convex hull and list convexes C^{P_n}. Construct a fuzzy regression model Y^{U_n} and Y^{L_n} using the convexes $C^{P_{(n)}}$ by LP.
Process_2	When we have new points of samples $p_{n+1}, ..., p_{n+q}$, if all q samples are included between Y^{L_n} and Y^{U_n}, then create $C^{P_{(n+w)}}$ which adding w samples to $C^{P_{(n)}}$. If any samples are not included in between Y^{L_n} and Y^{U_n}, then create $C^{P_{(n+w)}}$ by adding w samples to C^{P_n} and create $Y^{L_{n+w}}$ and $Y^{U_{n+w}}$ using vertices $C^{P_{(n)}}$ and q new samples.

On the other hand, there are several situations which dealing with fuzzy data. This type of data is considered as a problem of n samples with one-output and K -input interval values. It looks more difficult to solve because it consists of $n \times K$ products between the fuzzy coefficients and confidence intervals. Additionally, the most appropriate method while dealing with this situation is a vertices method [7].

This problem can be simplified into that of n samples with one-output interval value and 2^K vertices in K dimensions. Hence, we can solve this problem of $n \times 2^K$ samples with K attributes of input by the conventional method. The numbers of products which obtain from $n \times 2^K$ are becomes as potential vertices for convex hull polygon. We use this figure as an input of our proposed fuzzy dynamic regression model.

In the suggested approach, the selection of particular vertices for convex hull constructions concerns some related points, which were pointed as the particular rectangles that represent the loci of the membership function support of fuzzy data, see Figure 4.

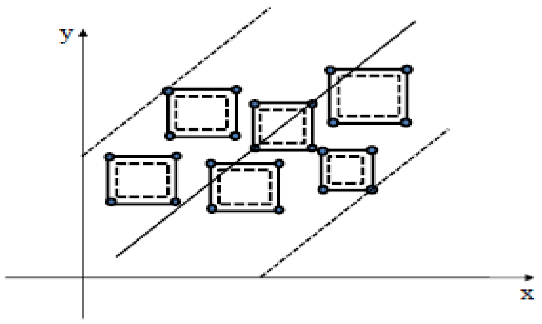

Figure 4: Example of rectangles that are loci of the membership functions

The vertex points of selected loci will become possible convex vertices and these points will connect to each other to create convex hull edges or boundaries. Consequently, the connected edges will construct a convex hull for selected samples of fuzzy data and basically, whole of analyzed data will fall inside constructed convex hull. Figure 5 illustrates the essence of the process.

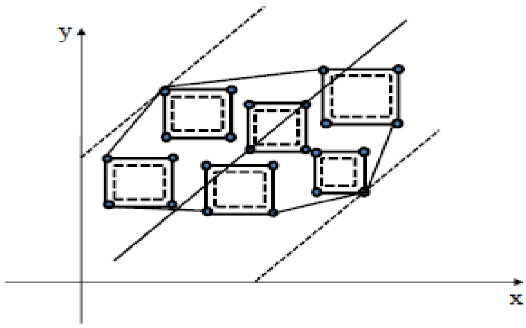

Figure 5: Convex hull polygon construction process

Therefore, based on the developed convex, the determination of the possibilistic regression to represent the distribution of sample data is easier because the slight changes (either decreasing or increasing) of the vertex points influence the model and these (that is the vertex points of only selected loci) must be considered for further LP formulation. In other words, we just consider selected vertex points in the analysis, which previously resulted in building up the convex hull structure.

A. An Efficient Formulation of the Fuzzy Linear Regression Model for Fuzzy Data

Basically, the proposed approach can be implemented by considering the outside of the locus hyper-rectangle points, which support a fuzzy set where their membership values become zero. The selected points potentially become vertices of a convex. Due to this selection step, the number of vertices considered for building a convex will decrease quite drastically. Therefore, the computing phase including LP and the convex hull approach has to be carefully revisited as far as the underlying complexity is concerned. Moreover, this feature may also reduce the time required to generate models of fuzzy linear regression. The incremental algorithm referred to as a Beneath-Beyond method provides the proposed approach with significant advantages. Moreover, this algorithm does not necessitate more additional computational time for selecting potential locus points, which become convex vertices as well as constructing a convex hull polygon.

Based on the original method of the determination of the convex hull, the process has to be realized using all given samples. A new method is developed to build a convex hull using the given calculated fuzzy data. The proposed algorithm consists of the following steps:

Step_1 : Form a set of fuzzy data.

Step_2 : Formulate product of fuzzy number optimization process for obtaining the total product of fuzzy numbers, $n \times 2^k$.

Step_3 : Determine the outsider vertex points given as the loci of the membership functions' graphs.

Step_4 : Perform the Beneath-Beyond algorithm to formulate the convex hull, H_0, using one of the selected vertex points that were chosen for building the convex hull.

Substep_4.1 : Connect each of the selected potential vertex points to construct convex edges.

Substep_4.2 : Connect constructed edges to create boundaries of a convex hull, H_β .

Substep_4.3 : Omit points that are not included in the convex hull, H_β .

Step_5 : Newly arrived data added to the initial analyzed sample of data

Step_6 : Newly added samples are large or processing time is adequate?
If Yes move to Step_2 otherwise proceed to Step_7

Step_7 : Output the solution and terminate the process.

Basically, there are two major optimization component called a fuzzy number and convex hull utilization optimization process, respectively. Iteration might occur while new samples of data arrive and are added to the remaining samples. There is a selective decision in dealing this condition; if newly added samples are huge or processing time is adequate then several previous steps should be restarted again, otherwise final process which concern with the use of convex hull vertices as a part of LP constraint portion in fuzzy linear regression formulation. Figure 6 offers a general view at all optimization activities realized here.
On the other hand, the choice of the Beneath-Beyond algorithm as a convex hull approach is legitimate here considering that no extra computing is required for the construction of the facet structure. This may reduce the computing time required to produce the best solution for an equivalent problem. In addition, the main concern that distinguishes the different variants of this algorithm is the way to search for the visible facets. Put it differently: we can find a visible facet among those added in the previous stage; therefore, we may simply search through all the latter facets until a visible one has been found. Then we examine adjacent facets and repeat the process on those that are visible.

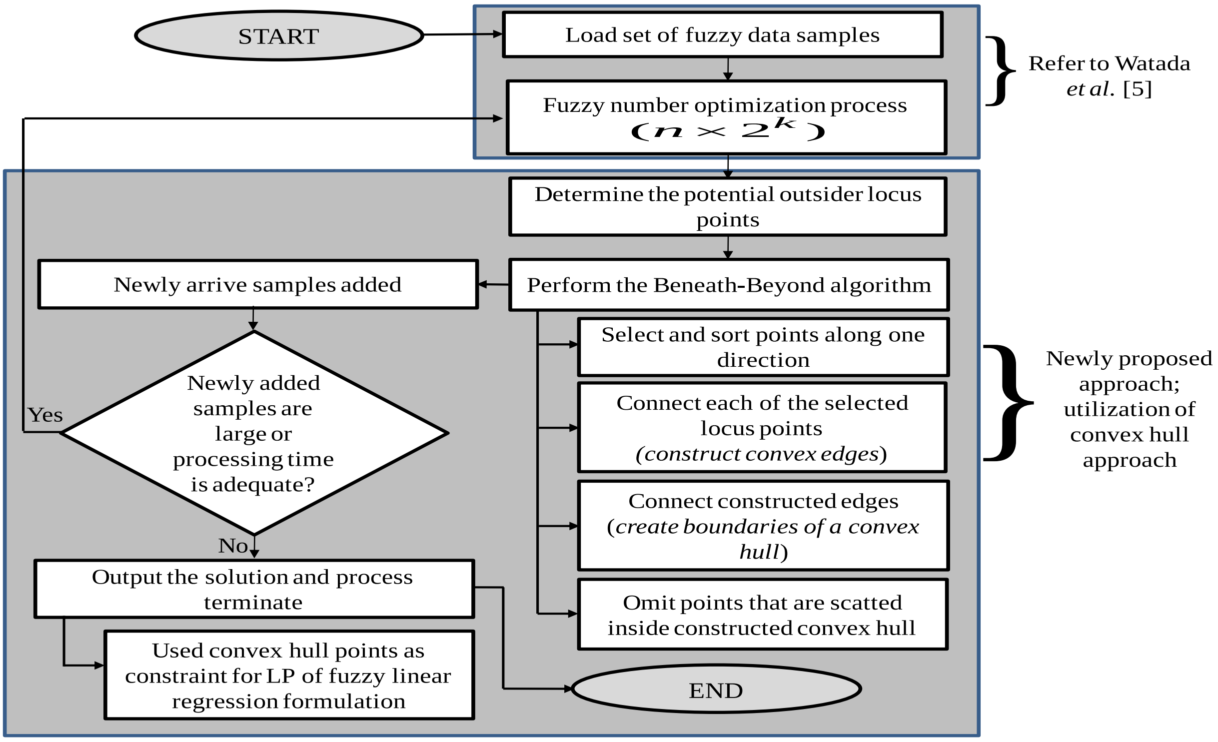

Figure 6: The overall optimization process

V. NUMERICAL EXAMPLE

Now let us examine another numerical example that concerns a larger number of samples. This sample concern about identification for the common factors that contribute to cardiovascular disease by following its development over a long period of time in a large group of participants. Here there are 100 data points, $n = 100$ and 12 inputs attribute, $(K = 12)$ of fuzzy numbers. Given the dimensionality of the problem, the possible number of all products of fuzzy numbers as estimated earlier is $100 * 2^{12} = 409{,}600$. This huge number of potential vertices may not be efficiently utilized when running fuzzy linear regression. As discussed in the previous example, the computed product of fuzzy numbers represents the loci of membership functions and its elements or locus points can be potentially selected to become the vertices of a convex polygon.

In order to minimize the computational complexity, we consider further adaptation of the convex hull approach. As shown in Examples 1, the implementation of this method is realized by selecting the outside plots of the loci that represent the graphs of fuzzy data. Selected points may become convex vertices and each of them will connect to each other. This will produce edges and finally these connected edges will result in a convex hull polygon. Therefore, this selection process will drastically reduce the number of vertex points under consideration. Following this strategy, as illustrated in Figure 6, the large number of fuzzy data that represent potential vertices is drastically reduced to 274 points $(P = 274)$. In other words, the load rate becomes reduced at the rate of $\Pr_P \approx 0.000669$ for which running the LP for this constructed convex hull becomes far more effective as well as leads to a significant time reduction. The obtained coefficients of the regression model (for $h = 0.50$) are included in Table 1.

Table 1.
Fuzzy Parameters for the Construction of Convex Hull, $\tilde{a}_i$ ($h = 0.50$)

Fuzzy Parameter	center m_i	spread c_i
$\tilde{a}_0$	0.0507	0.0024
$\tilde{a}_1$	0.0576	0.0027
$\tilde{a}_2$	0.3025	0.0268
$\tilde{a}_3$	-0.2010	0.0000
$\tilde{a}_4$	0.1357	0.0091
$\tilde{a}_5$	0.3881	0.0289
$\tilde{a}_6$	0.5350	0.0412
$\tilde{a}_7$	-0.0408	0.0000
$\tilde{a}_8$	-0.1660	0.0000
$\tilde{a}_9$	-0.1911	0.0000
$\tilde{a}_{10}$	0.0645	0.0031
$\tilde{a}_{11}$	0.1070	0.0068
$\tilde{a}_{12}$	-0.0484	0.0000

To present a real-time data analysis in which changes in sample size occurs over time, we add a number of samples, $n = 100$ to the initially analyzed samples. The total number of samples, which have same properties is equal to 200 and the total product of fuzzy number is $n \times 2^K = 200 \times 2^{12} = 819,200$. The obtained result shows that $P = 179$ locus points, which were initially selected as convex hull vertices are reused and $P = 268$ new locus points were added. Therefore, in total, $P = 447$ locus points were transformed to vertices to reconstruct a convex hull polygon. Moreover, the constraint portion of fuzzy linear regression formulation was utilizing these points for producing appropriate models. The load rate for this group of samples is decreased to $\Pr_P \approx 0.000546$. The lists of regression coefficients are present in Table 2.

Table 2.
Fuzzy Parameters for the Reconstruction of Convex Hull, $\tilde{a}_i$ ($h = 0.50$)

Fuzzy Parameter	Center m_i	Spread c_i
$\tilde{a}_0$	0.0319	0.0015
$\tilde{a}_1$	0.5169	0.0396
$\tilde{a}_2$	0.5333	0.0401
$\tilde{a}_3$	-0.2451	0.0000
$\tilde{a}_4$	-0.0696	0.0000
$\tilde{a}_5$	-0.1552	0.0000
$\tilde{a}_6$	0.0615	0.0023
$\tilde{a}_7$	0.1728	0.0125
$\tilde{a}_8$	-0.1050	0.0000
$\tilde{a}_9$	0.0153	0.0008
$\tilde{a}_{10}$	0.2089	0.0164
$\tilde{a}_{11}$	-0.2050	0.0000
$\tilde{a}_{12}$	0.1605	0.0117

Based on both numerical examples above, it is worth noting that we do not have to consider the complete feature vectors for building fuzzy linear regression models; just we utilize the selected vertices which are used for the construction of the convex hull. Therefore, this situation will lead to the decrease of computing load. On the other hand, with regard to computing overhead in the subsequent iteration, it will only consider the newly added samples of data together with the selected vertices of the previous convex hull. Thus, the LP realization becomes effective.

For that reason, this computing scenario will reduce the computation time as well as computational complexity because of the lower number of the feature vectors used for the subsequent processing of regression models. Overall, the computational complexity decreases which cuts down the computing time.

A. *Comparative Analysis*

The increase in sample size might cause computational difficulties in the implementation of the LP problem. Another problem might emerge when changes occur with regard to the variables themselves, thus the entire set of constraints must be reformulated. Therefore the computing complexity increases. The increase of computing complexity has been alleviated by the use of the proposed approach.

To highlight the main features of proposed fuzzy linear regression and ordinary (conventional) fuzzy linear regression models, we summarized the results (produced after second iteration) in Table 3.

Referring to the fuzzy linear regression as stated in above table, we can generalized here that, comparing of both implemented approach, almost similar results were obtained.

Table 3.
Ordinary and Proposed Fuzzy Linear Regression Models Obtained in the Two Numerical Examples (all data considered)

Regression Analysis Approach	Obtained Regression Models
Proposed of fuzzy linear regression model (*h=0.50*)	$Y = (0.0319, 0.0015) + (0.5169, 0.0396)\tilde{a}_1 + (0.5333, 0.0401)\tilde{a}_2 - (0.2451, 0.0000)\tilde{a}_3 - (0.0696, 0.0000)\tilde{a}_4 - (0.1552, 0.0000)\tilde{a}_5 + (0.0615, 0.0023)\tilde{a}_6 + (0.1728, 0.0125)\tilde{a}_7 - (0.1050, 0.0000)\tilde{a}_8 + (0.0153, 0.0008)\tilde{a}_9 + (0.2089, 0.0164)\tilde{a}_{10} - (0.2050, 0.0000)\tilde{a}_{11} + (0.1605, 0.0117)\tilde{a}_{12}$
Ordinary fuzzy linear regression model	$Y = (0.0320, 0.0016) + (0.5170, 0.0397)\tilde{a}_1 + (0.5333, 0.0401)\tilde{a}_2 - (0.2450, 0.0000)\tilde{a}_3 - (0.0697, 0.0000)\tilde{a}_4 - (0.1552, 0.0000)\tilde{a}_5 + (0.0614, 0.0022)\tilde{a}_6 + (0.1728, 0.0126)\tilde{a}_7 - (0.1051, 0.0000)\tilde{a}_8 + (0.0153, 0.0008)\tilde{a}_9 + (0.2088, 0.0163)\tilde{a}_{10} - (0.2051, 0.0000)\tilde{a}_{11} + (0.1605, 0.0118)\tilde{a}_{12}$

On the other hand, based on the experimental studies discussed, it becomes evident that the implementation of the proposed approach is efficient and may drastically reduce the complexity of the overall computing process. Table 4 offers some comparative insights into the performance of the ordinary regression approach and the proposed method.

Table 4.
Results of Comparative Analysis (Considered Convex Hull Vertices)

Regression Analysis Method	Potential Vertices Points	Considered Vertices Points	Newly Potential Vertices Points	Reconsidered Vertices Points
Proposed fuzzy linear regression analysis	409, 600	274	819, 200	447
Ordinary fuzzy linear regression analysis	409, 600	409, 600	819, 200	819, 200

All in all, the experiment showed that the implementation of this proposed approach made considered or reconsidered convex hull vertices fewer then recognized potential locus points which might be selected. In both examples studied here, we have witnessed a substantial improvement. Specifically, the reduction achieved in the first example is 99.34% for initial group of samples and 99.54% for second iteration, while in the second one; initially we have achieved 99.93% improvement and 99.95% for the next iteration.

As mentioned in Section 1, the time factor is crucial to the quality of the method. With this regard, Table 5 reports the pertinent details.

The implementation of the proposed approach is very much appropriate for fuzzy linear regression analysis in presence of fuzzy data especially for real-time data analysis. In order to show the difference of time usage between both implemented approaches, we report the detailed results in Table 5.

Table 5
Computing Time for Experiments Reported in the Study

Regression Analysis Method	Computing Time (*in seconds*)							
	Component 1	Component 2		Total	Component 1	Component 2		Total
	FNO	CHA	LPA		FNO	CHA	LPA	
	Initial Iteration				Second Iteration			
Proposed fuzzy linear regression analysis	00.35	00.91	00.33	01.59	00.39	01.35	00.41	02.15
Ordinary fuzzy linear regression analysis	00.00	00.00	04.45	04.45	00.00	00.00	07.26	07.26

***FNO: Fuzzy Number Optimization, CHA: Convex Hull-Based Approach, LPA: Linear Programming Approach*

The implementation of the proposed approach is very much appropriate for fuzzy linear regression analysis in presence of fuzzy data especially for real-time data analysis. In order to show the difference of time usage between both implemented approaches, we report the detailed results in Table 6.

Table 6
Computing Time for Reported Experiments

Iteration	Sample Size	Time Differences
First iteration	100 samples	02.86 seconds
Second Iteration	200 samples	05.11 seconds

Based on above table, we notice that the more newly acquired data have been added to the currently analyzed group, more significant time savings are reported. This observation has a far reaching implication for real-time data analysis.

Moreover, with regard to the real-time environment, we note that a dynamic data analysis will increase the size of data sets. This situation might cause computational difficulties in the implementation of the LPs. Other problems may arise when the variables themselves change meaning that the entire set of constraints must be reformulated. As a result, the computational complexity will increase. This problem can be handled efficiently through the proposed approach.

In summary, it becomes apparent that the quality of this proposed approach is refers to the computing time as well as the overall computational complexity.

VI. CONCLUSIONS

In this study, we have proposed an augmentation of the fuzzy linear regression for fuzzy data that relies on the Beneath-Beyond algorithm, one of the convex hull techniques for fuzzy healthcare dataset. As a result, the development of the model is based on the construction of related edges or boundaries by connecting the outside points, which in this way become convex vertices.

Our objective was to establish a practical approach to solving a real-time fuzzy linear regression analysis with a fuzzy data set by implementing a convex hull approach called Beneath-Beyond algorithm of any dimensionality. In general, the proposed approach becomes an effective alternative to realize real-time fuzzy linear regression in the presence of fuzzy data by connecting outside vertices points obtained from hyper-rectangles that are the loci of the membership functions.

Moreover, the number of vertices obtained by the convex hull does not extremely increase/decrease, therefore retaining the computing time relatively constant in spite of an increasing number of fuzzy data; this suggests that the method can be applied to large-scale systems in the real world, especially in real-time computing scenarios. The proposed approach does not lead to any computational overhead, as it focuses on the construction/reconstruction of the convex hull by connecting selected vertices points.

We have successfully performed extensive numerical healthcare datasets example which offered some evidence that this method performs like a randomized incremental algorithm that is output-sensitive to the number of vertices. In addition, this approach uses less space than most of the randomized incremental algorithms and executes faster for inputs with non-extreme points, especially when dealing with real-time data analysis. We anticipate that such fuzzy linear regression analysis could become an efficient vehicle for analyzing real-world data and for our future work; we plan to enhance this proposed approach for non-linear problems.

ACKNOWLEDGMENT

This work is sponsored by Universiti Tun Hussein Onn Malaysia under Tabung Penyelidikan ORICC, UTHM and by Gates IT Sdn Bhd.

REFERENCES

[1] C. Kahraman, A. Beşkese and F. T. Bozbura, "Fuzzy linear regression approach and applications", *Studies in Fuzziness and Soft Computing*, Springer-Verlag Berlin, Heidelberg, , Vol. 201, (2006), pp. 589-615.

[2] E. N. Nasibov, "Fuzzy least squares regression model based of weighted distance between fuzzy numbers", *Automatic Control and Computer Sciences*, Vol. 41, Number 1, (2007), pp. 10-17.

[3] P. Diamond, "Fuzzy least squares", *Information Science*, Vol. 46, (1988), pp. 141–157.

[4] J. S. Yuan, A. Reed, F. Chen F and C. N. Stewart Jr., "Statistical analysis of real-time PCR data", *BMC Bioinformatics*, Vol. 7, No. 85, (2006), pp. 1-12.

[5] J. Li, Y. Song and F. Simonot-Lion, "Providing real-time applications with graceful degradation of QoS and fault tolerance according to (m, k)-firm model", *IEEE Transactions on Industrial Informatics*, Vol. 2, Issue 2, (2006), pp. 112- 119.

[6] M. Lin, L. Xu, L. T. Yang, X.Qin, N. Zheng, Z. Wu. And M. Qiu, "Static security optimization for real-time systems", *IEEE Transactions on Industrial Informatics*, Vol. 5, No. 1, (2009), pp. 22-37.

[7] J. Watada, S. Wang and W. Pedrycz , "Building confidence-interval-based fuzzy random regression models", *IEEE Transactions on Fuzzy Systems*, Vol. 17, Issue 6, (2009), pp. 1273-1283.

[8] C. Bohm and H-P. Kriegel, "Determining the convex hull in large multidimensional databases", *International Conference on Data Warehousing and Knowledge Discovery (DaWaK)*, Munich, Germany, (2000), pp. 294-306.

[9] A. Abdalla and J. J. Buckley, "Monte Carlo methods in fuzzy linear regression", *Soft Computing*, Vol. 11, No. 5, (2007), pp. 991-996.

[10] R. Alex, "A new kind of fuzzy linear regression modelling and its combination with fuzzy inference", *Soft Computing*, Vol. 10, No. 7, (2006), pp. 618-622.

[11] B. Barber, D. P. Dobkin and H. Hupdanpaa, "The quickhull algorithm for convex hull", *ACM Transaction on Mathematical Software*, Vol. 22, No. 4, (1996), pp. 469-483.

[12] A. Bargiela, W. Pedrycz and T. Nakashima, "Multiple regression with fuzzy data", *Fuzzy Sets and Systems*, Vol. 158, No. 19, (2007), pp. 2169-2188.

[13] C. Bing-Yuan, "Regression and self-regression models with type-$(\cdot, c)$ fuzzy data", *IEEE Annual Meeting of the Fuzzy Information 2004 (NAFIPS '04)*, Vol. 1, No. 1, (2004), pp. 335- 338.

[14] J. Casillas, P. Martinez and A. D. Benitez, "Learning consistent, complete and compact sets of fuzzy rules in conjunctive normal form for regression problems", *Soft Computing*, Vol. 13, No. 5, (2009), pp. 451-465.

[15] Y-H. O. Chang and B. M. Ayyub, "Fuzzy linear regression methods-a comparative assessment", *Fuzzy Sets and Systems*, Vol. 119, No. 2, (2001), pp. 87-203.

[16] P. D'Urso, "Linear regression analysis for fuzzy/crisp input and fuzzy/crisp output data", *Computational Statistics and Data Analysis*, Vol. 42, Nos. 1-2, (2003), pp. 47-72.

[17] P. Diamond and R. Korner, "Extended fuzzy linear models and least squares estimates", *Computers and Mathematics with Applications*, Vol. 33, Issue 9, (1997), pp. 15-32.

[18] M. Drobics and J. Himmelbauer, "Creating comprehensible regression models (Inductive learning and optimization of fuzzy linear regression trees using comprehensible fuzzy predicates)", *Soft Computing*, Vol. 11, No. 5, (2007), pp. 421-438.

[19] Z. Emiris, "A complete implementation for computing general dimensional convex hulls", *International Journal of Computing Geometry and Applications*, Vol. 8, No. 2, (1998), pp. 223-249.

[20] D. H. Hong and H-C. Yi, "A note on fuzzy linear regression model with fuzzy input and output data for manpower forecasting", *Fuzzy Sets and Systems*, Vol. 138, No. 2, (2003), pp. 301-305.

[21] D. H. Hong, S. Lee, and H. Y. Do, "Fuzzy linear regression analysis for fuzzy input-output data using shape-preserving operations", *Fuzzy Sets and Systems*, Vol. 122, No. 3, (2001), pp. 513-526.

[22] J. J. Huang, G. H. Tzeng and C. S. Ong, "Fuzzy principle component regression (FPCR) for fuzzy input and output data", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 14, No. 1, (2006), pp. 87-100.

[23] J. Watada and W. Pedrycz, "A fuzzy linear regression approach to acquisition of linguistic rules", In: W. Pedrycz, A. Skowron and V. Kreinovich (eds.), *Handbook on Granular Commutation*, John Wiley & Sons, Chapter 32, (2008), pp. 719-740.

[24] N-H. Pan, H. Wang, M-D. Yang K-C and Hsu, "Estimates of bridge girder conditions based on fuzzy inspection data", *Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008*, Vol. 1, (2008), pp. 537-540.

[25] A.K. Nandi and F. Klawonn, "Detecting ambiguities in regression problems using TSK models", *Soft Computing*, Vol. 11, No. 5, (2007), pp. 467-478.

[26] H. Tanaka, S. Uejima and K. Asai, "Linear regression analysis with fuzzy model", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 12, No. 6, (1982), pp. 903-907.

[27] H. Wang and N-F. Pan, "A fuzzy linear regression model for predicting non-crisp variable", *Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008*, Vol. 1, 18-20, (2008), pp. 104–106.

[28] C. K. Kwong, Y. Chen, K. Y. Chan and H. Wong, "The hybrid fuzzy least-squares regression approach to modeling manufacturing processes", *IEEE Transactions on Fuzzy Systems*, Vol.16, No. 3, (2008), pp. 644–651.

[29] Y. Toyoura, J. Watada, M. Khalid and R. Yusof, "Formulation of linguistic regression model based on natural words", *Soft Computing*, Vol. 8, No. 10, (2004), pp. 681-688.

[30] H-F. Wang and R-C Tsaur, "Insight of a fuzzy linear regression model", *Fuzzy Sets and Systems*, Vol. 112, Issues 3, (2000), pp. 355-369.

[31] M-S. Yang and T-S. Lin, "Fuzzy least-squares linear regression analysis for fuzzy input-output data", *Fuzzy Sets and Systems*, Vol. 126, No. 3, (2002), pp. 389-399.

[32] Yao C-C. and Yu P-T., Fuzzy linear regression based on asymmetric support vector machines, Applied Mathematics and Computation, Vol. 182, No. 1 (2006) 175–193.

[33] H. Tanaka, S. Uejima and K. Asai, "Linear regression analysis with fuzzy model", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 12, No. 6, (1982), pp. 903-907.

[34] M. Sakawa and H. Yano, "Fuzzy linear regression analysis for fuzzy input-output data", *Information Science*, Vol. 63, No. 3, (1992), pp. 191-206.

[35] H-C. Wu, "Linear regression analysis for fuzzy input and output data using the extension principle", *An International Journal Computers and Mathematics with Application*, Vol. 45, (2003), pp. 1849-1859.

[36] Z. Emiris, "A complete implementation for computing general dimensional convex hulls", *International Journal of Computing Geometry and Applications*, Vol. 8, No. 2, (1998), pp. 223-249.

[37] R. Miller and Q. F. Stout, "Efficient parallel convex hull algorithm", *IEEE Transaction of Computers*, Vol. 37, No. 12, (1998), pp. 1605-1618.

[38] J. Watada, Y. Toyoura and S. G. Hwang, "Convex hull approach to fuzzy linear regression analysis and its application to oral age model", *Joint 9th International Fuzzy System Association World Congress and 20th North American Fuzzy Information Processing Society International Conference*, Canada, (2001), pp. 867-871.

[39] B. Barber, D. P. Dobkin and H. Hupdanpaa, "The quickhull algorithm for convex hull", *ACM Transaction on Mathematical Software*, Vol. 22, No. 4, (1996), pp. 469-483.