DOKUZ EYLÜL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

SYSTEM FOR EXTRACTING REFERENCE INTERVALS FROM ELECTRONIC HEALTH RECORDS

by
Doruk SOBAY

March, 2020
İZMİR

SYSTEM FOR EXTRACTING REFERENCE INTERVALS FROM ELECTRONIC HEALTH RECORDS

A Thesis Submitted to the
Graduate School of Natural and Applied Sciences of Dokuz Eylül University In Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Engineering, Computer Engineering Program

by
Doruk SOBAY

March, 2020
İZMİR

M.Sc THESIS EXAMINATION RESULT FORM

We have read the thesis entitled "**SYSTEM FOR EXTRACTING REFERENCE INTERVALS FROM ELECTRONIC HEALTH RECORDS**" completed by **DORUK SOBAY** under supervision of **PROF. DR. SÜLEYMAN SEVİNÇ** and we certify that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Prof. Dr. Süleyman SEVİNÇ

Supervisor

Dr. Öğr. Üyesi Özlem ALTAŞ

(Jury Member)

Prof. Dr. Nuri A BAŞOĞLU

(Jury Member)

Prof. Dr. Kadriye ERTEKİN

Director

Graduate School of Natural and Applied Sciences

ACKNOWLEDGMENTS

I would like to express my very great appreciation to Prof. Dr. Süleyman SEVİNÇ for his valuable and constructive suggestions during the planning and development of this research work. I would also like to thank Oktay YILDIRIM for his technical support, advice, and assistance in keeping my progress on schedule. My grateful thanks are also extended to Prof. Dr. Raif ÖNVURAL for his guidance and counseling about my entire career, experiences and for making me feel like family. Finally I wish to thank my parents Ali and Hülya, my brother Şevket Onat and my fiancee Büşra for their support and encouragement throughout my study. This study was conducted with sponsporship of Labenko Bilişim Inc. and all rights of the developed system belongs to Labenko Bilişim Inc.

Doruk SOBAY

SYSTEM FOR EXTRACTING REFERENCE INTERVALS FROM ELECTRONIC HEALTH RECORDS

ABSTRACT

In the reference interval studies of biochemical laboratory tests, samples are provided by at least 120 healthy volunteers for the subgroups of the population. The results that are gained by examining the samples are used to identifying the lower and upper bounds covering 95 percent of the population by statistical methods. This conventional method is based on certain rules and recommendations and the application of this method is onerous and costly. On the other hand, reference intervals that are defined for the adult patient group are used for the pediatric patient group. In this study, an integrated software system is developed and presented, which includes artificial intelligence-based learning algorithms that are empowered by data mining methods, allowing each health institution to perform its own pediatric reference interval studies with its own available data. By using this system, experts can examine the characteristics of the data, perform experiments with learning algorithms, and display results efficiently. In the tests that were conducted during the development phase of the system, the data that is published by Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER) and reference intervals that are published by CALIPER is used. The reference intervals that were generated by the developed system were compared with the reference intervals published by CALIPER, the results were examined. According to the results, it was observed that the reference intervals that were generated by the developed system are compatible with the reference intervals that are published by CALIPER. In addition, high-resolution reference intervals were generated for some biochemical tests.

Keywords: Artificial intelligence, pediatric reference intervals, machine learning, unsupervised learning

REFERANS ARALIKLARININ HASTA KAYITLARINDAN ELDE EDİLMESİ SİSTEMİ

ÖZ

Biyokimya testlerinin referans aralık çalışmalarında popülasyonun alt grupları için en az 120 sağlıklı gönüllüden numuneler alınır. Alınan bu numunelerin çalışılmasıyla elde edilen sonuçlar, istatistiksel yöntemlerle popülasyonun %95'ini kapsayan alt ve üst sınırların belirlenmesinde kullanılır. Bu klasik yöntem belirli kural ve önerilere dayalıdır. Bu yöntemin kullanılması oldukça zahmetli ve maliyetli olup, ayrıca çoğu zaman pediatrik hasta grubu için yetişkin hasta grubuna göre tanımlanmış referans aralıkları kullanılmaktadır. Bu çalışmada, her sağlık kurumunun kendi verileriyle kendi pediatrik referans aralık çalışmalarını yapabilmesine olanak sağlayan, veri madenciliği yöntemiyle güçlendirilmiş yapay zeka temelli öğrenme algoritmaları içeren entegre bir yazılım sistemi geliştirilmiş ve sunulmuştur. Bu sistem sayesinde uzmanlar verilerin karakteristiklerini inceleyebilir, öğrenme algoritmalarıyla deneyler gerçekleştirip sonuçları kolayca görüntüleyebilirler. Sistemin geliştirilme aşamasındaki testlerde Kanada'da klasik yöntemle yapılan bir referans aralık çalışmasının (CALIPER) açık erişimli verileri, sunduğu referans aralık sonuçları kullanılmış ve geliştirilen sistem ile yapılan testlerin sonucunda üretilen referans aralık çıktıları CALIPER çalışmasının sunduğu referans aralık çıktılarıyla karşılaştırılmış ve sonuçlar incelenmiştir. Karşılaştırma sonuçlarına göre geliştirilen sistemin pediatrik hasta grubu için ürettiği referans aralıklarının CALIPER çalışmasının sunduğu pediatrik referans aralık sonuçlarıyla uyumlu olduğu gözlenmiş ve bazı test verileri için daha yüksek çözünürlüklü referans aralıkları elde edilmiştir.

Anahtar kelimeler: Yapay zekâ, pediatrik referans aralığı, makine öğrenmesi, gözetimsiz öğrenme

CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER ONE
INTRODUCTION

Reference intervals, occasionally called "normative" or "expected" values, refer to observed values in a sample reference group by a range which is defined by a particular percentage. This particular percentage commonly defined as the central 95% .

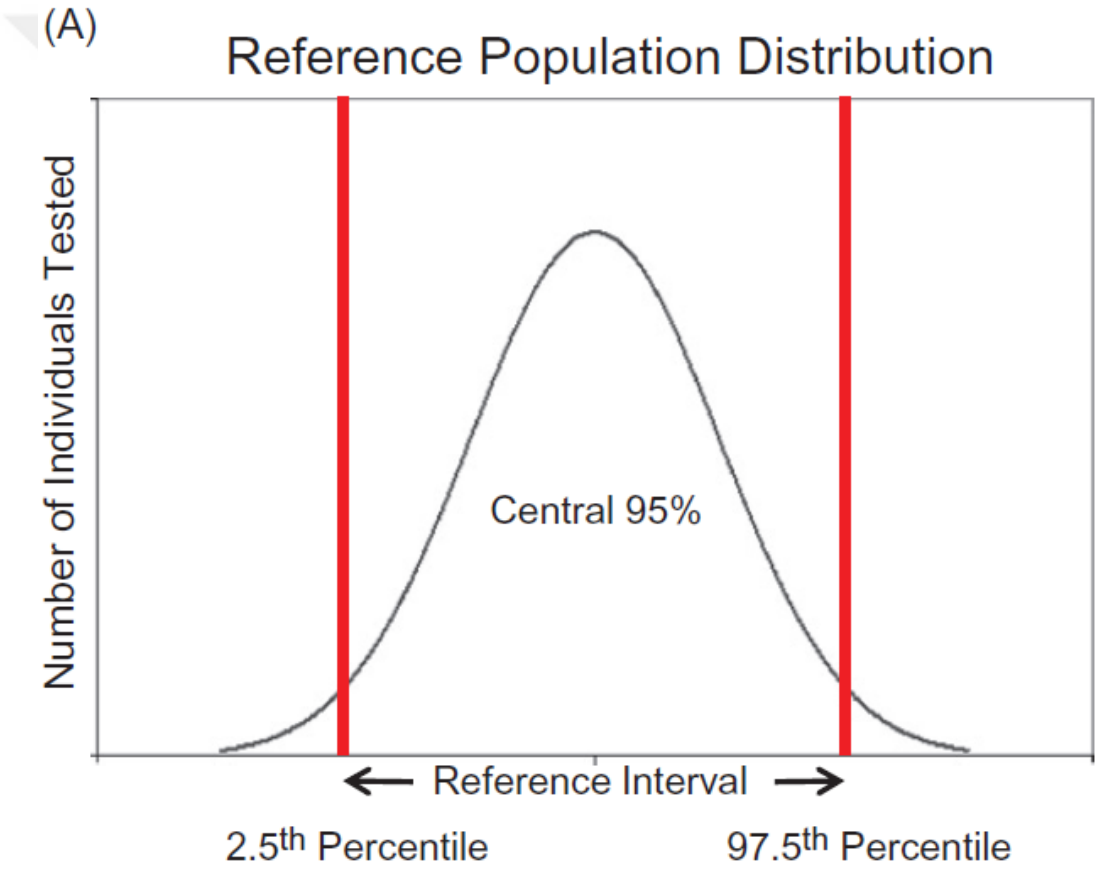

Figure 1.1 Reference intervals definition (Adeli et al., 2017)

Reference intervals and reference interval studies establish the basis of interpretation and evaluation of biochemical laboratory tests. Although reference intervals can be performed for adult patients, in the view of pediatrics patients they

still remain insufficient or inexistent for many analyses (Colantonio et al., 2012). Furthermore, many health institutions and organizations apply the reference intervals of the manufacturers for evaluation of biochemical test results. Each population in the different geographical regions may not have the same reference intervals. That's why it is recommended that each institution needs to generate its own reference intervals. In addition, to apply reference intervals for pediatric patients it is also recommended that detailing generated reference intervals.

In the conventional method, the reference intervals of the biochemical laboratory tests are determined by calculating the lower and upper limits covering by the particular percentage of the population, using statistical methods, from the results obtained by studying samples from at least 120 healthy volunteers for the subgroups of the population. This conventional method, based on rules and recommendations, is inconvenient and costly; It is even more difficult to apply in the pediatric patients.

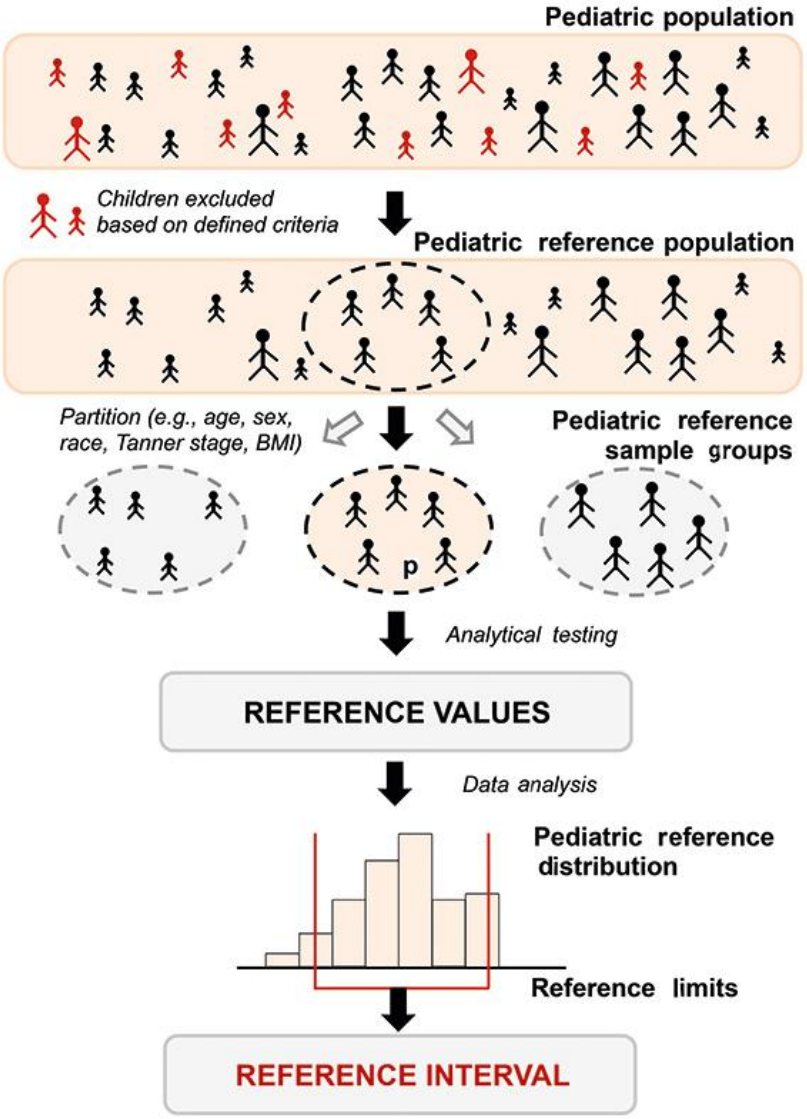

Figure 1.2 Convetional reference intervals study process (Gherasim, 2017)

According to recent statistical health data, the annual number of pediatric patients who contact in health institutions is huge. Test results of these pediatric patients are being gathered in the database of the health institution and they are not used for any other purpose most of the time. In this gathered data, there is a big amount of healthy group which could be used for reference interval studies.

In this study, it is aimed to develop and present an integrated software system, which includes artificial intelligence based learning algorithms and are empowered by data mining methods, allowing each health institution to perform its own pediatric reference interval studies with its own available data.

1.1 Background

1.1.1 Machine Learning

Machine learning is a mathematical study of algorithms and statistical models that affords computer systems the skill to learn by themselves and develop from knowledge without being programmed. Machine learning aims for the development of computer programs and algorithms that be able to access and work with data to improve themselves. The steps of learning starts with knowledge or data on the purpose of look for patterns in data and to improve decision making in the future based on the cases that are provided. Allowing computers to learn by themselves and to set an action without any human interaction or assistance is the essential purpose of machine learning. Machine learning algorithms are generally categorized as supervised and unsupervised learning.

1.1.1.1 Supervised Learning

Supervised machine learning algorithms are able to apply formerly learned action to new data by using tagged samples to predict future situations. Beginning from the evaluation of a known training dataset, the learning algorithm constructs a deductive function for forecasting the output values. The algorithm can procure targets for any other inputs after a convenient training and can also compare its output with the proper, desired output. Finally, to customize the model, it can detect errors.

1.1.1.2 Unsupervised Learning

Unsupervised machine learning algorithms are needed when the training data is neither classified nor tagged. Unsupervised learning analyzes how algorithms are able to deduct a function to specify a hidden structure from untagged data. The algorithm does not find out the proper output. It discovers the data and be able to make inferences from datasets to specify hidden structures from untagged data.

1.1.2 Gaussian Distribution

Gaussian distribution, also known as the normal distribution, is a probability distribution that has mean (μ) and standard deviation (σ) parameters and the data is distributed near the mean according to the standard deviation. It is showing that data near the mean are more frequent than the data far from the mean. In graph form, normal distribution appears a bell curve. The general form of probability density function of normal distribution is:

$$y = \frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma} \tag{1.1}$$

As is seen from the formula above, Gaussian distribution is defined by the parameters mean (μ) and standard deviation (σ).

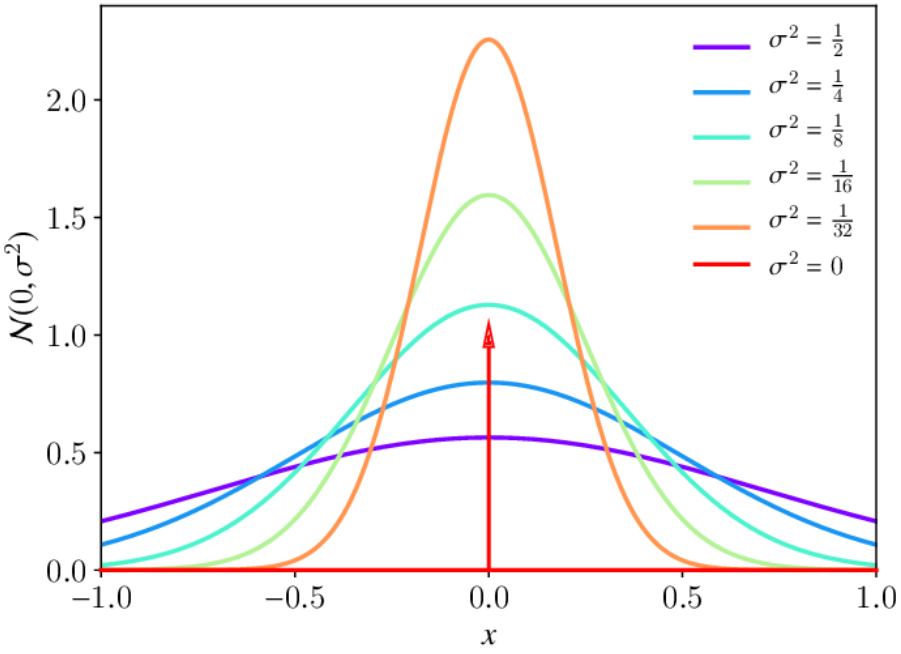

Figure 1.3 Gaussian distribution samples (Koschier et al., 2019)

1.1.3 Gaussian Mixture Models

Gaussian Mixture Models also known as the Mixture of Gaussians is a probabilistic model that supposes, by using unknown parameters data points, are computed from a mixture of a finite number of Gaussian distributions. One might conceive of the mixture of Gaussians as a generalization of the clustering of k-means to add knowledge of the covariance structures of data and the latent Gaussian centers.

1.1.4 Expectation Maximization (EM) Algorithm

In statistical models that depend on unobserved latent variables, the EM algorithm allows identifying the maximum-likelihood of models. The EM algorithm is an iterative process that performs an expectation (E) step and a maximization (M) step alternately. E step engenders a function for the prospect of the log-likelihood

evaluated utilizing the current estimate for the parameters. M step calculates the parameters that maximize the expected log-likelihood gained on the E step. These estimations are used to identify the distribution of the latent variables in the next E step.

1.1.5 Gaussian Similarity

Gaussian distributions that are distributed around mean value by standard deviation value may overlap when mean values close enough or standard deviation values big enough. This is the constitution of the Gaussian mixtures. Gaussian similarity can be defined as the ratio of the overlap area between 2 Gaussian distributions.

1.1.6 Reference Interval

The results gained from biochemical laboratory tests are evaluated by analyzing the reference intervals and are needed to be submitted in proper and reliable to the clinicians. Reference intervals are most commonly performed by using the Clinical and Laboratory Standards Institute (CLSI) and the International Federation of Clinical Chemistry (IFCC) guidelines (Sasse, 2000).

Each biochemical laboratory test has a reference interval that is established by the manufacturer. It is still unclear that these reference intervals properly reflect the desired intervals of related population. For this reason, every population needs to determine their own reference intervals and use these intervals instead of intervals provided by the manufacturer. Baadenhuijsen and Smit (1985) claim that the use of data of the health institution in the selection of reference individuals would produce more proper intervals that the use of healthy individuals.

By gathering reference individuals from an example population, a reference population could be created. The collection of reference individuals creates a crucial process in establishing reference intervals.

1.1.6.1 Conventional Methods in Reference Interval Determination

In the reference interval generation process, there are 2 types of classified methods that are being used. These methods are parametric and nonparametric methods. Parametric methods are used when the distribution is a Gaussian distribution, even if insufficient amounts of data available. On the other hand, nonparametric methods are used because of mostly detecting non-Gaussian distribution in biological data. To generate useful outputs, nonparametric methods commonly expect a huge amount of data.

1.1.6.1.1 Parametric Method. The basis of the parametric method is computing the mean and standard deviation values to specify the 95% confidence interval that includes the actual values. As mentioned above, the data must distribute in the form of Gaussian distribution. Alternatively, the data can be altered to distribute as Gaussian distribution (Linnet, 1987). The lower and upper boundaries are defined as 2.5% to 97.5%. The area between the boundaries specifies the 95% confidence interval, as well as the points constituting the ± 2 SD range of the mean specify the reference interval.

1.1.6.1.2 Nonparametric Method. The nonparametric method allows distinguishing the reference individuals effortless and conducting reference interval studies for each health institute by using their own databases. Due to the number of data for an accurate Gaussian distribution is at least 120, sizes that below this are not sufficient to apply the nonparametric method. As with the parametric method, 95% confidence interval is specified between lower and upper bounds. Following formulas are defined to compute the values of this interval:

$$\text{Lower value} = 0.025 * (n + 1) \quad (1.2)$$

$$\text{Upper value} = 0.975 * (n + 1) \quad (1.3)$$

where "n" stands for the number of data. The data is organized in ascending order in the first place. Outputs of the formulas above identify the index of lower and upper values of the reference interval. In addition, rational numbers are rounded to keep form (Lumsden & Mullen, 1978).

1.1.7 Z-score

Z-score also known as the standard score is a numerical measurement measured in terms of standard deviation that represents the distance of a point from the mean of a distribution. The value of the point is same as the value of mean when z-score of the point is 0. The value of the point is away from the mean as the amount of the value of standard deviation when z-score of the point is 1. The value of the point is above the mean when the z-score has a positive value. On the contrary, the value of the point is below the mean when the z-score has a negative value.

1.2 Literature Review

A method for forecasting reference intervals was proposed by Horn, Pesce, and Copeland (1998). This method aims applied to the data that has serious number of outliers. They presented a prediction interval that handles satisfying placing and scaling estimations. Particularly non-parametric, transformed, robust with non-parametric lower limit, and transformed robust methods were analyzed in this study. The reference intervals are estimated with and without outlier detection. In that way, this approach continuously generates upper reference intervals that are a more nearby place to the real underlying distributions. This study suggests a statistical analysis from shortened or incorrect data could be the perfect use for forecasting reference intervals. To detect the upper bound of the reference intervals, this proposed method could be used when the sample size is insufficient.

Jagarinec et al. (1998) performed a study that purposes to determine reference intervals of the 34 biochemical laboratory tests from local students aged between 8 to 18 in Zagreb and Croatia. After a serious number of medical inspections, a healthy group of students was selected by health professionals and that group was used as a source in this study. In the Faculty of Science of the University of Zagreb, gained data were used to evaluate by a software program that is called "Statistics". The variables of the group were identified by the features of the group and were evaluated by sex-specific and age-specific frequency distributions. Reference intervals were illustrated in the percentage of 2.5 to 97.5 of non-parametric distributions. The values out of the 3 standard deviations were excluded. Mann Whitney U-test was applied for forecasting the characteristics between sex and age in the groups. Jagarinec et al. (1998) indicate that the results gained from the study do not demonstrate serious difference with the related studies in USA, Canada, England, France, and Spain.

In the study of Kapelari et al. (2008), the significance of sex-specific and age-specific reference intervals was mentioned for evaluating thyroid hormone measurements in pediatric patients. It is expressed that working with thyroid hormone measurements is more effortless than TSH, free T3, and free T4 because of the insufficient number of studies. Thus, by using the International Classification of Diseases (ICD-10) codes, thyroid hormone measurements were classified by diagnostic categories. These measurements were gained from a pediatric group of a health institution. The members of the pediatric group that may alter thyroid function were excluded. And also those with eating disorders, pituitary disease or chromosomal anomalies were also excluded. The values of $p < 0.05$ were assumed sufficient in terms of statistics. Medians and percentages were defined and take as the reference intervals of each variable. To generate percentages and to make an evaluation in terms of statistics, statistical computer program The Social Sciences Statistical Package (SPSS) was used. The results display that the thyroid hormone levels show an alteration in the period of childhood so, in most cases, using reference intervals of adult patients for pediatric patients is not convenient. It is also mentioned that some dissimilarities were identified compared to previous studies. These dissimilarities may be caused by different races or unexplored geographical variations.

Generating reference intervals for hematological and biochemical tests is the purpose of the study of Humberg et al. (2010). The population of this study consists of local infants and children in Gabon. The data was gained from the group of the population who have gone regularly to Albert Schweitzer Hospital in Lambaréné, Gabon. Reference intervals were generated by using CLSI guidelines (Horowitz, 2010). The non-parametric method was applied to generate intervals and between the percentage range of 2.5 to 97.5 of data is used. To decide if the reference values need to be separated into subgroups like males and females, the standard normal deviate test was applied. This test measures the degree of dissimilarity in terms of statistics between the separated groups. The final reference intervals were compared with reference intervals of the European population. In conclusion, the values of

hemoglobin, hematocrit, red blood cell count and mean corpuscular volume were calculated lower. This study shows the importance of generating reference intervals for local populations and also shows that generated reference intervals could be applied for related populations in Central Africa.

Colantonio et al. (2012) mention that there is not sufficient information base about the effectiveness of age, sex, and ethnicity on reference intervals. To fulfill the lack of information base, a reference interval database should be constructed. In accordance with this purpose, a healthy group of children and adolescents were selected from a multiethnic population. This group examined a specific survey to be applied to exclusion. To generate age-specific and sex-specific reference intervals, full blood tests from 40 serum biochemical markers were gathered. The evaluation was also done by CLSI guidelines.

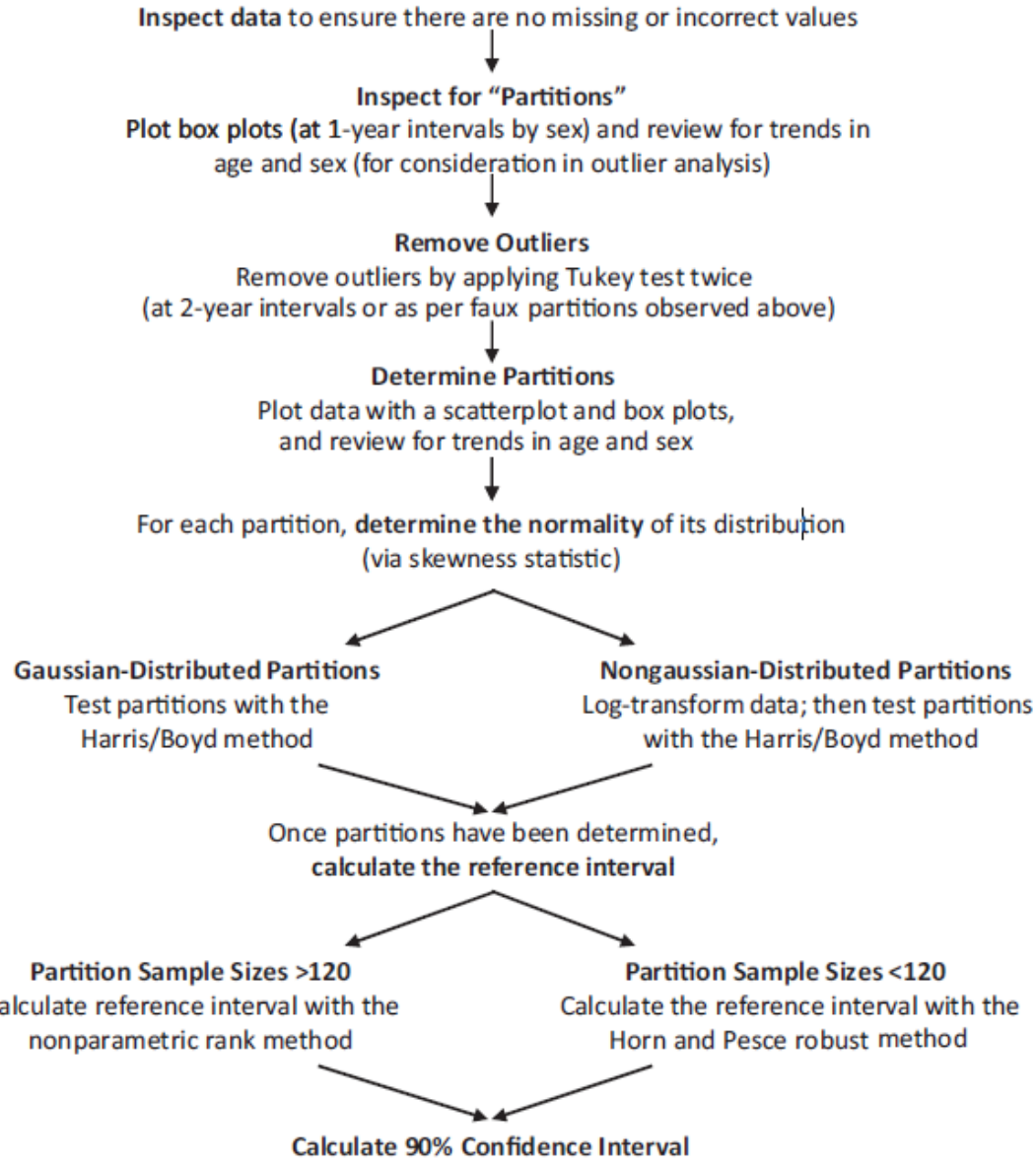

Figure 1.4 CALIPER data-analysis algorithm base on CLSI guidelines (Colantino et al., 2012)

In the study of Adeli et al. (2017), it is mentioned that because of the insufficient sex-specific and age-specific reference intervals for the healthy pediatric population, most clinical laboratories have to apply adult reference intervals to pediatrics in the field of pediatrics laboratory medicine. To overcome this problem, the Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER) started to collect healthy pediatrics data from the community. By using the data CALIPER developed an algorithm to determine correct sex-specific and age-specific pediatric reference intervals in consideration of CLSI guidelines. At the present time, clinical

laboratories in Canada and around the world use the reference interval database of CALIPER.

The Hoffmann indirect method was programmed to generate reference intervals and was used in the study of Katayev, Balciza, and Seccombe (2010) with the collected data from clinical laboratories in the United States for 5 analytes without applying any outlier or exclusion method. Chauvenet criteria were applied to determine and eliminate outliers. After applied the criteria, the data was modified to use the values from the linear portion of the cumulative frequency graph. Computer-driven Hoffmann indirect method generates pretty similar reference intervals with peer-reviewed ones. There are no meaningful variations in terms of statistics between generated upper and lower bounds and related published values. The generated reference intervals are a bit tighter when compared to the reference intervals of direct sampling methods. It is proven that this method is accurate and acceptable for forecasting the reference intervals of the data gained from the laboratory database.

In the other study of Katayev et al. (2015), a computer program was presented that uses a statistical data mining algorithm to determining of reference intervals for clinical laboratory tests. By applying a modified algorithm for data mining, 8 analytes from the laboratory database were used to generate reference intervals with dissimilar sex-specific and age-specific groups. The biochemical laboratory test values were gained from the database of the Laboratory Corporation of America for the parameters age, sex, and period of time. The sex-specific and age-specific populations were picked by applying indirect data mining methods and direct sampling techniques to generate reference intervals. The older released version of computer-driven Hoffmann indirect method was modified with several new features and improvements. After that, these outputs were compared to peer-reviewed studies by using direct sampling. There are no meaningful variations in terms of statistics between generated reference intervals in this study and related published values.

Hoffmann's approach in the computerized statistical algorithm is approved for generating reference intervals as an accurate and useful method.

Homocysteine and vitamin A-E biochemical test results are examined by Mersin University Hospital of Faculty of Medicine Clinical Biochemistry Laboratory in the study of Akbayir et al. (2011) and generated new sex-specific and age-specific reference intervals for Mersin territory. Homocysteine and vitamin A-E levels were separated into the subgroups by sex and expressive statistics were established for the ages of each variable. To identify and examine age subgroups of ages, Multivariate Adaptive Regression Splines (MARS) method was applied. Conventional parametric and non-parametric reference interval methods were selected according to the distribution. MedCalc statistical computer software was used to generate reference intervals. It is also claimed that the biochemical clinical laboratory of each health institution needs to perform its own reference interval studies for its own population of patients. To reduce and prevent complexity, high cost, and medical intervention, using indirect methods while generating reference intervals for each clinical laboratory could be beneficial. By considering this, the indirect method was applied to establish the reference intervals by sex and age in this study. The study proposes homocysteine and vitamin A levels need to be examined by sex and age, but vitamin E levels only need to be examined by sex. According to these outcomes, it is mentioned that using sufficient reference intervals for each clinical laboratory from its own population is crucial.

By using health institution data to generate the reference intervals, the procedure of selection of healthy groups was proposed in the study of Orekici Temel et al. (2015) by using the Bhattacharya method. Generating the reference intervals with data that mixed up healthy and unhealthy values may cause a complication when applying indirect methods, but the Bhattacharya method allows to create a healthy group from mixed data. To identify overlapping normal distributions, Bhattacharya proposes a graphical model. Although the Bhattacharya method has some strengths,

there are some weaknesses such as distribution of data. It is quite dependent data distribution to generate sufficient results and the data needs to be mostly sufficient for normal distribution.

In the study of Çaycı et al. (2015), specifying full blood reference intervals from clinical laboratory test results in the database of the health institution is the main purpose. To be able to make a comparison with the reference intervals of the manufacturer, a non-parametric percentage predicting method was applied to the clinical laboratory test and the SPSS was used for examining of the data in terms of statistics. In this study, the population consists of the people aged between 12 and 65 years old. Reference intervals were generated separately for subgroups of sex and age. According to the results, reference intervals of sub-parameters of sex and age groups aged between 0 and 12 need to be generated separately. There are serious differences between the reference interval of the manufacturer and the generated reference intervals in this study so, it was claimed that generated reference intervals from the population of health institutions by the indirect method are more sufficient for each clinical laboratory.

1.3 Methodology and Tools

In this study, HTML5, CSS3, and Javascript web technologies were used to develop a web application for user interactions. To construct a web server for the web application, the version v6.11.4 of Node.js was used because of features such as performance, scalability, accessibility, and extensive developer support environment. The version 3.6.3 of Python programming language was used to implement learning methods and construct a server to communicate Node.js web server. Python has a huge amount of statistical libraries to implement artificial intelligence techniques such as machine learning, deep learning, etc. Brackets, Notepad++, and Visual Studio Code were used as code editors. MongoDB that is a type of NoSQL database technologies was used to save data in form of JSON strings because of features such

as high performance, easy to use, and dynamic structure. MongoDB Community Server 4.0.3 and Robomongo Robo 3T 1.2.1 were used to perform database operations. Microsoft Windows 10 operating system was used to perform experiments and Google Chrome was used as testing environment.

CHAPTER TWO
IMPLEMENTATION

The implementation of this system consists of a combination of two separate modules: a web application tool (Eraslan, 2019) and a learning module. Web application tool provides some abilities with data mining techniques to the user and the main and most related ones with this study are;

- organizing and filtering data,
- displaying learning results,
- isplaying details of data, and
- feeding data to the learning module.

By providing these abilities, the web application tool covers data collection, data preparation, and model choosing steps of the machine learning process.

In the learning module, there are two types of clustering methods empowered with some artificial intelligence techniques. Method one is conducted by the GMM algorithm and includes some optional normalization techniques when necessary situations. This method can be also performed without applying any normalization process on the data. Method two is conducted by applying the measurement of Gaussian similarity to distributions. Basically both two learning methods are in the form of unsupervised learning.

After data related operations applied in the web application tool, the data is sent in JSON format from the web server to the python server via a socket connection. In the python server, the JSON data is parsed into the parameters and the learning method is determined by the “Method Controller” module of the python server. According to the identified learning method, sent learning parameters are distinguished. If the

parameters are not accomplished the requirements of the identified method, an error occurs and the python server sends a JSON error message. After parameters are distinguished, the identified learning method is fed with the data. The learning process is performed and evaluated results are sent back to the web server in JSON format for being saved and displayed in the web application tool.

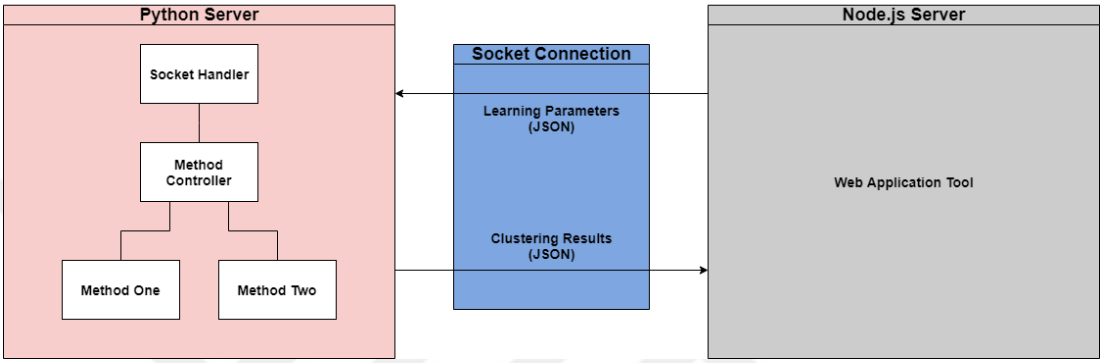

Figure 2.1 A high-level overview of proposed system

2.1 Material

The patients' data are confidential. Every health institution can carry out their own studies by using their own patients' data, but to use patients' data outside of the health institution is required to have some special permission. So, that means there needs to use public data in most studies like this. In this situation, CALIPER gives a chance by publishing biochemical analyte data publicly. In the testing period of the system, eleven different biochemical analytes of CALIPER' were commonly used. These analytes and SI units are;

- Albumin G – g/L,
- Alkaline phosphatase (ALP) – U/L,
- Alanine aminotransferase (ALT) – U/L,
- Calcium – mmol/L,

- Cholesterol – mmol/L,
- Creatinine-Jaffe - μmol/L,
- C-reactive protein (CRP) – mg/L,
- Immunoglobulin A (IgA) – g/L,
- Iron - μmol/L,
- Lipase – U/L, and
- Triglycerides – mmol/L.

2.2 Learning Parameters

The learning parameters are the parsed JSON data sent from web server of the web application tool to the python server. These parameters include the actual data to be used in learning methods, the number of requested/expected clusters, titles of the columns of data, method-specific data, and optional data.

Table 2.1 Learning parameters

Parameter	Method One	Method Two	Optional	Type
data	X	X		2D array
numCluster	X	X		number > 1
ratioSimilarity		X		number
sizeBatch		X	X	number
sizeBin	X	X	X	number
sizeThreshold	X	X	X	number
titleAxisX	X	X		text
titleAxisY	X	X		text
titleThreshold	X	X	X	text
typeMethod	X	X		text

2.3 Clustering Results

After the learning process is done, outputs of the process are sent to the web server in JSON format to be persisted. These parameters primarily include plotted HTML string, label values, mean values, and standard deviation values.

Table 2.2 Clustering results

Parameter	Method One	Method Two	Optional	Type
htmlText	X	X		text
listLabel	X	X		2D array
listMean	X	X		1D array
listStd	X	X		1D array
others	X	X		any

2.4 Python Libraries

NumPy is the essential library to perform scientific computations in Python language. On the other hand, NumPy also allows defining arbitrary data-types and using as powerful multi-dimensional storage of generic data.

Scikit-learn (sklearn) is a library that allows performing machine learning techniques. It supplies several tools to:

- select-fit-evaluate model,
- organize data before using it.

It also assists in supervised and unsupervised learning techniques. sklearn.mixture.GaussianMixture is a class that is designed for predicting parameters of Gaussian Mixture distribution.

SciPy is a library that allows scientific and technical computing in Python language just like the NumPy library. It includes a bunch of packages for science and engineering areas such as linear algebra, integration, interpolation, etc. Packages of SciPy that were used in the implementation phase of this study;

- scipy.stats is the package that includes a huge number of statistical functions such as probability distributions.
- scipy.stats.norm is a function of a normal continuous random variable.
- scipy.integrate is the package that includes integration functions.
- scipy.integrate.quad is a function to compute a definite integral.

Random is a library that allows generating a number by defined range. By implementing a pseudo-random number generator, it achieves this.

Matplotlib is a library that allows plotting in Python. In the implementation of this study, the color set of Matplotlib was used.

Bokeh is the library that aims to improve the viewing experience with more interaction with the user. Bokeh CDN is the mode that imports Bokeh JS and CSS from the library. file_html() is a function that generates the HTML output from the Bokeh figure. The output is in the form of string.

2.5 Method One

In this method, the GMM clustering algorithm that is a commonly used algorithm of clustering algorithms constitutes the basis of the learning process. Before

performing the learning process, this method allows applying the normalization process on the data. Normalization process has two different methods:

- normalization by outliers,
- normalization by z-score.

Normalization by the outliers method is a scaling process and the purpose of this process is scaling down the distances between the numerical values. The items of the data are recalculated by the maximum and the minimum values of the data.

Normalization by z-score is a method that calculates the z-score of each item of the data with means and standard deviations.

The learning and clustering process starts with the definition of the model. The constructor of GMM algorithm has some basic parameters to define the model and perform prediction. Although it is a form of unsupervised learning, it allows specifying the number of clusters that is desired. In that way, the first parameter of the model is the cluster number. After the model is defined, with the EM algorithm of the GMM the estimation of model parameters is started via a function called "fit()" by feeding the data. In this situation, the model was trained and the next step is the prediction process. By using "predict()" function, labels for the data are predicted via trained model. At the end of the prediction process, the variables labels, means, and covariances are gained.

2.5.1 Python Programming

In this section, some code snippets are illustrated. These code snippets constitute the basis of the implementation phase of the method one.

2.5.1.1 Normalization Methods

2.5.1.1.1 Normalization By Outliers. This method takes a 1D array of values that desired to be normalized as a parameter. As mentioned above it uses the maximum and the minimum values of the parameter and returns the updated parameter.

```
def __normalizeByOutliers(arg):
    maxValue = max(arg)
    minValue = min(arg)
    size = len(arg)

    for i in range(size):
        nextValue = arg[i]
        nextValue = (nextValue - minValue) / (maxValue - minValue)
        arg[i] = nextValue

    return arg
```

Figure 2.2 Code snippet of normalization by outliers

2.5.1.1.2 Normalization By Z-score. This method takes a 1D array of values that desired to be normalized as a parameter. As mentioned above it applies z-score formula to the values of parameter and returns the updated parameter. The formula of z-score as follows:

$$z = \frac{x - \mu}{\sigma} \tag{2.1}$$

where;

- x symbolizes the value,

- μ symbolizes mean, and
- σ symbolizes standart deviation.

```python
import numpy as np

def __normalizeByZedScore(arg):
    mean = np.mean(arg)
    std = np.std(arg)
    size = len(arg)

    for i in range(size):
        nextValue = arg[i]
        nextValue = (nextValue - mean) / std
        arg[i] = nextValue

    return arg
```

Figure 2.3 Code snippet of normalization by z-score

2.5.1.2 Model Creation and Prediction Proces. To apply GMM, the sklearn library provides a package. This package allows creating-sampling-estimating the models. GaussianMixture object of the package implements the EM algorithm to fit models. As mentioned above the object has two main functions called "fit()" that is implemented as GaussianMixture.fit and "predict()" that is implemented as GaussianMixture.predict. GaussianMixture.fit function conducts the learning phase from data while GaussianMixture.predict function conducts the assigning phase of each sample to the Gaussian that it most probably belongs to.

```python
from sklearn.mixture import GaussianMixture

def __estimateModel(args):
    nComponents = args["numCluster"]
    data = args["data"]

    gmm = GaussianMixture(n_components=nComponents)

    estimation = gmm.fit(data)
    labels = estimation.predict(data)
    means = gmm.means_

    return labels, means
```

Figure 2.4 Code snippet of model creation and prediction process

2.6 Method Two

In this method, a new clustering technique was applied to generate reference intervals. This technique can identify the subgroups of the data and generate reference intervals among them. The fundamental principle of this method is to identify subgroups that are similar to each other by Gaussian similarity and combine them by specific conditions. To fulfill this technique, the method conducts an iterative process. Before the iterative process, the method;

- divides data to the maximum number of subgroups by age,
- compares the size -the number of items- of each subgroup to the batch size parameter,
    - if the size smaller than the batch size, combine subgroup with the closest subgroup.

After that, the method fires the iterative process. The statistical Gaussian distribution of each subgroup is calculated. Besides of the similarity ratio parameter, there is a variable called "current ratio" that is ready to use with default value 1 at runtime. Starting from the first Gaussian distribution, the ratios of the overlapped area between the current distribution with previous and next distributions are calculated separately. Distributions with the biggest overlapped ratio are determined and the overlapped ratio is compared with the current ratio. If the overlapped ratio is equal to the current ratio, subgroups that related to distributions are combined, a new subgroup is created and its distribution calculated, and the value of the current ratio is set to the default value. If the overlapped ratio is smaller than the current ratio, the value of the current ratio is decreased .01 and iteration starts over.

At the end of each iteration, the number of distributions and the current ratio is checked with the number of clusters parameter and similarity ratio parameter. The iterations stop;

- if the number of distributions equals to the number of clusters, or
- if the current ratio smaller than the similarity ratio.

2.6.1 Python Programming

In this section, some code snippets are illustrated. These code snippets constitute the basis of the implementation phase of the method two.

2.6.1.1 Organization of Data By Batch Size

This section covers organizing the initial version of the data before the learning process. The number of samples of each subgroup must be at least the specified batch size for this study. To fulfill this requirement, after sorted data by age values the numbers of samples of each age value are counted and held. According to the comparison of counts and batch size, some age groups get combined and create a new group. At the end of this process, subgroups are distinguished.

```
def __organizeByBatch(args):
    sizeBatch = args["sizeBatch"]
    liSortedByAge = [...] #2D array of sorted arg [[age-1, result-1], ... , [age-n, result-n]]
    liCountsOfAges = [...] #2D array of age counts of liSortedByAge [[age-1, count-1], ... , [age-n, count-n]]
    liInitial = []
    sizeLiSorted = len(liSortedByAge)
    sizeLiCounts = len(liCountsOfAges)

    i = 0

    while i < range(sizeLiCounts):
        nextItemLC = liCountsOfAges[i]
        ageLC = nextItemLC[0]
        countLC = nextItemLC[1]
        liTemp = []
        flag = 0

        for j in range(sizeLiSorted):
            nextItemLS = liSortedByAge[j]
            ageLS = nextItemLS[0]
            resultLS = nextItemLS[1]

            if ageLS == ageLC:
                liTemp.append(resultLS)

        if countLC < sizeBatch:
            for j in range(i + 1, sizeLiCounts):
                nextItemInnerLC = liCountsOfAges[j]
                ageInnerLC = nextItemInnerLC[0]
                countInnerLC = nextItemInnerLC[1]

                if countLC < sizeBatch:
                    countLC = countLC + countInnerLC

                    for k in range(sizeLiSorted):
                        nextItemInnerLS = liSortedByAge[k]
                        ageInnerLS = nextItemInnerLS[0]
                        resultInnerLS = nextItemInnerLS[1]

                        if ageInnerLS == ageInnerLC:
                            liTemp.append(resultInnerLS)
                    i = j
                    flag = 1
                else:
                    break

        liInitial.append([ageLC, liTemp])

        if flag == 0:
            i = i + 1
```

Figure 2.5 Code snippet of organization of data by batch size

2.6.1.2 Calculation of Overlap Area. This section handles a method applying the Gaussian similarity technique mentioned in the introduction section. The parameters of this method are means and standard deviations of Gaussian distributions. Basically, it computes the integrals of the Gaussian distributions by mean values and standard deviation values. Limit values of the integrations are computed with ± 2 times standard deviation values of the mean values.

```python
import scipy as scipy

def __calculateOverlapArea(args):

    # integral function
    def minF1F2(x, m1, m2, std1, std2):
        f1 = scipy.stats.norm(m1, std1).pdf(x)
        f2 = scipy.stats.norm(m2, std2).pdf(x)

        return min(f1, f2)

    mean1 = args["mean1"]
    mean2 = args["mean2"]
    std1 = args["std1"]
    std2 = args["std2"]

    pointStart = 0
    pointEnd = 0

    if mean1 < mean2:
        pointStart = mean1 - 2 * std1
        pointEnd = mean2 + 2 * std2
    else:
        pointStart = mean2 - 2 * std2
        pointEnd = mean1 + 2 * std1

    return scipy.integrate.quad(minF1F2, pointStart, pointEnd, args=(mean1, mean2, std1, std2))[0]
```

Figure 2.6 Code snippet of calculation of overlap area

2.6.2 Flowcharts

This section includes flowcharts of the main flow of method two and code snippets that were mentioned in previous section.

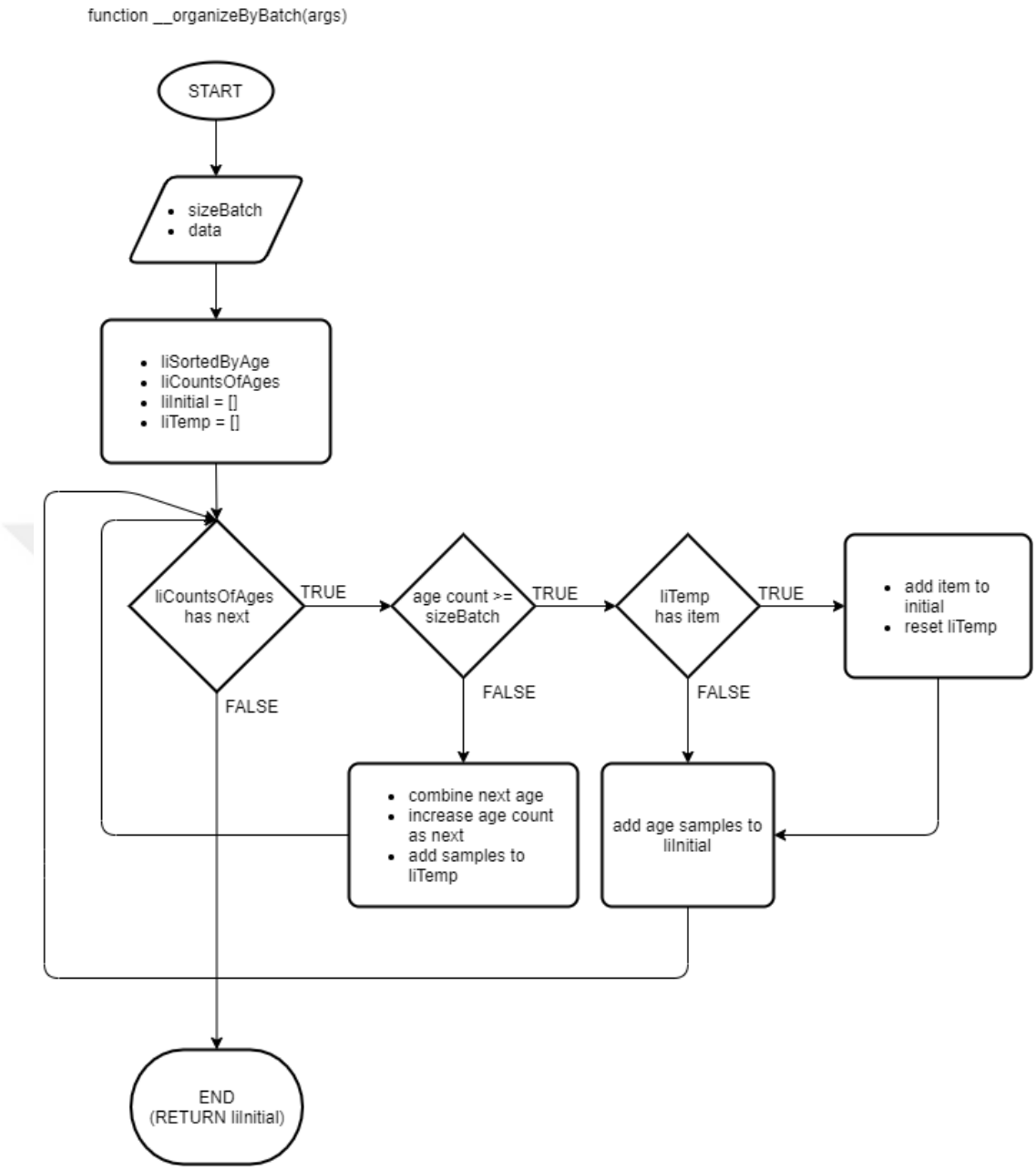

Figure 2.7 Flowchart of organizing of data by batch size

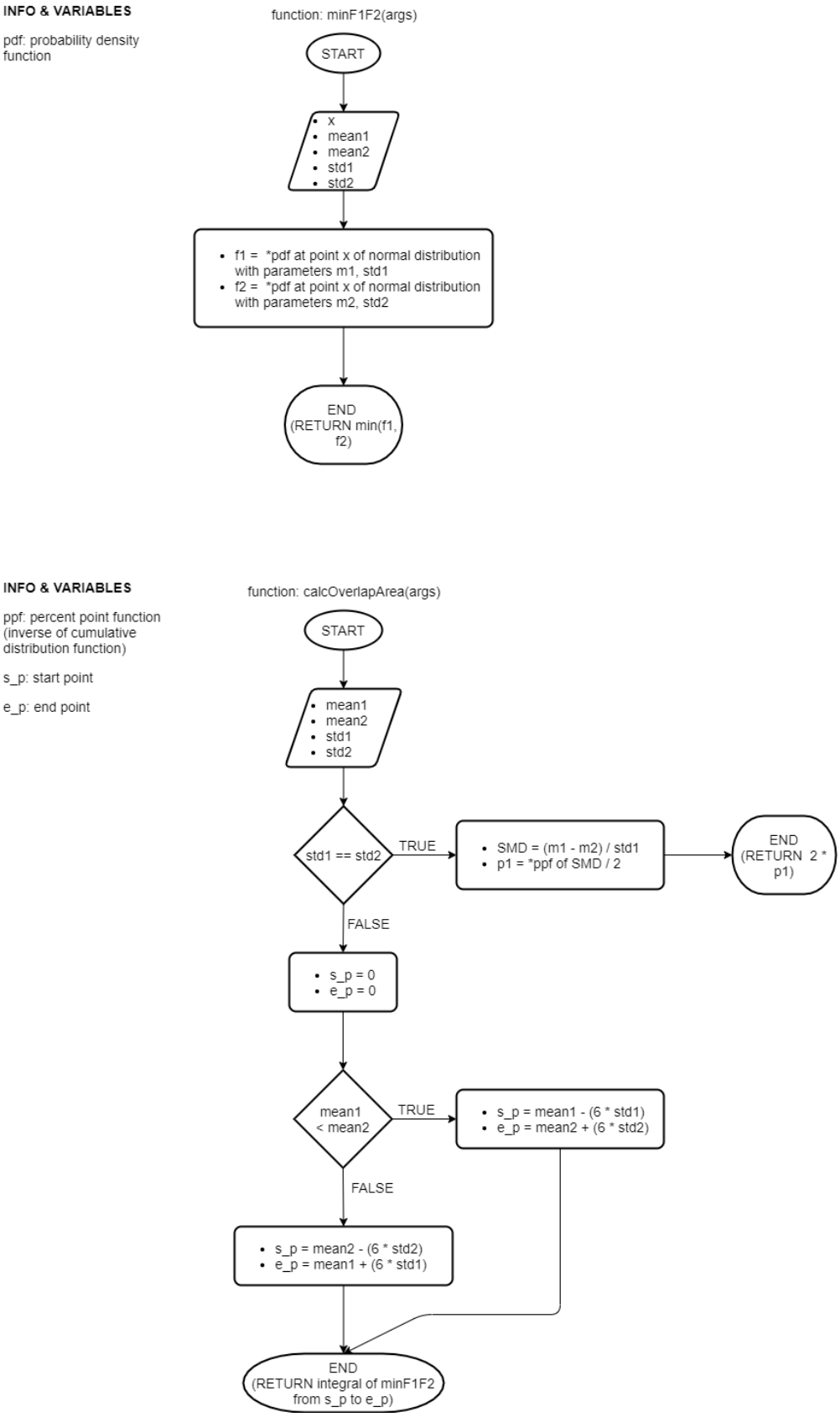

Figure 2.8 Flowchart of calculation of overlap area

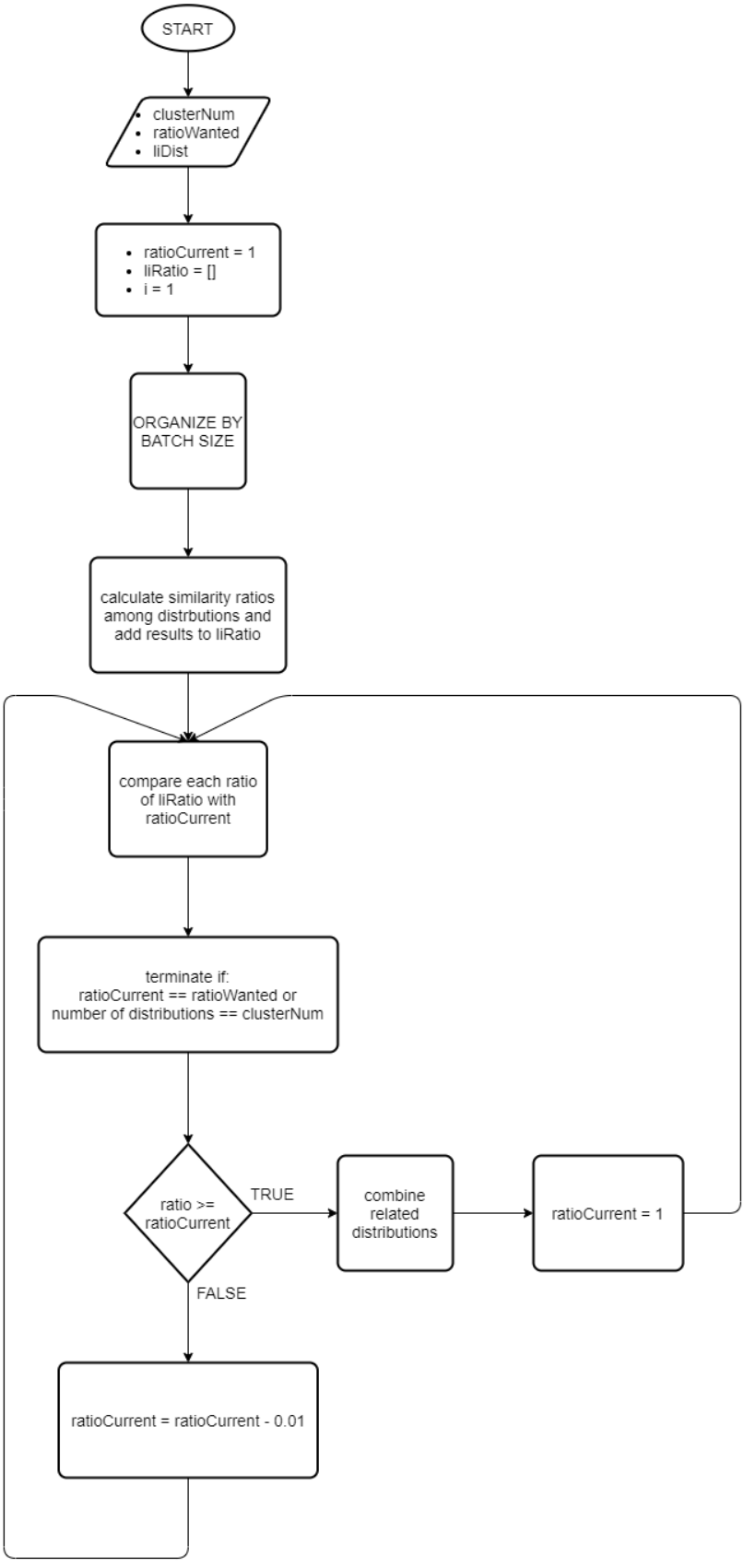

Figure 2.9 Flowchart of main flow of method two

2.7 Plotting

Both two methods generate HTML text outputs to illustrate visual graphics. These HTML outputs include Gaussian distributions of subgroups and indicate mean points of each subgroup. Subgroups and mean points are illustrated in different random colors to improve visual quality.

2.7.1 Python Programming

Parameters of the plotting method include width-height attributes of the graph, titles of the axis, and the final version of Gaussian distributions.

```python
import numpy as np
import random

from matplotlib import colors as mcolors

from bokeh.plotting import figure
from bokeh.resources import CDN
from bokeh.embed import file_html
from bokeh.models.annotations import Title


__liColor = dict(**mcolors.TABLEAU_COLORS)
__liColorName = []

for colorObj in __liColor.items():
    __liColorName.append(colorObj[0])

def __drawHtml(args):
    titleAxisX = args["titleAxisX"]
    titleAxisY = args["titleAxisY"]
    titleOfGraph = args["titleGraph"]
    width = args["width"]
    height = args["height"]
    liDistributions = ["liDist"]
    sizeLiDistributions = len(liDistributions)

    p = figure(title=titleAxisY, title_location="left", plot_width=width, plot_height=height)
    p.add_layout(Title(text=titleAxisX, align="center"), "below")

    for i in range(sizeLiDistributions):
        liNextDistribution = liDistributions[i]
        sizeLiNextDistribution = len(liNextDistribution)

        liAge = []
        liResult = []

        for j in range(sizeLiNextDistribution):
            nextPoint = liNextDistribution[j]

            liAge.append(nextPoint[0])
            liResult.append(nextPoint[1])

        meanAge = np.mean(liAge)
        meanResult = np.mean(liResult)

        index = random.randint(10, len(__liColorName))
        color = __liColor[__liColorName[index]]

        # points
        p.circle(liAge, liResult, size=10, color=color)

        index = random.randint(10, len(__liColorName))
        color = __liColor[__liColorName[index]]

        # means
        p.circle(meanAge, meanResult, size=20, color=color)

        return file_html(p, CDN, titleOfGraph)
```

Figure 2.10 Code snippet of plotting

CHAPTER THREE
EXPERIMENTAL RESULTS

In this section, experiments that were performed with this system are covered. Both method one and method two were used to generate reference intervals with the data of biochemical analytes published publicly by CALIPER study (Colantonio 2012). The number of clusters is needed to be defined in both methods before starting the learning process and it is certainly hard to determine how many different clusters the data includes in the first place. To overcome this difficulty, the number of clusters identified by the CALIPER study was used for each analyte. 11 biochemical tests that are used often were selected by biochemistry experts for the testing phase of these experiments. According to mean and standard deviation values of the distribution of data, the confidence intervals that are defined as 95% were calculated by the learning methods. The calculated values and generated reference intervals of each cluster are illustrated in tables. Outputs of method one and method two were compared with the CALIPER study individually to validation and also were compared with each other. By comparing the outputs of the methods, the performances were examined.

3.1 CALIPER

The following tables illustrate age groups, gender, and reference interval values of 11 analytes published publicly by the CALIPER study. The reference interval values listed as maximum and minimum separately male and female subgroups. In the notation of tables days and years are specified as age information where the abbreviations:

- "d" stands for days,
- "yr" stands for year, and
- "yrs" stands for years.

Table 3.1 CALIPER albumin g reference intervals

Albumin-G (g/L)	Female Min	Male Min	Female Max	Male Max
0 -< 15 d	33		45	
15 d -< 1 yr	28		47	
1 -< 8 yrs	38		47	
8 -< 15 yrs	41		48	
15 -< 19 yrs	40	41	49	51

Table 3.1 shows that the CALIPER study specifies the reference intervals of the test Albumin G as 5 different clusters by age for both female and male patients. It also is seen that gender is a determinant factor in this test for the age of 15 and the ages that greater than 15.

Table 3.2 CALIPER alp reference intervals

ALP (U/L)	Female Min	Male Min	Female Max	Male Max
0 -< 15 d	90		273	
15 d -< 1 yr	134		518	
1 -< 10 yrs	156		369	
10 -< 13 yrs	141		460	
13 -< 15 yrs	62	127	280	517
15 -< 17 yrs	54	89	128	365
17 -< 19 yrs	48	59	95	164

Table 3.2 shows that the CALIPER study specifies the reference intervals of the test ALP as 7 different clusters by age for both female and male patients. It also is seen that gender is a determinant factor in this test for the age of 13 and the ages that greater than 13.

Table 3.3 CALIPER alt reference intervals

ALT (U/L)	Female Min	Male Min	Female Max	Male Max
0 -< 1 yr	5		33	
1 -< 13 yrs	9		25	
13 -< 19 yrs	8	9	22	24

Table 3.3 shows that the CALIPER study specifies the reference intervals of the test ALT as 3 different clusters by age for both female and male patients. It also is seen that gender is a determinant factor in this test for the age of 13 and the ages that greater than 13.

Table 3.4 CALIPER calcium reference intervals

Calcium (mmol/L)	Female Min	Male Min	Female Max	Male Max
0 -< 1 yr	2.13		2.74	
1 -< 19 yrs	2.29		2.63	

Table 3.4 shows that the CALIPER study specifies the reference intervals of the test Calcium as 2 different clusters by age for both female and male patients. It also is seen that gender is not a determinant factor in this test.

Table 3.5 CALIPER cholesterol reference intervals

Cholesterol (mmol/L)	Female Min	Male Min	Female Max	Male Max
0 -< 15 d	1.2	1.1	3.23	2.82
15 d -< 1 yr	1.66		6.13	
1 -< 19 yrs	2.9		5.4	

Table 3.5 shows that the CALIPER study specifies the reference intervals of the test Cholesterol as 3 different clusters by age for both female and male patients. It

also is seen that gender is not a determinant factor in this test for the 15 days and the ages that greater than 15 days.

Table 3.6 CALIPER creatinine-jaffe reference intervals

Creatinine-Jaffe (µmol/L)	Female Min	Male Min	Female Max	Male Max
0 -< 15 d	37		93	
15 d -< 1 yr	28		47	
1 -< 4 yrs	34		48	
4 -< 7 yrs	39		57	
7 -< 12 yrs	46		61	
12 -< 15 yrs	50		71	
15 -< 17 yrs	52	58	76	92
17 -< 19 yrs	53	61	78	97

Table 3.6 shows that the CALIPER study specifies the reference intervals of the test CreatineJaffe as 8 different clusters by age for both female and male patients. It also is seen that gender is a determinant factor in this test for the age of 15 and the ages that greater than 15.

Table 3.7 CALIPER crp reference intervals

CRP (mg/L)	Female Min	Male Min	Female Max	Male Max
0 -< 15 d	0.3		6.1	
15 d -< 15 yrs	0.1		1	
15 -< 19 yrs	0.1		1.7	

Table 3.7 shows that the CALIPER study specifies the reference intervals of the test CRP as 3 different clusters by age for both female and male patients. It also is seen that gender is not a determinant factor in this test.

Table 3.8 CALIPER iga reference intervals

IgA (g/L)	Female Min	Male Min	Female Max	Male Max
0 -< 1 yr	0		0.3	
1 -< 3 yrs	0		0	
3 -< 6 yrs	0.3		1.5	
6 -< 14 yrs	0.5		2.2	
14 -< 19 yrs	0.5		2.9	

Table 3.8 shows that the CALIPER study specifies the reference intervals of the test IgA as 5 different clusters by age for both female and male patients. It also is seen that gender is not a determinant factor in this test.

Table 3.9 CALIPER iron reference intervals

Iron (µmol/L)	Female Min	Male Min	Female Max	Male Max
0 -< 14 yrs	2.8		22.9	
14 -< 19 yrs	3.5	5.5	29	30

Table 3.9 shows that the CALIPER study specifies the reference intervals of the test Iron as 2 different clusters by age for both female and male patients. It also is seen that gender is a determinant factor in this test for the age of 14 and the ages that greater than 14.

Table 3.10 CALIPER lipase reference intervals

Lipase (U/L)	Female Min	Male Min	Female Max	Male Max
0 -< 19 yrs	4		39	

Table 3.10 shows that the CALIPER study specifies the reference intervals of the test Lipase as 1 cluster by age for both female and male patients. It also is seen that gender is not a determinant factor in this test.

Table 3.11 CALIPER triglycerides reference intervals

Triglycerides (mmol/L)	Female Min	Male Min	Female Max	Male Max
0 -< 15 d	0.93		2.93	
15 d -< 1 yr	0.6		2.92	
1 -< 19 yrs	0.5		2.23	

Table 3.11 shows that the CALIPER study specifies the reference intervals of the test Triglycerides as 3 different clusters by age for both female and male patients. It also is seen that gender is a determinant factor in this test for the 15 days and the ages that greater than 15 days.

3.2 Method One vs CALIPER

In the experiment phase of method one, data of females and males were tested individually to compare with the CALIPER study. The outputs of each test case were examined by biochemistry experts. To generate meaningful outputs, new test cases were created with different cluster numbers, the different sizes of data, and the data that was normalized in addition to the current test cases. At the end of several testing and examining processes, it is observed that the method one that was empowered with GMM had some disadvantages. Meaningful clusters and reference intervals for some biochemical analytes with the data gained from the CALIPER study could not had been generated. In the testing of Calcium and Albumin G analytes, the method one generated the best outputs.

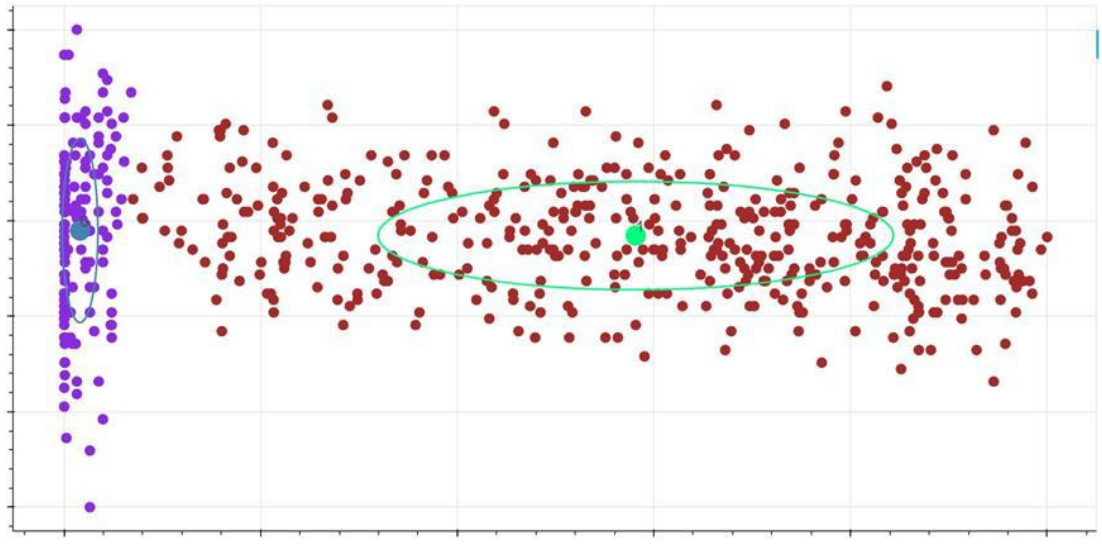

Figure 3.1 Method one calcium female graph

Table 3.12 Method one calcium female reference intervals

	Result (mmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 1 yr	2.4626	0.1449	2.1728	2.7524
1 -< 19 yrs	2.4514	0.0857	2.28	2.6229

In the figure 3.1, horizontal axis represents ages while vertical axis represents results. Table 3.12 shows the outputs of method one for the Calcium test of female patients. Method generated identical age ranges and close reference interval values with CALIPER study for female data.

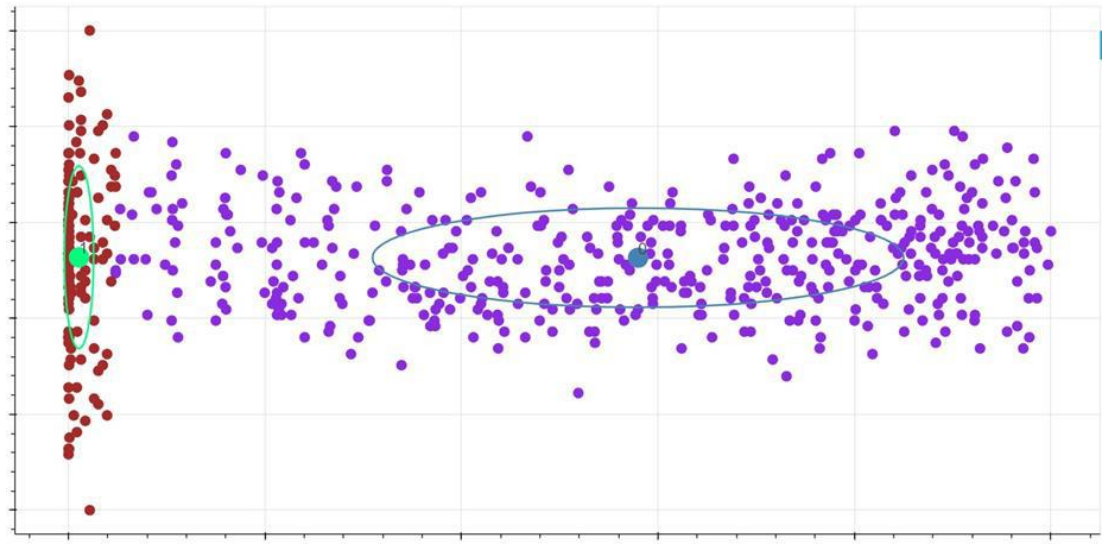

Figure 3.2 Method one calcium male graph

Table 3.13 Method one calcium male reference intervals

	Result (mmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 9 m	2.4537	0.1624	2.1288	2.7785
1 -< 19 yrs	2.4526	0.0889	2.2750	2.6302

In the figure 3.2, horizontal axis represents ages while vertical axis represents results. Table 3.13 shows the outputs of method one for the Calcium test of male patients. Method detected a bit difference in the one of the clusters and generated close reference intervals with CALIPER study for male data as with female data.

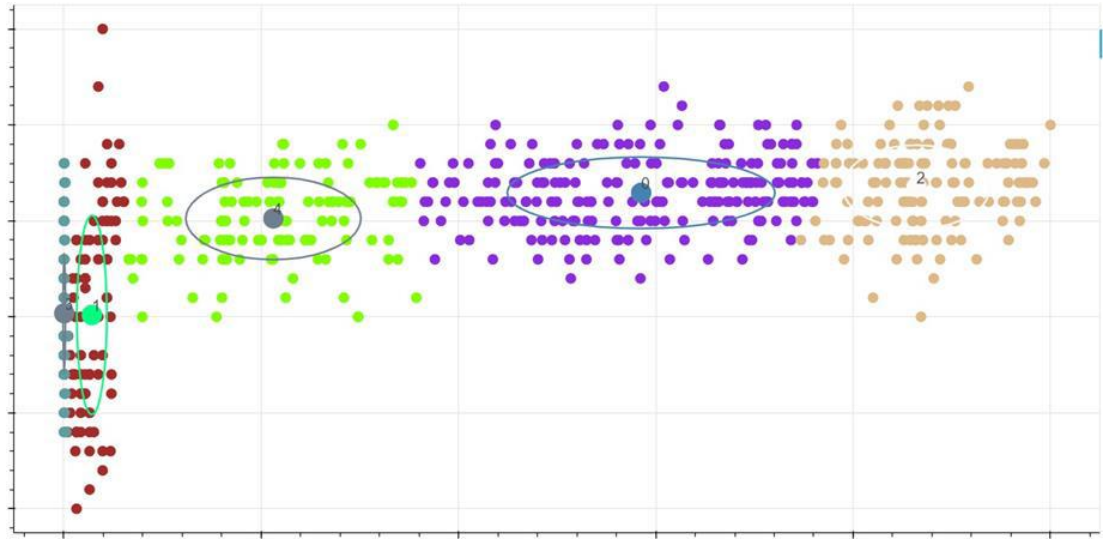

Figure 3.3 Method one albumin g female graph

Table 3.14 Method one albumin g female reference intervals

	Result (g/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 14 d	38.1760	3.1264	31.9231	44.4288
0 -< 13 m	38.3851	5.2054	27.9743	48.7958
14 m -< 7 yrs	43.1230	2.1599	38.8031	47.4428
7 -< 15 yrs	44.4734	1.8173	40.8389	48.1080
15 -< 19 yrs	44.7941	2.2033	40.3875	49.2008

In the figure 3.3, horizontal axis represents ages while vertical axis represents results. Table 3.14 shows the outputs of method one for the Albumin G test of female patients. Method shows that ages smaller than 1 year can be examined more detailed. Reference interval values are close with CALIPER study for female data.

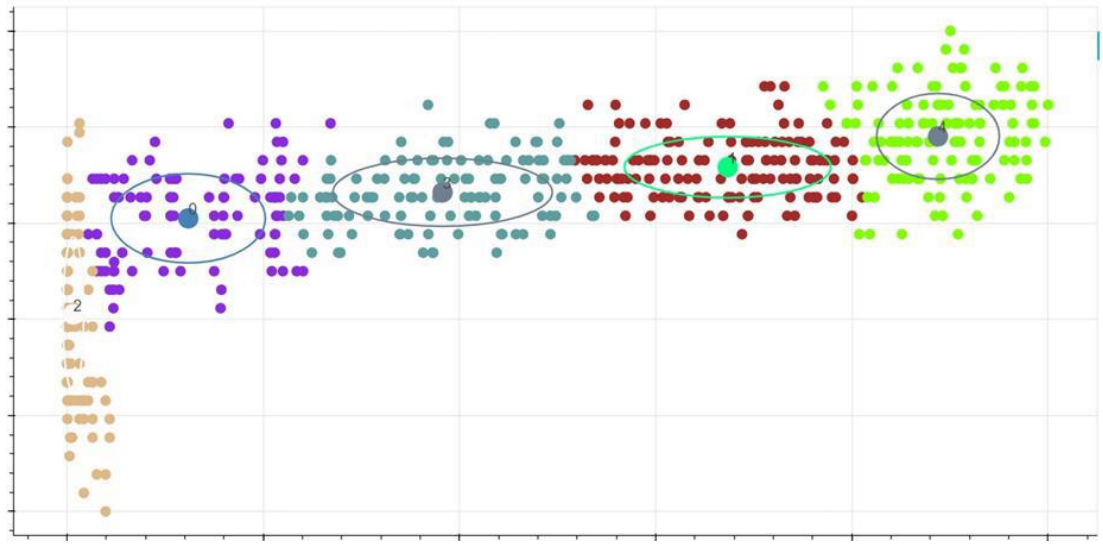

Figure 3.4 Method one albumin g male graph

Table 3.15 Method one albumin g male reference intervals

	Result (g/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 6 m	36.8152	4.2346	28.3461	45.2843
6 m -< 5 yrs	41.9242	2.4051	37.1140	46.7345
5 -< 10 yrs	43.2609	1.7648	39.7312	46.7905
10 -< 15 yrs	44.6383	1.6035	41.4312	47.8454
15 -< 19 yrs	46.2705	2.3260	41.6184	50.9226

In the figure 3.4, horizontal axis represents ages while vertical axis represents results. Table 3.15 shows the outputs of method one for the Albumin G test of male patients. Method shows that newborns up to 6 months can be examined more detailed.

In conclusion, this method generated different age ranges between female and male pediatric patient population for Albumin G biochemical test but reference interval values for both tests are compatible with CALIPER study.

3.3. Method Two vs CALIPER

Method two generated more suitable and meaningful reference intervals than method one. As with the experiment phase of method one, the data of females and males were tested individually to compare with the CALIPER study, in the experiment phase of method two. To generate meaningful outputs, new test cases were created with different cluster numbers, the different size of data, and different similarity ratios in addition to the current test cases. After the outputs of each test case were examined by biochemistry experts, the optimal similarity ratio was identified as 80% and cluster numbers were determined as identified in the CALIPER study. According to experiments that were performed by optimal parameters, compatible results with CALIPER study were generated and also high-resolution results were generated in some analytes.

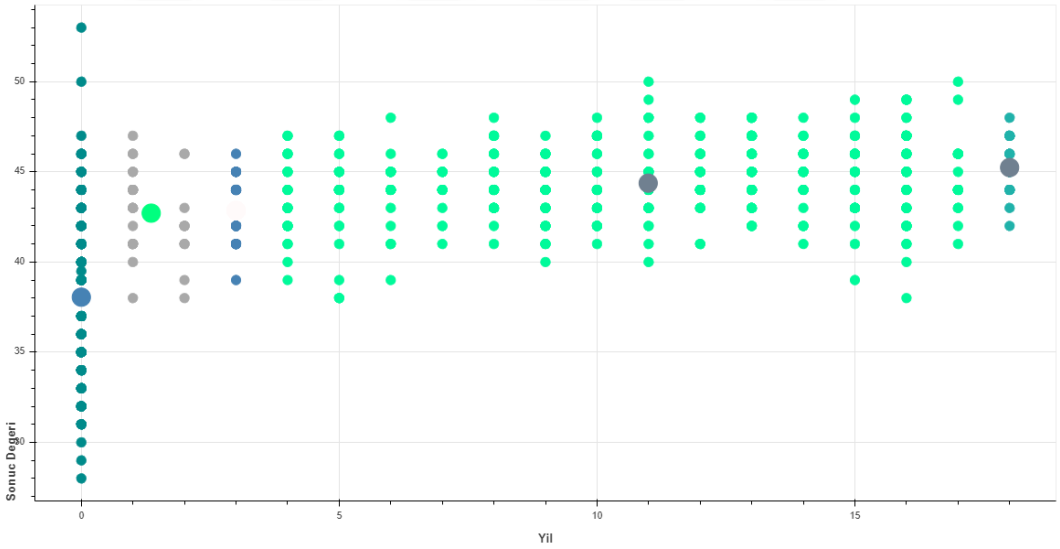

Figure 3.5 Method two albumin g female graph

Table 3.16 Method two albumin g female reference intervals

	Result (g/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	38.0505	4.0932	29.8642	46.2369

Tablo 3.16 continues

1 -< 3 yrs	42.7097	2.3167	38.0762	47.3432
3 yrs	42.8571	1.6194	39.6183	46.0959
4 -< 18 yrs	44.3724	2.0798	40.2128	48.5320
18 yrs	45.2381	1.4444	42.3494	48.1268

In the graph 3.5, horizontal axis represents ages while vertical axis represents resuts. Table 3.16 shows the outputs of the method two for Albumin G test of female patients. Method two generated same number of clusters and close reference interval values with the CALIPER study.

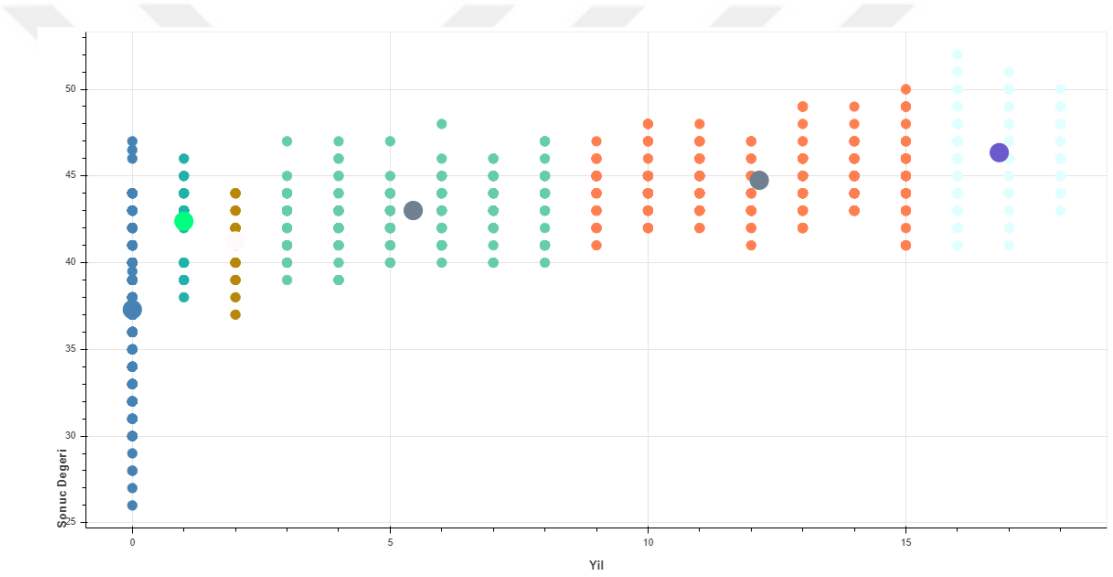

Figure 3.6 Method two albumin g male graph

Table 3.17 Method two albumin g male reference intervals

	Result (g/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	37.2956	4.2275	28.8406	45.7506
1 yr	42.3889	2.2395	37.9099	46.8679
2 yrs	41.25	1.9462	37.3577	45.1423
3 -< 9 yrs	43.0076	1.9009	39.2057	46.8095
9 -< 16 yrs	44.7487	1.8074	41.1339	48.3635
16 -< 18 yrs	46.3474	2.3250	41.6974	50.9974

In the figure 3.6, horizontal axis represents ages while vertical axis represents resuts. Table 3.17 shows the outputs of the method two for Albumin G test of male patients. On the contrary of female results, method two generated more detailed clusters and reference intervals than female results and CALIPER study.

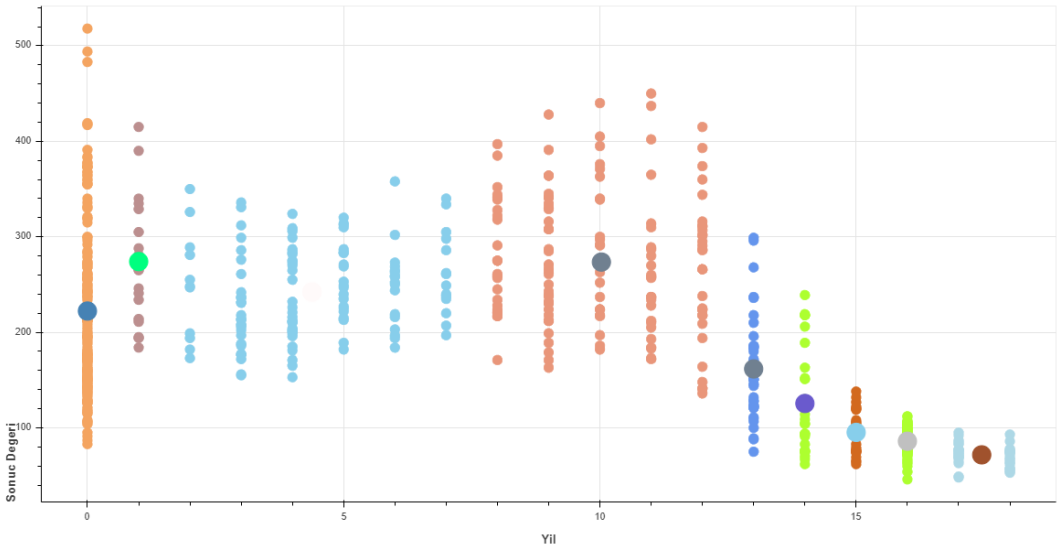

Figure 3.7 Method two alp female graph

Table 3.18 Method two alp female reference intervals

	Result (U/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	222.2711	88.5238	45.2235	399.3187
1 yr	274.0556	65.6476	142.7603	405.3508
2 -< 7 yrs	242.2033	47.9037	146.3959	338.0106
7 -< 13 yrs	273.5468	70.5703	132.4062	414.6873
13 yrs	161.6923	53.4070	54.8782	268.5065
14 yrs	125.5172	52.5512	20.4148	230.6169
15 yrs	95.25	22.5486	50.1529	140.3471
16 yrs	85.96	15.7657	54.4285	117.4915
17 -< 19 yrs	71.725	11.2405	49.2440	94.2060

In the figure 3.7, horizontal axis represents ages while vertical axis represents resuts. Table 3.18 shows the outputs of the method two for ALP test of female patients. Method two generated more detailed clusters and reference intervals than male results and CALIPER study.

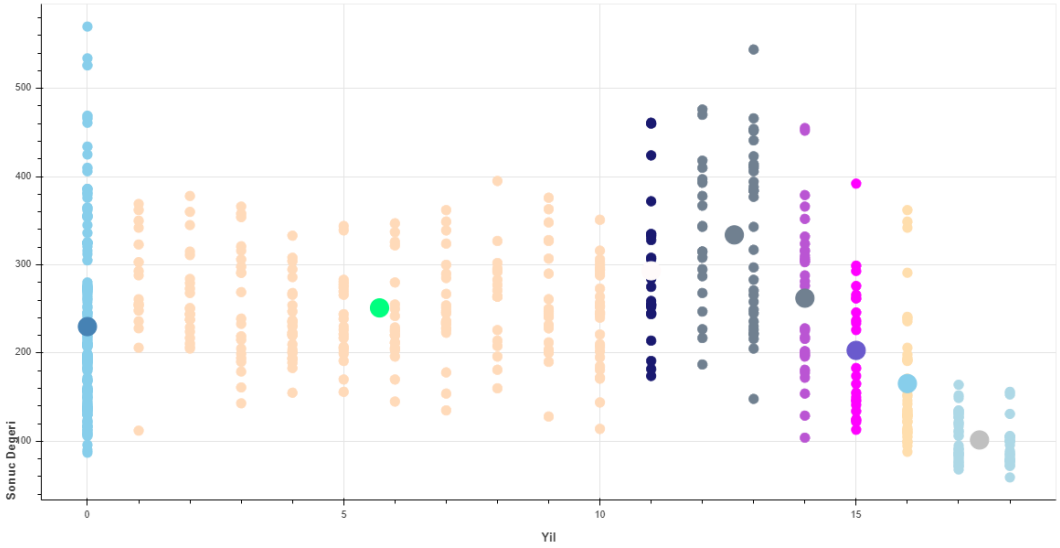

Figure 3.8 Method two alp male graph

Table 3.19 Method two alp male reference intervals

	Result (U/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	229.9338	102.4992	24.9355	434.9322
0 -< 11 yrs	251.1318	56.3498	138.4321	363.8315
11 yrs	293.2	75.2861	142.6278	443.722
12 -< 14 yrs	333.8868	91.3537	151.1794	516.5942
14 yrs	262.1818	84.0416	94.0985	430.2651
15 yrs	202.9583	71.7864	59.3855	346.5312
16 yrs	165.375	69.5204	26.3342	304.4158
17 -< 19 yrs	101.4444	25.7701	49.9042	152.9847

In the figure 3.8, horizontal axis represents ages while vertical axis represents resuts. Table 3.19 shows the outputs of the method two for ALP test of male patients. Method two generated more detailed clusters and reference intervals than CALIPER study.

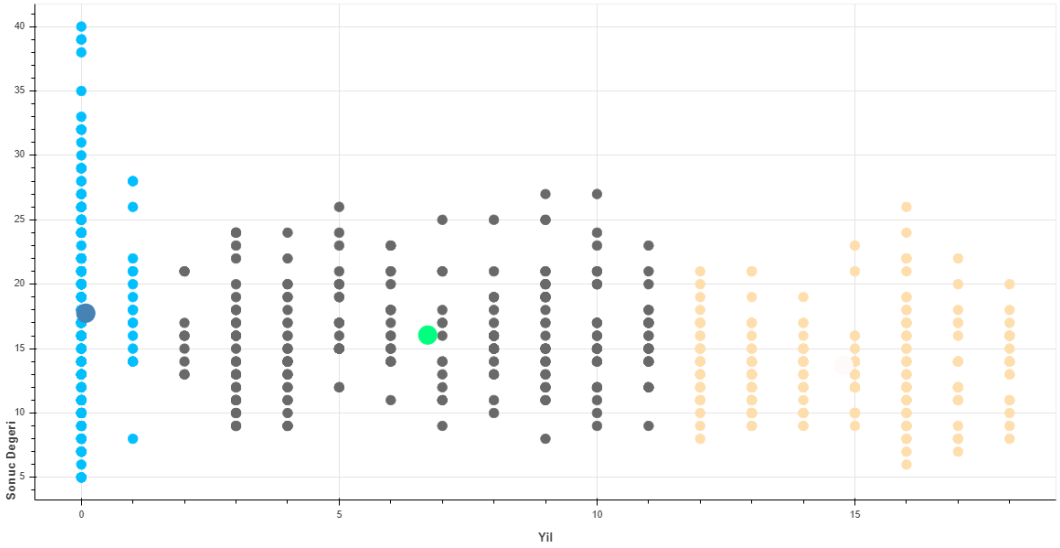

Figure 3.9 Method two alt female graph

Table 3.20 Method two alt female reference intervals

	Result (U/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 2 yrs	17.7488	6.8613	4.0263	31.4713
2 -< 12 yrs	16.0515	3.8787	8.2940	23.8090
12 -< 19 yrs	13.6842	3.4525	6.7793	20.5891

In the figure 3.9, horizontal axis represents ages while vertical axis represents resuts. Table 3.20 shows the outputs of the method two for ALT test of female patients. Method two generated same number of clusters and close reference intervals with the CALIPER study.

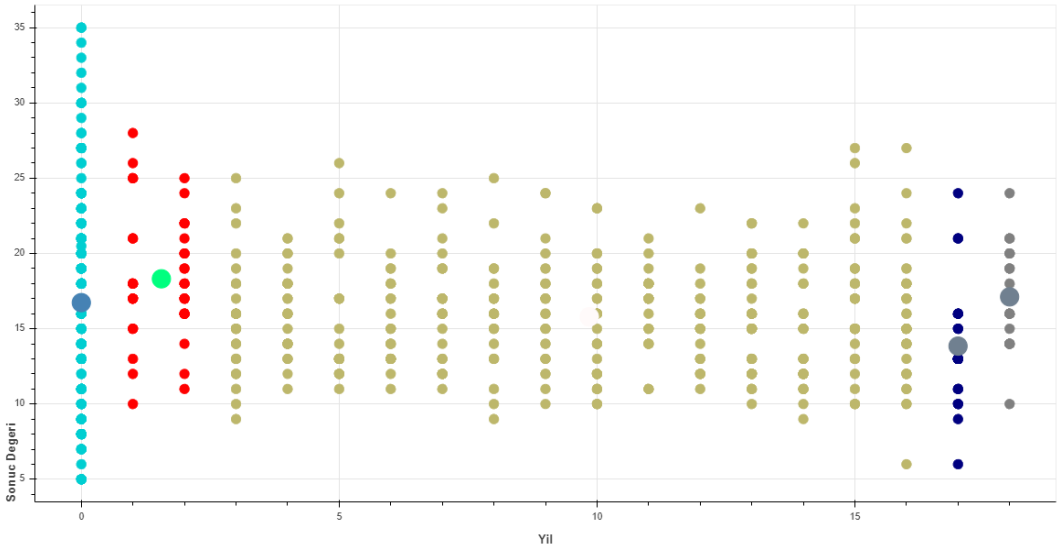

Figure 3.10 Method two alt male graph

Table 3.21 Method two alt male reference intervals

	Result (U/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	16.7327	6.8207	3.0914	30.3740
1 -< 3 yrs	18.3158	4.2246	9.8665	26.7651
3 -< 17 yrs	15.7930	3.6742	8.4447	23.1413
17 yrs	13.8462	3.7179	6.4104	21.2819
18 yrs	17.1177	3.4280	10.2617	23.9736

In the figure 3.10, horizontal axis represents ages while vertical axis represents resuts. Table 3.21 shows the outputs of the method two for ALT test of male patients. Method two generated more detailed clusters and reference intervals than female results and CALIPER study.

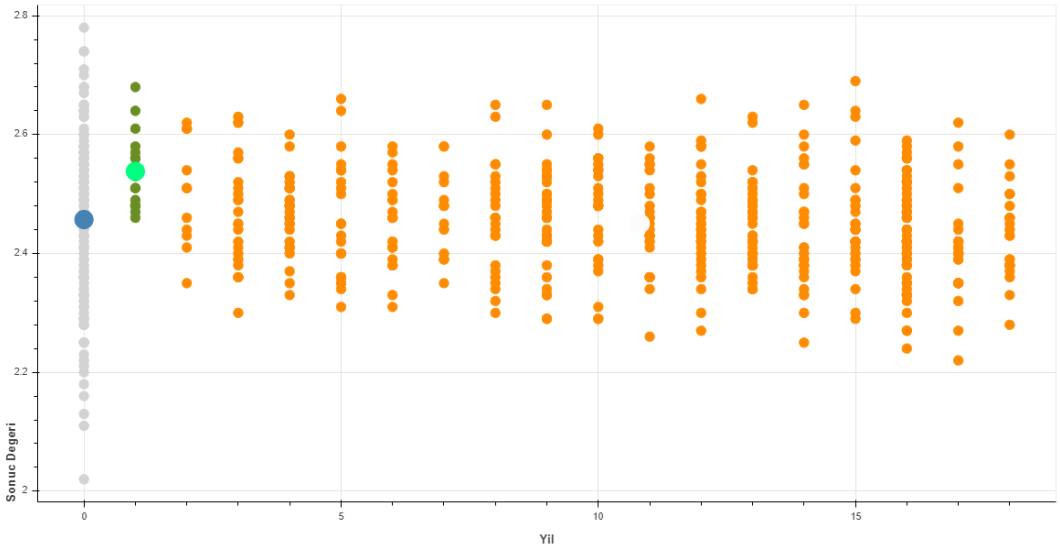

Figure 3.11 Method two calcium female graph

Table 3.22 Method two calcium female reference intervals

	Result (mmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	2.4570	0.1456	2.1658	2.7482
1 yr	2.5381	0.0634	2.4114	2.6649
2 -< 19 yrs	2.4501	0.0859	2.2784	2.6219

In the figure 3.11, horizontal axis represents ages while vertical axis represents resuts. Table 3.22 shows the outputs of the method two for Calcium test of female patients. Method two generated more detailed clusters and reference intervals than male results and CALIPER study.

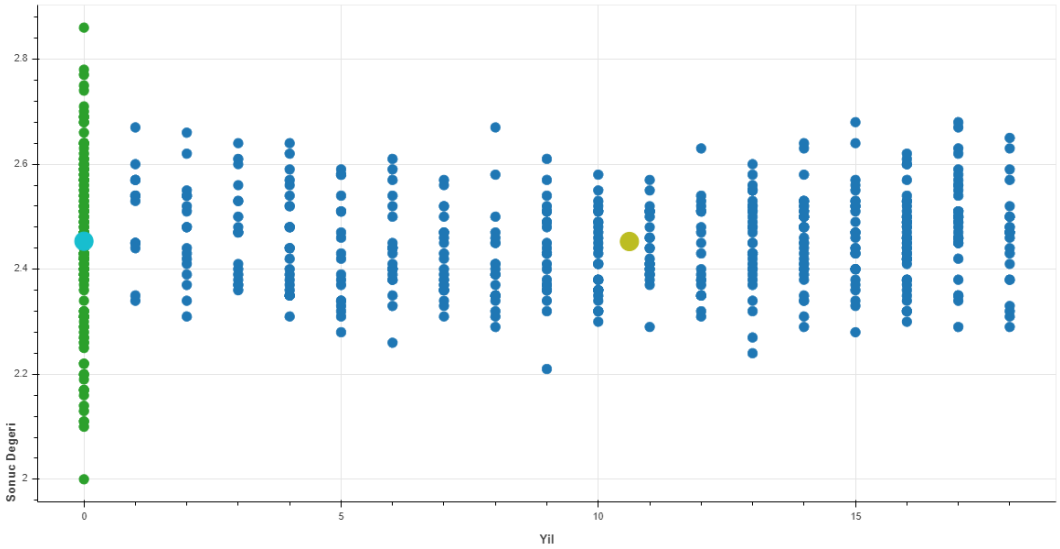

Figure 3.12 Method two calcium male graph

Table 3.23 Method two calcium male reference intervals

	Result (mmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	2.4532	0.1612	2.1309	2.7756
1 -< 19 yrs	2.4527	0.0890	2.2747	2.6307

In the figure 3.12, horizontal axis represents ages while vertical axis represents resuts. Table 3.23 shows the outputs of the method two for Calcium test of male patients. Method two generated same number of clusters and close reference intervals with the CALIPER study.

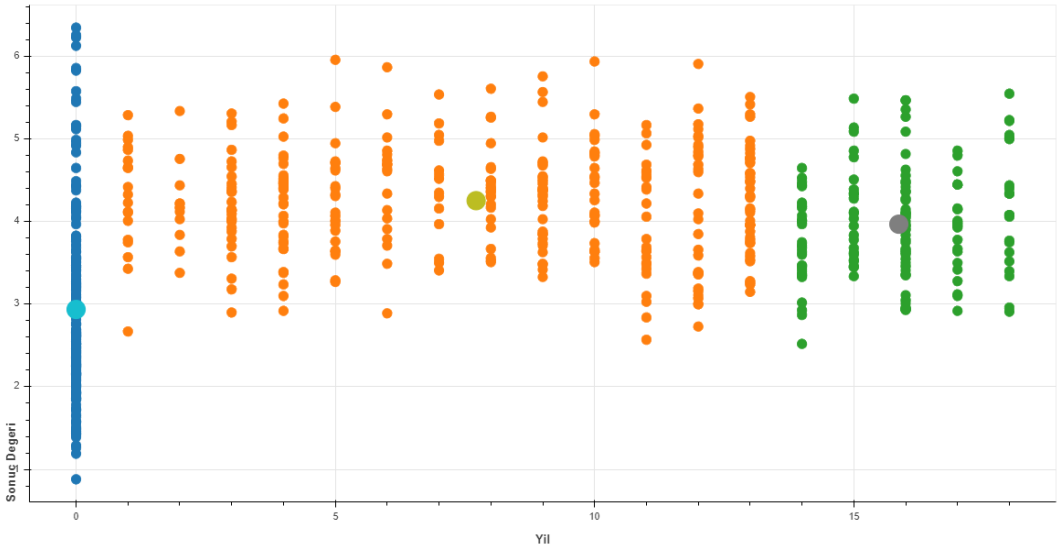

Figure 3.13 Method two cholesterol female graph

Table 3.24 Method two cholesterol male reference intervals

	Result (mmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	2.9371	1.1060	0.7251	5.1490
1 -< 14 yrs	4.2543	0.6492	2.9559	5.5526
14 -< 19 yrs	3.9694	0.6469	2.6756	5.2632

In the figure 3.13, horizontal axis represents ages while vertical axis represents resuts. Table 3.24 shows the outputs of the method two for Cholesterol test of female patients. Method two generated same number of clusters and close reference intervals with the CALIPER study.

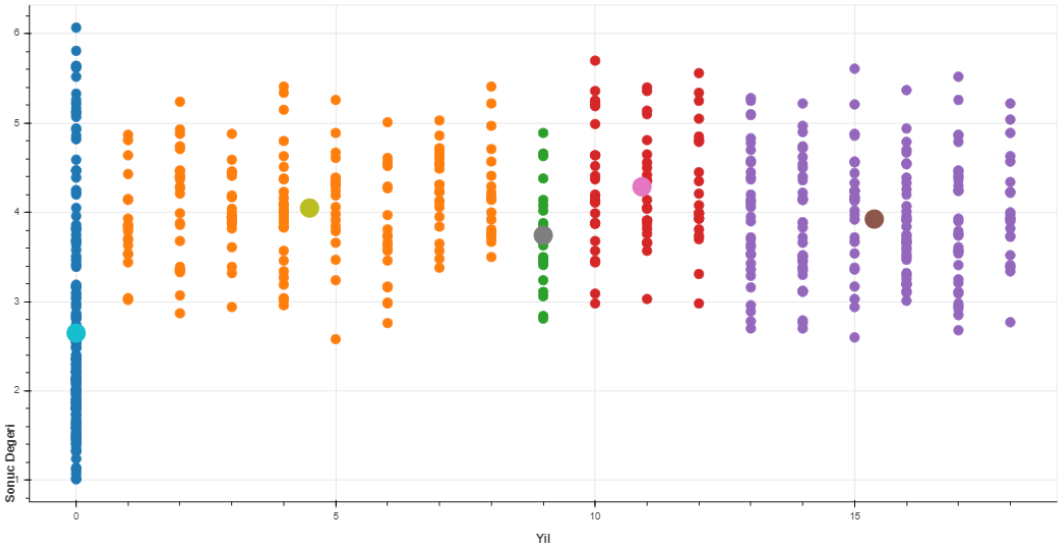

Figure 3.14 Method two cholesterol male graph

Table 3.25 Method two cholesterol male reference intervals

	Result (mmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	2.6484	1.1732	0.3019	4.9948
0 -< 9 yrs	4.0485	0.5759	2.8967	5.2003
9 yrs	3.7430	0.5475	2.6481	4.9380
10 -< 13 yrs	4.2868	0.6618	2.9632	5.6103
13 -< 19 yrs	3.9253	0.6641	2.5971	5.2535

In the figure 3.14, horizontal axis represents ages while vertical axis represents resuts. Table 3.25 shows the outputs of the method two for Cholesterol test of male patients. Method two generated more detailed clusters and reference intervals than female results and CALIPER study.

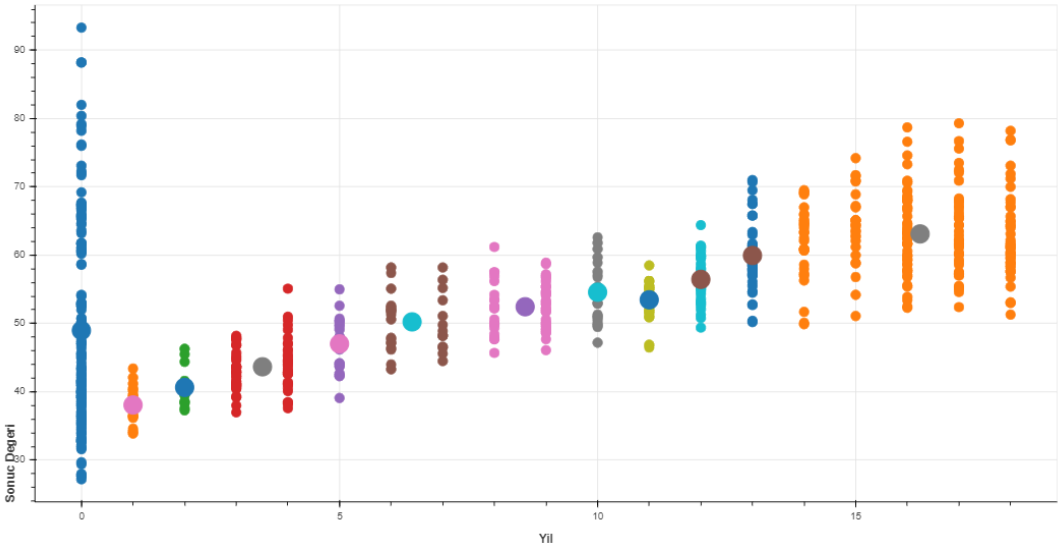

Figure 3.15 Method two creatinine-jaffe female graph

Table 3.26 Method two creatinine-jaffe female reference intervals

	Result (µmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	48.9842	14.6961	19.5920	78.3764
1 yr	38.075	2.8021	32.4708	43.6792
2 yrs	40.65	2.7789	35.0921	46.2079
3 -< 5 yrs	43.6651	3.5740	36.5171	50.8131
5 yrs	47.0231	3.4915	40.0402	54.0060
6 -< 8 yrs	50.2375	4.1030	42.0316	58.4434
8 -< 10 yrs	52.4509	3.5013	45.4483	59.4534
10 yrs	54.5783	4.2032	46.1719	62.9847
11 yrs	53.4667	2.7559	47.9550	58.9784
12 yrs	56.4567	3.6031	49.2504	63.6629
13 yrs	59.9548	4.9938	49.9672	69.9423
14 -< 19 yrs	63.1108	5.9747	51.1614	75.0602

In the figure 3.15, horizontal axis represents ages while vertical axis represents resuts. Table 3.26 shows the outputs of the method two for Cholesterol test of male

patients. Method two generated more detailed clusters and reference intervals than male results and CALIPER study.

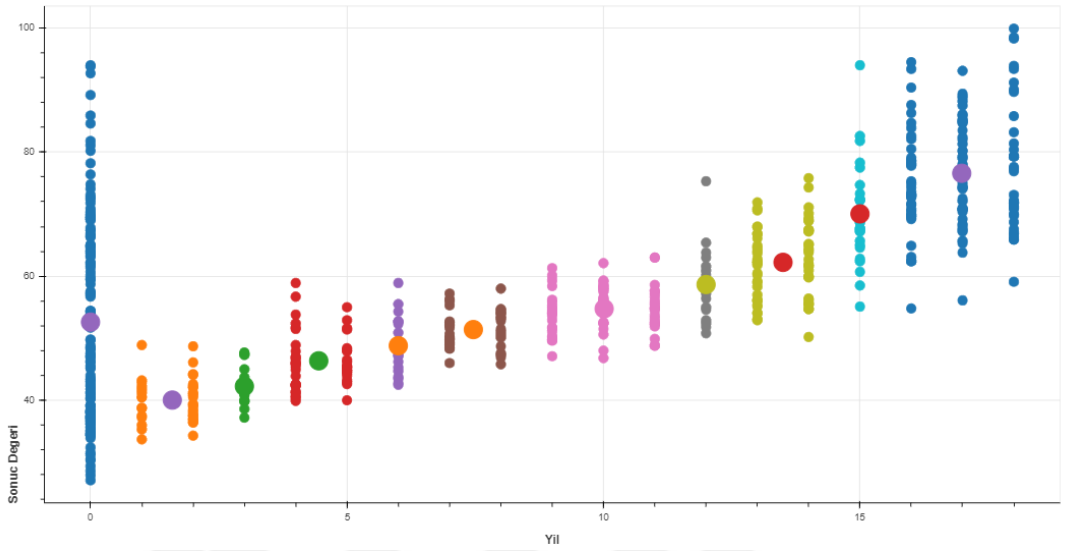

Figure 3.16 Method two creatinine-jaffe male graph

Table 3.27 Method two creatinine-jaffe male reference intervals

	Result (µmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	52.5952	16.2603	20.0745	85.1159
1 -< 3 yrs	40.0324	3.5411	32.9502	47.1147
3 yrs	42.2522	2.5968	37.0586	47.4458
4 -< 6 yrs	46.3571	4.3881	37.5810	55.1333
6 yrs	48.795	4.4447	39.9056	57.6844
7 -< 9 yrs	51.3952	3.0067	45.3818	57.4084
9 -< 12 yrs	54.7539	3.5771	47.5999	61.9081
12 yrs	58.6526	5.6343	47.3840	69.9213
13 -< 15 yrs	62.2266	5.8512	50.5242	73.9289
15 yrs	70.0429	7.7963	54.4502	85.6355
15 -< 19 yrs	76.5816	9.0478	58.4861	94.6771

In the figure 3.16, horizontal axis represents ages while vertical axis represents resuts. Table 3.27 shows the outputs of the method two for Cholesterol test of male patients. Method two generated more detailed clusters and reference intervals than CALIPER study.

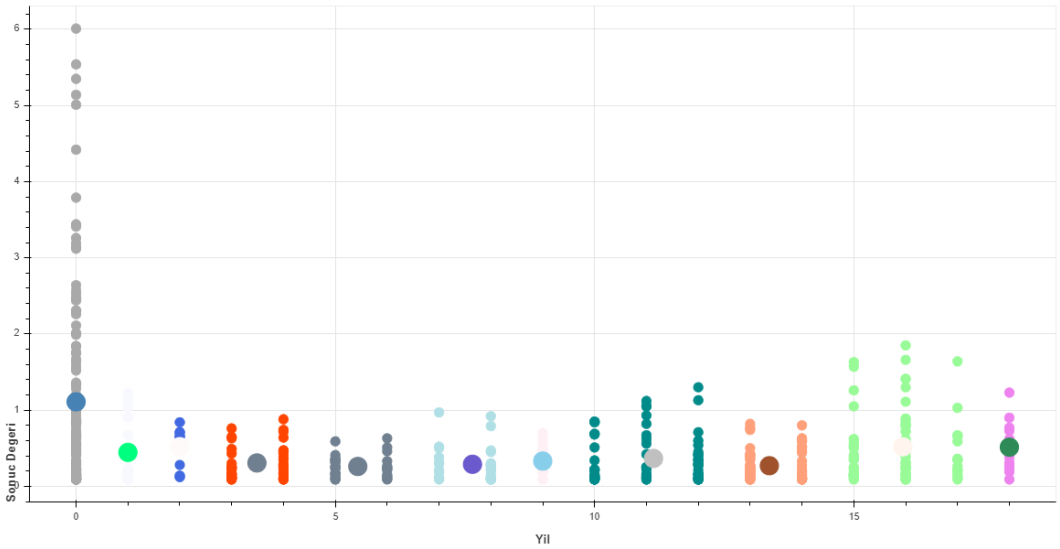

Figure 3.17 method two crp female graph

Table 3.28 Method two crp female reference intervals

	Result (mg/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	1.1090	1.2627	1.2627	3.6344
1 yr	0.4453	0.4073	0.4073	1.26
2 yrs	0.52	0.2397	0.0405	0.9995
3 yrs	0.3063	0.2207	0.2207	0.7478
4 -< 7 yrs	0.26	0.1492	0.1492	0.5538
7 -< 9 yrs	0.2887	0.2375	0.2375	0.7637
9 yrs	0.3308	0.1579	0.0149	0.6467
10 -< 13 yrs	0.3656	0.3142	0.3142	0.9940
13 -< 15 yrs	0.27	0.1993	0.1993	0.6685
15 -< 18 yrs	0.5176	0.4517	0.4517	1.4210
18 yrs	0.5131	0.2823	0.2823	1.0778

In the figure 3.17, horizontal axis represents ages while vertical axis represents resuts. Table 3.28 shows the outputs of the method two for CRP test of female patients. Method two generated more detailed clusters and reference intervals than mae results and CALIPER study.

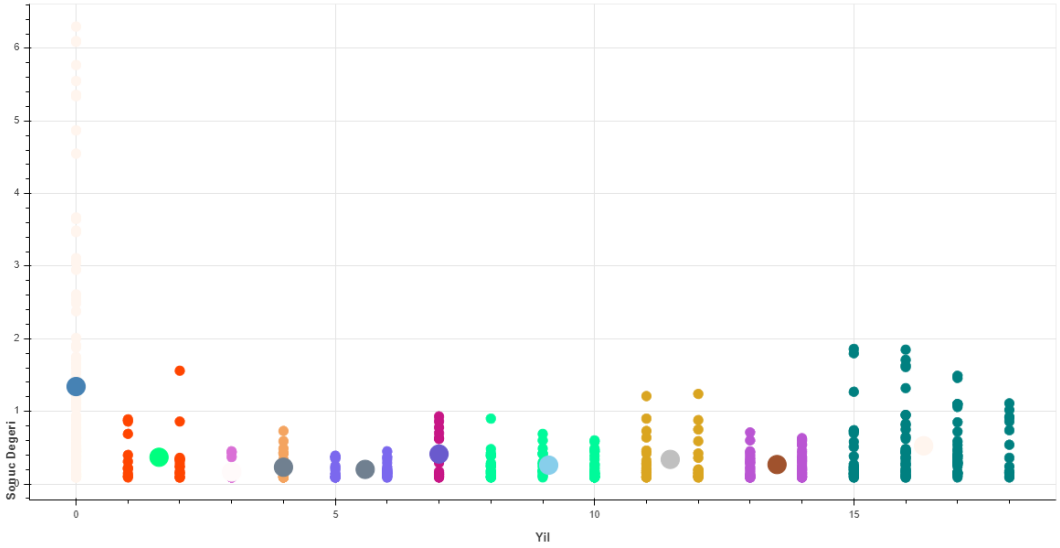

Figure 3.18 Method two crp male graph

Table 3.29 Method two crp male reference intervals

	Result (mg/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	0.3417	1.5294	1.5294	4.4005
1 -< 3 yrs	0.3676	0.3433	0.3433	1.0542
3 yrs	0.1647	0.1286	0.1286	0.4219
4 yrs	0.2304	0.1806	0.1806	0.5917
5 -< 7 yrs	0.2006	0.1049	0.1049	0.4103
7 yrs	0.4118	0.2783	0.2783	0.9684
8 -< 11 yrs	0.2587	0.1778	0.1778	0.6141
11 -< 13 yrs	0.3384	0.3125	0.3125	.9633
13 -< 15 yrs	0.2671	0.1739	0.1739	0.6149
15 -< 19 yrs	0.5256	0.4522	0.4522	1.43

In the figure 3.18, horizontal axis represents ages while vertical axis represents resuts. Table 3.29 shows the outputs of the method two for CRP test of male patients. Method two generated more detailed clusters and reference intervals than CALIPER study.

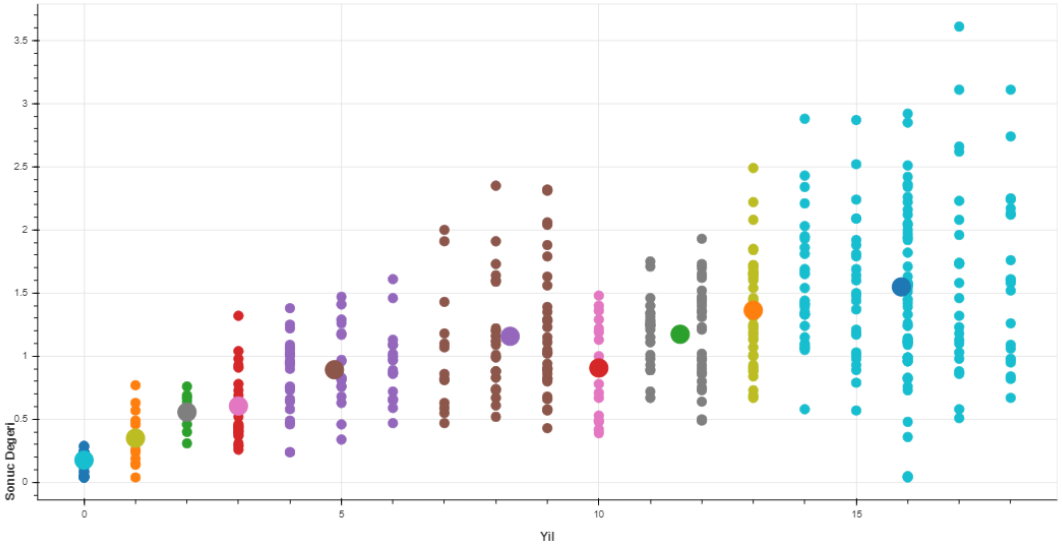

Figure 3.19 Method two iga female graph

Table 3.30 Method two iga female reference intervals

	Result (g/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	0.1773	0.0624	0.0526	0.3020
1 yr	0.3507	0.2006	0.3507	0.7519
2 yrs	0.557	0.1321	0.2929	0.8211
3 yrs	0.6055	0.2774	0.0506	1.1603
4 -< 7 yrs	0.8925	0.3034	0.2858	1.4993
7 -< 10 yrs	1.1574	0.4933	0.1707	2.1441
10 yrs	0.9067	0.3291	0.2484	1.5649
11 -< 13 yrs	1.1736	0.3426	0.4885	1.8587
13 yrs	1.3632	0.41	0.5432	2.1831
14 -< 19 yrs	1.5485	0.6404	0.2678	2.8292

In the figure 3.19, horizontal axis represents ages while vertical axis represents resuts. Table 3.30 shows the outputs of the method two for IgA test of female patients. Method two generated more detailed clusters and reference intervals than male results and CALIPER study.

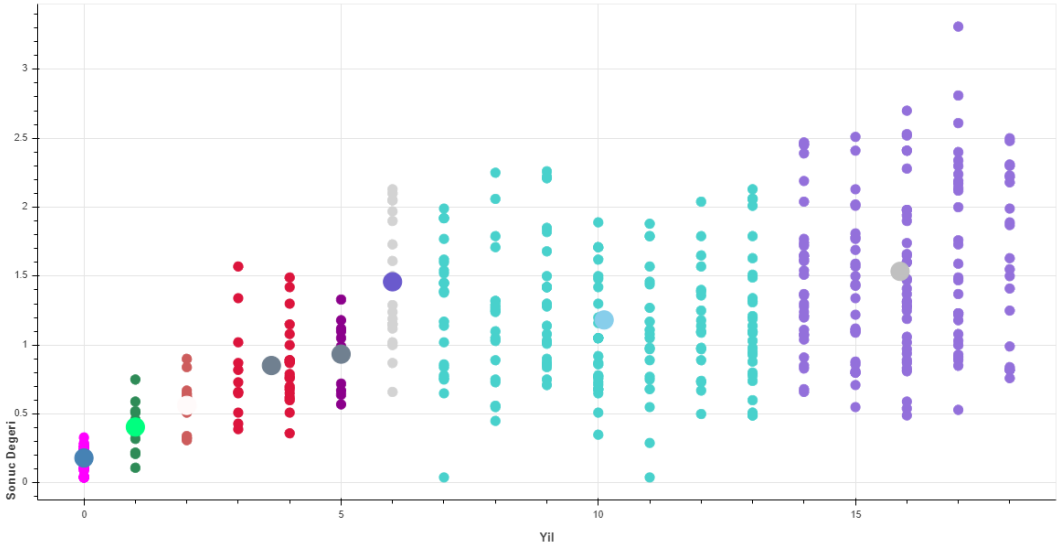

Figure 3.20 Method two iga male graph

Table 3.31 Method two iga male reference intervals

	Result (g/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	0.1809	0.0776	0.0258	0.3360
1 yr	0.4055	0.1781	0.0492	0.7617
2 yrs	0.5623	0.1864	0.1895	0.9351
3 -< 5 yrs	0.8516	0.3147	0.2222	1.4810
5 yrs	0.9343	0.2124	0.5096	1.3590
6 yrs	1.459	0.4448	0.5694	2.3486
7 -< 14 yrs	1.1822	0.4660	0.2502	2.1143
14 -< 19 yrs	1.5346	0.5907	0.3532	2.7159

In the figure 3.20, horizontal axis represents ages while vertical axis represents resuts. Table 3.31 shows the outputs of the method two for IgA test of male patients. Method two generated more detailed clusters and reference intervals than CALIPER study.

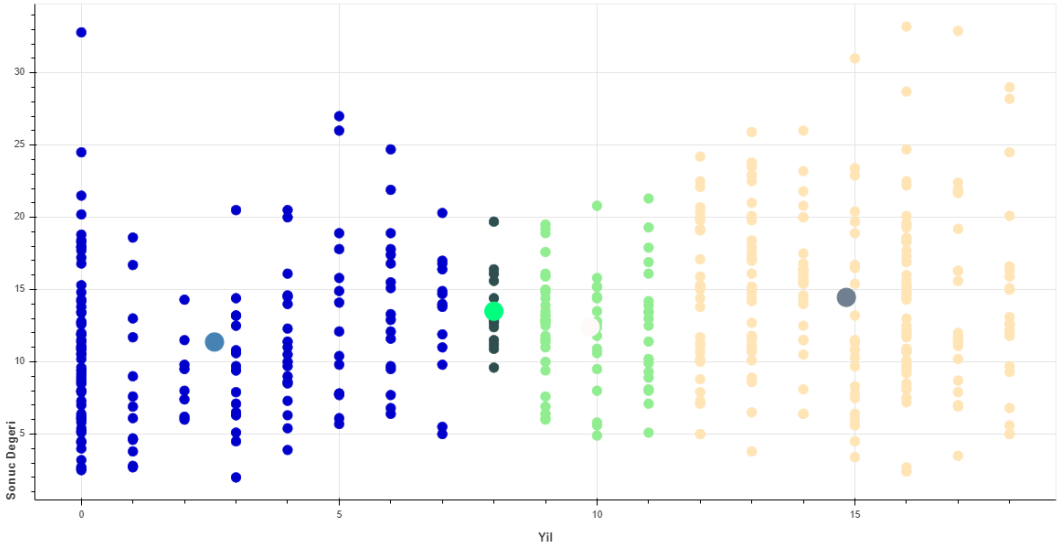

Figure 3.21 Method two iron female graph

Table 3.32 Method two iron female reference intervals

	Result (µmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 8 yrs	11.3709	5.4526	0.4657	22.2760
8 yrs	13.5	2.4114	8.6771	18.3229
9 -< 12 yrs	12.4104	3.9438	4.5229	20.2980
12 -< 19 yrs	14.4592	5.9476	2.5641	26.3544

In the figure 3.21, horizontal axis represents ages while vertical axis represents resuts. Table 3.32 shows the outputs of the method two for Iron test of female patients. Method two generated more detailed clusters and reference intervals than CALIPER study.

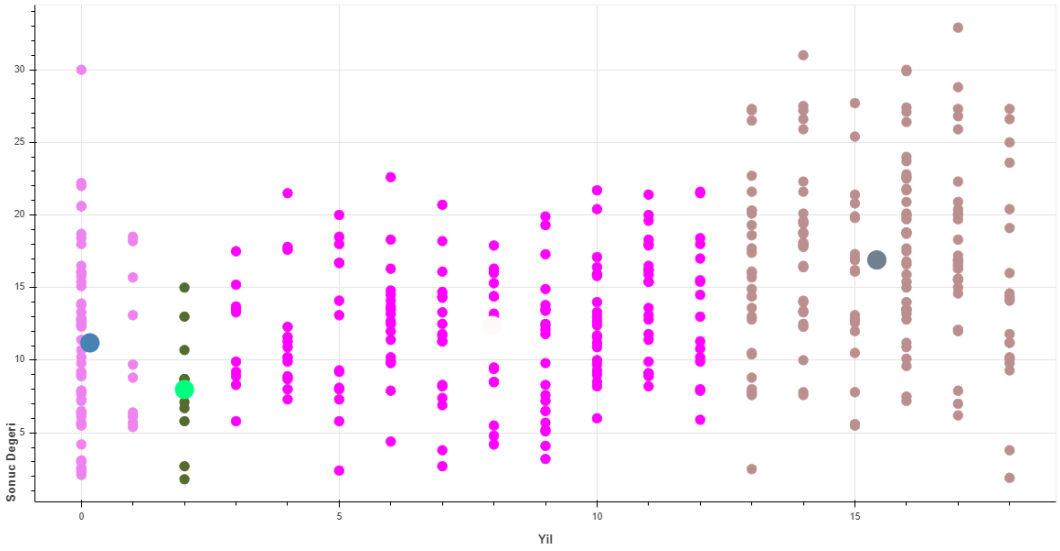

Figure 3.22 Method two iron male graph

Table 3.33 Method two iron male reference intervals

	Result (µmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 2 yrs	11.1933	5.9110	5.9110	23.0153
2 yrs	7.9818	3.7695	0.4429	15.5207
3 -< 13 yrs	2.3925	4.4330	3.5266	21.2585
13 -< 19 yrs	16.9024	6.2140	4.4745	29.3303

In the figure 3.22, horizontal axis represents ages while vertical axis represents resuts. Table 3.33 shows the outputs of the method two for Iron test of male patients. Method two generated more detailed clusters and reference intervals than CALIPER study.

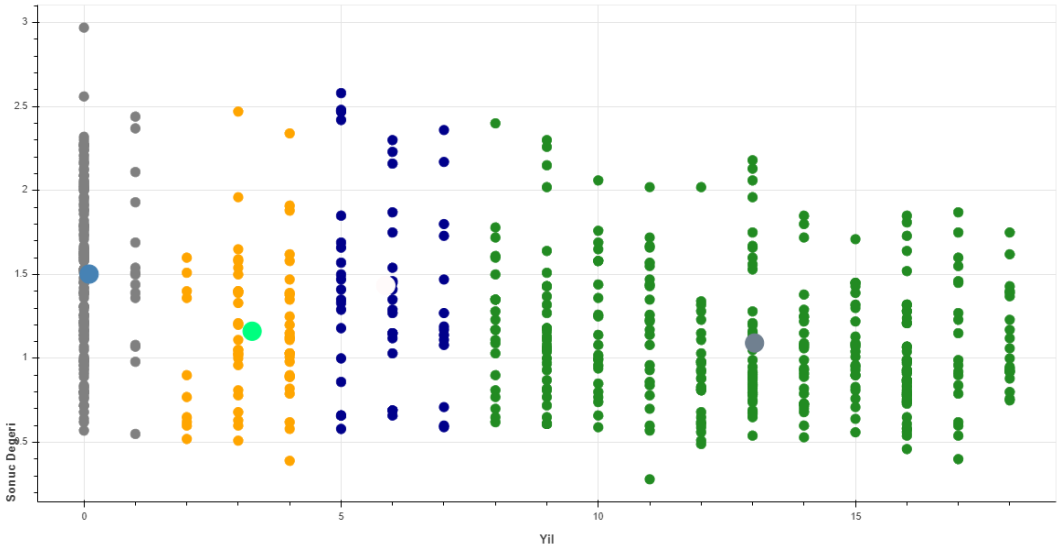

Figure 3.23 Method two triglycerides female graph

Table 3.34 Method two triglycerides female reference intervals

	Result (mmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 2 yrs	1.5021	0.4877	0.5267	2.4774
2 -< 5 yrs	1.1617	0.4338	0.2942	2.0293
5 -< 8 yrs	1.4341	0.5580	0.3181	2.55
8 -< 19 yrs	1.0917	0.3884	0.3150	1.8685

In the figure 3.23, horizontal axis represents ages while vertical axis represents resuts. Table 3.34 shows the outputs of the method two for Triglycerides test of female patients. Method two generated more detailed clusters and reference intervals than CALIPER study.

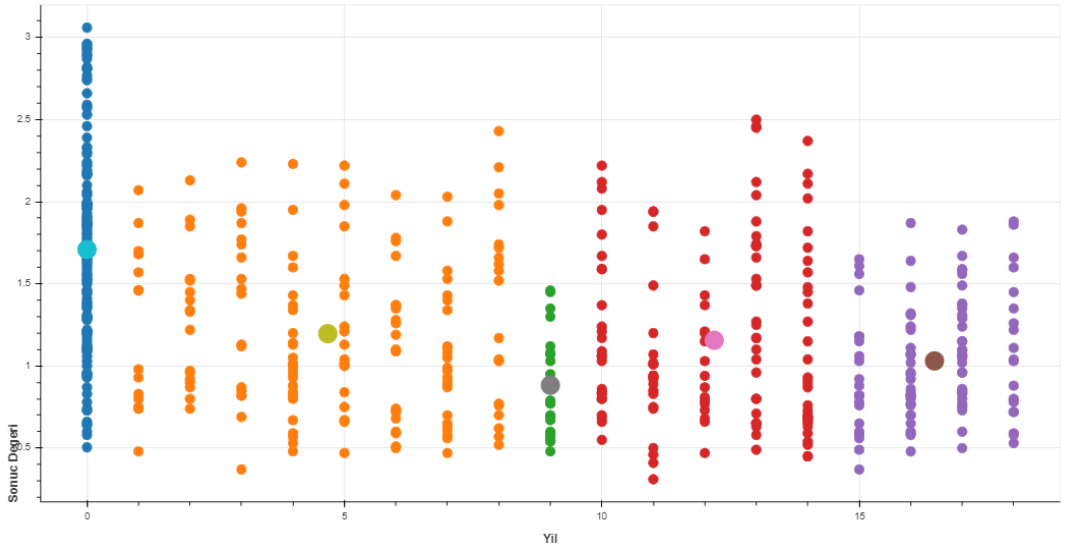

Figure 3.24 Method two triglycerides male graph

Table 3.35 Method two triglycerides male reference intervals

	Result (mmol/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 yr	1.7090	0.6174	0.4742	2.9438
0 -< 9 yrs	1.1968	0.4818	0.2331	2.1604
9 yrs	0.8829	0.3090	0.2648	1.5009
10 -< 15 yrs	1.1564	0.5266	0.1032	2.2097
15 -< 19 yrs	1.0309	0.3581	0.3148	1.7471

In the figure 3.24, horizontal axis represents ages while vertical axis represents resuts. Table 3.35 shows the outputs of the method two for Triglycerides test of male patients. Method two generated more detailed clusters and reference intervals than female results and CALIPER study.

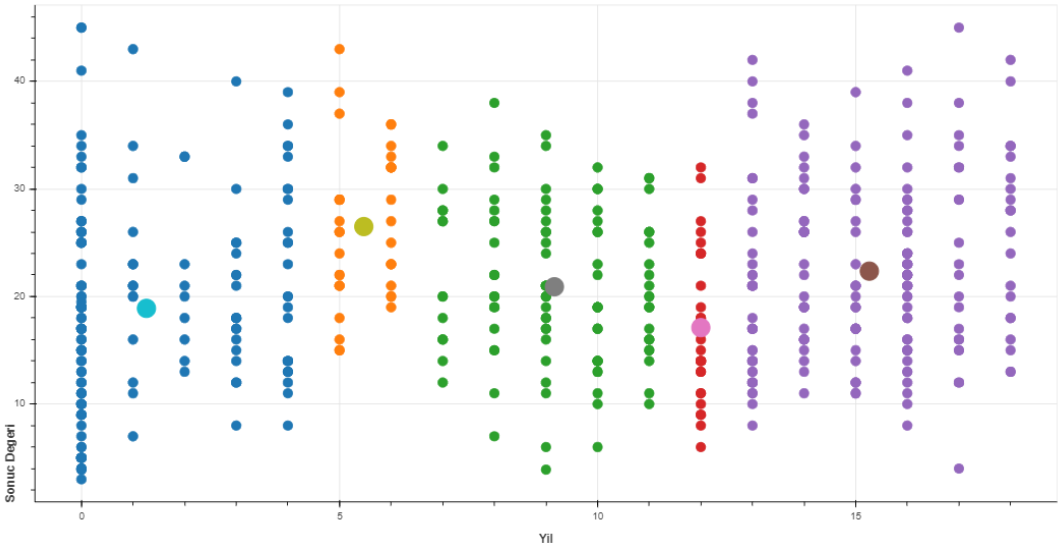

Figure 3.25 Method two lipase female graph

Table 3.36 Method two lipase female reference intervals

	Result (U/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 5 yrs	18.8953	9.3462	0.2030	37.5877
5 -< 7 yrs	26.5	7.0429	12.4141	40.5859
7 -< 19 yrs	21.3676	7.9085	5.5508	37.1847

In the figure 3.25, horizontal axis represents ages while vertical axis represents resuts. Table 3.36 shows the outputs of the method two for Lipase test of female patients. Method two generated more detailed clusters and reference intervals than CALIPER study.

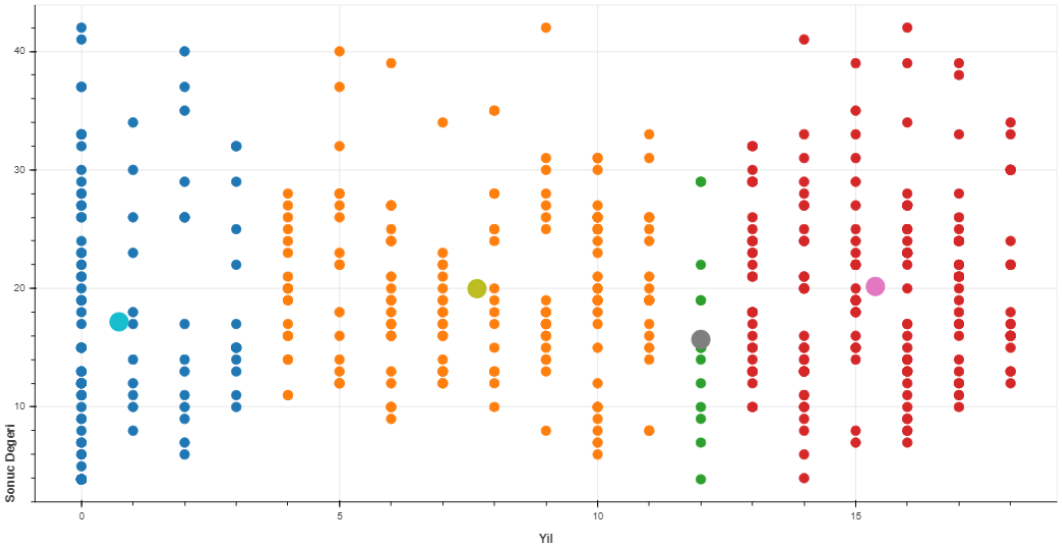

Figure 3.26 Method two lipase male graph

Table 3.37 Method two lipase male reference intervals

	Result (U/L)			
Age Range	**Mean**	**Std**	**Min**	**Max**
0 -< 4 yrs	17.1822	10.3303	10.3303	37.8429
4 -< 12 yrs	19.9810	7.0589	5.8631	34.0989
12 yrs	15.7	6.5344	2.6311	28.7689
13 -< 19 yrs	20.1676	7.7959	4.5759	35.7593

In the figure 3.26, horizontal axis represents ages while vertical axis represents resuts. Table 3.37 shows the outputs of the method two for Lipase test of male patients. Method two generated more detailed clusters and reference intervals than female results and CALIPER study.

3.2.1 A Visual Sample of The Steps for The Combination of Distributions

In this section, the combination process that is mentioned in the implementation section of 2 Gaussian distributions is presented visually. ALT female test iterations were used to provide the visual sample. There are 13 iterations for the ALT female

test. Current distributions of each iteration step and Gaussian similarity ratios between each 2 of distributions are displayed in the following figures. Optimal similarity ratio were used in this sample.

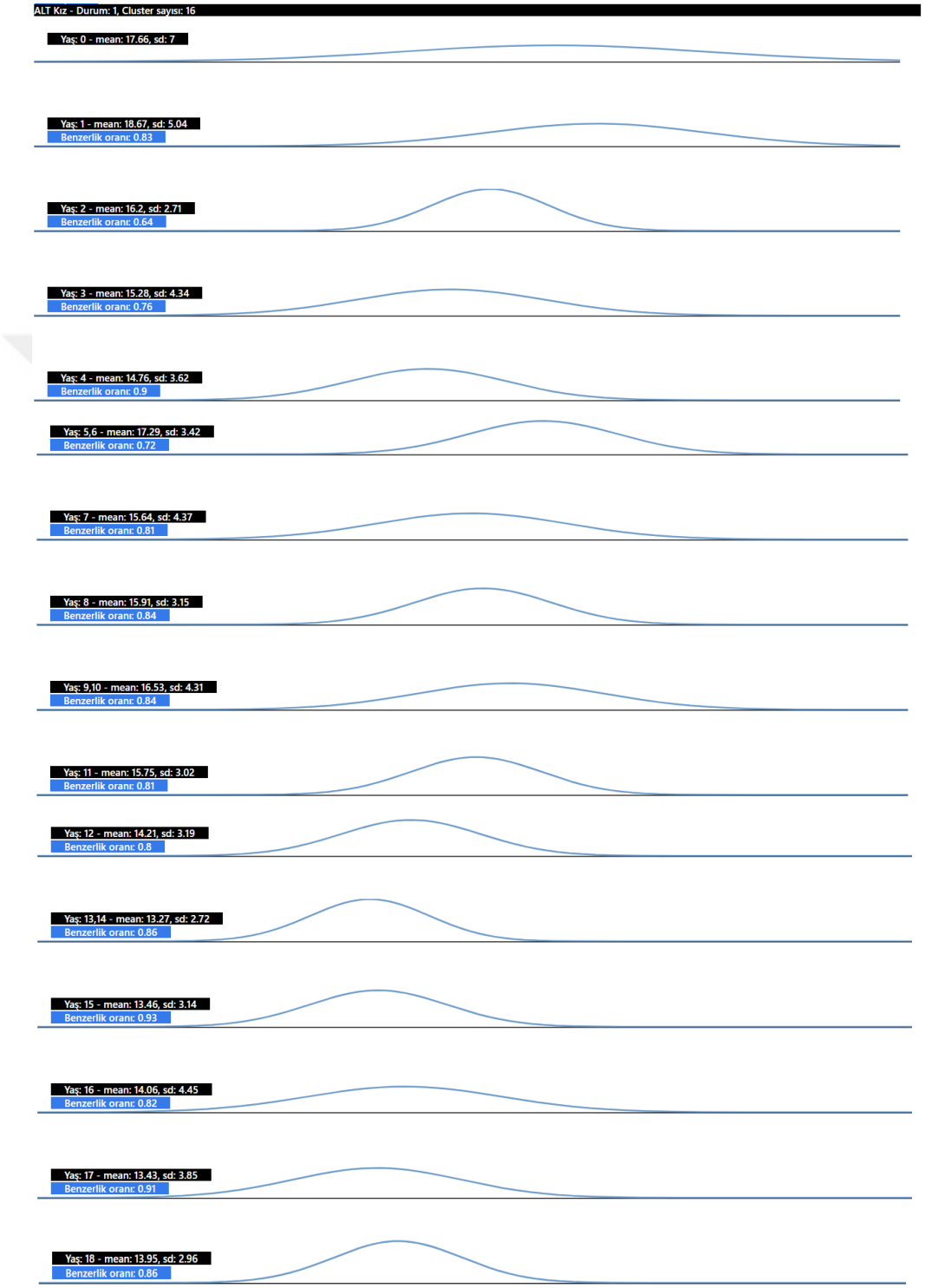

Figure 3.27 Learning iteration 1 of alt female test

In the figure 3.27 the initial distributions by age of the data are displayed. The highest similarity ratio is 93% between age groups 13-14 and 15. That means this subgroups are combined and a new subgroup will be created.

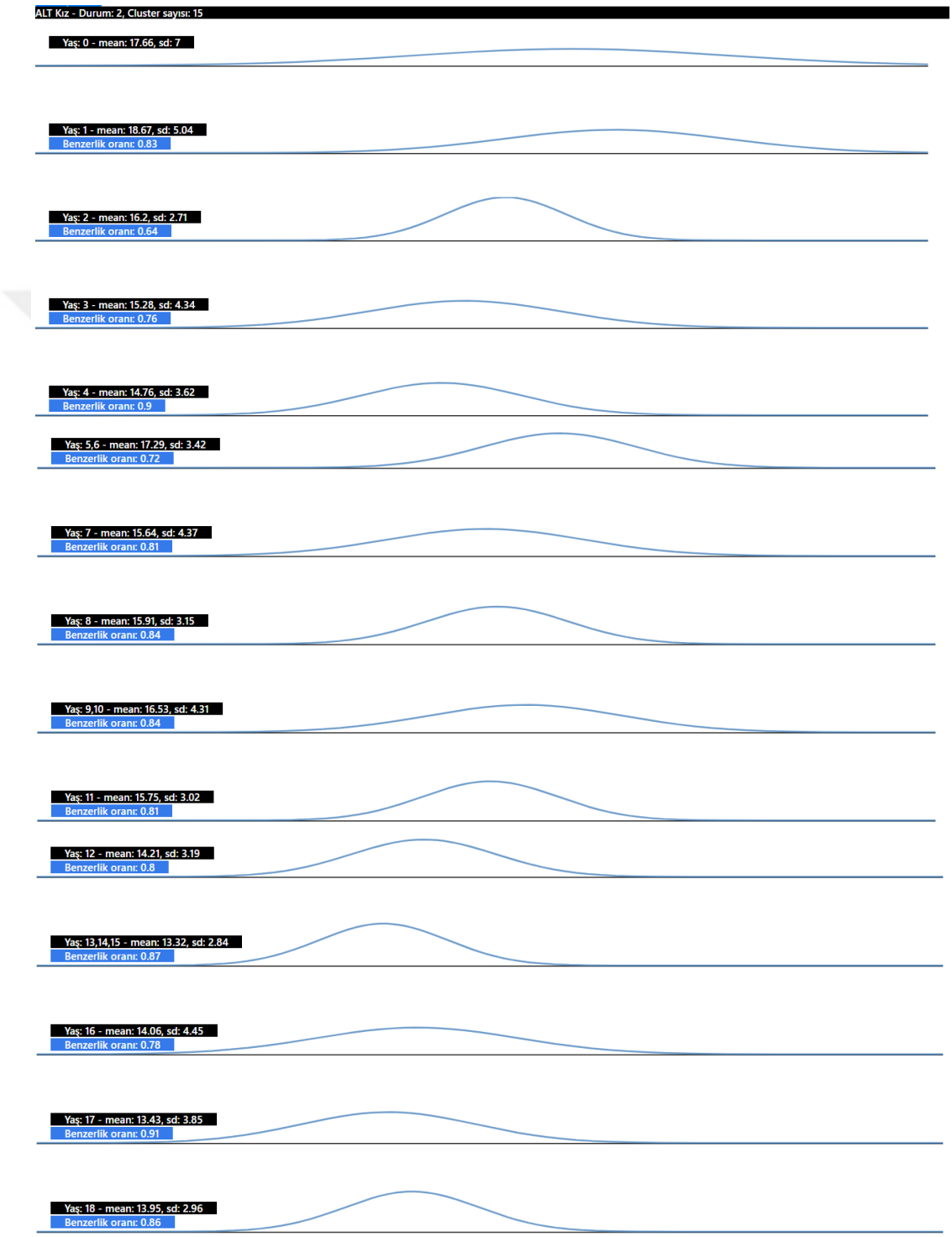

Figure 3.28 Learning iteration 2 of alt female test

In the figure 3.28 it is seen that a new subgroup was created 13-14-15by combining subgroups in the previous iteration. The highest similarity ratio is 91% between age groups 16 and 17 so, a new subgroup will be created by combining these subgroups.

At the end of the iterations, the final version of the combination steps will be presented. In that way, remaining clusters and their confidence intervals will define the generated reference intervals and subgroups.The final version of the ALT female iterations as follows:

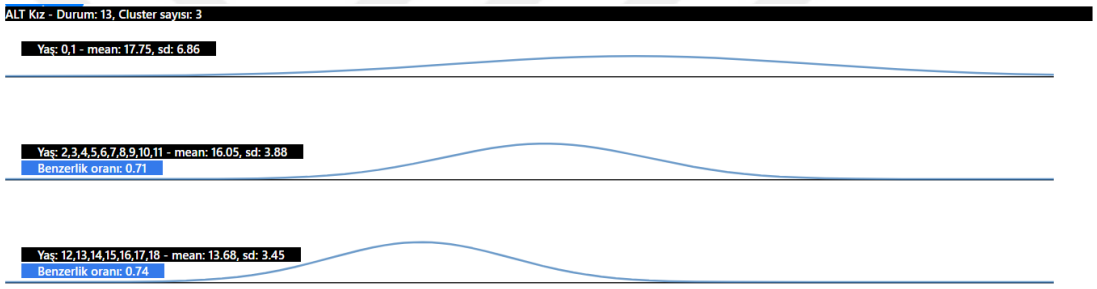

Figure 3.29 Final learning iteration of alt female test

According to the figure 3.29, method two generated 3 different clusters in the final. Reference intervals for each cluster will be computed with calculated mean and standard deviation values.

CHAPTER FOUR
CONCLUSION AND FUTURE WORK

Reference interval studies are not cost-effective for each health institution. Due to most health institutions use the manufacturers' results, determining the suitable reference intervals for pediatrics may be complicated. Because of the cost issues, most of the time health institutions can not perform a study to generate new reference intervals for pediatrics. To overcome these difficulties, this study purposes present an integrated software system to biochemistry experts and health institutes to perform their own pediatric reference intervals studies with their own patient data. Through the implemented learning methods that empowered with artificial intelligence techniques, high-resolution results are generated for subgroups of the data and reference intervals. In this way, reference intervals are generated from the available patient data without requiring any medical intervention. Besides that, the cost is minimized as much as possible, because of not requiring expensive kits or new staff member to examine results. Through the provided web application tool, data preprocessing operations are performed with data mining techniques and features such as displaying learning results, visualizing statistics of the dataset are presented.

In the literature review, it is obvious that there are lots of studies are conducted by researchers in determining reference intervals. On the other hand, most of these studies do not cover pediatric reference intervals. Especially, to propose a solution to the lack of pediatric reference studies, this kind of integrated system with data mining and artificial intelligence techniques is not available in the literature.

The software system that is proposed in the study has a flexible modulation hierarchy so, different learning methods both supervised and unsupervised can be developed and implemented as a new solution type. On the other hand, an auto-learning feature for available methods will be developed and implemented in the future. Instead of defining the initial number of clusters in the first place, the auto-

learning method will determine the best possible number of clusters. In addition, this study is a preliminary study of Ph.D. research (Yıldırım, 2019).

REFERENCES

Adeli, K., Higgins, V., Trajcevski, K., & White-Al Habeeb, N. (2017). The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper. *Critical Reviews in Clinical Laboratory Sciences, 54*(6), 358-413.

Akbayir, S., Balci Fidanci, S., Sen, F., Yurtsever Bakir, A., Orekici Temel, G., & Unal, N., et al. (2011). Mersin bölgesinde homosistein, vitamin A ve vitamin E düzeylerine ait referans aralıklarının belirlenmesi. *Mersin Üniversitesi Saglik Bilim Dergisi, 4*(1), 7-11.

Baadenhuijsen, H., & Smit, J. C. (1985). Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure. *Clinical Chemistry and Laboratory Medicine, 23*(12), 829-39.

Bokeh developers (2019). *Quickstart*. Retrieved January 28, 2020, from https://docs.bokeh.org/en/latest/docs/user_guide/quickstart.html

Bokeh developers (2019). *Embedding Bokeh Content*. Retrieved January 28, 2020, from https://docs.bokeh.org/en/latest/docs/user_guide/embed.html

Bokeh developers (2019). *bokeh.resources*. Retrieved January 28, 2020, from https://docs.bokeh.org/en/latest/docs/reference/resources.html

Colantonio, D., Kyriakopoulou, L., Chan, M., Daly, C., Brinc, D., Venner, A., et al. (2012). Closing the gaps in pediatric laboratory reference intervals: A CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. *Clinical Chemistry, 58*(5), 854-868.

Çayci, T., Kurt, Y. G., Honca, T., Taş, A., Özgürtaş, T., Agilli, M., et al. (2015) Hastane bilgi sistemindeki kayıtlı hasta sonuçlarından tam kan referans aralıklarının tayini. *Gulhane Tıp Dergisi, 57*, 111-117.

Eraslan, D. (2019). *Experimenting with some data mining techniques to establish pediatric reference intervals for clinical laboratory tests*. Master Thesis, Dokuz Eylül University, İzmir

Gherasim, C., La'ulu, S.L., Wyness, S.P., & Straseski, J.A. (2017), *Pediatric reference intervals: tailoring reference intervals to the target population*. Retrieved January 28, 2020, from https://www.clinlabint.com/detail/clinical-laboratory/pediatric-reference-intervals-tailoring-reference-intervals-to-the-target-population

Horn, P. S., Pesce, A. J., & Copeland, B. E. (1998). A robust approach to reference interval estimation and evaluation. *Clinical Chemistry, 44*(3), 622–631.

Horowitz, G. (2010). *Defining, establishing, and verifying reference intervals in the clinical laboratory*. Wayne, PA: Clinical and Laboratory Standards Institute.

Humberg, A., Kammer, J., Mordmüller, B., Kremsner, P., & Lell, B. (2010). Haematological and biochemical reference intervals for infants and children in Gabon. *Tropical Medicine & International Health, 16*(3), 343-348.

Hunter, J., Dale, D., Firing, E., & Droettboom, M. (2012). List of named colors. Retrieved January 28, 2020, from https://matplotlib.org/3.1.0/gallery/color/named_colors.html

Jagarinec, N., Flegar-Meštrić, Z., Šurina, B., Vrhovski-Hebrang, D., & Preden-Kereković, V. (1998). Pediatric reference intervals for 34 biochemical analytes in urban school children and adolescents. *Clinical Chemistry and Laboratory Medicine, 36*(5), 327-337.

Kapelari, K., Kirchlechner, C., Högler, W., Schweitzer, K., Virgolini, I., & Moncayo, R. (2008). Pediatric reference intervals for thyroid hormone levels from birth to adulthood: a retrospective study. *BMC Endocrine Disorders, 8*(15), 1-10.

Katayev, A., Balciza, C., & Seccombe, D. (2010). Establishing Reference Intervals for Clinical Laboratory Test Results. *American Journal of Clinical Pathology, 133*(2), 180-186.

Katayev, A., Fleming, J., Luo, D., Fisher, A., & Sharp, T. (2015). Reference intervals data mining. *American Journal of Clinical Pathology, 143*(1), 134-142.

Koschier, D., Bender J., Solenthaler B., & Teschner (2019). *Smoothed particle hydrodynamics techniques for the physics based simulation of fluids and solids*. Retrieved January 28, 2020, from https://www.researchgate.net/ publication./334535945_Smoothed_Particle_Hydrodynamics_Techniques_for_the_Physics_Based_Simulation_of_Fluids_and_Solids

Linnet, K. (1987). Two-stage transformation systems for normalization of reference distributions evaluated. *Clinical Chemistry, 33*(3), 381-386.

Lumsden, J. H., & Mullen, K. (1978). On establishing reference values. Canadian Journal of Comparative Medicine: *Revue Canadienne de Medecine Comparee, 42*(3), 293–301.

NumPy community (2019). *What is NumPy?*. Retrieved January 28, 2020, from https://numpy.org/doc/1.17/numpy-user-1.17.0.pdf

Orekici Temel, G., Ovla, H. D., Derici Yildirim, D., & Kalafat, H. (2015). Hastane biyokimyasal verilerinden sağlıklı alt grubun belirlemesi: Bhattacharya prosedürü. *Düzce Üniversitesi Sağlık Bilimleri Enstitüsü Dergisi, 5*(2), 28-31.

Python Software Foundation (2018). *Generate pseudo-random numbers*. Retrieved January 28, 2020, from https://docs.python.org/3.6/library/random.html

Sasse, E. (2000). *How to define and determine reference intervals in the clinical laboratory*. Wayne, PA: NCCLS.

Scikit-learn developers (2019). *Getting Started*. Retrieved January 28, 2020, from https://scikit-learn.org/stable/getting_started.html

Scikit-learn developers (2019). *sklearn.mixture.GaussianMixture*. Retrieved January 28, 2020, from https://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html

The Scipy community (2019). *SciPy Tutorial*. Retrieved January 28, 2020, from https://docs.scipy.org/doc/scipy/reference/tutorial/general.html

The Scipy community (2019). *Statistical functions (scipy.stats)*. Retrieved January 28, 2020, from https://docs.scipy.org/doc/scipy/reference/stats.html

The Scipy community (2019). *Integretion and ODEs (scipy.integrate)*. Retrieved January 28, 2020, from https://docs.scipy.org/doc/scipy/reference/integrate.html

The Scipy community (2019). *Integretion and ODEs (scipy.integrate.quad)*. Retrieved January 28, 2020, from https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.quad.html#scipy.integrate.quad

The Scipy community (2019). *Statistical functions (scipy.stats.norm)*. Retrieved January 28, 2020, from https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norm.html