ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES

PhD THESIS

Emre KILINÇ

DESIGN OF A SOIL TEXTURE ANALYSIS DEVICE BASED ON ULTRASOUND SENSORS AND MACHINE LEARNING METHODS

DEPARTMENT OF COMPUTER ENGINEERING

ADANA-2022

ABSTRACT

PhD THESIS

DESIGN OF A SOIL TEXTURE ANALYSIS DEVICE BASED ON ULTRASOUND SENSORS AND MACHINE LEARNING METHODS

Emre KILINÇ

ÇUKUROVA UNIVERSITY
INSTITUTE OF NATURAL AND APPLIED SCIENCES
DEPARTMENT OF COMPUTER ENGINEERING

Supervisor : Assoc. Prof. Dr. Umut ORHAN
Year: 2022, Pages: 84
Juries : Assoc. Prof. Dr. Umut ORHAN
: Assoc. Prof. Dr. Zekeriya TÜFEKÇİ
: Assoc. Prof. Dr. Sami ARICA
: Assoc. Prof. Dr. Ali İNAN
: Asst. Prof. Dr. Mehmet SARIGÜL

In this thesis, a digital soil texture analysis system is designed and introduced, which can be an alternative to the traditional hydrometer method used to find the proportional distributions of sand, silt and clay minerals in the soil. Traditional methods have many disadvantages such as being completely mechanical, needing expert control and laboratory. Considering today's advanced technologies and innovations, it has become inevitable to design a computerized forecasting system. The system has been redesigned using a 3D-printed container with ultrasound sensors. The system makes predictions by interpreting the changes in the intensity of the sound signals passed through the soil-water mixture placed in a closed container with machine learning methods. The changes in these signals, which are obtained by utilizing the sedimentation properties of sand, silt and clay particles in the soil-water mixture at different rates, were recorded on the computer, and computerized estimation steps were applied to the data. By using Support Vector Regression and Multi-Layer Perceptron architectures, the success of machine learning methods have been compared against traditional hydrometer results of the sample soils. Considering the 10% margin of error accepted in the standard hydrometer method, it has been seen that the proposed machine learning supported automated texture analyzer produced acceptable results. Thus, a computerized soil texture analyzer, which can predict the percentages of sand, silt and clay in the soil-water mixture in a closed container, using machine learning methods, is independent of expert supervision and laboratory environment, has a high portability, and can work with less material, has been presented in detail.

Keywords: Soil Texture Analysis, Time series, Support Vector Regression, Multi Layer Perceptron, Computerized Estimation, Soil Texture Classification

ÖZ

DOKTORA TEZİ

ULTRASES SENSÖRLERİ VE MAKİNE ÖĞRENMESİ TABANLI TOPRAK DOKU ANALİZ CİHAZI TASARIMI

Emre KILINÇ

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Danışman : Doç. Dr. Umut ORHAN
Yıl: 2022, Sayfa: 84
Jüri : Doç. Dr. Umut ORHAN
: Doç. Dr. Zekeriya TÜFEKÇİ
: Doç. Dr. Sami ARICA
: Doç. Dr. Ali İNAN
: Dr. Öğr. Üyesi Mehmet SARIGÜL

Bu tezde, toprak bünyesindeki kum, silt ve kil minerallerinin oransal dağılımlarını bulmak için kullanılan geleneksel hidrometre metoduna alternatif olabilecek dijital bir toprak tekstür analiz sistemi tanıtılmıştır. Toprak bünyesini belirlemede kullanılan hidrometre, pipet vs. gibi geleneksel ve oturmuş yöntemler, tamamen mekanik olmaları, elle yapılmaları, uzman kontrolüne ve laboratuvara ihtiyaç duymaları gibi birçok dezavantaja sahiptir. Günümüz gelişen teknolojileri göz önünde bulundurulunca, bilgisayarlı bir tahminleme sisteminin tasarlanması kaçınılmaz olmuştur. 3B tasarlanmış ve yazdırılmış bir kap kullanılarak, makine öğrenmesi metodlarının da yardımıyla veriyi yorumlayarak tahminleme yapabilen bir cihaz ortaya konmuştur. Sistem, kapalı bir kap içerisine konmuş toprak-su karışımından geçirilen ses sinyallerinin şiddetinde meydana gelen değişimleri makine öğrenmesi metodları ile yorumlamaktadır. Toprak su karışımındaki kum, silt ve kil parçacıklarının farklı hızlarda çökelme özelliklerinden yararlanılarak elde edilen bu sinyallerdeki değişimler bilgisayarda kaydedilerek, bilgisayarlı tahminleme adımları uygulanmıştır. Makine öğrenmesi metodu olarak Destek Vektör Regresyonu ve Çok Katmanlı Algılayıcı mimarileri kullanılarak, eldeki örnek toprakların hidrometre sonuçlarına karşılık makine öğrenmesi metodlarının başarıları karşılaştırılmıştır. Standart hidrometre metodunda kabul edilen 10% hata payı da dikkate alındığında önerilen makine öğrenmesi destekli otomatize tekstür analizörünün kabul edilebilir sonuçlar ürettiği görülmüş. Böylece kapalı bir kap içerisine koyulan toprak-su karışımındaki kum, silt ve kil yüzdelerini makine öğrenmesi metodları ile tahminleyebilen, uzman denetiminden ve laboratuvar ortamından bağımsız, taşınabilir olma potansiyeli yüksek, daha az madde ile çalışabilen bilgisayarlı bir toprak tekstür analizörü detaylarıyla birlikte ortaya konmuştur.

Anahtar kelimeler: Toprak Tekstür Analizi, Zaman Serileri, Destek Vektör Regresyonu, Çok Katmanlı Algılayıcı, Bilgisayarlı Tahminleme, Toprak Textür Sınıflandırması

EXTENDED ABSTRACT

In this study, a digital soil texture analyzer, the foundations of which were prepared before by using light sensors, the working principles of this device and the success results obtained are presented in detail.

Soil is formed by the erosion of big boulders, coarse rock and stone fragments by external factors such as wind, flood, etc. and their decomposition into minerals of different sizes. Soil is composed of 50% air and 50% solid matter. While 45% of the solid part consists of inorganic materials (dead animal and plant wastes), 5% consists of inorganic materials, namely sand, silt and clay. The physical structure of the soil, also called soil texture, is determined by the proportional distribution of these inorganic minerals. In other words, soil texture is named according to the percentage distribution of sand, silt and clay particles in the soil.

It is of great importance to determine the texture of the soil before any operation such as agriculture, mining, etc. are carried out on it. Although there are many methods in the literature for determining soil texture, traditional methods such as pipette or hydrometer are the most accepted ones. In the hydrometer method, which is frequently used by soil analysis laboratories, soil sample is first kept in sodium hexametaphosphate (calgon) solution to separate the adhered particles. After waiting for 1 day in this way, it is decanted to a 1000 ml laboratory beaker and filled up with distilled water. After thorough mixing, the glass hydrometer rod is immersed into mixture at certain time intervals and the values obtained with the help of the markings on the rod are recorded. After these values are mathematically processed with certain coefficients depending on the mixture temperature, the sand, silt and clay percentages of the soil are determined. Although it is an old and well-established method, it has many disadvantages such as doing all the steps manually, needing expert control, and being dependent on the laboratory environment. Especially considering the advances in today's technology and artificial intelligence, the lack of computerized systems in this field is felt to a great extent. Although there are many technology-based applications in the literature, there is no mobile measuring device designed with simple and inexpensive parts that directly analyzes sand, silt and clay. In order to eliminate this deficiency, a measurement device has been tried to be implemented with the help of laser lights, light dependent resistor (LDR) sensors and computer with a limited number of soil samples. Although the study is promising, it needs radical changes and innovations due to reasons such as the inability of light to reach LDR sensors for a long time due to clay particles suspended in the mixture. For this reason, a new analysis device having different sensor types based on same working princple with a different structure, detailed working principles of the device and the success rates of the obtained results have been proposed.

The name of the newly developed device is briefly referred to as USTA by taking the capital letters of the words "**U**ltra**S**ound Penetration Based Digital Soil **T**exture **A**nalyzer" for ease of representation. In order to experiment with sound sensors, which will be used instead of light sensors in the first design of the USTA, measuring cups of various volumes and sensor slots facing each other at various heights were created using 3D printer. According to the test results examined, it has been seen that the most suitable location for the sensors to be placed was 2 ~ 3 cm below the soil-water mixture surface. In order to prevent leakage problems, the measuring cup was printed with the help of a 3D printer at 100% density without any gaps. In order for the emitted sound wave to penetrate the mixture in the container at the maximum level, the sensors were attached to the sensor slots on the outer surface of the container with cyanoacrylate adhesive and integrated with the container. In addition, the measuring cup is designed to be closed with a screwable cap to prevent any physical intervention into the container. The DS18B20 digital temperature sensor was glued on the inner wall of the container and the temperature values were recorded simultaneously during the experiments. Sensors were driven with a Rigol DS1104Z model oscilloscope, one as a transmitter and the other as a receiver, and the changes in the intensity of the sound waves in the form of a sinus graph depending on the precipitation in the container were transferred to the MATLAB software and recorded on the computer at 2 Hz frequency for 2 hours.

80 soil samples, the contents of which were determined by the traditional Bouyoucos-hydrometer method, provided by Çukurova University Department of Soil Science and Plant Nutrition, were measured with this system, and the dataset of all soils was collected on the computer, and then feature extraction and machine learning-based estimations were made. Since the data of each soil show similar behavior as a general pattern, 10 points, which were thought to be the most meaningul points, were selected in the time series for each row throughout the entire dataset. These selected features were given as input to the Support Vector Regression (SVR) and Multi-Layer Perceptron Neural Network (MLPNN) machine learning methods, which are frequently used in the literature, and the sand, silt and clay contents were estimated. Estimation results were presented comparatively using the Mean Absolute Error criterion. According to the results of the experiments carried out with SVR, 47 soils for sand, 40 soils for silt and 41 soils for clay, the error rate was at most 10%. In the experiments conducted with the MLPNN structure with 10 neurons in its single hidden layer, it was estimated that 54 soils for sand, 48 soils for silt, and 48 soils for clay were below 10% error rate. During the experiments, it was observed that the error rates were high in some soil samples. After detailed examinations, it was understood that the data of these soils were very few, therefore, machine learning methods could not learn these soil types well enough during the training. Soils that did not have enough samples in the dataset were excluded from the dataset and the same experiments have been repeated with SVR and it has been observed that the error rates have dropped to 7.21%, 9.64% and 10.80% for sand, silt and clay, respectively. Classification experiments were also carried out in parallel with the increase in the estimation success by increasing the samples belonging to the same soil in the dataset. The

experiments were carried out on some soil classes with the SVM, K-nearest Neighbors, Decision Table and Random Forest methods, which are frequently used in the literature, and the total success rates are presented comparatively along with the confusion matrix.

Finally, by interpreting the change in the sound signals passed through the soil-water mixture in a closed container, an automated analyzer which, can predict the sand, silt and clay ratios within an acceptable margin of error, is capable of work with less materials, consists of cheap and easy-to-find material, is independent from expert calculations and laboratory environment and has potential to be mobile, has been presented in details.

GENİŞLETİLMİŞ ÖZET

Bu çalışmada daha önce ışık sensörleriyle gerçeklenerek temelleri atılmış olan, dijital bir toprak textür analiz cihazı, bu cihaza ait çalışma prensipleri ve elde edilen başarı sonuçları detaylarıyla beraber sunulmuştur. Toprak, iri kaya ve taş parçalarının rüzgar, sel vb gibi dış etkenler taraından aşındırılması ve farklı büyüklüklerdeki minerallere ayrıştırılmasıyla oluşmaktadır. Toprağın %50'si hava %50'si ise katı maddeden oluşmaktadır. Katı kısmın %45'i organik maddelerden (ölü hayvan ve bitki atıkları) oluşurken %5'i ise inorganik maddelerden, yani kum, silt ve kilden oluşmaktadır. Toprak tekstürü de denen toprağın fiziksel yapısını bu inorganik maddelerin oransal olarak dağılımları belirlemektedir. Bir diğer deyişle toprak tekstürü, toprak içerisindeki kum, silt ve kil parçacıklarının yüzde olarak dağılımlarına göre isimlendirilmektedir.

Üzerinde tarım, ziraat, madencilik vb gibi işlemler yapılmadan önce toprağın tekstürünün belirlenmesi büyük önem arz etmektedir. Toprak tekstürünün belirlenmesi için literatürde birçok yöntem bulunmasına rağmen, pipet veya hidrometre gibi geleneksel yöntemler en çok kabul gören yöntemlerdir. Toprak analiz laboratuvarları taraından sıklıkla kullanılan hidrometre yönteminde bir miktar toprak önce sodyum hegzametafosfat (kalgon) çözeltisinde bekletilerek yapışmış parçacıkların ayrılması sağlanır. Bu şekilde 1 gün bekletildikten sonra 1000 ml'lik laboratuvar mezürüne aktarılıp üzeri saf su ile doldurulur. İyice karıştırıldıktan sonra belirli zaman aralıklarında hidrometre çubuğu daldırılarak çubuk üzerindeki işaretler yardımıyla elde edilen değerler kaydedilir. Bu değerler ortam sıcaklığına göre belirli katsayılarla matematiksel işlemlerden geçirildikten sonra toprağın kum, silt ve kil yüzdeleri belirlenir. Eski ve oturmuş bir yöntem olsa da tüm adımların elle yapılması, uzman kontrolüne ihtiyaç duyması, laboratuvar ortamına bağımlı olması gibi birçok dezavantajı bulunmaktadır. Özellikle günümüz teknolojisi ve yapay zeka alanındaki ilerlemeler göz önüne alındığında, bu alanda bilgisayarlı sistemlerin eksikliği büyük oranda hissedilmektedir. Literatürde teknoloji tabanlı birçok uygulama var olsa da doğrudan kum, silt ve kil analizi yapan, basit ve ucuz parçalarla tasarlanmış mobil bir ölçüm cihazı bulunmamaktadır. Bu eksikliği gidermek amacıyla daha önce sınırlı sayıda toprak örneğiyle lazer ışıkları, ldr sensör ve bilgisayar yardımıyla bir ölçüm cihazı gerçeklenmeye çalışılmıştır. Çalışma her ne kadar umut vaadetse de ışığın, havada asılı kalan kil parçacıklarından dolayı uzun süre ldr sensörlere ulaşamaması gibi nedenlerden dolayı köklü değişikliklere ve yeniliklere ihtiyaç duymaktadır. Bu sebeple yine aynı çalışma prensibine dayalı farklı sensör tipleri ve farklı yapıda, 3B yazıcı yardımıyla üretilen yeni bir analiz cihazı ve beraberinde cihazın çalışma prensipleri ve elde edilen sonuçların başarı oranları ortaya konmuştur.

Geliştirilen yeni cihazın ismi temsil kolaylığı açısından "**U**ltra**S**es penetrasyonu tabanlı dijital toprak **T**ekstür **A**nalizörü" kelimelerindeki büyük harfler alınarak kısaca USTA ismiyle anılmaktadır. USTA'nın ilk tasarımında ışık sensörleri yerine kullanılacak olan ses sensörleri ile denemeler yapmak için, çeşitli hacimlerde ölçüm kabı ve bu kabın üzerine çeşitli yüksekliklerde karşılıklı denk gelecek şekilde yerleştirilmiş sensör yuvaları oluşturulmuştur. İncelenen deney sonuçlarına göre sensörlerin yerleştirilebileceği en uygun konumun toprak-su karışım yüzeyinin 2 ~ 3 cm altı olduğu görülmüştür. Sızıntı problemlerini önlemek için ölçüm kabı 3B yazıcı yardımı ile 100% doluluk oranında ve üzerinde herhangi bir boşluk olmayacak şekilde yazdırılmıştır. Yayılan ses dalgasının kap içerisindeki karışıma maximum düzeyde penetre olabilmesi için sensörler, kabın dış yüzeyinde bulunan sensör yuvalarına siyanoakrilat yapıştırıcı ile yapıştırılarak kap ile bütünleştirilmiştir. Ayrıca kap içerisine herhangi bir fiziksel müdahaleyi engellemek için ölçüm kabı çevirmeli kapak ile kapatılabilecek şekilde tasarlanmıştır. Kabın iç çeperine DS18B20 dijital sıcaklık ölçer sabitlenmiş ve deneyler sırasında eş zamanlı olarak sıcaklık değerleri de kaydedilmiştir. Karşılıklı yerleştirilen sensörler Rigol DS1104Z model osiloskop ile biri verici diğeri alıcı olarak sürülmüş ve kap içerisindeki çökelmeye bağlı olarak değişen sinüs grafiği formundaki ses dalgalarının şiddetindeki değişimler MATLAB yazılımına aktarılarak bilgisayarda 2 Hz frekansında 2 saat boyunca kaydedilmiştir.

Çukurova Üniversitesi Toprak Bilimi ve Bitki Besleme Bölümü tarafından sağlanan, içerikleri geleneksel Bouyoucos-hidrometre metodu ile belirlenmiş 80 adet toprak numunesi tasarlanan bu sistem ile ölçülerek tüm topraklara ait veriseti bilgisayarda toplanmış, sonrasında özellik çıkarma ve makine öğrenmesi tabanlı tahminlemeler yapılmıştır. Her bir toprağın verisi birbirinden farklı olsa da genel örüntü olarak benzer davranış gösterdikleri için tüm veri seti için zaman serilerindeki (her bir satır) en anlamlı noktalar olarak 10 adet nokta öznitelik olarak seçilmiştir. Seçilen bu öznitelikler literatürde sıklıkla kullanılan Support Vector Regression ve Multi-Layer Perceptron Neural Network makine öğrenme metodlarına sırasıyla girdi olarak verilerek kum, silt ve kil içerikleri tahmin edilmiş, tahminleme başarıları Ortalama Mutlak Hata kriteri kullanılarak karşılaştırmalı olarak sunulmuştur. SVR ile gerçekleştirilmiş deney sonuçlarına göre kum için 47, silt için 40, kil için de 41 adet toprakta hata oranı en fazla %10 olmuştur. Mimari olarak tek gizli katmanında 10 neuron bulunan YSA yapısıyla yapılan deneylerde ise kum için 54, silt için 48, kil için 48 tane toprak 10% hata oranının altında tahminlenmiştir. Deneyler esnasında bazı toprak numunelerinde hata oranlarının yüksek çıktığı görülmüş, detaylı incelemeler sonucunda bu topraklara ait verilerin çok az olduğu, bu sebeple eğitim esnasında makine öğrenmesi metodlarının bu toprak türlerini yeterince öğrenemediği anlaşılmıştır. Verisetinde yeterince örneği bulunmayan topraklar verisetinden çıkarılarak aynı deney SVR ile tekrarlanmış ve hata oranlarının kum, silt ve kil için sırasıyla 7,21%, 9,64% ve 10,80% seviyelerine düştüğü gözlemlenmiştir. Verisetinde aynı toprağa ait örneklerin artırılmasıyla tahminleme başarısının da artmasına paralel olarak sınıflandırma deneyleri de gerçekleştirilmiştir. Literatürde sıklıkla kullanılan SVM, K-nearest Neighbours, Decision Table ve Random Forest metodları ile bazı toprak sınıfları üzerinde sınıflama deneyleri yapılmış, hata matrisi ile beraber toplam başarı oranları karşılaştırmalı olarak sunulmuştur.

Sonuç olarak kapalı bir kap içerisindeki toprak-su karışımından geçirilen ses sinyallerindeki değişimi yorumlayarak kum, silt ve kil oranlarını kabul edilebilir hata payı içerisinde tahminleyebilen, daha az materyal ile çalışabilen, kolay bulunabilir ve ucuz parçalardan oluşan, bilirkişi hesaplamalarından ve laboratuvar ortamından bağımsız, mobil ve otomatize bir ölçüm cihazı tüm ayrıntılarıyla beraber ortaya konmuştur.

ACKNOWLEDGEMENT

This thesis was completed under the supervision of Umut ORHAN to whom i would like to express my sincere thanks for his help and efforts. I would like to thank to valuable academicians Assoc. Prof. Dr. Zekeriya TÜFEKÇİ, Asst. Prof. Dr. Ahmet AYDIN, Prof. Dr. Ayfer TORUN, Assoc. Prof. Dr. Sami ARICA and MSc. Ferhat ALBAYRAK for their help and support throughout the study.

I would also like to thank TÜBİTAK for supporting this work within the scope of the project numbered 118O162.

CONTENTS **PAGE**

LIST OF TABLES **PAGE**

LIST OF FIGURES **PAGE**

XIX

LIST OF ABBREVIATIONS

Laser	: Light Amplification by Stimulated Emission of Radiation
ADC	: Analog to Digital Converter
LGB	: Laser Guided Bouyoucos
LED	: Light Emitting Diode
KNN	: K-nearest Neighbors
SVM	: Support Vector Machine
SVR	: Support Vector Regression
MLP	: Multilayer Perceptron
MLPNN	: Multilayer Perceptron Neural Network
ANN	: Artificial Neural Network
ISSS	: International Soil Science Society
pH	: Potential Hydrogen
$NaPO_3$	: Sodium Hexametaphosphate
LDR	: Light-Dependent Resistor
USDA	: United States Department of Agriculture
ISO	: International Organization for Standardization
USTA	: Ultrasound Penetration Based Digital Soil Texture Analyzer
PLA	: Polylactic Acid
ABS	: Acrylonitrile Butadiene Styrene
V	: Voltage
Vpp	: Peek-to-Peek Voltage
LM	: Levenberg-Marquardt
MAE	: Mean Absolute Error
MSE	: Mean Squared Error
ML	: Machine Learning
PSD	: Particle Size Distribution

1. INTRODUCTION

Our need for land, which is used in almost every aspect of our lives, is basically no different from that in the early ages. Soil is used as a basic component in agriculture and agricultural activities for food production, in construction works to build shelters, and in many areas imaginable thanks to materials science. Soil is formed by the erosion and disintegration of rocks in the earth's crust over many years. Large rocks are exposed to various factors and begin to break down from the upper parts. Soil is formed as a result of the decay of organic materials on the surface over time and combining with minerals separated from the rocks. Variation of the types and proportions of the inorganic components that make up the soil determines the physical structure, also known as the soil texture.

Soil needs structure analysis according to its physical properties in order to use it more efficiently. Soil texture is an important property used in soil classification. Some properties such as drainage, water retention capacity, aeration, erosion susceptibility, organic matter content, cation exchange capacity, pH buffering capacity, hydraulic conductivity can be determined based on the soil texture. Some of these characteristics are important values in plant growing. A soil suitable for growing plants is 50% solid matter and 50% void by volume. While solids contain 45% organic matter on average, they contain 5% inorganic matter. The void part consists of 25% air and 25% water. Seen in this way, soil is a mixture of solid, liquid and gaseous substances. The gas part of the soil is composed of air, the liquid part is water, while the solid part is composed of organic substances and minerals. The organic part of the soil consists of microorganisms, dead plant and animal wastes.

According to the particle size, minerals are divided into three groups called sand, clay and silt. Sand is easily visible to the naked eye. Silt and especially clay minerals are very small particles. While the average size of the clay particles is less than 0.002 mm, the particles between 0.002 mm and 0.02 mm are called silt particles, and the particles from 0.02 mm to 2 mm are called sand particles.

Soil texture is determined by the percentage of sand, silt and clay particles in it. Although there are many methods used for composition determination, the most common methods are hydrometer and sieving methods. Due to the fact that these traditional methods are cumbersome, tiring and voulnerable for errors, studies on soil analyzer models with various technology adaptations are still being carried out.

In previous studies initiated to address the aforementioned concerns, computerized devices has been designed, first supported by light emitting diode (LED) light (Gök and Orhan, 2015; Kılınç and Orhan, 2016), and then laser-guided sensors (Orhan and Kılınç, 2020). Although promising results were achieved in these devices, the fact that light cannot pass through the soil-water mixture for a long time (such as a few days), especially for some clay types, caused the designed device to not be able to measure in those soil types. It has been envisaged that this problem could be overcome only by replacing the sensor in the new designed device, and it was decided to use ultrasound, which is also a cheap sensor.

In this study, scaling has been made in the hydrometer experimental setup, a portable system has been designed by using a 3D printer, an electronic measurement mechanism based on the penetration of ultrasound sensors has been developed on this system, and finally, the tests have been carried out on the soil samples of which their texture were determined by traditional hydrometer method. In the tests, the responses of the device to temperature, mixing, dissolution and time have also been analyzed.

2. SOIL TEXTURE ANALYSIS

2.1. Litarature Review and Hydrometer Method

In this section, the fundamental principles underlying the particles, particle sizes, particle distributions and relationships between particles, which have an important place in the soil texture, and the methods used to measure these materials are discussed. Soil analysis is the determination of the proportions of sand, silt and clay particles in the soil. Soil classification can be done with the determined ratios. The beginning of soil classification dates back to the work of (Atterberg, 1916). In his work, Atterberg was influenced by Ritter von Rittinger's work on sieves and (Odén, 1915) application of Stoke's law to soil science. Soil is named according to the particles with higher percentage when classifying.

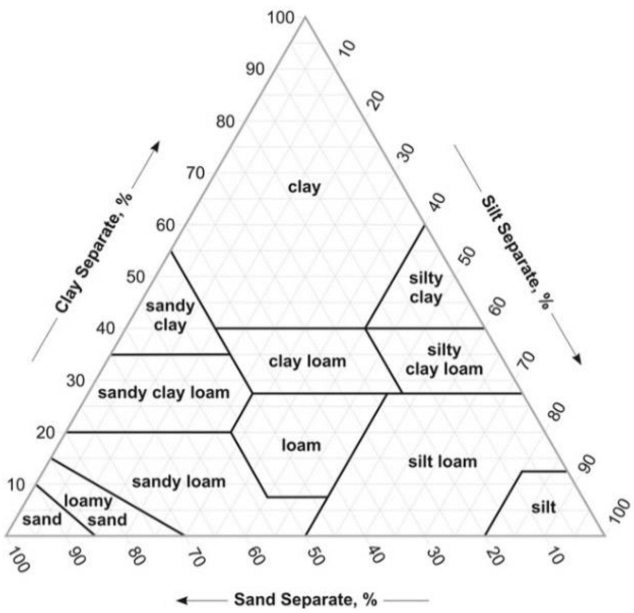


Figure 2.1. Texture triangle

As seen in Figure 2.1., soil is divided into 12 classes according to the percentage distribution of inorganic particles.

Today, there are more than 400 methods for soil texture analysis (Barth and Sun, 1985; Syvitski, 1991). However, most of these methods require lots of equipment for analysis. For this reason, methods such as, pipette and hydrometer, which are acceptable methods and require less equipment than other methods, are used more frequently. In mechanical analysis, which is another method, particles passing through a 2mm sieve are counted. Mechanical analysis has been standardized by the ISSS (International Society of Soil Science, 1928).

Particles in the soil can be in many different forms. For this reason, there is no common decision on analyzes and analysis results. In addition, external factors such as equipment used for analysis, environment, temperature also affect the sensitivity of the analysis. For this reason, the sensitivity of the analysis varies according to the environment in which the analysis is carried out, and factors such as the error possibilities of the equipment used. In addition, even within the same region, soil samples taken from more than one place may differ from each other. For this reason, the soil sample taken from the region to be analyzed must have the characteristics of the region. The prerequisites for the analysis are that the soil sample is taken from the soil suitable for sampling and that it is sufficient in terms of the particles it contains. Samples taken in accordance with the preconditions are called “repeatable” and these samples are also suitable for post-analysis experiments.

Particle shape and size are essential for particle classification in soil analysis. Since it can only be represented by a single parameter, experts classify particles based on their spherical size. Requirement for expert supervision, high error rate and laboratory environment required in traditional soil analysis methods make soil analysis difficult. For this reason, there are many technological studies in the literature (Grout et al., 1998; Ringrose-Voase and Bullock, 1984; Stabinger et al., 1967; Tyler and Wheatcraft, 1992). However, the equipment used in these studies is expensive and scarce. In the recent past, it has been seen that with the help of cheap and easily available sensors, with the right method and hard work, smooth, realistic and reliable data can be produced and soil analysis can be made by processing these data.

The characterization process is done by looking at the amount of large particles. Particles between 10 cm and 1 mm are considered large for soil analysis methods. Although there are sieving methods for calculating the amounts of these particles, the most preferred method is the direct measurement method. The direct measurement method represents the measurement made with an equipment such as a ruler, caliper or scale without any intervention. (Laxton, 1980) made estimations with a photographic technique, (Shiozawa and Campbell, 1991) with the help of particle radius and geometric standard deviation, and (Nadeau, 1985) with the help of a microscope. (Allen et al., 1996) compiled the most used microscope studies. (Smart and Tovey, 1982) used electron microscopy to examine particles smaller than 20µm.

With the direct measurement method, full analysis of the particles can be made. However, the preparation time for analysis is very long (Griffiths, 1967). And during the preparation period, the organic substances in the soil dissolve and the soil content changes. The unequal particle size is another problem for direct measurement. While particles of equal size can be easily examined with a microscope, the measurement differences between the axes will make a big difference for particles smaller than 5µm (Nadeau, 1985). Also, the error rate depends on the vertical size of the particle. As the vertical dimension increases, the difference between the sections to be taken will increase and so will the error rate (Allen et al., 1996).

The sieving method, which is one of the traditional methods, is a primitive but at the same time one of the most effective methods. The sieves used to determine the type of soil have square or circular holes between 125 mm and 5 µm. Circular holed sieves measure the particle according to the radius, while square holed sieves measure the particles according to the distance between its two parallel faces and the diagonal distance in the particle. The sieving method can be done on wet or dry soil. While there are tools for wet sieving, there is a method called air jet for dry sieving (AFNOR, 1979). Square hole sieves are used for particles below 2 mm. Electronically shaped sieves are used for particles below 37 µm (ISO, 1973). Sieves made with metal plates are used to separate large particles because they have larger holes. It is still a method used for soil analysis today, as they can produce good results with the right method and sufficient care. However, the sieving method is not recommended for particles below 30 µm, as it is a tiring method for smaller particles. The majority of the errors made in the sieving method are due to the use of the sieves by shaking in only one direction. Putting a large amount of soil samples on the sieve and sieving using sieves with faulty holes are other errors in the sieving method (Head, 1992; Metz, 1985). In addition, the difficulty of cleaning and maintaining the sieves is another disadvantage of the sieving method.

The sedimentation (precipitation) method is the most widely used traditional soil analysis method today. The sedimentation method is based on the movement of the particles in the suspension towards the bottom of the container over time with the effect of gravity. Sieving method followed by the sedimentation method is the most commonly used method for particle size analysis. Before the sedimentation method, the soil sample to be analyzed is dried and sieved. Afterwards, the soil sample is mixed with a solution and estimation is made on the particles sinking to the bottom of the container. The solution used is sodium hexametaphosphate ($NaPO_3$), a.k.a. calgon, which has a pH value of 9.5, and is alkaline and may rarely differ in some solutions (Klute, 1986). The solution used for soils containing volcanic ash should be acidic (Maeda et al., 1977).

In the pipette method, density method and centrifuge method, which are all developed based on the sedimentation method, it is assumed that the particles behave according to Stokes' law.

For the pipette method, also called macro pipette, the soil to be analyzed is placed in a container and a suspension is created by adding water to it. The mixture is taken from the upper part of this suspension at certain time intervals and dried. Particle ratios are estimated according to the differences of the particles measured from the dried parts. The pipette method is generally used after sieving particles smaller than 63µm (AFNOR, 1979; BSI, 1981; ISO, 1973). The United States Department of Agriculture (USDA, 1996) and Agriculture Canada (Carter, 1993) prefer the pipette method for soil analysis. The pipette method is very laborious for particles smaller than 2µm. (Burt et al., 1993), who wanted to avoid this disadvantage, conducted a micropipette study and presented the results of their study in comparison with the macropipette.

Density method, which is another method, is a method developed with the logic of precipitation. The density is lower when the soil contains more air or organic matter in its content. On the contrary, if inorganic substances (such as iron, manganese, titanium) are high, the density is high. Bouyoucos-hydrometer method is the most used method among the density methods. In the preparation step of the hydrometer method, the soil sample is kept in the stove oven at approximately 100ºC for 1 day, and then it is separated from large particles by passing through a 2mm sieve. 40gr sodium hexametaphosphate (calgon) is mixed with 1 liter of distilled water and left for 1 day. 50-100 gr soil sample is mixed with 125ml sodium hexametaphosphate and left for 1 day for adhered particles to be separated. After mixing for 5-10 minutes in a mixer, it is placed in a 1000 ml laboratory beaker and distilled water is added up to the 1000 ml line and mixed thoroughly. 40 seconds after the suspension is prepared and mixed, the hydrometer is immersed into the beaker and the reading is recorded. This value represents the sum of the clay and silt values in the soil sample. The second reading is done after 60 seconds, the third reading is done after 1 hour and the last reading is done after 2 hours. The temperature value of the suspension is also recorded at each reading operation. The value obtained with the last reading gives the clay ratio. The silt ratio is calculated by the difference between the first read value and the last read value. The sand ratio is also obtained by subtracting the calculated values from 100. The hydrometer method has been standardized by (ISO, 1973).

The reason why the Bouyoucos-hydrometer method is the most commonly used method in soil analysis is that it is simple and requires less equipment. However, the hydrometer method has many disadvantages. The most common error in the hydrometer method is the wrong reading of the hydrometer value (Head, 1992). Excess organic matter in the suspension is the main reason for this error. Even foaming formed by a suitable oxidation method affects the hydrometer value. Another reason is the presence of soluble salt in the soil sample. Salts dissolved in the suspension affect the results of the hydrometer method as they change the density. In addition, the valve part of the hydrometer can cause errors (Allen et al., 1996). (Gee and Bauder, 1986) showed that 40gr soil sample for 1 liter suspension would give good results in their study. There are also studies conducted to accelerate the analysis time in the hydrometer method (Sur and Kukal, 1992). In order to improve the method and eliminate the need for expert control, (Stabinger et al., 1967) used an ultrasonic technique, (Orhan and Kılınç, 2020) used lasers and light dependent resistors (LDR). In both of these studies, the applicability of the hydrometer method by digitizing has been defended. In another study, a significant relationship has been established between the amount of particles with a density less than 2 µm in the suspension (Smith, 2000).

Another method created in the light of the sedimentation method is the centrifugation method. Centrifuge is the artificial realization of strong gravity with tubes connected to a high-speed motor. The centrifuge method is generally used for particles smaller than 1µm. The centrifugation study about time and the sedimentation of particles by (Tanner and Jackson, 1948) have been accepted as the standard for small particles by (Avery and Bascomb, 1982) and (USDA, 1996).

2.2. Laser Guided Bouyoucos

In order to overcome the known problems of the hydrometer method, (Orhan and Kılınç, 2020) designed a new electronically assisted device called Laser Guided Bouyoucos (LGB) by sticking to the preliminary preparations and measurement standards of the hydrometer method used for soil texture analysis. With the computerized estimations made with this device, it has been revealed how close to the results of hydrometer analysis can be achieved. In the study, as in the hydrometer experiment, the soil to be analyzed is first dried. Then it is sieved, mixed with calgon, left for 1 day and poured into the measuring beaker of the LGB device. The LGB test setup is shown below.

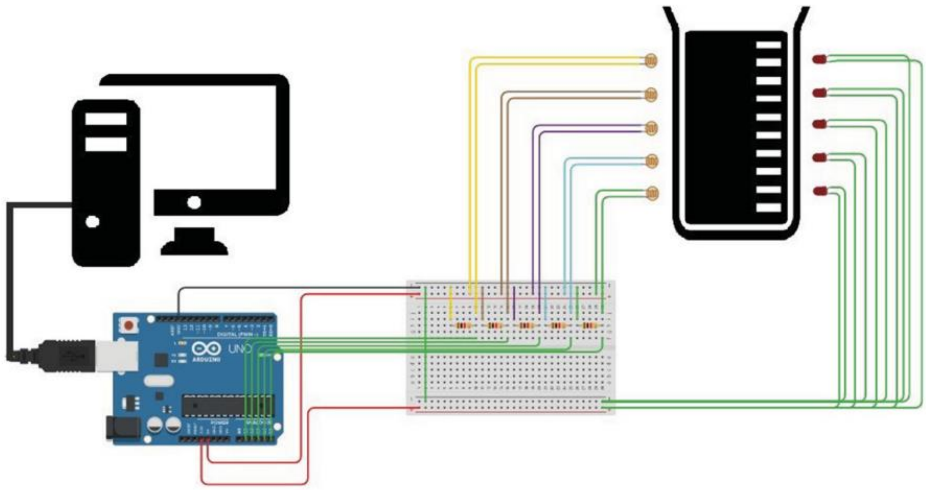


Figure 2.2. LGB test setup (Orhan and Kılınç, 2020)

As seen in Figure 2.2, laser light source and LDR sensors were placed at 5 different levels on two opposite surfaces of the 1,000 ml beaker, passing through beaker's center on the circular axis. The light intensities coming to the LDR sensors have been recorded with the help of a microcomputer at a frequency of 2Hz (2 samples per second). In the study, the most efficient and consistent results have been obtained with the measurements made at the top of the soil-water mixture. The basic logic of the study is based on the correlation of the changes in light intensity with the sedimentation over time with the lasers and LDR sensors placed opposite to each other, taking advantage of the difference in the sedimentation rates of sand, silt and clay in the soil-water mixture. With the analysis of the obtained 2-hour time series, estimation has been made on the proportions of sand, silt and clay in the soil. The time series obtained from a tested soil sample are given in Figure 2.3.

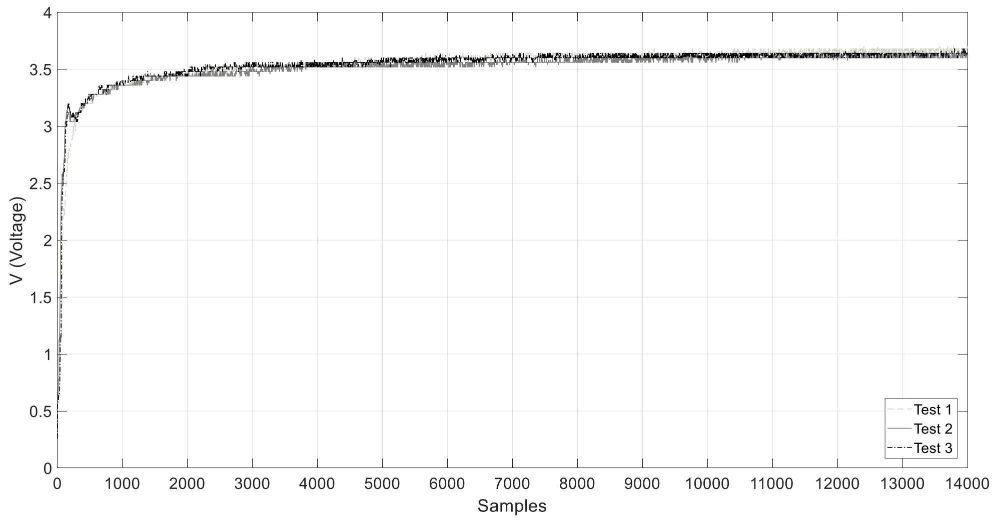


Figure 2.3. Example of the time series obtained from a tested soil sample.

When Figure 2.3 was examined, it has been observed that as the sand, silt and clay particles in the soil precipitated, the intensity of light coming to the sensor increased and the slope of the curve changed at certain intervals. The resulting signal traditionally appears to have exponential curve form.

The starting point of the LGB technique is the elimination of errors caused by human interventions (manual measurements) during hydrometer experiments. In addition, the idea that a more durable, more sensitive and fully automated system can be designed by making use of today's technology has been the biggest motivation point of the project. Although the collected data is processed by computer in the LGB study, the measurement steps were made completely by adhering to the preliminary steps of the hydrometer method (using a beaker, mixing with a rod, etc.). Therefore, in addition to the special problem that the wavelength of the light cannot reach the opposite side at all due to some clay types suspended in the mixture, it still has problems to be solved such as the equipment volume does not allow mobility and the experiment takes a long time.

3. MATERIALS AND METHODS

In this section, general information about the data, tools and algorithms used in the study is given. While the algorithms used have been taken from the known studies in the literature, the hardware has been developed by considering the experimental results obtained during the studies.

3.1. Ultrasound Penetration Based Digital Soil Texture Analyzer (USTA)

The developed device has been named as USTA by choosing the capital letters in the "**U**ltra**S**ound penetration-based digital soil **T**exture **A**nalyzer" for ease of representation, and will be referred as such hereinafter. For a better understanding, the explanation has been made by taking into account the chronological order of the studies and the actions taken according to the difficulties encountered.

In the previous study, with 5 laser and LDR sensor pairs placed opposite each other, light could not penetrate the mixture and reach the LDR sensors because the clay particles were suspended in the mixture for a very long time. Because of this situation, the proposed system, although promising in principle, requires serious modifications. The first change that was considered here was to use sensors that work with a different principle instead of light sensors. Although there are more technological options such as x-ray source, sound sensors, which are thought to perform the same job in a simpler and easier way, have been the first choice. It is thought that sound waves with a wide frequency range will travel through the particles suspended in the mixture, albeit crashing. In this direction, the first sensors to be tested were sound sensors operating at 40KHz frequency, which can be used as receivers and transmitters, and are shown in the picture below.

Figure 3.1. Ultrasound sensors operating at 40KHz frequency

Each of the ultrasonic sensors seen in Figure 3.1 operate at a frequency of 40Khz. The same sensor can generate sound waves with a frequency of 40KHz or collect sound waves with a frequency of 40KHz and convert them into a digital signal. Therefore, on the system to be designed, two separate sensors, one for the receiver and one for the transmitter, are placed facing each other. The control of the sensors is provided by the Aduino device, an inexpensive and easily available microcomputer. The transmitter sensor, driven by square waves of 5V intensity at a frequency of 40KHz produced via Arduino, converts the square waves it receives into an analog signal by vibrating and spreads it to the environment. The receiver sensor, which is also controlled by the Arduino device, converts these analog audio signals into digital signals thanks to the structure called piezo element and transmits them to the Arduino device. The Arduino device, which is used in the control of the sensors, transfers the intensity of the digital signal reaching it from the receiver sensor to the computer as electrical intensity (volts) at a speed of 2Hz per second (two measurements per second). This scalar data transferred to the computer is stored in a text file to be used in estimation methods.

When the output of the Arduino device is given directly to the sensors to produce a square wave of 5V and 40KHz frequency, it has been observed that the intensity read from the receiver sensor is very low (~10mV).

After it was seen that the existing ADC (analog to digital converter) circuit of the Arduino device was not enough to read the sound intensity, an external amplifier circuit had to be used. The schematic of the designed circuit is shown in the figure below.

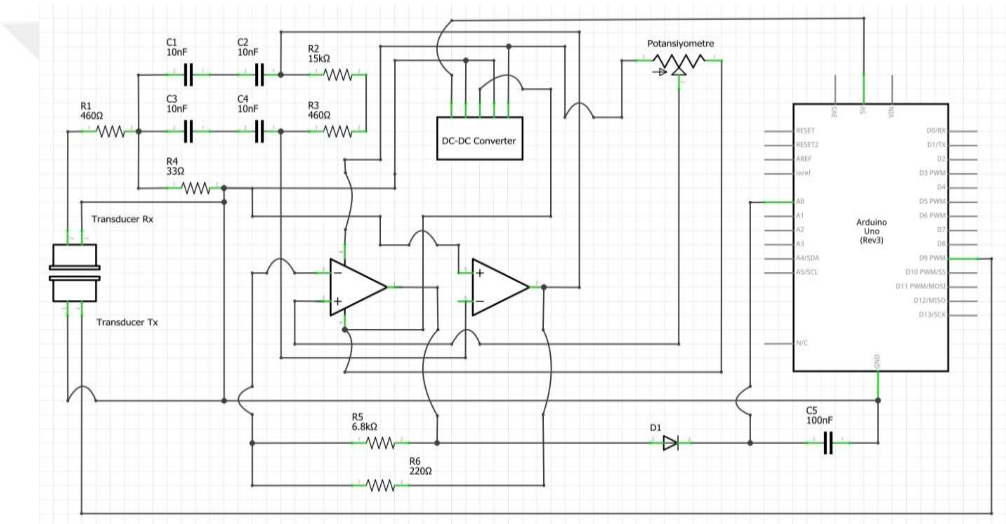


Figure 3.2. Schematic drawing of the designed diode supported amplifier circuit

The amplifier circuit, whose design is given in Figure 3.2, is implemented and the signal is strengthened. Since it will only work with sound waves of 40KHz frequency, a narrow-band filter has been designed and added to the circuit, which filters out sound waves outside this frequency. The signals are amplified by approximately 465 times with this designed amplifier circuit and pulled to the 0V-5V range. In the figure below, the values of the square wave sent to the transmitting sensor and the sine signal reaching the receiver sensor after passing through the narrow-band filter are shown.

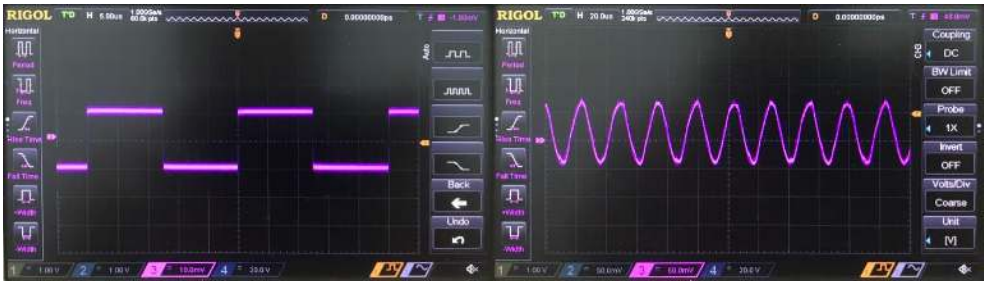


Figure 3.3. The square wave (left) fed to the transmitter transducer and the sine wave (right) emanating from the narrow-band filtered and amplified circuit of the sensed signal.

The wave, which has already been amplified by a certain amount by passing through the filter, is amplified by approximately 30 times by passing through the TL082CP operational amplifier. The graph of the amplified sine wave by the filter and then the TL082CP operational amplifier is shown below.

Figure 3.4. Incoming signals after non-inverting amplifier.

After all these processes, the graph of the sinus signal reaching the receiver sensor in distilled water is shown in Figure 3.4. Since the device is based on the principle of measuring the time change of the sound signal passing through the soil-water mixture, it is quite unnecessary to record all the points of the repetitive sinus signals exactly as in the figure. Instead, since recording the V values of only the peaks at certain time intervals (for example, once in half a second) will make both obtaining and processing of this data set more effective, a signal envelope detector circuit that measures only the peaks has been designed. The output obtained when the sine circuit shown in the figure above is given to the designed signal envelope detector circuit is as follows.

Figure 3.5. Output of the sine signal after the signal is passed through envelope detector.

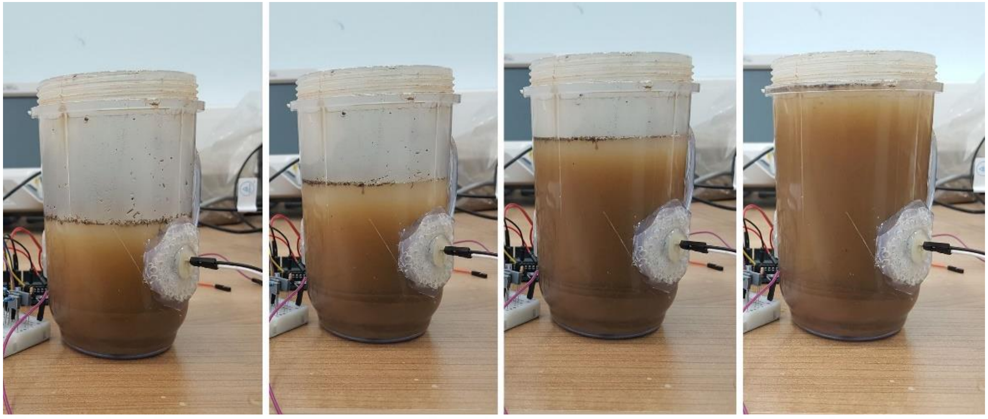
As seen in Figure 3.5, the designed signal envelope detector circuit always shows the peak of the sinus signal. Thus, the maximum value of the measurement signal can be stored directly at desired time intervals. The signal in the figure shows the magnitude of the sinus signal coming to the receiver sensor in distilled water after passing through narrow-band filter, TL082CP operational amplifier and signal envelope detector circuit, respectively. The signal line seen in the picture increased with the effect of the settling particles in the mixture. These changes have been transferred to the computer at 2Hz and recorded in a file.

Another consideration before starting to create the actual dataset was where to place the sensors. In the system based on the sedimentation principle, since the particles in the soil-water mixture will settle from the mixture surface to the bottom, measurements were made with 4 different soil samples in order to analyze the height at which the sensors should be placed from the bottom of the container. The contents of the soils used in these measurements are given in the table below.

Table 3.1. Clay, silt and sand ratios in the soil samples used in the experiments.

Soil sample #	**Sand (%)**	**Silt (%)**	**Clay (%)**
#1	22,5	50,0	27,5
#2	83,2	4,2	12,6
#3	59,7	13,7	26,5
#4	46,7	30,8	22,4

As seen in Table 3.1, the content of sand, silt and clay ratios of the four selected soils were chosen relatively different from each other. After the soils were determined, two opposing holes were drilled on an empty plastic container, large enough for the sensors to enter tightly, to analyze where the sensors would be placed. Receiver and transmitter sensors were placed in these holes, and they were sealed with hot silicone to prevent leakage. Here, since the slots where the sensor pair is placed are prepared by drilling a hole in the container and it is almost impossible to disassemble it and place it at another level, the level of soil-water mixture has been increased. Therefore, the measurement capability of the sensor pair at different heights, starting from the bottom, has been tried to be simulated by using a soil-water mixture of 240 ml at the 1st level, 300 ml at the 2nd level, 360 ml at the 3rd level and 420 ml at the 4th level. Then, each of the soils given in the table was prepared separately by adhering to the preliminary steps in the hydrometer experiments and by proportioning the soil-water amounts to the total volume. Figure 3.6 shows the measurement of soil #4 at four different levels.



(a) Level 1 (b) Level 2 (c) Level 3 (d) Level 4

Figure 3.6. Measurement of soil #4 in (a) 240ml, (b) 300ml, (c) 360ml, (d) 420ml volumes at 4 different levels.

In Figure 3.6, the measurement of the soil #4 at 4 different levels is seen. For the experiments at each level, the amounts of 50 g soil/125 ml calgon water/1000 ml distilled water in the standard hydrometer tests were re-proportioned according to the test levels. For example, for 240 ml level 1, all quantities are multiplied by a coefficient of 240/1000 (0.24). Accordingly, all soils for the level 1 experiment were prepared using a mixture of 12 g soil and 30 ml calgon water. Here, since the point where the sensor pair is placed is fixed and it is almost impossible to remove and re-place it manually to another level, the level of soil-water mixture has been increased. In other words, the measurement capability of the sensor pair at different heights, starting from the bottom, was tried to be simulated by using a soil-water mixture of 240 ml at level 1, 300 ml at level 2, 360 ml at level 3, and 420 m at level 4. In the figure below, the graph of the signals recorded at level 1 for the soil #4 is shown.

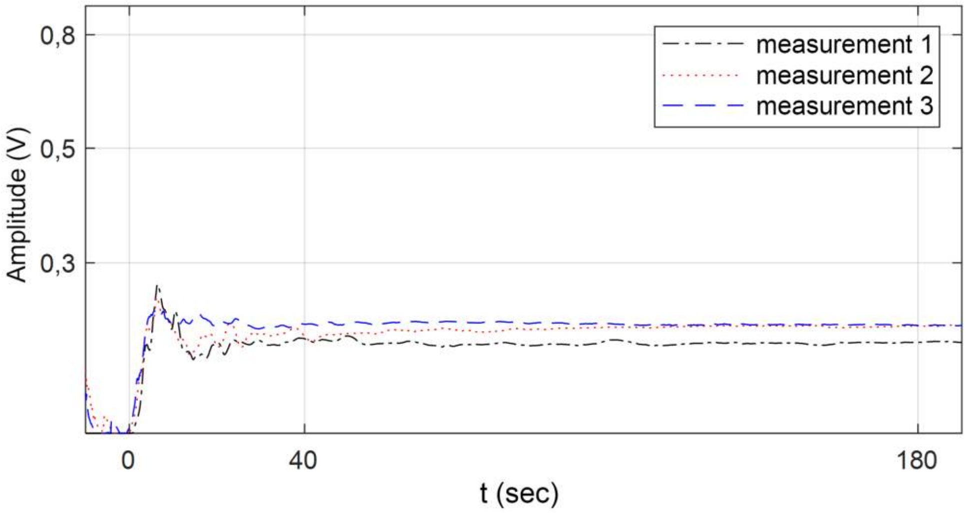


Figure 3.7. Signals recorded at level 1 for soil #4.

As seen in Figure 3.7, the experiments of the same soil were repeated at least 3 times to eliminate measurement errors that may arise. The experiments were carried out by pouring the soil samples prepared in accordance with the preliminary steps in the hydrometer experiments, into the container in Figure 3.6 and filling them with distilled water up to the level and mixing them. In addition, since the measurements were made to observe the behavior of the sensor pairs at different heights, they were recorded for only 3 minutes at a constant room temperature of 20°C. In the figure, the x axis shows the time in seconds, and the y axis shows the value in volts of the read signal. The zero point on the Y axis represents the moment when the soil-water mixture put into the container is thoroughly mixed. In other words, the mixing process continues until the amplitude axis drops to the point 0. The intensity of the signal measured with the effect of the particles that started to precipitate from the moment the mixing was finished, started to increase as expected. The following figure shows the tests of the 4 soils in the table for 4 different levels.

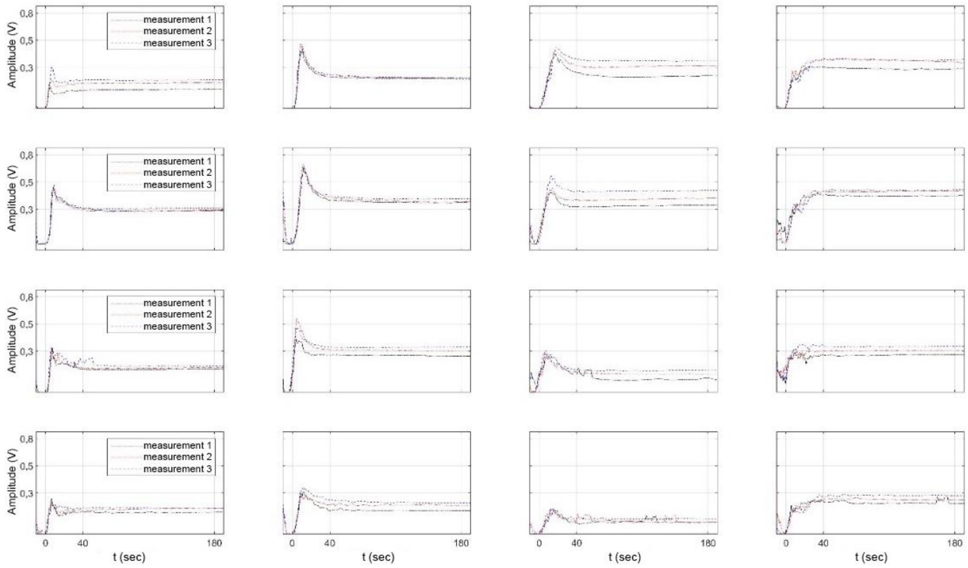


Figure 3.8. Measurement results at 4 different levels for four soil samples

In Figure 3.8, line 1 represents the test result performed for level 1, level 2, level 3 and level 4 of the soil. When the graph is examined in detail, it is striking that the values read by the receiver sensor are higher for soils with high sand content. As the particles in the cross-section between the transmitter and the receiver sensor settle towards the bottom (they are pulled down between the sensors), the intensity of the signal reaching the receiving sensor becomes higher. From this point of view, it can be said whether the soil is roughly sandy or not, even by looking at the amplitude value of the signal against time. Another remarkable point is that in the measurements performed at level 1 and level 2, the up and down points of the graph can be seen more clearly in a wider range. In the experiments performed at level 3 and level 4, it is seen that it is difficult to distinguish the contents of the soil in the experiment due to the accumulation of the particles in the section between the receiver and the transmitter, towards the bottom of the container. Therefore, while conducting the experiments, the amplitude value at the 40th second when the sand detection was performed, as in the traditional hydrometer method, has been taken into account. In a simple linear regression formula in the form of Y=aX+b, the amplitude value at the 40th second was given as input, and the Y value, namely the sand ratio, was tried to be estimated as the output. a and b coefficients were found by solving the equation programmatically on the computer with the least squares method, and absolute errors were calculated for all levels by making estimations with these coefficients. Absolute errors for levels 1, 2, 3 and 4 were calculated as 9, 10, 17, 16, respectively. Considering all these results, the ideal height to place the sensors was determined as the point that would be approximately 3 ~ 4 cm below the soil-water mixture surface in the container. After the determination of the points where the sensors will be placed, the same estimation method was applied for the signals in the range of 0sec - 80sec and the absolute error results against time for the sand ratio estimation are presented as follows.

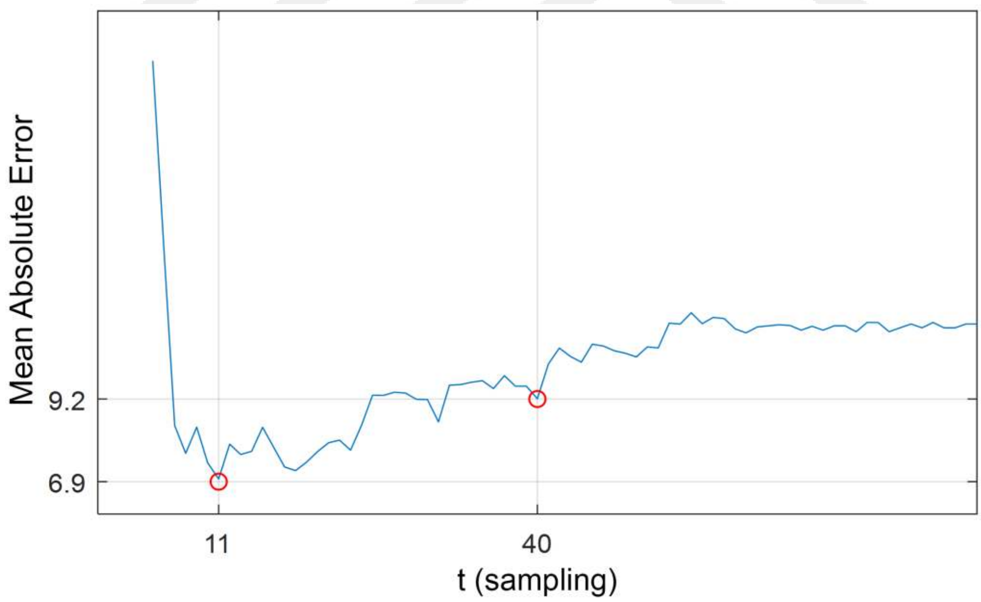


Figure 3.9. Sand ratio estimation errors at 0 sec – 80 sec time range in experiments for level 1.

The first and perhaps the most important point that draws attention in the figure is that the absolute error reached the minimum value at the 11th second. Since another goal of the work is to make the estimations in a shorter time, this result was promising for the continuation of the study. In particular, the small amount of substance used showed that the final analysis device considered could be implemented in a smaller and portable structure. The estimated sand values in each replicate of the 4 soils tested at level 1 are given in the table below for the 11th and 40th seconds.

Table 3.2. Sand ratio estimation values of soil samples used in context of shortening the measurement time experiments.

Soil #	**Repeats**	**Sand (%)**	**Estimation at 11^{th} second**	**Estimation at 40^{th} second**
#1	1	22,5	25	27
	2		36	44
	3		45	51
#2	1	83,2	84	82
	2		84	78
	3		84	83
#3	1	59,7	45	41
	2		48	44
	3		59	59
#4	1	46,7	45	39
	2		38	41
	3		44	47
Mean Absolute Error			**6,9**	**9,2**

Although the error rates do not seem acceptable when the table is examined, it clearly shows the modifications and design changes that need to be addressed for the continuation of the study. It has been seen that the duration of the experiments, the precision of the measurements and the main factors affecting the experiment should all be handled separately and examined in detail. The first of these factors is that the error caused by human in the placement of the sensors should be eliminated. In other words, the container roughly designed in Figure 3.6 should be designed more professionally and the sensors should be aligned in perfect precision without causing any slippage. For this, the container described above has been redesigned in 3D in the computer environment. The 3D drawing of the container is shown in the Figure 3.10.

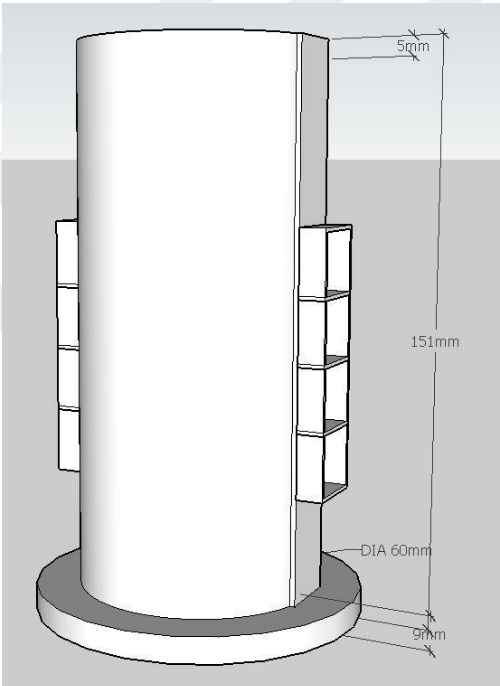


Figure 3.10. First design of the container in computer environmet.

The computerized redesign of the container adapted for the experiments in Figure 3.6 is shown in Figure 3.10. It is designed as cylindrical to help the soil-water mixture to be placed in the container to be distributed uniformly. There are 8 opposing slots on both surfaces of the container, which is cast with the Flashforge

Creator Pro 3D printer, where the 40KHz sensors can fit perfectly without any gaps. Although this way it is guaranteed that the sensor centers will line up exactly, the inside of the slots is still drilled by hand, using the housing walls as guides. Problems such as the leaking of water from the adhesion points of the sensor have been solved by fixing the sensors using hot silicone. Also, the signals circulating over the container as well as the soil-water mixture (causing vibration in the container) reaching the receiving sensor and creating interference had to be solved. For this, the first thing that came to mind was to isolate the container from the signal by preventing the sensor from directly contacting the PLA printed container. The second container designed is as follows.

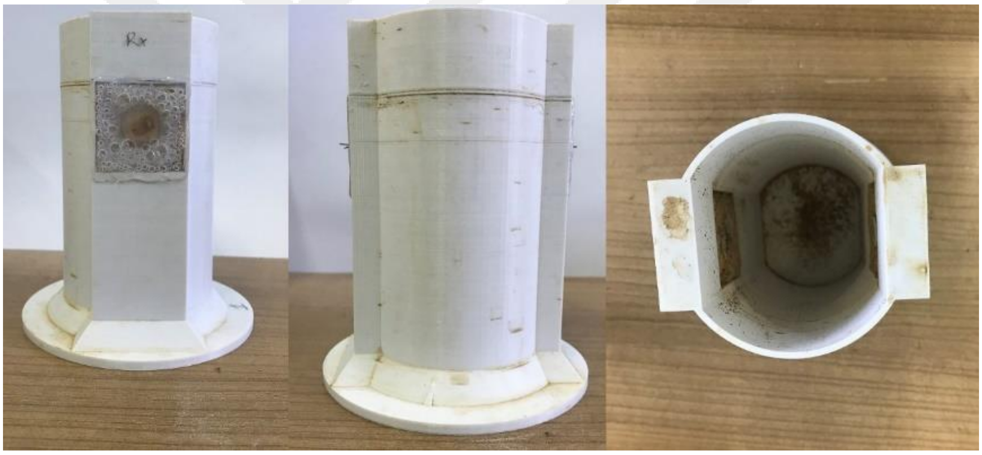


Figure 3.11. Second major desing of the container that does not allow direct contant with ultrasound sensors.

As can be seen in Figure 3.11, in the second design, the ultrasound sensors were first placed inside the insulation foam and then mounted together with the foam using hot silicone to the container which is designed with square holes. Soil-water mixtures were simply analyzed in each new measuring cup designed and realized, and it was checked whether the sensors worked stably. Although in the second designed cup, signal transmission over the body was prevented, problems were encountered with the alignment of the sensors and the insulating foam wearing down in water over time. The new design printed in order to eliminate these problems is in monoblock structure and is as shown in the figure below.

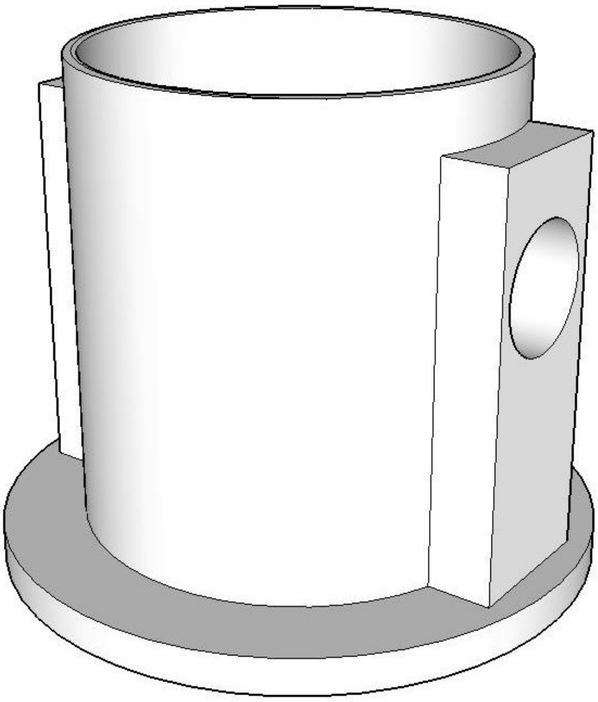


Figure 3.12. 3rd major design of the container body.

As seen in the figure, the designed container has a solid structure. That is, there are no holes on it, but there are slots on both sides where the sensors will be placed (aligned). These slots are designed at a height of approximately 3 cm below the soil-water mixture surface of 250ml volume. Thus, the ultrasound sensors are placed in the slots on the container with hot silicone and contact the container from the outside. Since there is no gap on the container, leakage problems are solved.

There should be no air gap between the sensors and the container when they come into contact with the container through the slots. Because if the sound waves produced in the sensors encounter a less dense environment to propagate, they tend to disperse without penetrating the water in the container by changing the medium. It has been noticed that the hot silicone, which fills the air gap between the sensor and the container, significantly dampens the sound waves emitted from the sensor due to its soft structure. Instead, cyanoacrylate-based adhesive is used to glue the sensors to the housing without an air gap. The cyanoacrylate adhesive wraps the gaps on porous surfaces such as PLA and ABS and integrates the sensors with the container by welding them. Thus, the waves released from the transmitting sensor penetrate directly into the soil-water mixture without encountering any less dense medium. On the other hand, during the 3D printing process, the body of the container is printed at 100% density setting so that there is no gap between the printed layers, thus reducing the stray sound wave problem as much as possible.

When the main problems in the design of the containers were solved by these counter measures, problems of inconsistency in the measurements, which were thought to be related to the sensors, were detected. Although 40KHz ultrasound sensors produced signals close to the expected graphic in a short time (first few minutes), they are greatly affected by particles adhering to the inner surface of the container (especially the area where the sensors are located) and fluctuations in the water. This makes it impossible to interpret the signals collected during the experiment due to inconsistencies in certain time segments. Examples of these inconsistent measurements are shown in Figure 3.13.

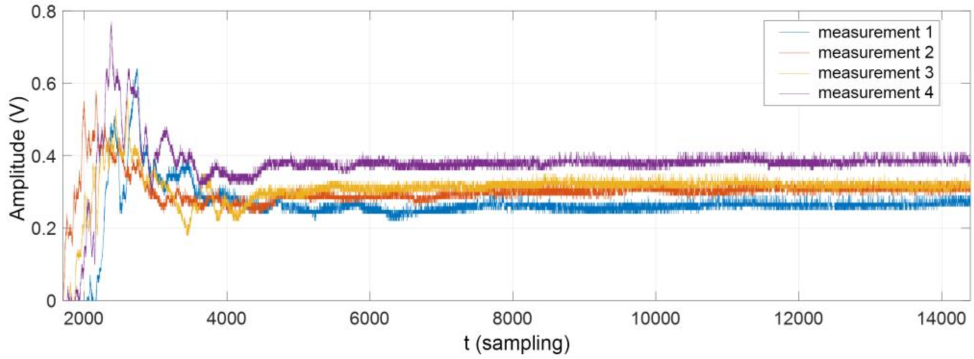


Figure 3.13. Inconsistency problem encountered in 40 KHz ultrasound sensors in 3D printed monoblock cup.

As seen in Figure 3.13, the data collected with the 40 KHz ultrasound sensor causes inconsistency in some time intervals. Although the bumpy curves created by the ripples and reflected vibrations in the water seem to be correctable with filters, the processes become very laborious and cause a lot of time loss as they require a lot of pre-processing in terms of workload in the long run. Losses in the number of data points above a certain value make it almost impossible for machine learning methods to interpret properly later on. In order to test whether the problem is frequency related, a few more easily available 40 KHz sensors were purchased, but the result did not change. On the other hand, when some applications are examined, it has been determined that low ultrasound frequencies such as 40 KHz are preferred in air or for long distances underwater and high ultrasound frequencies such as 1 MHz are preferred in liquids or close penetration required mediums such as tissue like textures. (Sherman and Butler, 2007; Vogelaar and Golombok, 2017). In this context, it was deemed appropriate to test 1 MHz sensors in new experiments. The 1 MHz sensors used are shown in Figure 3.14.

Figure 3.14. Ultrasound sensors operating at 1 MHz frequency.

Each of the sensors in the figure can be used separately as a receiver and transmitter and is driven by 1 MHz square wave. A 3D-printed measuring cup with a slot suitable for a 1MHz ultrasound sensor is given in the figure below.

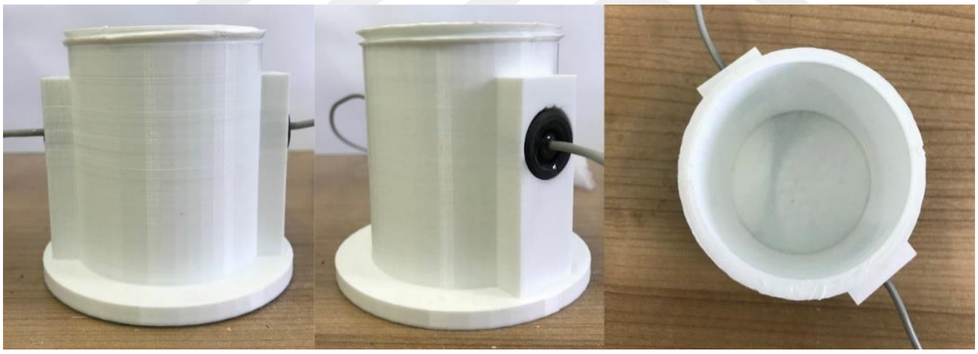


Figure 3.15. Adapted edition of USTA's 3rd design suitable for 1 MHz ultrasound sensors.

As seen in the figure, there are 1MHz sensors glued to the outer surface of the container with cyanoacrylate adhesive, with oval surfaces facing each other, on both sides of the cylindrical body of the container. Since these sensors are driven with 1 Mhz square wave, structural changes have been made in the electronic part of the system. It has been observed that the Arduino device, which produces the square wave and transfers the incoming signal to the computer, is insufficient to operate high-frequency sensors such as 1 MHz and collect the incoming signals at the same time. Therefore, a professional (industrial) Rigol DS1104Z model oscilloscope was used to avoid any doubt in the measurements and the computerized estimation steps to be made afterwards. The transmitter sensor is driven by an adjustable signal source on the oscilloscope with a square wave at a frequency of 1MHz, and the sound waves collected in the receiving sensor are read by the same oscilloscope and transferred to the MATLAB software installed on the computer. MATLAB software can directly access the oscilloscope with the created connection object. In this way, any delay or electronic failure that may arise from the hardware during the experiments was prevented, and a fast and stable data flow from the oscilloscope to the computer was ensured. These signal values collected as vectors in the MATLAB program were recorded in separate files for each experiment.

Another problem that we encountered in the first experiments with the oscilloscope and MATLAB software and that we needed to solve was how to mix the soil-water mixture. One of the first solutions that comes to mind is to mix the soil-water mixture in the container with a glass rod, as in the hydrometer experiment. However, since the particles are radially distributed towards the walls of the container during the mixing process, it is very difficult to obtain a random distribution in the container. In order to randomly shake the soil-water mixture in the container, the container was designed once again to have a screwable cap structure. The visual of the redesigned container is presented as follows.

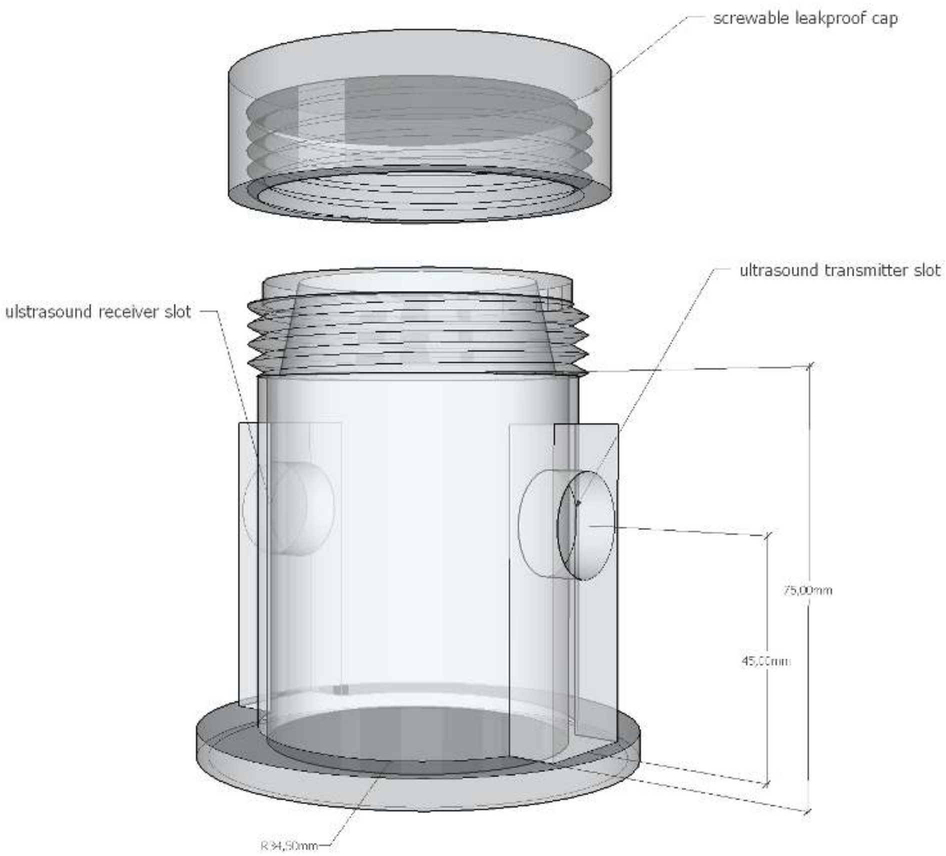


Figure 3.16. USTA's latest body design with screwable (sealed) cap support.

As seen in the figure, the container, which is designed to have a total internal volume of around 300 ml, is designed with a grooved mouth so that it can be closed with a screwable cap. In order to record the temperature of the mixture simultaneously with the intensity of the sound signals, a DS18B20 model digital temperature sensor is fixed on the inner wall of the container. The completed version of the device, together with the sealed cap, which is then printed separately, is given in figure 3.17 below.

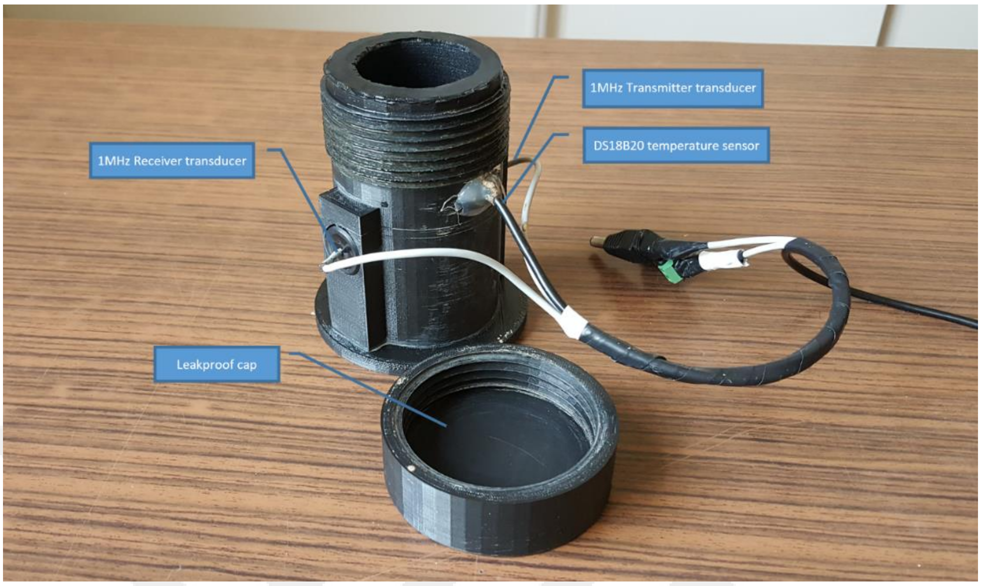


Figure 3.17. The final 3D-print of the USTA's body.

As seen in the figure, the temperature sensor, which is passed through a small hole just below the groove designed for the lid, is placed on the inner surface of the container in such a way that it does not prevent the sensors from seeing each other. The temperature information coming from the temperature sensor is recorded simultaneously with the audio signals. Signals from temperature and ultrasound sensors are recorded simultaneously. The actual experiments with 80 soil samples were carried out using this latest designed container. In Figure 3.18, graphs of 6 repeated measurements of a soil sample performed at 25°C (temperature analysis and 25°C standard are explained in detail in section "5.2. Effect of Temperature Change") are given.

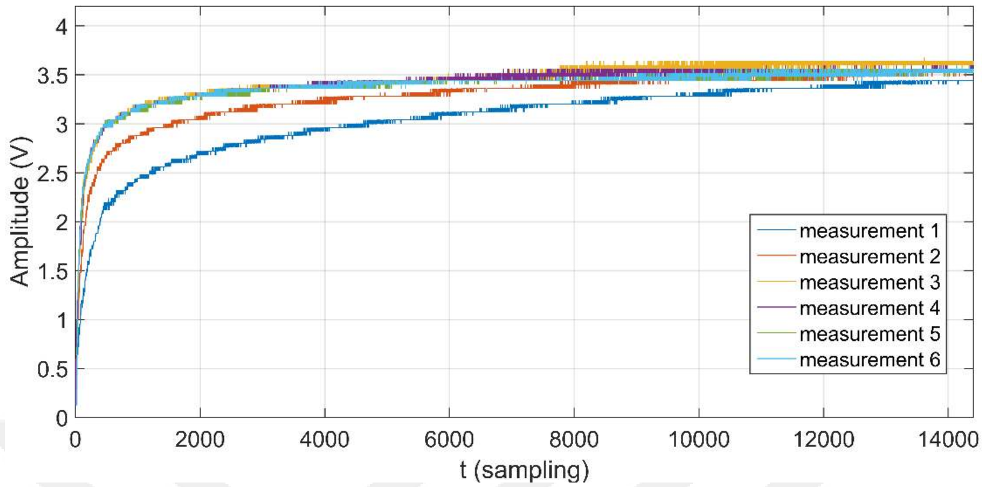


Figure 3.18. Measurement values of a soil sample performed with USTA's final design.

As seen in Figure 3.18, the first two measurements display a different pattern from the other measurements. When the measurements are examined, it is seen that the first two measurements have lower amplitudes for the same time period than the other measurements. This situation was interpreted as the soil-water mixture in the measuring cup was not ready with the effect of shaking. For this reason, the 1st and 2nd measurements shown in the figure were eliminated and the remaining measurements (when the mixture was ready) were included in the data set for that soil test. This practice was continued for all of the experiments and the measurements where the graphs overlapped were included in the dataset. For each soil test, the experiments were repeated until at least 3 overlapping measurements were obtained to eliminate errors that may arise in the computerized estimation step. As an example, 4 repetitive measurement data of a soil is shown in Figure 3.19.

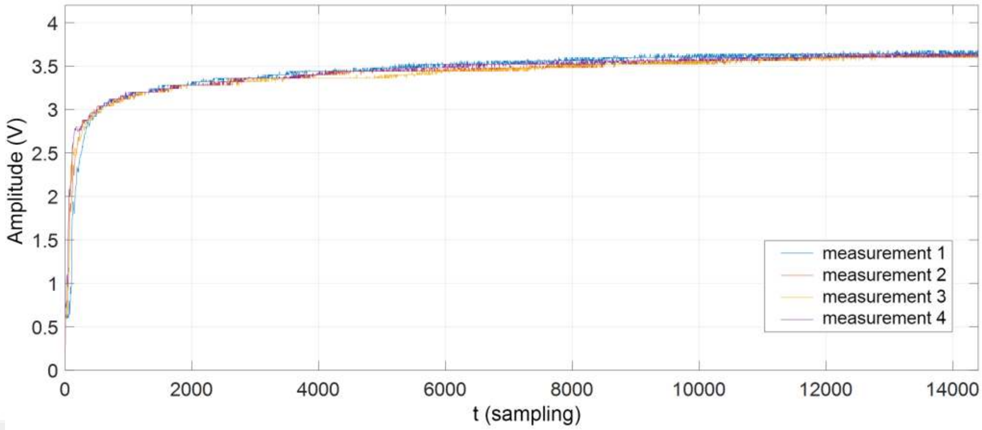


Figure 3.19. Measurement amplitude (V) values of a sample soil against time (2Hz x 120 min) inside the measuring cup.

Figure 3.19 shows the markings for the collected signal of a sample soil. As a result of this process, 4 rows and 14,400 columns of data belonging to this particular soil has been obtained. All these processes have been repeated for each of the 80 soils included in the dataset.

4. SOIL TEXTURE ESTIMATION USING ML AND USTA

In the next part of the study, the estimation algorithms applied to the datasets prepared with the signals obtained from the soil experiments (time series with electrical measurements) are explained. The general approach followed in engineering applications aiming to make predictions from time series consists of three steps. In the first step, a.k.a. preprocessing step, in order to work with healthier data in the next steps, the signs are cleared of errors and if necessary, they are cleaned from noise with various filters. Then, the highly representative features that will best summarize the signs are extracted. Thanks to this step, time series containing many samples are reduced to a few scalar numbers at once, thus reducing the computational complexity of the algorithm to be used in the next step. In the last step, machine learning methods were applied on the dataset. Artificial intelligence methods capable of data-driven learning are often referred to as machine learning algorithms in the literature. It can be said that these studies are gathered under two objectives in scope of our study: 1) estimation of sand, silt and clay ratios in soils (regression), 2) determining which group the soils belong to in the texture triangle (classification).

First of all, the signals in the dataset were examined one by one and faulty measurements caused by electronic equipment were corrected if possible. Otherwise, the experiment was ignored and repeated. Rigol DS1104Z, which is generally a reliable oscilloscope, contains some electrical filters and it has been determined that the signals do not need to be cleaned from a different noise, probably due to the effect of these filters. In this preprocessing step, it has been determined that the dataset generally consists of time series with the same characteristic. In order to both reveal the characteristics of the signs and apply them in feature extraction studies, firstly, the curve fitting method was analyzed. Then, using support vector regression and multilayer neural network methods, which can be used in both regression and classification, the results are presented under separate headings.

4.1. Curve Fitting

The curve fitting method is based on the chronological ordering of time stamped signals and finding the most suitable function for this ordering. The curve functions found can be exponential, trigonometric, exponential or linear. A quadratic exponential function is represented by equation (1).

$$y = a.e^{bx} + c.e^{dx} \tag{4.1}$$

The main purpose of curve fitting is to find the coefficients a, b, c and d in the example of equation (1). Generally, minimum error values are used to find the most suitable function among the candidate functions that are thought to represent the curve being studied well. Criteria such as mean absolute error, maximum absolute error, mean square error between the values calculated using the candidate function and the values in the real data are determined and the curve function with the lowest deviation is decided according to the desired criterion. Traditional mathematical models such as gauss-newton and levenberg-marquardt (LM) are generally used when finding the coefficients in the function. An example curve fitting application is given in Figure 4.1.

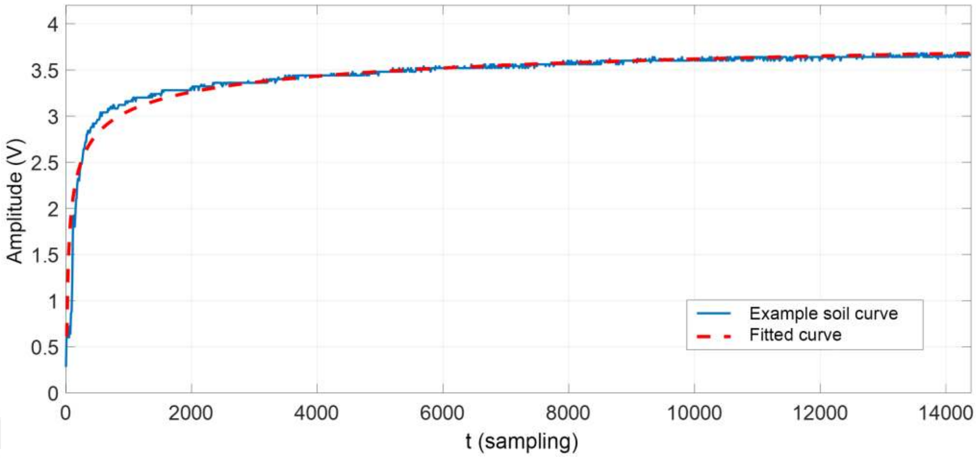


Figure 4.1. An example application of cuve fitting over a time series signal vector in the dataset.

As shown in Figure 4.1, the closest curve to the sign with a certain margin of error can be found with the curve fitting method depending on the data, and the mathematical function of this curve can be reached. It has been determined that the most suitable one among many functions for the soil curve selected as an example in the figure is the second order exponential function ($a.e^{bx} + c.e^{dx}$). Thanks to the fitted curve, the amplitude value of the curve at any time t can be easily found within the 95% confidence interval.

4.2. Support Vector Regression

One of the most frequently used methods in many applications since its foundation in the 1960s is Support Vector Machines (SVM) due to its success in high-dimensional feature space. The algorithm, which was first developed by (Vapnik and Lerner, 1963), took its final form known today with the studies at the AT&T Bell laboratory (Boser et al., 1992; Cortes and Vapnik, 1995; Guyon et al., 1992; B Schölkopf et al., 1995; Bernhard Schölkopf et al., 1996). Thanks to its extraordinary success in many different problems, it has become one of the standard methods in machine learning. (Bernhard Schölkopf and Smola, 2002;

Smola and Schölkopf, 2004) studied the use of SVM for regression and introduced the concept of Support Vector Regression (SVR). The SVR is defined based on a linear or non-linear "kernel" function and is solved by the quadratic programming approach given by equation (2).

$$\min_{w,b,\xi,\xi^*} \frac{1}{2} ||w||^2 + C(e^T\xi + e^T\xi^*)$$
$$s.t.\ \ (\phi(\mathrm{A})\mathrm{w} + \mathrm{eb}) - Y \le e\varepsilon + \xi\ ,\ \ \xi \ge 0e$$
$$Y - (\phi(\mathrm{A})\mathrm{w} + \mathrm{eb}) \le e\varepsilon + \xi^*, \xi^* \ge 0e \tag{4.2}$$

The variable indicated by C in equation (2) is a predefined value and provides a balance between errors and regression. A represents the training examples and Y represents the regression outputs. On the other hand, the expressions ξ and ξ^* are artificial examples, while the expression e is the unit vector. Equation (2) transforms into equation (3) using the Lagrangian multipliers a and a^*.

$$\max_{a,\ a^*} - \frac{1}{2} (a^* - a)^T K(A,\ A^T)(a^* - a) + Y^T(a^* - a) + \varepsilon e^T(a^* + a)$$
$$s.t.\ \ e^T(a^* + a) = 0$$
$$0e \le a\,, a^* \le Ce \tag{4.3}$$

In order to solve the problem that is the subject of the SVR, the goal becomes to find the b threshold and a^* by solving equation (3). This solution is also represented by equation (4).

$$f(x) = \sum_{i=1}^{n} (a^* - a)\, K(x_i, x) + b \qquad (4.4)$$

Here, the kernel function denoted by $K(x_i, x)$ is the representation of $\phi(x_i) \times \phi(x)$ vectoral product in hyperspace. Here, the ϕ function can be selected from functions such as linear, polynomial, exponential, diametric, depending on the problem.

4.3. Multi-Layer Perceptron Neural Network

Perhaps the most important step to which machine learning owes its popularity today is the emergence of artificial neural networks (ANNs). In simple terms, the projection of an artificial neuron, which mathematically mimics the learning process of a nerve cell (neuron) in the human brain, with a biological nerve cell is given in Figure 4.2.

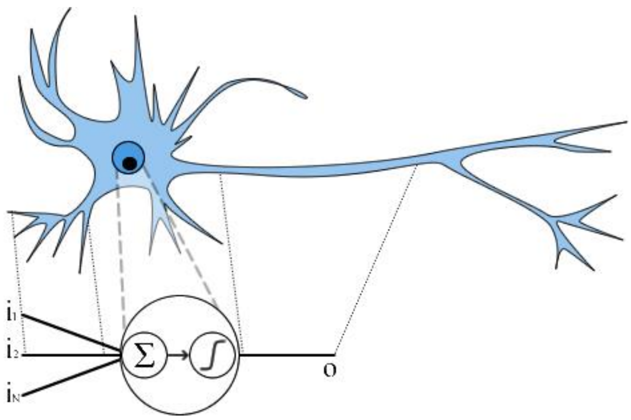


Figure 4.2. Visual demonstration of biological and mathematical (logical) neuron.

The biological neuron shown in Figure 4.2 and its mathematical imitation actually have the same functional steps (collection of external data, evaluation of this collected data and transmission to other neurons). There is a serious neural network communication in the brain, which consists of many biological neurons. ANN cells can also produce solutions to problems by interacting with each other, just like neurons in the brain. Multi-Layer Perceptron (MLP) architecture, which is a type of ANN, includes at least one hidden layer in addition to the input layer and output layer. A standard MLP structure is given in Figure 4.3.

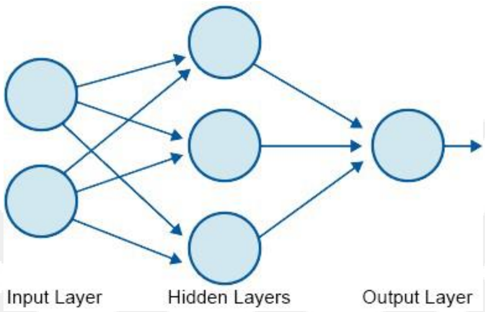


Figure 4.3. Visual demonstration of a standart MLPNN structure.

The architecture shown in Figure 4.3 is also referred to as a feedforward hierarchical model in the ANN literature. This structure has achieved great success in numerous applications thanks to this architecture it uses. In the input layer, the data subject to the problem is accepted. Incoming values are multiplied by various weights and transmitted to the middle layer, also known as the hidden layer. Increasing the number of hidden layers and the number of neurons in the hidden layers often increases the computation time along with the success of ANN. The output layer produces the final output by processing the weighted variations of the values calculated in the middleware. Another detail that affects the success as much as the architecture in all ANNs is the activation function. Activation functions are used to produce the output value of the neuron by applying the weighted sum of the values given to the neurons as input. While it is accepted that a simple threshold control is made for this task in biological neurons, many nonlinear equations can be selected as activation functions in ANN models.

Finally, it is known that by changing the algorithm used for training in ANNs, the success of the system and the calculation time spent in training change. Although the first known successful ANN training algorithm is back-propagation, a faster version of it, the LM algorithm, is often used today.

5. RESULTS

In the light of the methods and tools described in the previous section, primarily the soil samples used, soil experiments with USTA and data collection phase are explained. Then, some factors that may affect the time series measured by USTA (the effect of temperature change, the effect of water-soluble soil components, the effect of agitation applied to the soil-water mixture) were inspected. Finally, regression and classification studies are presented for the estimation of sand, silt and clay values within the scope of estimation with machine learning.

5.1. Data Collection

In our study, around 100 samples were obtained for the testing of the developed USTA device, and some records could not be used due to some problems. Therefore, there are a total of 80 soil samples in the dataset that can be used without any problems. The texture results of soils were determined using the traditional hydrometer method by Çukurova University Soil Science and Plant Nutrition Department faculty member Prof. Dr. Ayfer TORUN, who also worked as a researcher in the funded project of which this thesis takes part. Samples were collected by taking certain amount of soil mass from approximately 30 cm below the soil surface and at a distance of 2 m from each sample collection point. Complete list and texture information of the soils can be found at appendix.

It is seen that the majority of the samples are clayey due to the fact that soils collected from the Çukurova region, where arable lands are located. As a result, soils belonging to some classes are more numerous than other soils in the dataset. However, as variable as soil samples were tried to be included in the dataset to cover all texture classes to ensure a balanced dataset distribution. Then, the soils were labeled among themselves, separated into samples and made ready for measurement. Soils are named according to the particles with a higher percentage when classifying, and there is a texture triangle with 12 predetermined class segments for this. All the soils supplied for our study are placed in the soil texture classification triangle and presented in Figure 5.1.

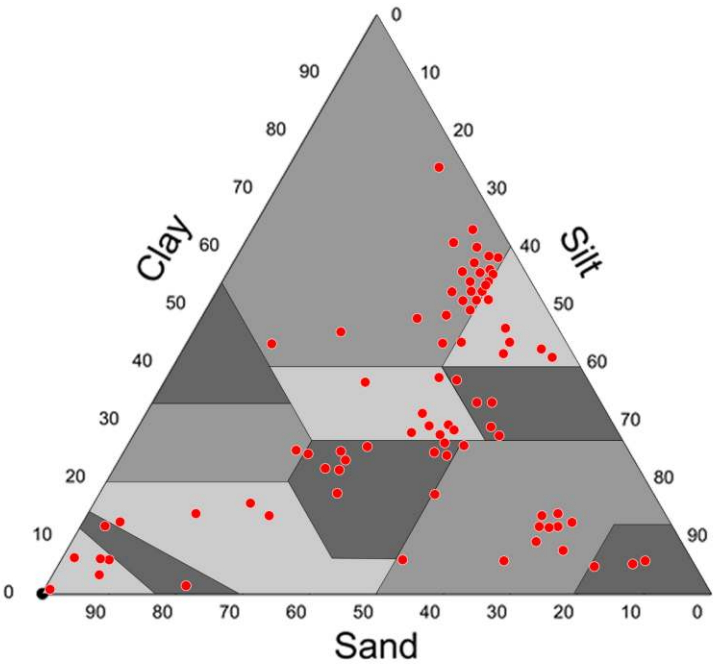


Figure 5.1. Visual distribution of 80 soil samples over the soil texture triangle.

The dataset includes various soils with sand content in the range of 2% - 99%, silt content in the range 1% - 87%, and clay content in the range 1% - 74%. Sample soil belonging to only one of the 12 classes in the texture triangle could not be obtained. Since the samples were collected from the arable Çukurova region, most of them belong to the clayey class.

Due to the physical properties of the USTA, the whole content has been reduced to ¼ in order to be able to test with less soil sample. For example, 50 g soil sample used in the hydrometer method was reduced to 12.5 g and success was desired with smaller samples. In this way, the soil sample and calgon water required for texture analysis were used more sparingly.

In the first stage of the process, 12.5 grams of soil sample is weighed with the help of a precision scale and mixed with a prepared 30 ml calgon-water solution in a tube. This mixture is kept for at least 16 hours as in the standard hydrometer method, and any adherent particles in the mixture are separated. The mixture is then decanted into the container without leaving any residue in the tube and filled with distilled water up to the 250 ml line. After the screw cap is closed, the container is shaken randomly for at least 20 s and left on a flat surface. The experimental setup prepared is given in Figure 5.2.

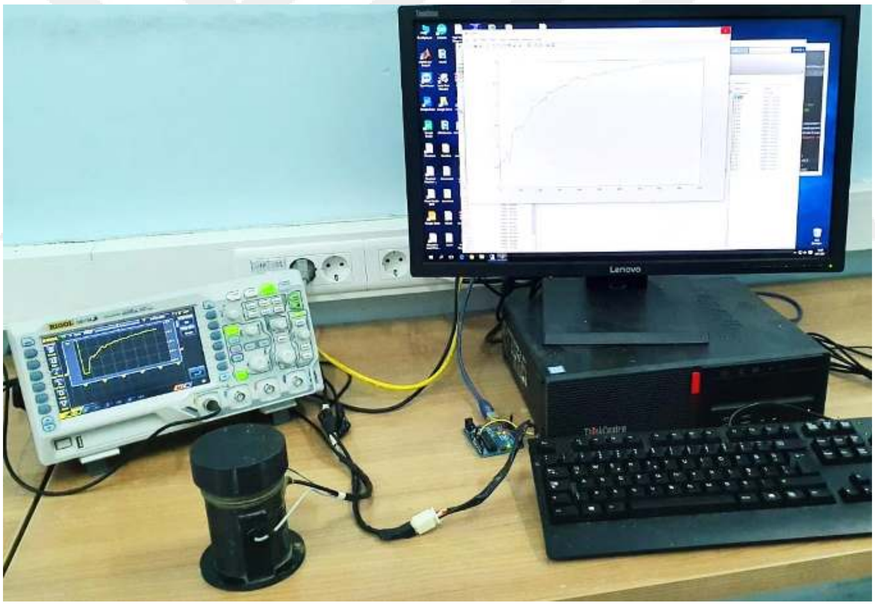


Figure 5.2. Oscilloscope and computer supperted experimental setup.

As shown in Figure 5.2, the transmitting sensor is connected to the output channel of the Rigol DS1104Z oscilloscope and it is ensured that it receives a constant square wave of 5 V amplitude and 1 MHz frequency. Likewise, the receiver sensor was connected to the oscilloscope input without being connected to any intermediate circuit element and the sine wave was read in real time. These read values were sent to the MATLAB software running on the computer via the USB interface of the oscilloscope. The peak (Vpp) value in each half-second sine wave received through the codes prepared on Matlab was recorded for 2 hours. In addition, the temperature values of the soil suspension during the test were recorded with the same time and sampling rate with the help of the temperature sensor in the soil and the Arduino.

Due to the high number of particles suspended in the soil-water mixture at the beginning of the recording, almost no ultrasound waves reach the receiver. Therefore, the shaking process should continue until the Vmax value is seen as zero on the oscilloscope screen. Although a reliable function is prepared for temperature correction by performing temperature effect analysis, the ambient temperature is tried to be kept constant at 25 °C by means of an air conditioner. The ultrasound waves reaching the receiving sensor by penetrating the soil-water mixture from the transmitting sensor in the container are shown in Figure 5.3 as a representative.

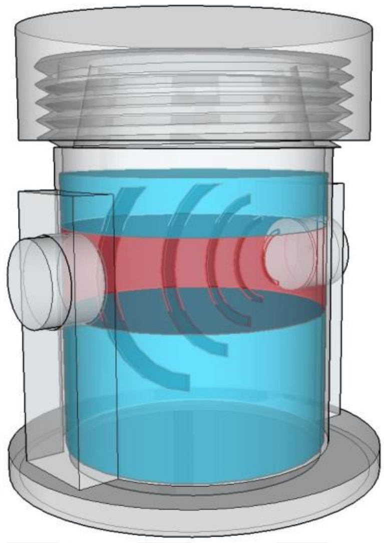


Figure 5.3. Visual representation of propagation of sound waves from transmitting sensor to receiving sensor inside the measurement cup (USTA).

As seen in Figure 5.3, ultrasound waves passing through the region highlighted with red are transferred to the computer from the first channel of the same oscilloscope through sensors placed 3 cm below the mixing surface. Since the entire measurement time (determination of sand, silt and clay) takes 2 hours in the standard hydrometer method, the signals were collected for 2 hours in all experiments. These collected signals are passed through the median filter and are purified from artifacts caused by noise and electrical contact problems. As a result of the experiment, which lasted for 120 minutes with 2 measurements per second (2Hz), 1 row of 14,400 columns of data belonging to a single measurement of a soil is formed. The same experiment was repeated many times to eliminate measurement errors that may arise.

5.2. Effect of Temperature Change

The traditional hydrometer test is carried out in the literature based on a temperature of 20°C. However, since it can be difficult to keep the mixture temperature unchanged at 20°C for a long time, correction is applied for measurements at different temperatures than this temperature. During the hydrometer experiments, the temperature is included in the calculation by correcting with the formula (Current Temperature-20) x 0.36 during each notation process. Based on this, 5 soil samples were selected among the soil samples and experiments were carried out at two different ambient temperatures, 20°C and 25°C. Although the amplitude values for all soils vary in certain intervals, the same patterns were measured for the same soils. Therefore, it is thought that the desired temperature forms of the graph can be found by using a temperature-dependent multiplier. In fact, when each measurement data is considered as a vector consisting of 14,400 columns, if the measurement vector at 25°C is named as A and the measurement vector at 20°C is named as B, the vectors can be correlated with the solution of a linear system of equations such as $Ax = B$. This approach was applied to the measurement curves of the soil sample shown in Figure 3.19 and a coefficient of 0.91 was determined. Figure 5.4 shows the vector at 25°C and the curve of the "corrected measurement" vector calculated by multiplying this coefficient.

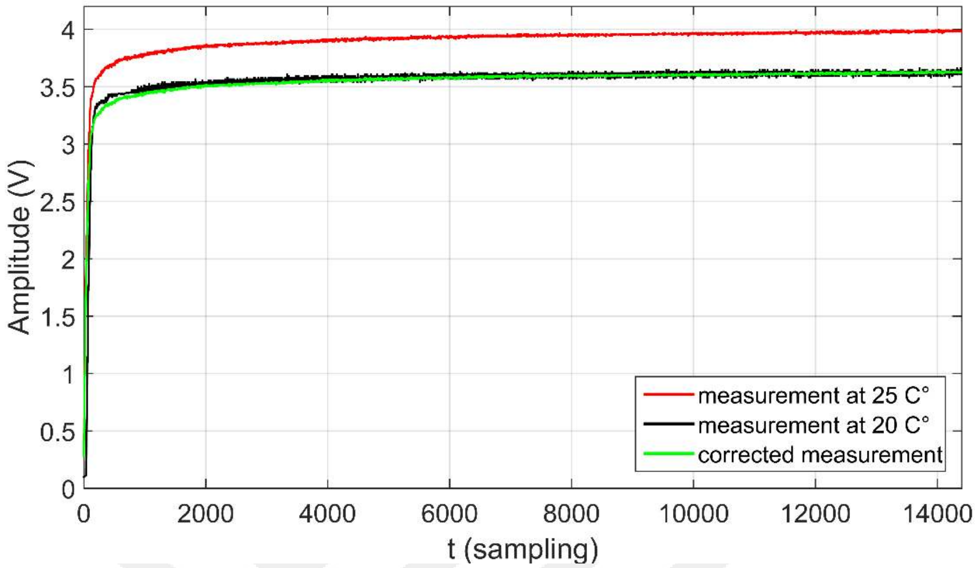


Figure 5.4. Curve of the a soil sample in 20°C and 25°C ambient along with the corrected measurement of the same soil sample.

When Figure 5.4 is examined, it is seen that when the measurement values of the soil at 25°C are multiplied by the coefficient of 0.91, it gives almost the same result as the measurement values performed at 20°C. The coefficients calculated for the 5 selected soils are given in order in the table below.

Table 5.1. Correction coefficients found to transform soil curves performed under 25°C to equivalent of 20°C measurement.

Soil #	Coefficient
#1	0,91
#2	0,91
#3	0,92
#4	0,94
#5	0,94

When Table 5.1 is examined, the measurement data performed at 25°C, when multiplied by the coefficient of 0.92, match with the measurements performed at 20°C. These calculations are based on the assumption that the room temperature always remains constant at 20°C and 25°C with air conditioning. However, due to the operating principle of the air conditioner, it is expected to keep the room around ±1°C of the set temperature. The slight deviations in Table 5.2 may be due to this. Therefore, a more detailed temperature analysis was planned. Accordingly, the measurements of distilled water only (without soil), which was first cooled to 9°C and then heated to 38°C, were made in the USTA. Experiments of warming distilled water at 9°C to room temperature and cooling of distilled water at 38°C to room temperature were repeated several times, and the behavior of sound waves was observed by examining the graph obtained. Figure 5.5 shows the variation of the measured signs depending on the temperature.

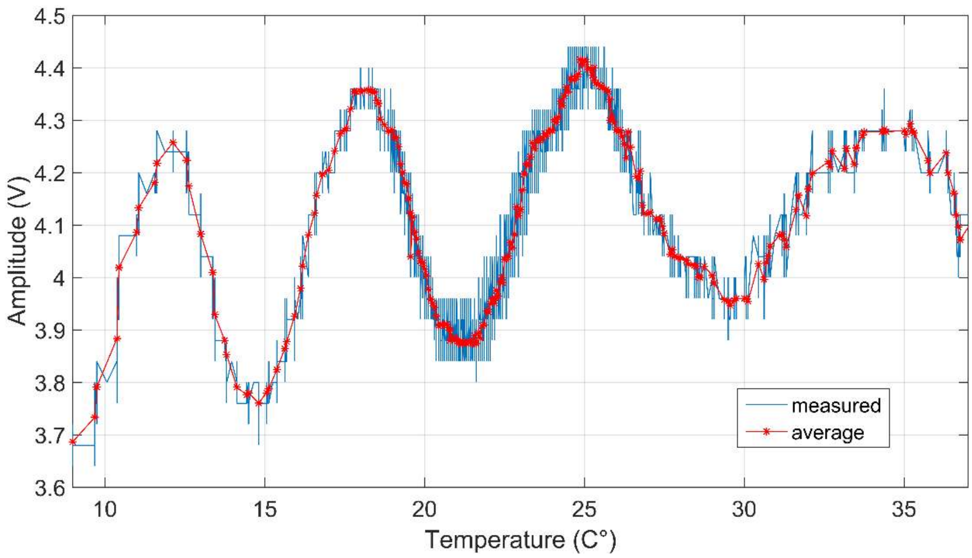


Figure 5.5. Variation of amplitude of sound waves in distilled water between 9°C and 38°C temperatures measured with USTA.

Figure 5.5 shows the temperature-dependent change of ultrasonic sensor data performed with distilled water in the range of [9°C to 38°C]. Although the graph looks like a sinusoidal curve, it was not possible to determine a simple sinusoidal curve fitting equation. Therefore, it was decided to use the average values shown in red in the figure and to set the measurement standard as 25°C, which indicates the peak of this curve. After this stage, the corrections were made according to the 25°C temperature. For example, to find the calibrated value of an x value measured at 15°C at 25°C, the calculation of $4.4 \times (x/3.8)$ is applied. Here, the value of 4.4 represents the amplitude of the curve at 25°C, while the value of 3.8 represents the amplitude of the curve at 15°C. In addition, the highest amplitude value was observed at 25°C, eliminating the need to strengthen the signal.

5.3. Effect of Water-Soluble Substances

One of the standards in the hydrometer test requires that the soil to be analyzed is kept in water with calgon for at least 16 hours. In order to understand the role of this issue in the design of the USTA, which tries to produce as practical solutions as possible, the soils prepared with and without calgon were analyzed and compared with the USTA with a simple approach. Since only the deviation will be examined in the experiments, only 1-hour recordings were taken. The curves obtained in these experiments are presented in Figure 5.6.

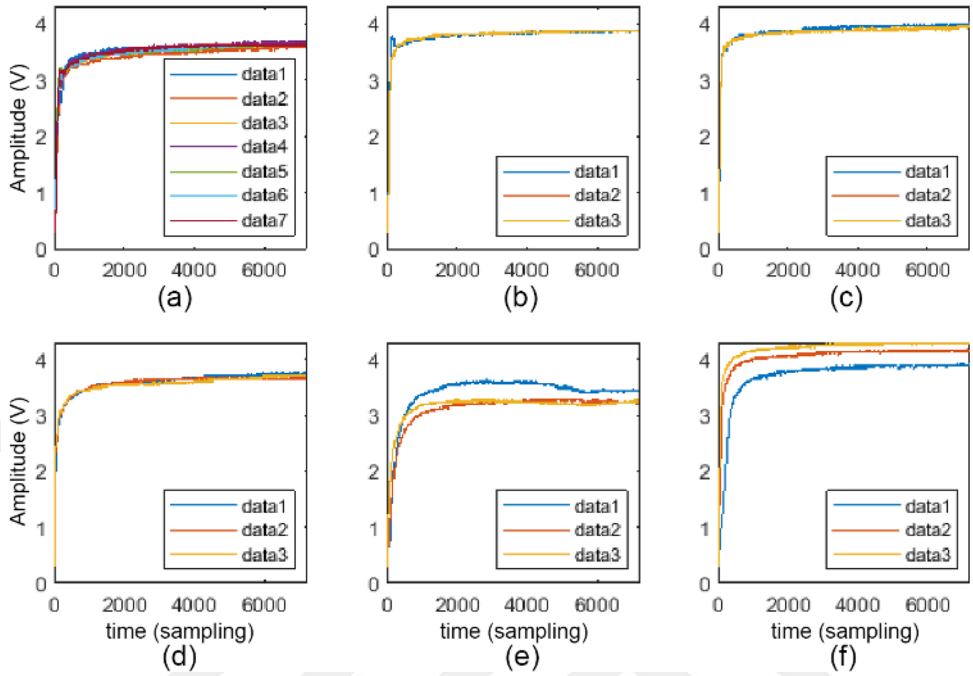


Figure 5.6. Measurement of some soils with (a, b, c) and without (d, e, f) sodium hexamethaphosphate (calgon) solution for 1 hour.

Figure 5.6 shows 3 soil tests with calgon in the upper row (a, b, c) and without calgon in the lower row (d, e, f). Each column describes repeated experiments on the same soil (a-d, b-e, c-f). Although the similarity of the test curves of the same soil with and without calgon is particularly interesting in the graphs shown with a and d, the constant change of the curves in the experiments of the other two soils without calgon reveals the inconsistency in the test standard. The fact that these experiments also support the standard in traditional hydrometer tests has revealed the necessity of “waiting the soil in the calgon for 1 day for USTA as well”.

On the other hand, it is known that all ultrasound sensors (including those used within USTA) are seriously affected by the density of the environment in which they are used. This means that the curves obtained during the experiments within the scope of our study should be considered in this context. Because water-soluble soil components can seriously affect the signals measured from the soil-water mixture, even the curves of two different soils with similar textures may not match.

In order to measure how the USTA is affected by water-soluble soil components, two artificial test materials called salty and sugary distilled water were prepared. In the first of the experiments, 10 g of sugar and in the second, 10 g of salt dissolved in two separate distilled water mixtures were prepared. These two mixtures were prepared by mixing the sugar and salt directly into the container of the USTA device, in which distilled water was pre-placed, and the change in the measured sign in the whole process is shown in Figure 5.7.

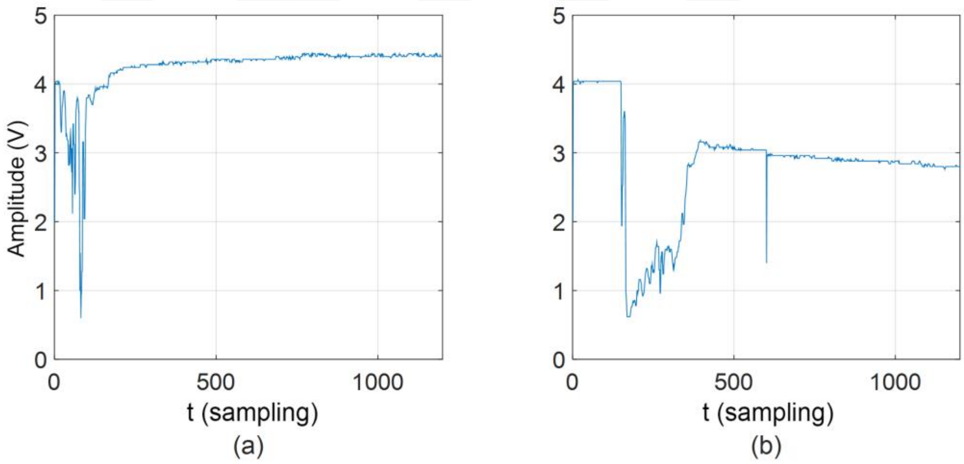


Figure 5.7. Curve of amplitude change of sound waves in sugary water (a) and salty water (b) made with USTA.

As seen in Figure 5.7, the amplitude value changes when any soluble substance is added to the distilled water in the container. In both graphs (a and b), the non-stationary part in which the amplitude is very close to zero represents the mixing moment. Accordingly, while the intensity of the ultrasound waves reaching the receiving sensor increased in the first experiment with sugar added (a), the wave intensity decreased in the second experiment with salt (b), on the contrary.

Therefore, if there are water-soluble substances in the soil-water mixture (which is generally known), it is not known to what extent these substances will affect (increase or decrease) the measurement result. This problem also exists in the traditional hydrometer method as well. But there is a slight difference. Due to the application in the hydrometer test, all water-soluble components are included in the clay component because they increase the density of the water. However, in the USTA experiment, it is not possible to predict this since some water-soluble components increase the amplitude while others decrease it. Therefore, it is a natural consequence that the USTA and hydrometer estimates are slightly different due to water-soluble substances.

There is a known approach for performing the hydrometer method without errors for water soluble substances. Accordingly, the soil sample is first purified from organic matter by burning it with hydrogen peroxide. Then, in order to remove the stickiness, the lime in it is neutralized with the help of acidic chemicals. The soil, which is separated from the hydrogen peroxide and acid substances added at the end, is called inorganic soil. Error-free results are only possible with soil samples that have undergone all these pre-treatments without skipping any of them. However, since these processes must be done sequentially and each process takes a very long time, the test of a soil sample can take up to 1 whole month. As it can be understood from this, hydrometer methods are based on a certain compromise. It is known that serious concessions are made in determining the percentages of sand, silt and clay, since it is important for soil science to correctly classify the soil rather than the percentage distribution of the particles. From this point of view, some time-consuming and compelling processes, which are thought to not affect the estimation results much in the USTA experiments, have been neglected.

5.4. Preprocessing Steps and Feature Extraction

In the light of the analyzes described above, the final shape of the USTA device was determined, the temperature correction function was prepared, it was determined that it was beneficial to keep the mixture in water with calgon for 1 day and there was no need to use a mixer in mixing phase. After this stage, soil experiments with data preparation were carried out and 2 hours long (containing 14.400 samples) repetitive markers were provided for each soil sample. After this preparation, an inspection process, also called the preprocessing step, was run and artifacts caused by noise and electrical contact problems, which are rarely observed in the signals, were corrected with a median filter.

In multi-parameter data such as time series and images, the data is simplified by applying a feature extraction step after the preprocessing step and before applying machine learning methods. At this stage, different features can usually be selected according to the available data. In Figure 5.8, the properties extracted from the test curve obtained from a soil are given.

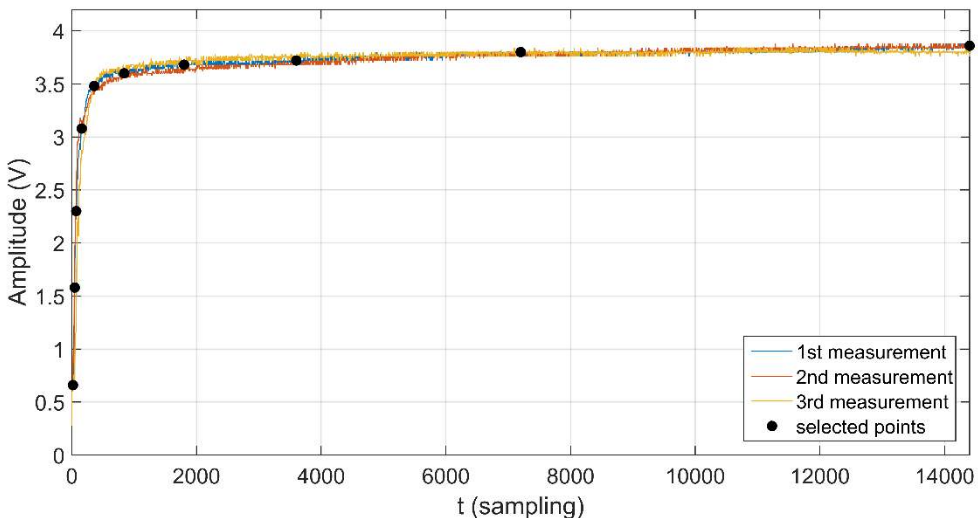


Figure 5.8. Extracted features (points) from a measurement of a soil sample.

When Figure 5.8 is examined, it is observed that the graphs behave like the combination of two curves with different characteristics. It is known that at the beginning of the experiment the particles of sand and silt in the soil-water mixture precipitated. At those moments, the density of large particles between the transmitter and receiver sensor decreases. This causes the intensity of the waves reaching the receiving sensor and thus the amplitude of the recorded signals to increase rapidly. After the settling of these large granular structures is completed, the rate of precipitation, and thus the amplitude increase in the recorded signals, slows down sharply due to the fact that only clay particles remain suspended in the mixture. t = {10 sec, 20 sec, 40 sec, 80 sec, 3 min, 7 min, 15 min, 30 min, 60 min, 120 min} points were determined as the best features to extract. It is thought that especially the rapid rise and return part at the beginning of the measurement contains important information in estimating the soil content. For this reason, data points were selected at frequent intervals from these parts, while data points were selected sparsely after the point where the graph started to be stationary. The datasets to be given to machine learning methods are ready by applying this feature extraction step to all signs.

Due to the nature of computational experiments to be made with machine learning methods, the data should be divided into two parts named training and testing by applying a cross validation method. For this stage, the "leave-one-out" validity was preferred, considering that the data could be processed quickly enough. Accordingly, training experiments were conducted as much as the number of soil samples in the data set, and a soil sample was selected for testing in each experiment and spared from the data. Therefore, for a data with N samples, training is carried out with N-1 samples in each training experiment and testing is provided with the remaining single sample. After all the experiments are finished, N test results are obtained and the average of all these results is used for overall success.

5.5. Estimation with Support Vector Regression

Two regression methods (SVR, MLP), which are successful and popular in the literature, are used in this section, which aims to estimate the sand, silt and clay ratios of the soils tested with USTA. Experiments with USTA, data preprocessing and feature extraction steps from signs, and separation of data into training and test sets using the cross validation method are applied as described in the previous sections. Regression analyzes were performed for each of the three soil components separately. Therefore, in the datasets prepared separately for each soil component, the output information is in the form of a single attribute showing the ratio of the relevant soil component. In other words, the input part of the main dataset is prepared with 10 features extracted from the signals obtained from all regression experiments of 80 soils, the sand ratios for the sand regression experiment are added to these inputs and the machine learning method is applied. The same data preparation is then repeated for silt and clay.

In our regression analysis study, firstly, the SVR method, which does not need parameter adjustment, works faster than modern regression methods such as ANN, and does not need to repeat the training due to the optimization method it uses, was used. In Figure 5.9, the mean absolute error (MAE) values in the sand, silt and clay estimations of 80 soils calculated by SVR are given.

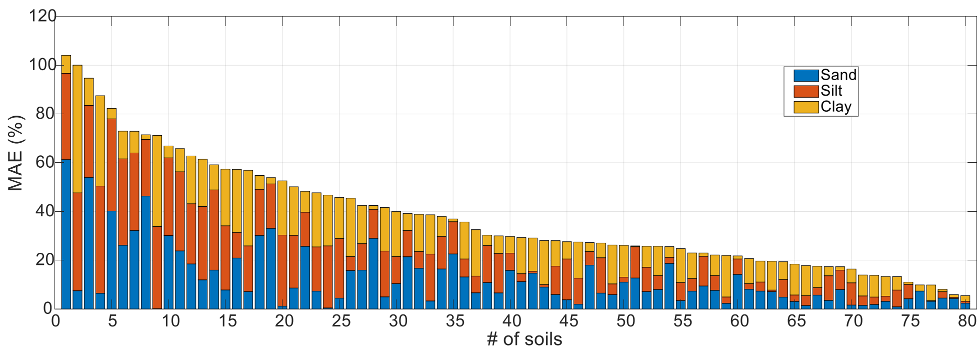


Figure 5.9. Results as MAE values of estimated sand, silt and clay ratios using SVR method on the dataset.

In Figure 5.9, the absolute error rates obtained as a result of the experiments performed for all soils with the SVR are shown separately for sand, silt and clay. Considering the 10% error margin, which was also taken into account in the hydrometer tests, there are 47 soils for sand, 40 for silt and 41 for clay, with a maximum error rate of 10%. Another noticable point in the figure is that the component estimates of some soils are realized with very high error. When these soils with a high error rate were examined, it was understood that there were unique samples in the data set, so it was determined that high error estimation for these soils was a natural result.

5.6. Estimation with Multi-Layer Perceptron Neural Network

Log-sigmoid activation function and levenberg-marquardt training algorithm were used in the ANN network used for analysis, and the whole training-testing process was repeated 50 times starting from the beginning due to random starts. In order to determine the optimal value of the number of hidden layers and the number of neurons in the hidden layers, a trial and error analysis was carried out and the mean absolute error (MAE) values obtained are given in Figure 5.10.

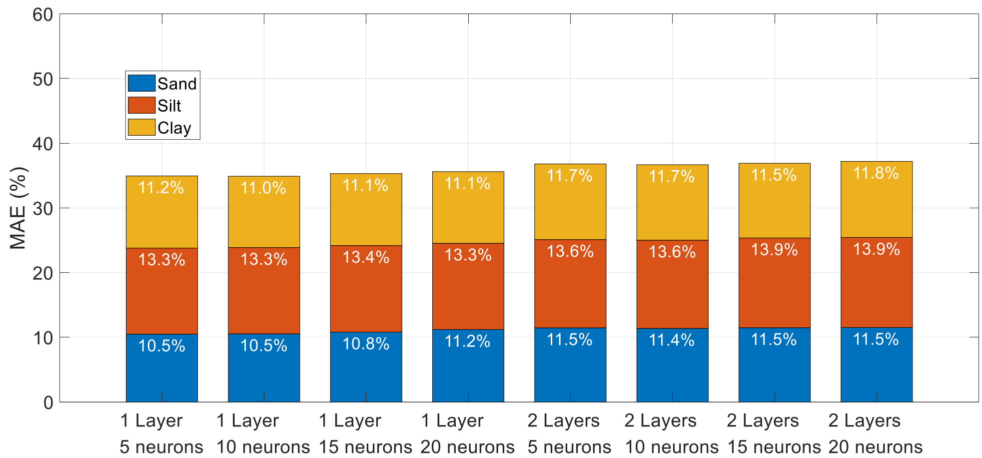


Figure 5.10. MAE values obtained from tests with varying hidden layers having varying neurons in each according to the used ANN architecture.

Figure 5.10 shows the MAE values produced by the repeated experiments on the basis of sand, silt and clay with the ANNs having the various number of hidden layers and the number of neurons in each hidden layer. Experiments were performed separately with ANN constructs with 5, 10, 15 and 20 neurons in 1 hidden layer and 5, 10, 15 and 20 neurons in each of 2 identical hidden layers. Accordingly, it was observed that the architecture with 10 neurons in a single hidden layer produced the lowest MAE. Further increasing of layers or neurons on layers has not produced any improved result. For this reason, ANN structure with 10 neurons in single hidden layer is used in the presentation of the experimental results. All these processes were carried out separately for each of the sand, silt and clay components of 80 soils, and the MAE values obtained in the tests are given in Figure 5.11.

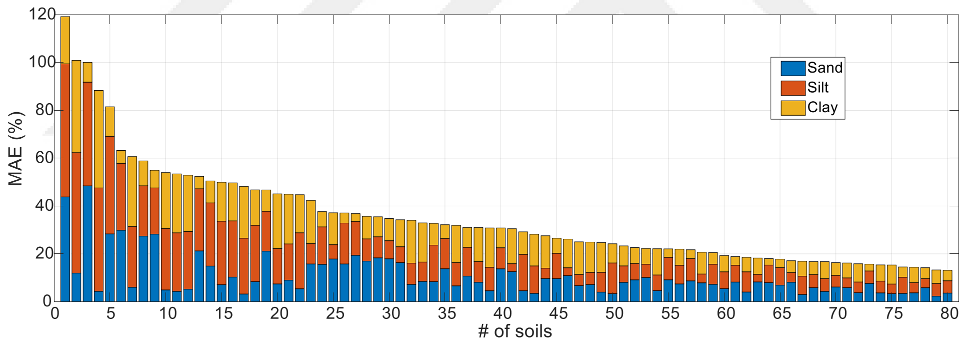


Figure 5.11. Results as MAE values of estimated sand, silt and clay ratios using ANN architecture having single hidden layer and 10 neurons on it.

Figure 5.11 shows the MAE values in the test estimates with ANN of 80 soil samples in the dataset. Estimation errors represent the absolute difference between the percentages of sand, silt and clay in soils tested with hydrometer method and the percentages obtained as a result of the estimation. Accordingly, there are 54 soils for sand, 48 for silt, and 48 soils for clay, with an estimation error of less than 10%.

The fact that all methods produce high errors in the same soil samples proves that the errors are due to the low class diversity in the dataset. The estimation errors of sand, silt and clay ratios obtained by two different machine learning methods used in Table 5.3 are given in terms of the mean absolute error (MAE), mean square error (MSE) and R^2 criteria.

Table 5.2. MAE, MSE and R2 values of SVR and ANN methods.

Method	**Sand (%)**			**Silt (%)**			**Clay (%)**		
	MAE	**MSE**	**R^2**	**MAE**	**MSE**	**R^2**	**MAE**	**MSE**	**R^2**
SVR	12,42	299,68	0,52	13,51	310,29	0,10	11,69	221,22	0,38
MLP	10,53	183,24	0,81	13,34	302,86	0,29	11,02	177,07	0,52

According to the averages of the values in Table 5.2, it was determined that the MLP method produced the lowest error. While the R^2 method by its nature, seems to be successful in the training set in regression models that are non-linear and contain too many independent variables, it performs poorly in the test data. In comparisons across Table 5.2, it can be said that the MSE values coincide with the MAE values. In terms of the amount of information it provides, ease of interpretation and consistency, the use of the MAE criterion was preferred throughout the study.

If a situation similar to the one encountered during testing is not encountered during training, the machine learning method, by its nature, forces itself to simulate this new situation with any pattern it has learned before, and therefore inconsistent (sometimes good, sometimes bad) or highly erroneous results are produced. When the analyzes we made in the study are examined, it can be said that the highest test error occurs in soils that are unique in training. For this reason, a new dataset (subset) named D66 was prepared by excluding the unique soil samples from the dataset (D80), which has 80 soil samples, and the estimation experiments were repeated with the SVR method, which is fastest regression method in the study, and the obtained MAE values are shown in Figure 5.12.

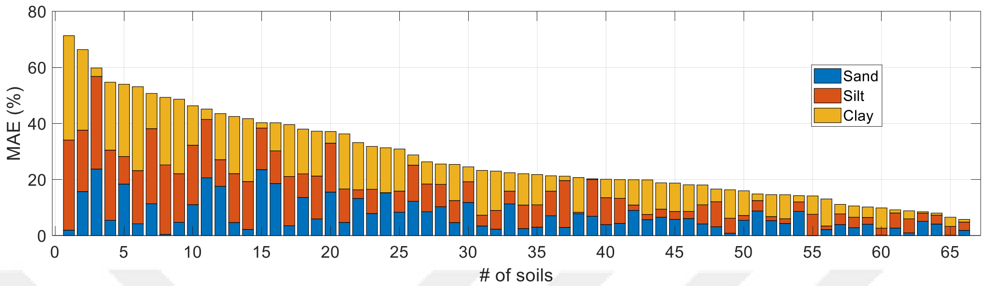


Figure 5.12. Results of sand, silt and clay estimation errors on D66 dataset of using SVR method.

In Figure 5.12, the results of the SVR experiment with soils belonging to classes with at least 3 measurement samples in the dataset are shown. As a result of the experiment, the averages of sand, silt and clay MAE values for the soils included in the experiment were calculated as 7.21%, 9.64% and 10.80%, respectively. On the other hand, if it is remembered that the sum of the first two components is subtracted from the number 100 for the third component in the traditional hydrometer method, it can be said that the sum of the ratios is guaranteed to be 100%. The fact that the sum of the estimations made separately in the USTA experiments produced values very close to 100% created the impression that machine learning methods learned the patterns of soil measurement curves well.

Considering all the estimation experiments and classification experiments carried out together, it is thought that learning methods will produce more accurate and more sensitive results when the samples in the dataset are diversified and reproduced in a balanced way. Based on this idea, a study was conducted on how increasing the similarity of the tested soil in the training set would change the estimation errors, using the available data, and the results are presented in Figure 5.13.

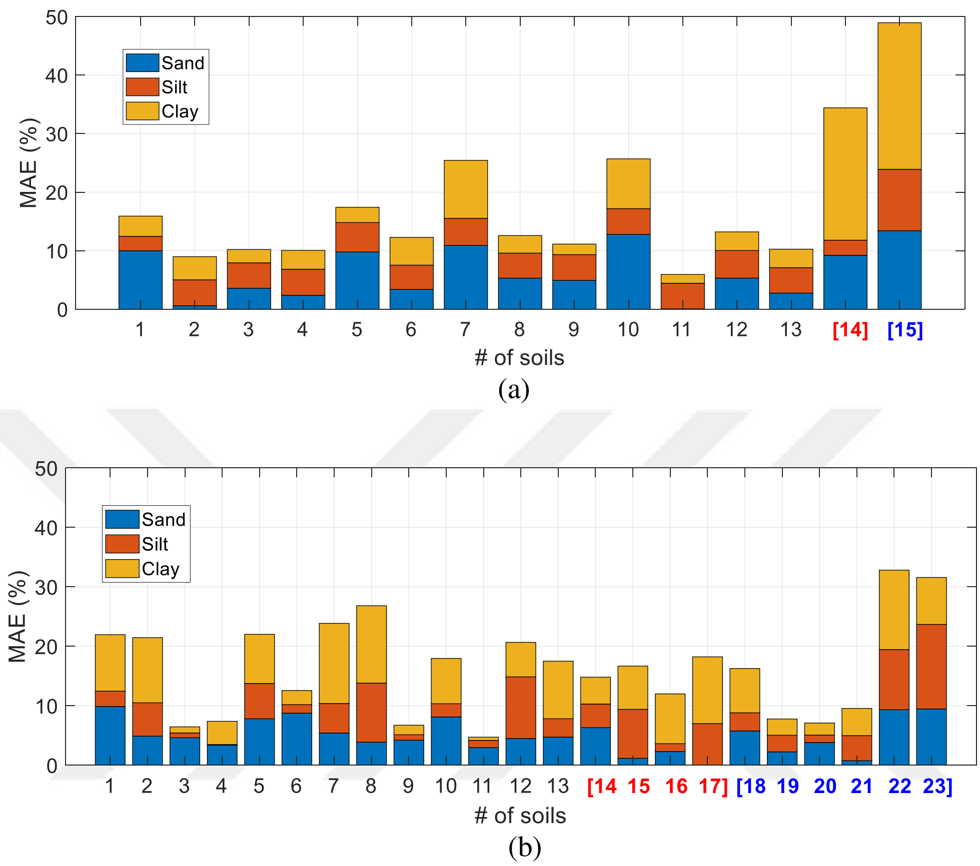


Figure 5.13. The effect of diversifying the data on MAE values. (a) showing the dataset having single sample of two different classes in square brackets, and (b) showing the reduce in MAE values after populating the classes with more sample.

In Figure 5.13 (a), the estimation experiment carried out by adding 1 soil sample belonging to the "clay loam" class (sample 14) and 1 soil sample belonging to the "silty loam" class (sample 15) among 13 soils belonging to the "clay" class, is shown. When Figure 5.13 (b) is examined, it has been observed that the estimation errors of these classes have decreased significantly with the increase of the samples belonging to the "clay loam" (red colored group) and "silty loam" (blue colored group). Initial estimation rates 9.2%, 4.6% and 22.2% for sand, silt and clay, respectively, were dropped to 6.3%, 4.0% and 4.5% with the participation of three more soil samples from the same class for soil sample 14. Likewise, the error rates of soil measurement 15, whose estimation error for sand, silt and clay were 13.4%, 10.5% and 25.1%, respectively, decreased to 5.8%, 3.1% and 7.5% with the participation of five more soil samples from the same class in the experiment. Therefore, it can be said that almost all of the estimations with high error rate in the first figure are due to the unbalanced class distribution in the dataset. Considering the classification analysis explained in next subsection, it can be said that diversifying the dataset with soil samples belonging to different classes will increase both classification and estimation success. Calculated criteria to compare the success of the SVR method on the whole dataset (D80) and the diversity verified dataset (D66) are presented in Table 5.3.

Table 5.3. MAE, MSE and R^2 values obtained for sand, silt and clay estimations on D66 and D80 datasets with SVR methods.

	Sand (%)		**Silt (%)**		**Clay (%)**	
	D66	**D80**	**D66**	**D80**	**D66**	**D80**
MAE	7,21	12,42	9,64	13,51	10,80	11,69
MSE	84,71	299,87	154,78	310,45	182,96	221,22
R^2	0,48	0,48	0,12	0,10	0,38	0,37

When Table 5.3 is examined, it is observed that the results of the experiments performed with classes with three or more soil samples from the same class in the dataset have improved significantly. Considering all three evaluation criteria, the error rates of D66 are lower than D80.

5.7. Determination of Texture Class by Classification

It is known that many particles are included in the wrong class in hydrometer measurements because they are not perfectly spherical. However, since the soil sample needs to be pre-processed for a very long time in order for the hydrometer analysis to give a perfect result, the texture class of the soil at hand may be much more important for soil scientists rather than the percentage distribution of the particles. As a result of the hydrometer test, if the soil belongs to the predicted texture class, percentage errors can often be ignored in order to save time. What is mentioned here is the classification of the tested soil, that is, determining which of the 12 texture classes it belongs to. At this point, the classification ability of the proposed device, which is another potential area of use, has been tested. Although our main goal in our study was to develop a device that detects the sand, silt and clay content of the soil as percentages, it has been observed that the device has a high classification potential as well as its predictive ability during the development stages.

The dataset obtained with the device consists of at least 3 repeated measurements of 80 soil samples and includes 280 measurements in total. However, on the basis of texture class, since soil samples belonging to some classes cannot be obtained, unfortunately, the number of measurements in certain classes is quite low. In the table below, the distribution of the soils used in the dataset by classes is given.

Table 5.4. Number of soils according to texture classes.

Texture Class	**Number of Soils**
Clay	24
Clay Loam	8
Silty Clay Loam	6
Silty Clay	6
Loam	9
Silty Loam	13
Loamy Sand	4
Sandy Loam	4
Sandy Clay Loam	1
Silt	2
Sand	3

It is clearly seen in Table 5.4 that the number of soils belonging to the clay or classes with high clay content in our dataset is higher than the soils of other texture classes. We stated in the previous parts of our article that the reason for this is that the soils used in the experiments consisted of natural soils collected from Çukurova region and its surroundings. However, in the classification experiments, the measurement data of the soils belonging to the first 4 classes shown in the table were used one by one, first in equal numbers and then increased. Classes with 2 or less soil samples were not used in the texture class.

Firstly, 4 soils from clay, clay loam and silty clay loam classes and 3 soils from silty clay classes were determined, and a sub-dataset of 15 samples was created by randomly selecting one from repeated measurements of these soils. This sub-dataset used in classification has been tried to be classified using some classifier methods. One of the measurements in the dataset was given for testing purposes and the others as a training set were given as input to the methods with the leave-one-out method. The classification results of each method were analyzed with a confusion matrix showing the predicted classes versus the actual classes. Accordingly, the estimation results of the first sub-dataset with the SVM method are as follows.

Table 5.5. Confusion matrix of classification process using SVM method with 4 soils from clay, clay loam and silty clay loam classes, 3 soils from silty clay class.

	Clay	Clay Loam	Silty Clay Loam	Silty Clay
Clay	1	3		
Clay Loam		3	1	
Silty Clay Loam		1	2	1
Silty Clay		1		2

Each row in Table 5.5 shows which classes the soils of that class were included in the testing phase. In this notation, also called error matrix representation, the intersection of rows and columns representing the same class, in other words the results on the diagonal, show the soil samples classified correctly. The red colored ones show the samples that were incorrectly included in another class. For example, while 3 soils from the "clay loam" class were correctly classified as they should be, 1 of them was included in the "silty clay loam" class. Since these two classes are adjacent classes in the texture triangle, it can be understood that the incorrectly classified sample is not included in a class far from the correct class. It can be interpreted that as the sensitivity and the number of sample data are increased, the misclassifications will decrease rapidly. The success rates obtained as a result of testing the same sub-dataset with 4 different classification methods are given in the table below on class basis and in total.

Table 5.6. Soil classification of 15-sample sub-dataset using various classification methods.

	Clay	Clay Loam	Silty Clay Loam	Silty Clay	Overall Success
SVM	1/4 (25 %)	3/4 (75 %)	2/4 (50 %)	2/3 (66 %)	53 %
K-nearest Neighbours	2/4 (50 %)	3/4 (75 %)	2/4 (50 %)	2/3 (66 %)	60 %
Decision Table	0/4 (0 %)	3/4 (50 %)	1/4 (25 %)	0/3 (0 %)	26 %
Random Forest	1/4 (25 %)	3/4 (75 %)	2/4 (50 %)	2/3 (66 %)	53 %

The total success rates obtained with very few randomly selected samples are shown in Table 5.6. The aim here was to see whether the time series obtained as a result of the experiments could be used in classification, rather than the success of the classifiers. The results of the second classification experiment, in which the "clay" class was repeated by increasing the number of samples belonging to other classes by a few more samples, are as in the table below.

Table 5.7. Soil classification of 18-sample sub-dataset using various classification methods.

	Clay	Clay Loam	Silty Clay Loam	Silty Clay	Overall Success
SVM	5/7 (71 %)	3/4 (75 %)	2/4 (50 %)	2/3 (66 %)	66 %
K-nearest Neighbours	2/7 (28 %)	3/4 (75 %)	2/4 (50 %)	2/3 (66 %)	50 %
Decision Table	4/7 (57 %)	2/4 (50 %)	0/4 (0 %)	0/3 (0 %)	33 %
Random Forest	5/7 (71 %)	3/4 (75 %)	2/4 (50 %)	2/3 (66 %)	66 %

When Table 5.7 is examined, it can be seen that the success in that class gradually increased with the increase in the number of samples belonging to the "clay" class. While the success of the clay column in Table 5.7 varies between 25 ~ 50%, it can be seen in Table 4 that this success increased to the range of 57 ~ 71%. In the third sub-dataset, which was prepared to reveal more clearly whether this increase in success is a coincidence of only the "clay" class, the samples belonging to the "clay" class were decreased, while the samples belonging to the "clay loam" class were increased and the balance was changed, and the results are presented in the table below.

Table 5.8. Classification of 18-sample sub-dataset (reduced clay class, increased clay loam class) using various classification methods.

	Clay	Clay Loam	Silty Clay Loam	Silty Clay	Overall Success
SVM	1/4 (25 %)	6/7 (85 %)	2/4 (50 %)	2/3 (66 %)	61 %
Knearest neighbours	2/4 (50 %)	6/7 (85 %)	2/4 (50 %)	2/3 (66 %)	66 %
Decision Table	1/4 (25 %)	5/7 (71 %)	1/4 (25 %)	0/3 (0 %)	38 %
Random Forest	1/4 (25 %)	6/7 (85 %)	2/4 (50 %)	2/3 (66 %)	61 %

When the Table 5.8. is examined, it is clearly seen that the success rates of the "clay" class with reduced sample numbers decreased, while the success rates of the "clay loam" class with increased sample numbers increased. The results of the fourth sub-dataset containing 25 measurements, including the maximum number of samples that can be used, are as shown in the table below.

Table 5.9. Classification of 25-sample sub-dataset using various classification methods.

	Clay	Clay Loam	Silty Clay Loam	Silty Clay	Overall Success
SVM	9/9 (100 %)	6/7 (85 %)	4/6 (66 %)	2/3 (66 %)	84%
K nearest neighbours	5/9 (55 %)	6/7 (85 %)	4/6 (66 %)	2/3 (66 %)	68%
Decision Table	7/9 (77 %)	5/7 (71 %)	4/6 (66 %)	0/3 (0 %)	64 %
Random Forest	9/9 (100 %)	6/7 (85 %)	4/6 (66 %)	2/3 (66 %)	84 %

When Table 5.9 and the previous tables are examined respectively, it is seen that the success rate increases for the classes with increasing sample data (“clay”, “clay loam” and “silty clay loam”), while the success decreases for the classes with less sample data. It is observed that there is a noticeable increase in total success, especially with the increase in the total number of data. Below is the graph of the success rate of the 3 classes, in which the success increases significantly as the number of samples in the dataset increases.

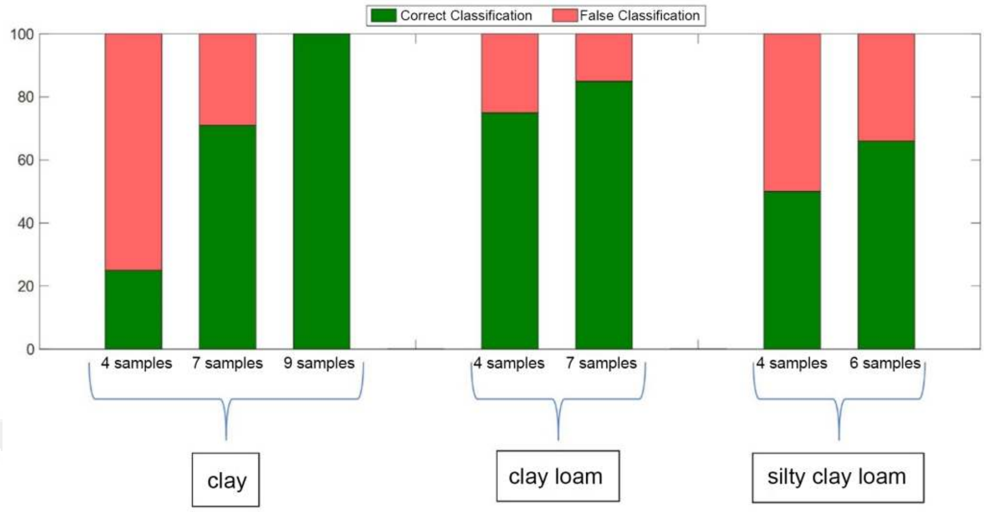


Figure 5.14. Increasing texture classification success of clay, clay loam and silty clay loam classes depending on the number of samples belonging to respective classes

Figure 5.14 visualizes the increasing classification success rates due to the gradual increase of the data sample in the “clay”, “clay loam” and “silty clay loam” classes. The x-axis represents the increase in the number of samples belonging to the same class during classification experiments, and the y-axis represents the percent success rate. Columns represented in green indicate the correct classification rate, and columns represented in red indicate the rate of misclassification. The "silty clay" class included in the datasets was not included in the image because it did not have enough sample diversity to show variation. Experiments with a small number of samples are promising for future studies in the development of the device. Accordingly, it is obvious that soil measurements belonging to different classes in the dataset should be increased by diversifying. In parallel, the number of experiments should be increased significantly. In conclusion, although it is still early to say anything definitively for the classification process, it can be clearly said that the device can classify as the measurements of soils belonging to different texture classes are increased.

6. CONCLUSIONS AND DISCUSSIONS

Traditional analysis methods, such as the hydrometer method, are still frequently used in the physical analysis of soils in areas requiring tillage, such as agriculture, construction and mining. Although these methods are old and therefore have well-established methodologies, they are very laborious and time-consuming because they are performed manually. Since all stages are carried out manually under the supervision of an expert, they are highly prone to human error. Considering today's technology, together with all these disadvantages of traditional methods, it has become inevitable to switch to computerized calculation in areas where soil is in question, as in all areas. Especially as software designed using smart algorithms develops, the cost of computation and the time spent decrease, and the need for manpower decreases to a minimum. In this context, with our study on the determination of the sand, silt and clay content of the soil, the preliminary steps of a system design that will provide serious benefits to the literature are presented.

In our study, the proportions of inorganic substances in the soil-water mixture in the 3D designed container were tried to be determined with the help of ultrasound sensors, with the same principle as in a pioneering study carried out with light sources, and many experiments were carried out for the determination of texture. During the experiments, the effects of water-soluble substances on the prediction, the effects of temperature on the experiments, the shortening of the experiment period were discussed and the results were presented comparatively. In addition, classification experiments have been carried out to determine the texture classes of soils by using algorithms such as SVM, K-nearest Neighbors, Decision Table and Random Forest, which are frequently used in the literature, and the results are presented in comparative tables. One of the most important issues that stood out during the experiments was that the estimation times could be shortened even below the standard time by taking advantage of the benefits of computerized estimation. Although the shortening of the estimation time is seen as an important advantage, it has not been completed due to the high processing workload and is considered as a future study.

In the studies carried out within the scope of the design, a special analysis was made for the sensor positions and it was determined that the measurement of the sensors 3 cm below the soil-water mixture gave the most successful results. In the regression studies, the mean absolute error remained at the level of 5% and the target was achieved, especially for the samples where reliable data were obtained. It has been determined that the temperature seriously affects the results and therefore the measurements need to be corrected with a function, and it has also been determined that keeping the soil in water with calgon provides serious confidence. Within the scope of the study, a total of 80 samples, each weighing at least 100 gr, were obtained. Experiments were re-evaluated by using "examples with sufficient number for training" and it was determined that the success was at the 5% level. During the studies, the design of the USTA device was changed many times due to different problems and the 3D printing process has been repeated over and over.

Although there are many methods focused on finding the particle size distribution (PSD), the most important skill of the proposed device is that it is used with real soils. By “real”, it means it is used with the soils with unknown parameters in it (water soluble substances, etc.). One can infer that it would be quite easy to find the size distribution of pre-treated soil (artificial particles or soils with organic substances removed.). But naturally, that would not be usable for in-situ applications. In the comparative results, machine learning methods, which are frequently used in the literature, such as SVR and MLPNN, were used. As a result, an automated and mobile device designed with 3D printing, controlled by a microcomputer, capable of predicting with the smart systems installed on it, has been introduced, and the design details and working principle are clearly presented.

REFERENCES

AFNOR., 1979. Granulome´trie. Analyse granulome´trique des poudres fines sur tamiseuse a` de´pression d'air. Doc. NF-X-11- 640; and subsequent document: 1981b: NF-X-11-630. Paris: Association Française de Normalisation.

Allen, T., Davıes, R., and Scarlett, B., 1996. Particle Size Measurement. (2 vols.). London: Chapman and Hall.

Atterberg, A., 1916. Die Klassifikation der humusfreien und der humusarmen Mineralboden Schwedens nach den Konsistenzverhaltnissen derselben. Int. Mitt. Bodenkd., 6, 27–37.

Avery, B. W., and Bascomb, C. L., 1982. Soil survey laboratory methods. Harpenden: Soil Survey of Great Britain [England and Wales].

Barth, H. G., and Sun, S. T., 1985. Particle size analysis. Analytical Chemistry, 57(5), 151–175. https://doi.org/10.1021/ac00282a009

Boser, B. E., Guyon, I. M., and Vapnık, V. N., 1992. A Training Algorithm for Optimal Margin Classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory (pp. 144–152). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/130385.130401

BSI., 1981. Code of Practice for Site Investigation, Standard BS5930; and subsequent Standards: 1986: BS410; 1990: BS1377 (Part 2); 1993: BS3406 (Part 4); 1994a: BS3406 (Part 5); 1994b: BS 3406 (Part 6); 1998: BS7755. London.

Burt, R., Reınsch, T. G., and Mıller, W. P., 1993. A micro-pipette method for water dispersible clay. Communications in Soil Science and Plant Analysis, 24(19–20), 2531–2544. https://doi.org/10.1080/00103629309368975

Carter, M. R., 1993. Soil sampling and methods of analysis. Boca Raton: Lewis Publishers.

Cortes, C., and Vapnık, V., 1995. Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Gee, G. W., and Bauder, J. W., 1986. Particle-size Analysis. In A. Klute (Ed.), Methods of Soil Analysis (pp. 383–411). https://doi.org/https://doi.org/10.2136/sssabookser5.1.2ed.c15

Gök, H., and Orhan, U., 2015. Topraktaki Kum Oranının Bilgisayar Destekli Tespiti. In 23. Sinyal İşleme ve İletişim Uygulamaları Kurultayı (pp. 1178–1181). Adana.

Grıffıths, J. C., 1967. Scientific Method in Analysis of Sediments. New York: McGraw-Hill.

Grout, H., Tarquıs, A. M., and Wıesner, M. R., 1998. Multifractal Analysis of Particle Size Distributions in Soil. Environmental Science & Technology, 32(9), 1176–1182. https://doi.org/10.1021/es9704343

Guyon, I., Boser, B. E., and Vapnık, V., 1992. Automatic Capacity Tuning of Very Large VC-Dimension Classifiers. In Advances in Neural Information Processing Systems 5, [NIPS Conference] (pp. 147–155). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Head, K. H., 1992. Manual of soil laboratory testing. Vol. 1, Soil classification and compaction tests (2nd ed.). London: Pentech.

INTERNATIONAL SOCIETY OF SOIL SCIENCE., 1928. The Study of Soil Mechanics and Physics. Report of Commission I. In 1st Int. Congr. Soil Sci (pp. 359–404). Washington, DC.

ISO., 1973. Guide to the Choice of Series of Preferred Numbers and Series Containing More Rounded Values of Preferred Numbers. Doc. ISO-497-1973-(E) (1st ed.); and subsequent documents: 1977: ISO- 387-(E) (1st ed.); 1998: ISO-11277-1-(E) (1st ed.). Geneva, CH.

Kılınç, E., and Orhan, U., 2016. Estimation of sand ratio in soil by LDR sensors and linear regression. In 1st International Mediterranean Science and Engineering Congress (pp. 1335–1342). Adana.

Klute, A. (Ed.)., 1986. Methods of Soil Analysis, Part I. Physical and Mineralogical Methods. (2nd ed.). Madison, WI: Am. Soc. Agron.

Laxton, J. L., 1980. A method for estimating the grading of boulder and cobble grade material (size-analysis). (Vol. 80).

Maeda, T., Takenaka, H., and Warkentın, B. H., 1977. Physical properties of allophane soils. Adv. Agron., (29), 229264.

Metz, R., 1985. The importance of maintaining horizontal sieve screens when using a Ro- Tap. Sedimentology, 32, 613–614.

Nadeau, P. H., 1985. The Physical Dimensions of Fundamental Clay Particles. Clay Minerals - Clay Mıner (Vol. 20). https://doi.org/10.1180/claymin.1985.020.4.06

Odén, S., 1915. Eine neue Methode zur mechanischen Bodenanalyse. Int. Mitt. Bodenanal, 5, 257–311.

Orhan, U., and Kılınç, E., 2020. Estimating soil texture with laser-guided Bouyoucos. Automatika, 61(1), 1–10. https://doi.org/10.1080/00051144.2019.1654283

Rıngrose-Voase, A. J., and Bullock, P., 1984. The automatic recognition and measurement of soil pore types by image analysis and computer programs. Journal of Soil Science, 35(4), 673–684. https://doi.org/10.1111/j.1365-2389.1984.tb00624.x

Schölkopf, B, Burges, C., and Vapnık, V., 1995. Extracting Support Data for a Given Task. In KDD.

Schölkopf, B., Burges, C., and Vapnık, V., 1996. Incorporating invariances in support vector learning machines. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, & B. Sendhoff (Eds.), ICANN 96 (pp. 47–52). Berlin, Heidelberg: Springer Berlin Heidelberg.

Schölkopf, B., and Smola, A. J., 2002. Learning with Kernels. MIT press.

Sherman, C. H., and Butler, J. L., 2007. Electroacoustic Transduction. In Transducers and Arrays for Underwater Sound (pp. 31–75). Springer New York. https://doi.org/10.1007/978-0-387-33139-3_2

Shıozawa, S., and Campbell, G., 1991. On the calculation of a mean particle diameter and standard deviation from sand, silt and clay fractions. Soil Science, 152, 427–431.

Smart, P., and Tovey, N. K., 1982. Electron microscopy of soils and sediments : techniques. Clarendon Press.

Smıth, K. A., 2000. Soil and Environmental Analysis: Physical Methods, Revised, and Expanded. Soil and Environmental Analysis (2nd ed). CRC Press. https://doi.org/10.1201/9780203908600/Soıl-Envıronmental-Analysıs-Keıth-Smıth

Smola, A. J., and Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing, 14(3), 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88

Stabınger, H., Leopold, H., and Kratky, O., 1967. Eine neue Methode zur Präzisionsmessung der Dichte von Flüssigkeiten - Kurze Mitteilung. Monatshefte Fur Chemie - Monatsh Chem (Vol. 98). https://doi.org/10.1007/BF00899963

Sur, H. S., and Kukal, S. S., 1992. A Modified Hydrometer Procedure For Particle Size Analysis. Soil Science, 153(1), 1–4.

Syvıtskı, J. P. M., 1991. Principles, methods, and application of particle size analysis. (J. P. M. Syvitski, Ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511626142

Tanner, C. B., and Jackson, M. L., 1948. Nomographs of Sedimentation Times for Soil Particles Under Gravity or Centrifugal Acceleration. Soil Science Society of America Journal, 12(C), 60–65. https://doi.org/https://doi.org/10.2136/sssaj1948.036159950012000C0014x

Tyler, S. W., and Wheatcraft, S. W., 1992. Fractal Scaling of Soil Particle-Size Distributions: Analysis and Limitations. Soil Science Society of America Journal, 56(2), 362. https://doi.org/10.2136/sssaj1992.03615995005600020005x

USDA., 1996. Soil Survey Laboratory Methods Manual. Soil Survey Investigations Report No. 46, Version 3.0. Washington D.C.

Vapnık, V., and Lerner, A., 1963. Pattern recognition using generalized portrait method. Automation and Remote Control, (24), 774–780.

Vogelaar, B., and Golombok, M., 2017. Simplified Ultrasonic Damage Detection in Fluid-Filled Pipes. Journal of Nondestructive Evaluation, 36(4), 63. https://doi.org/10.1007/s10921-017-0443-3

CURRICULUM VITAE

Emre KILINÇ, I finish Halil Çiftçi Anatolian High School there. I studied Computer Engineering between 2010 and 2014 in Eskişehir Anadolu University. After a short work in private sector (OBASE software), i started MSc in Çukurova University in 2014. I am currently working as Research Assistant and looking forward to achive my PhD title upon the completion of this thesis.

APPENDIX

Table. Detailed textural information of 80 soil samples.

Soil #	Sand (%)	Silt (%)	Clay (%)	Texture Class
1	6.4	38.1	55.3	Clay
2	5.5	39.9	54.5	Clay
3	6.1	39.7	54.1	Clay
4	8	39.8	52	Clay
5	6.1	39.7	54	Clay
6	2.8	39.2	57.9	Clay
7	6.8	39.9	53.2	Clay
8	9.8	39.6	50.4	Clay
9	7.1	39.9	52.9	Clay
10	6.4	39.8	53.6	Clay
11	6.4	39.7	53.8	Clay
12	5.5	39.8	54.5	Clay
13	11.4	39.5	49	Clay
14	9.7	39.7	50.5	Clay
15	7.5	39.8	49.1	Clay
16	9.2	35.2	55.6	Clay
17	5.1	39	55.9	Clay
18	5.1	34.9	60	Clay
19	3.9	37.8	58.3	Clay
20	4	33.1	62.9	Clay
21	5.1	39.7	55.2	Clay
22	12.5	35.1	52.3	Clay
23	32.5	22	45.4	Clay
24	43.8	13	43	Clay
25	30.8	41.3	27.9	Clay Loam
26	27.6	41.3	31.1	Clay Loam
27	27.4	43.5	29.1	Clay Loam

28	24.1	47.4	28.5	Clay Loam
29	33.2	29.9	36.7	Clay Loam
30	24.6	46.4	28.9	Clay Loam
31	28.8	46.4	24.7	Clay Loam
32	20.4	42.2	37.3	Clay Loam
33	89.6	7.1	3.3	Sand
34	92	1.8	6.2	Sand
35	99.9	0	0.1	Sand
36	48.1	27.6	24.3	Sandy Clay Loam
37	60.8	23.4	15.8	Sandy Loam
38	69.7	16.1	14.2	Sandy Loam
39	59.1	27.2	13.7	Sandy Loam
40	82.1	5.8	12.1	Sandy Loam
41	6.8	87	6.2	Silt
42	9.1	85.3	5.6	Silt
43	10.3	48	41.7	Silty Clay
44	8.2	48.2	43.6	Silty Clay
45	3.3	55.7	41	Silty Clay
46	4.2	53.4	42.4	Silty Clay
47	16.2	40	43.7	Silty Clay
48	7.7	46.4	45.8	Silty Clay
49	18.5	38.3	43.2	Silty Clay Loam
50	20.1	32.2	47.7	Silty Clay Loam
51	3.7	22.6	73.7	Silty Clay Loam
52	18.2	52.7	29	Silty Clay Loam
53	18.2	48.5	33.1	Silty Clay Loam
54	16.2	50.6	33.1	Silty Clay Loam
55	24	50.2	25.8	Silty Loam
56	18.6	70	11.4	Silty Loam
57	15.4	79.6	5	Silty Loam

58	15.9	70.2	13.9	Silty Loam
59	21.5	69.3	9.2	Silty Loam
60	18.5	68	13.5	Silty Loam
61	27.9	66	6.1	Silty Loam
62	17.2	71.2	11.6	Silty Loam
63	18.3	74.2	7.5	Silty Loam
64	32.7	50.1	17.2	Silty Loam
65	19.6	68.8	11.6	Silty Loam
66	43.5	50.5	6	Silty Loam
67	14.5	73	12.5	Silty Loam
68	26.6	47.1	26.3	Loam
69	46.7	31.5	21.8	Loam
70	28.3	46.2	25.5	Loam
71	47.1	35.4	17.5	Loam
72	43.1	33.6	23.3	Loam
73	27.6	48.5	23.9	Loam
74	44.8	33.7	21.5	Loam
75	43.1	32.2	24.7	Loam
76	38.3	35.7	25.9	Loam
77	84	3.6	12.4	Loamy Sand
78	87	7	6	Loamy Sand
79	77.8	20.6	1.6	Loamy Sand
80	88	6	6	Loamy Sand