ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

DIAGNOSIS OF THYROID DISEASE VIA SUPPORT VECTOR MACHINES

Msc. THESIS

Nuri KORHAN

Department of Mechatronics Engineering

Mechatronics Engineering Programme

DECEMBER 2015

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

DIAGNOSIS OF THYROID DISEASE VIA SUPPORT VECTOR MACHINES

M.Sc. THESIS

Nuri KORHAN
518121046

Department of Mechatronics Engineering

Mechatronics Engineering Programme

Thesis Advisor: Asst. Prof. Dr. Gülay ÖKE

DECEMBER 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

DESTEK VEKTÖR MAKİNELERİ İLE TİROİD HASTALIKLARI TANISI

YÜKSEK LİSANS TEZİ

Nuri KORHAN
518121046

Mekatronik Mühendisliği Anabilim Dalı

Mekatronik Mühendisliği Programı

Tez Danışmanı: Yrd. Doç. Dr. Gülay Öke

ARALIK 2015

Nuri Korhan, a **M.Sc.** student of ITU **Graduate School of Science Engineering and Technology** student ID 518121046, successfully defended the **thesis** entitled "**DIAGNOSIS OF THYROID DISEASE VIA SUPPORT VECTOR MACHINES**", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Asst. Prof. Dr.Gülay ÖKE GÜNEL** İstanbul Technical University	.....
Jury Members :	**Asst. Prof. Dr.Gülay ÖKE GÜNEL** İstanbul TechnicalUniversity	.....
	Asst. Prof. Dr. Hatice KÖSE İstanbul Technical University	.....
	Asst. Prof. Dr.Figen ÖZEN Haliç University	.....

Date of Submission : 25 November 2015
Date of Defense : 25 December 2015

To my Family

FOREWORD

I am really grateful to my supervisor Asst. Prof. Dr. Gülay ÖKE GÜNEL for her management and support. There is no way that I can repay the goods of her guidance. I am truely in debt and I will always be.
I would like to thank to my friends for their encouragement and being with me all the time.
I would like to offer my gratitude to my friend Esra KAYA for sharing the first dataset that led me to start studying on classification.
Finally i would like to express my special thanks to all members of my lovely family for their endless support of any kind.

November 2015 Nuri KORHAN

TABLE OF CONTENTS

Page

ABBREVIATIONS

SVM	**:** Support Vector Machines
PNN	**:** Probabilistic Neural Networks
KKT	**:** Karush-Kuhn-Tucker
GA	**:** Genetic Algorithms
FS	**:** Feature Selection
T3RU (F1)	**:** T3-Resin Uptake Test
T4 (F2)	**:** Total Serum Thyroxin
T3 (F3)	**:** Total Serum Triiodothyroine
TSH (F4)	**:** Thyroid-Stimulating Hormone
MAD-TSH (F5)	**:** Maximal absolute difference of TSH after injection of 200 micrograms of thyrotropin-releasing hormone as compared to the basal value
ROC	**:** Receiver Operating Characteristics

SYMBOLS

$\mathbf{w}$	: Normal vector to the hyperplane
$\mathbf{x}$	: Input vector
$\mathbf{b}$	: Bias
$\mathbf{C}$	: Penalization coefficient of the slack variable
$\boldsymbol{\xi}$	: Slack Variable
$\boldsymbol{\alpha}$	: Lagrangian multiplier
$\mathbf{y}$	: Output vector
$\boldsymbol{\gamma}$	: Positive regularization parameter
$\boldsymbol{\mu}$	: Mean vector
$\mathbf{S}$	: Total scatter matrix

LIST OF TABLES

Page

LIST OF FIGURES

Page

DIAGNOSIS OF THYROID DISEASE VIA SUPPORT VECTOR MACHINES

SUMMARY

The thyroid gland is the organ that is located on the anterior side of our neck. The duty of thyroid gland is to produce and stock thyroid hormone and to regulate the metabolism by transferring thyroid hormone to the blood when it is needed.
When thyroid gland does not secrete enough amount of hormones, pituitary gland makes thyroid gland produce more hormone by increasing TSH secretion. Shortage of the hormones of thyroid gland is called hypothyroid.
If thyroid gland secretes too much hormones, TSH hormone, secreted from pituitary gland, decreases. That is, the more T3 and T4 hormones are in our blood, the less is the TSH hormone. This condition is called hyperthyroid.
All cells in our body are affected by thyroid hormones. The growth of human in mother's womb, after birth and all metabolism functions are controlled by thyroid hormones. There is almost no organ or cell that is not affected by thyroid hormones. Therefore early diagnosis of thyroid disease is undoubtedly important.

Support Vector Machine (SVM) is an important learning machine that is based on a search for an optimal separating hyperplane that is able to separate the samples of two different classes.

This research aims to construct a system based on classification via SVM for diagnosis of thyroid diseases. Since it is the matter of decision, the diagnosis of diseases can be predicted by classifiers. Having been a quite popular classification algorithm, SVM is among the best classifiers to deal with this duty.
Training and testing data consists of test results of 215 different people taken from a machine learning repository. 35 samples of hyperthyroid patients , 30 samples of hypothyroid patients and 150 healthy samples are used. Input space of data consists of 5 different inputs (T3RU, T4, T3, TSH and MAD-TSH).

Feature subset selection is used to increase the accuracy of the corresponding classifier. By using feature selection, one can easily decrease the number of features. In this thesis Fisher Score Algorithm is used to perform this preprocessing procedure. The most important features in classification are obtained by Fisher Score Algorithm. Then the most important features are used to train the network.

Another important feature in an SVM classifier is the selection of parameters values used in classification. Parameters are what make a classifier manipulatable in order to successfully classify the patterns. One of the most important parameters in SVM is slack variable. Slack variable is used when the data is not linearly or non-linearly separable (this could happen when there is noise or when it is really impossible to

separate the patterns). Using slack variable prevents SVM to create useless optimally separating hyperplanes. The weight of slack variable is adjusted by a coefficient called Soft Margin constant "C". Normally parameter C is adjusted by the user considering the characteristics of dataset. However, it is not easy to decide which value for C is optimal. This problem is handled by one of the most popular optimization algorithms called "Genetic Algorithm". Genetic Algorithm is an optimization algorithm that is inspired by the nature of evolutionary process. In this thesis, an expert system is developed for the diagnosis of thyroid diseases, by combining these three methods.

DESTEK VEKTÖR MAKİNELERİ İLE TİROİD HASTALIKLARI TANISI

ÖZET

Destek Vektör Makineleri Günümüzde en popüler Makine Öğrenmesi yöntemleri içerisinde yer alır. 1964 yılında Vladimir Vapnik tarafından alt yapısı oluşturulan bu yöntem günümüzde medikal analizden görüntü işlemeye, el yazısı tanımadan oyun programlarına kadar hemen her yerde kullanılır hale gelmiştir.

Optimal Ayırıcı Düzlem esasına dayanan Destek Vektör Makineleri yöntemi yöntemi öncelikle veriyi lineer olarak iki sınıfa ayrılabilecek şekilde daha yüksek boyutlu bir uzaya taşır. Eğer verinin içerisindeki iki farklı sınıf bu uzayda ayrılabiliyorsa iki veri kümesinin arasına her iki sınıfın en yakın üyelerine eşit uzaklıkta olacak şekilde bir ayırıcı düzlem yerleştirilir.

Bu şekilde öğrenme işlemini gerçekleştirdikten sonra sınıflandırılması istenecek herhangi bir örneği başarılı bir şekilde sınıflandırır. Eğer ilgili veri lineer olarak ayrılamıyor ise "esnek marjin" adı verilen bir yöntemle verinin mümkün olduğu kadar az hata ile sınıflandırılması sağlanır. Bu parametre kullanıcı tarafından veriye özgü ayarlanır.

Tiroid bezi boynumuzun ön tarafında bulunan salgı bezidir. Tiroid bezinin görevi tiroid hormonlarını üretmek, depolamak ve gerektiğinde kana vererek metabolizmayı ayarlamaktır. Bu hormonları üretirken gıda ve su yardımı ile aldığımız iyot mineralini kullanır.

Tiroid bezinin bozukluğundan kaynaklanan çeşitli hastalıklar vardır. Bunlardan en sık görülen hipertiroidizm ve hipotiroidizm hastalıklarıdır. Hipertiroidizm tiroid bezinin aşırı çalışması durumudur. Graves hastalığı veya benzeri durumlardan sonra ortaya çıkar. Hipotiroidizm tiroid bezinin az çalışması durumudur. Bu durum genellikle iyot eksikliği sonucu ortaya çıkar.

Tiroid hastalıklarını teshis edebilmek için hastalardan kan örneği alınarak çeşitli hormon seviyesi değerleri incelenir.

Genetik algoritmalar, herhangi bir sürekli veya süreksiz fonksinunun aldığı en düşük veya en yüksek değerleri bulmak için kullanılabilecek stokastik bir optimizasyon yöntemidir. İlk olarak John Holland tarafından 1975 yılında ortaya atılmıştır.

Öğrenme ve Test etme verileri 215 farklı kişinin test sonuçlarından elde edilmiştir. İlgili verileri elde edebilmek için "machine learning repository" adlı internet sitesinin veri tabanı kullanılmıştır. Verileri dağılımı: "Hipertiroid=35 sonuç, Hypotiroid=30 sonuç ve sağlıklı sonuç sayısı =150" şeklindedir. Tanıyı koyabilmek için kullanılan giriş verileri 5 adet (T3RU, T4, T3, TSH and MAD-TSH), çıkış ise bir adettir (0=sağlıklı,1=hipertiroid, 2=hipotiroid).

Öznitelik Seçimi (Feature Selection), Sınıflandırma uygulamalarında ve veri madenciliğinde oldukça önemli bir yere sahiptir. Yüksek Boyutun Zararlılığı (Curse of Dimensionality) olarak adlandırılan, veri çok boyutlu olduğunda sınıflandırmanın çok zor, uzun zaman alan ve verimsiz olmasına sebebiyet veren problemin çözümünde oldukça etkilidir.

Bu çalışmada öznitelik seçimi yöntemi olarak Fisher Skoru algoritması kullanılmıştır. Fisher 'in Sınıflandırma için ortaya attığı ilk yöntemden türetilen bu algoritma öznitelik seçiminde oldukça başarılıdır.

Bu çalışmada tiroid hastalıklarının destek vektör makineleri aracılığı ile tanısı amaçlanmıştır. Bu tanıyı mümkün olan en yüksek doğrulukla yapabilmek için Genetik Algoritmalar ve öznitelik seçimi yöntemi olarak Fisher skoru" yöntemi kullanılmıştır. Bu özellikler hastaların kan testi sonuçları sonuçlarından elde edilir. Bunlar T3 resin- test (Serumda doymamış trioglobulin ölçümü),T4 (Thyroxine), T3 (Triiodothyronine), TSH (tiroid uyarıcı hormon), ve MAD-TSH(vucut normal haldeyken ölçülen TSH değeri ile kana 200 migrogram tirotropin enjekte edildikten sonra ölçülen TSH değerinin farkı) olmak üzere toplamda 5 adettir.

Bu çalışma üç aşamadan oluşan altı deney sonucu içerir. Altı adet deney olmasının sebebi Fisher skoru kullanarak elde öğrendiğimiz oransal olarak daha önemsiz verileri giriş verilerinden (Input Data) oluşturduğumuz yeni giriş verileri ile yeni deneyler yapıp Öğrenme Ağı'nın sonuçlarını nasıl etkilediğini bulmak.

İlk aşama Giriş ve çıkış verilerini (kan değerleri ve hastalık durumu) çapraz doğrulama (cross-validation) yöntemini kullanılarak rastgele bir şekilde öğrenme ve test verileri olarak ayırıp hastalık teşhisi yapabilen bir Destek Vektör Ağı oluşturmak (Öğrenme verileri ile ag'ı eğitme işlemi gerçekleştirilirken test verileri yardımı ile ağ'ın ne sınıflandırma becerisi ölçülür.) ve bu ağı bir fonksiyon içerisine yerleştirmek. Bu fonksiyonun giriş değeri yumuşak marjin katsayısı, çıkış değeri ise Öğrenme Ağı'nın yüzde cinsinden doğru çıkış verme oranıdır.

İlk aşamada oluşturulan fonksiyonların optimum değerini (en yüksek doğruluk oranını verdiği değer) hangi giriş değerinde (yumuşak marjin katsayısı) aldığını öğrenmek için ikinci aşamada Genetik Algoritmalar yardımı ile optimizasyon gerçekleştirildi.

İkinci aşamada bulunan optimum marjin katsayısı değeri üçüncü aşamada yeni öğrenme ağları oluşturmak için kullanıldı. Her bir dataset için ayrı ayrı oluşturulan altı farklı destek vektör sınıflandırıcısında en iyi sonucu verdiği parametre değeri (C katsayısı) kullanılanılmak üzere tasarlandı.

Üçüncü ve son aşamada her bir hastalığın hormon değerlerine bakılarak tahmin edilmesi işlemini gerçekleştirmek için tasarlanan altı farklı network 10'ar defa test edilip gerekli sonuçlar (doğruluk, hassasiyet ve özgüllük değerleri) elde edildi. Elde edilen sonuçlar kıyaslanarak hangi hastalığın teşhisi için hangi öznitelik değerlerini içeren veri seti kullanılması gerektiği belirlenmiştir. Ayrıca Genetik Algoritmalar yardımıyla optimizasyon işlemi yaparken dikkat edilmesi gereken bazı parametrelerin değişimini gösteren grafikler incelenerek yorumlanmıştır.

1. INTRODUCTION

1.1 Research's Motivation

Support Vector Machine (SVM) is one of the most fascinating learning methods and sometimes gives such remarkable results that encourage any researcher to study this learning procedure. In this thesis, a system based on classification via SVM for diagnosis of thyroid disease is presented.

The purpose of this study is to use SVMs as a classification tool and to search for methods to increase the efficiency of SVMs via feature selection and parameter optimization.

The objectives of this thesis are

- To diagnose thyroid disease via SVM with high accuracy.
- To search for the similar problems in literature and to have an idea about the solving techniques.
- To test the algorithms with different experiments and to analyze the results of the experiments.
- To see the effects of feature selection on SVM classifiers and how to improve the accuracy of SVM via feature selection.
- To analyze the effects of optimization via Genetic Algorithm on classification success of SVM classifiers.

In order to increase the efficiency of the corresponding classifier, the number of features have been decreased by feature selection. To perform this preprocessing procedure Fisher Score Algorithm has been used to obtain the most important features. Then the most important features are used to train the SVM. After training and testing the network a couple of times it has been noticed that the performance depends not only on the feature subsets of the corresponding pattern but also on parameters of SVM. To have a better performance (in terms of accuracy) it has been decided to find the optimal value of a parameter "C" (penalization coefficient of the

slack variable) by one of the most famous optimization technique called “Genetic Algorithm”.

1.2 Literature

SVM is considered as a reliable tool still but it has some unpredictability in data classification. Its performance seems to suffer when the amount of data increases and also with the addition of noise in the data. This problem can be solved by using robust SVM that has nonlinear separation and multiclass classification. SVM is being widely used and preferred in data mining [1-4].

(SVM) [5], although first proposed as a classification algorithm, can be used in many fields such as clustering [6], density estimation [7] or regression [8]. This method is based on predicting the best function that separates the data.

SVM is a learning machine that can be used to create regression rules (functions). First proposed in 1960s, it became very popular, and suggested in classification in 1990s. since then researchers have been working on this learning machine and trying to improve its learning ability. It has been protecting its importance in machine learning techniques since 1995 and it seems that, in the future, it is going to serve as one of the most important techniques. It is based on Vapnik’s “Statistical Learning Theory” .

1.2.1 Overview of SVM

The journey of SVM [9], although the term “support vector” was first mentioned in Vapnik’s work in 1965, dates back to 1936. Presentation of the first algorithm for pattern recognition is published by fisher in 1936[10]. In 1950, N. Aronszajn presented “The Theory of reproducing kernels “[11].

In 1957 Frank Rosenblatt first used the simplest kind of separating hyperplane (with the activation function it is generally a surface but without the activation function we can say that it is a hyperplane)-a linear classifier-and called it the perceptron (that is the most basic feed-forward neural network). In 1963 Vapnik introduced the first support vector algorithm as a pattern recognition method. Then in 1964, SVMs was introduced as a linear classifier by Vladimir Vapnik and Chervonenkis [12].

Now-a-day, data classification is considered as one of the most important steps in data mining processes because of the increasing amount of data. The main objective of data classification is to assign a group to unclassified data. SVM is considered as a classification tool in machine learning and pattern recognition methods and used in many areas including data mining.

The purpose of SVM is to find such an optimal hyperplane that separates the data as accurately as possible i.e. to find such a surface that is as distant as possible from the samples of each class. The crucial thing about this kind of classification is to select samples which are closest to the other class. Those samples are called "Support Vectors". The fact that the separating hyperplane is optimum makes the algorithm to have the highest generalization ability. In order for the separating hyperplane to be optimum it must be located in the middle of the two hyperplanes that are representing each class. By joining support vectors of each class we construct the two hyperplanes. And in the middle of these two hyperplanes we place the optimal separating hyperplane.

To find the optimal separating hyperplane, first we formulize the problem then we solve it by using quadratic programming [13]

1.2.2 Applications of SVM

SVMs are used in the text categorization [14], image classification [15], many kinds of cancer classification [16-19], bioinformatics [20], data mining [1-4]. Also in medical diagnosis SVM has a great importance and has many applications.

1.2.3 Applications of SVM in medical diagnosis

SVM has been studied many times in medical diagnosis for general purposes (e.g. [21-23]) it is also used in [24] for hepatitis disease diagnosis combined with simulated annealing, used in [25] for diagnosis of chronic kidney disease, diagnosis of heart disease [26] ,diagnosis of liver disease [27] etc.

SVM has been particularly used in diagnosis of thyroid diseases It is also used in combination with various feature selection methods and optimization techniques [10, 28-32].

In [28], SVM and probabilistic neural networks (PNN) are used in diagnosis of thyroid disease, and feature selection is performed by using genetic algorithms. In

[29], SVM is compared with several artificial neural networks (including ANFIS) when diagnosing thyroid disease. In [31] and [32], diagnosis of thyroid disease is performed by using SVM structure in a multiprocessor system on chip. Being successfully diagnosed by a field programmable gate array, diagnosis via SVM deserves to be studied more to create more reliable systems.

1.3 Outline of Research

After this brief introduction about SVM, in the second section thyroid disease and its symptoms have been explained.

In the third section an extensive explanation of the SVMs has been given.

In the fourth section Fisher score and genetic algorithms are introduced.

In the fifth section the experiments with their results are presented.

In the sixth section concluding remarks are given.

2. THYROID DISEASES

The thyroid gland is the organ that is located on the anterior side of our neck. The duty of thyroid gland is to produce and stock thyroid hormone and to regulate the metabolism by transferring thyroid hormone to the blood when it is needed.

The thyroid gland is the endocrine gland situating in the medial of the neck. The average weight of thyroid gland is 20 grams and it is as big as a walnut. As it is seen in figure 2.1 it situates on the anterior side of the neck and under the skin and looks like a butterfly. The wings of the thyroid are named as right and left. Each lobe is 4-5 cm long and 1-2 wide [33-35].

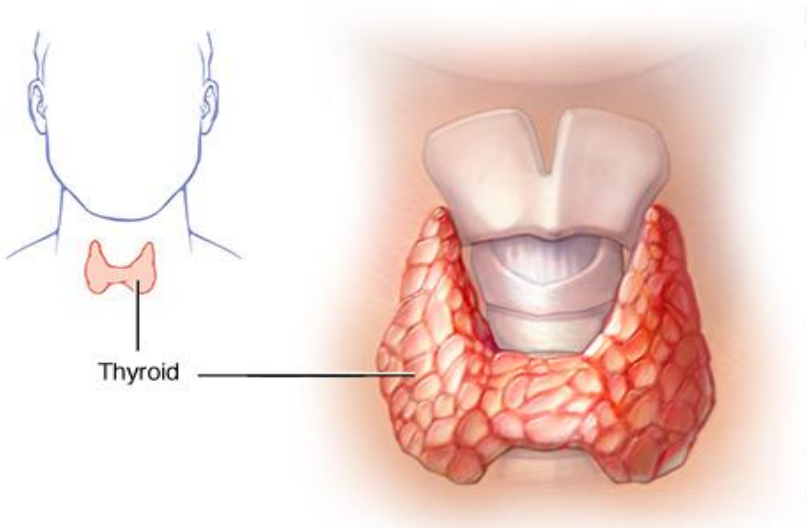


Figure 2. 1 : Thyroid gland [36].

The thyroid gland is on the right posterior of the trachea bulge and moves up and down during swallowing so the doctors ask their patients to swallow when examining the patients [37].

Thyroid gland is the organ which produces thyroid hormone from iodine minerals, taken by food and water. After passing from intestine to the blood, iodine taken from water and food comes to the thyroid gland, and it is used for producing thyroid hormone. The iodine, enters in thyroid gland, and produces the hormones named T3

and T4 by reacting with amino acid named tyrosine. There are 4 iodine molecules in T4 hormone structure, 3 iodine molecules in T3 hormone. Tyrosine amino acid is provided by the food with protein nutrient that we consume. As it is seen, it is needed to take enough protein and iodine into body by food and water in order to produce enough thyroid hormone. After being produced in thyroid gland, T3 and T4 hormones enters all of the body organs an cells then become effective[33].

All cells in our body are affected by thyroid hormones. The growth of human in mother's womb, after birth and all metabolism functions are controlled by thyroid hormones. There is almost no organ or cell that is not affected by thyroid hormones [37].

Two types of hormones are secreted from thyroid gland. While 80% of the thyroid hormones secreted are T4 , 20% secreted are T3, but the hormone entering the cells and the effective one is T3. T4 hormones do not enter the cells, so T4 hormones transform into T3 hormones in our body, especially in liver and in other organs. If this transform fails, T3 is not formed and the thyroid hormones do not show their effects. The T4 and T3 hormones circulate by engaging to some proteins. The thyroid hormones engaging to these proteins are named as Total T4 and Total T3. There are very few hormones in blood circulating without engaging to any proteins .These hormones are named as Free T3 and Free T4. Free T3 and Free T4 hormones are in balance with Total T3 and Total T4 hormones so the best experiment that reflects the working condition of thyroid gland is Free T3 and Free T4 hormones.

The functioning of thyroid gland is controlled by the pituitary gland in the bottom of our brain. Pituitary gland secretes hormone named TSH and this hormone reaches thyroid gland by blood and makes it produce thyroid hormones. TSH hormones not only help thyroid gland to keep iodine, but also they help to produce the thyroid hormones [33, 37, 38].

When thyroid gland secretes not enough amount of hormones, pituitary gland makes thyroid gland produce more hormone by increasing TSH secretion. Shortage of the hormones of thyroid gland is called hypothyroid. In this case, while TSH hormone rate is higher than normal, T3 and T4 hormones are lower than normal in our blood.

If thyroid gland secrete too much hormones, TSH hormone, secreted from pituitary gland, decreases. That is, the more T3 and T4 hormones are in our blood, the less is the TSH hormone. This condition is called hyperthyroid.

2.1 Hypothyroid

Hypothyroid disease generally occurs in conditions that are related to the autoimmune reasons (autoimmune disease is a disease that alienates the organism to one of its own tissues with an unknown reason and occurs as a result of attempts to remove this tissue from the body by its immune system), thyroid gland performs less because of being damaged. In case of being composed Goiter, causes Hashimoto's thyroiditis (It is the inflammation occurred in autoimmune process.)

Although the risk of hypothyroidism exists at all ages, it increases with age and after the age of 60, 2% - 4% hypothyroidism is observed.

Since it is seen frequently, diagnosed easily, checked up by inexpensive methods and cured easily, even if there is no complaint nowadays it is recommended to check TSH (ultrasensitive TSH) once in five years through blood measurement after the age of 35 and after the age of 60 every other year. Adding TSH measurement to the examinations done in almost all kinds of endocrinological disorders is important to diagnose this frequently seen disease[38, 39].

2.1.1 Symptoms of hypothyroidism

The most frequent symptoms in women at fertility age are menstrual irregularities. As well as being directly related to the hypothyroid, delayed menstruating and amenorrhea can also occur as a result of hyperprolactinemia stemming from TRH hormone's, which increases as a result of hypothyroidism, alerting prolactin hormone secretion.

Although no symptoms can be seen in hypothyroid disease, symptoms and signs are related to the decrease of body metabolism and the function of the organs slow down. Constipation (deceleration of intestine function), intolerance to the cold and deceleration of body temperature (slowing down of temperature with the deceleration of metabolism), deceleration of mental functions (forgetfulness, tendency to sleep, clumsiness, speaking slowly) getting tired easily, low pulse (related to the less

functioning of heart, anemia (related to the decreasing production of blood) increasing level of cholesterol (related to consuming little cholesterol) water retention, carpal tunnel syndrome related to the nerve entrapment are among the common ones[38].

In addition to the symptoms above, heart enlargement, slowing reflexes, muscle weakness and depression can be found among the examination findings [38, 39].

2.1.2 Diagnosis of hypothyroidism

After TSH hormone measurement is found high, the low measurement of sT4 measurement verifies the diagnose. Generally, in this process, along with starting the treatment without advanced examination, in some cases it is determined whether it is autoimmune or not, with measurement of anti-thyroid antibody [38].

2.2 Hyperthyroid

Hyperthyroid means the thyroid gland is producing more thyroid hormone than the body needs as a result of over-functioning. Increasing of thyroid hormone in blood because of different reasons is called thyrotoxicosis. The symptoms of these two thyroids are similar. The level of TSH in blood decreases. The most common reason of this is known as Graves' Disease. Hot nodule is one of the other reasons of hyperthyroid. Also uncontrolledly taken thyroid hormone, thyroid after birth, some inflammatory thyroid diseases and extreme iodine consumption are among the causes of Hyperthyroidism disease. Hyperthyroid is a damaging toxic case that causes disorder and complaint [39].

2.2.1 Symptoms of hyperthyroid

The most prominent symptoms of hyperthyroid are aggression, palpitation, rapid weight loss , reasonless trembling hands, weak muscle , increasing body temperature, sweating much, hair shedding, moisturized skin, fragile nails, fast function of intestine, menstrual irregularity, decreasing menstrual bleeding, swelling eyes, diplopia [39].

2.2.2. Causes of hyperthyroid

The clearest reason of hyperthyroid is Graves' disease. However, what causes Graves' disease is unknown. Any kind of distortion in the immune system causes thyroid gland to start over-production of hormones which leads the over-production of proteins. Another reason of hyperthyroid is the over-functioning nodules. Apart from thyroid gland, hot nodules regularly produce lots of thyroid hormone and transfer it to the blood. In case of inflammation of the thyroid gland, taking excessive iodine and the damaged thyroid gland can be accepted as other reasons of hyperthyroid rarely seen. Unrestrainedly and excessively taken thyroid hormone, given in hyperthyroid treatment, is also a reason of hyperthyroid [39].

2.2.3. Diagnosis of hyperthyroid

As hyperthyroid is a clear disease, it is easy to diagnose. The general examinations done considering the symptoms are enough to diagnose. The thyroid hormones and the level of TSH in blood are monitored for exact diagnosis. Generally both of the T3 and T4 or just one of them is high. In case of Graves' disease the thyroid antibody increases. Data about nodules are obtained from thyroid scintigraphy. However thyroid uptake test gives information about all the functions of thyroid gland [39].

3. SUPPORT VECTOR MACHINES

The purpose of statistical learning is to search for a way that we can estimate outputs from inputs. To be more precise, outputs are properties of objects and inputs are measurable features of the corresponding object. The outputs are all about decision while the inputs are about indication. In most of the cases the assumption is that the feature subsets (input variables, measurements) are all known while only few of outputs are known. Determination or estimation process is based on the fact that the outputs are dependent on inputs. In pattern recognition, this is an estimation of a function that is able to separate objects based on previous examples.

Support Vector Machines (SVM) [40] is one of the most important learning algorithms. It is famous for its capability of handling high dimensional data, powerful theoretical foundations and generalization performance. This chapter demonstrates a brief description of SVM.

3.1 Linear SVMs

Classification in a linearly separable data is the most fundamental task of SVMs. On the other hand it is not always the case. Due to noise or some other factors the data might be linearly non-separable or non-separable at all. We will now explain what to do in both of these situations. Of course the first case is linearly separable case.

3.1.1 Classifying linearly separable data

Let $((x_1,y_1),...,(x_n,y_n))$ be the training dataset where xi are the input vectors representing the observations (i.e. instances) and $y_i \in$ [-1,1] are labels. The purpose of support vector learning is to find a hyperplane (see figure 3.1) which is capable of separating positive samples from negative samples where the distance between the hyperplane and nearest positive and negative samples is maximum. To make it easy to understand we introduce a variable called “margin”. Margin is the summation of the shortest distances of positive and negative samples to the separating hyperplane.

From the definition of margin we can revise the purpose of SVM as "the search for the hyperplane that has the largest margin". The main aim in looking for a largest margin is based on the assumption that the hyperplane that has the largest margin has probably less sensitivity to the noise [41].

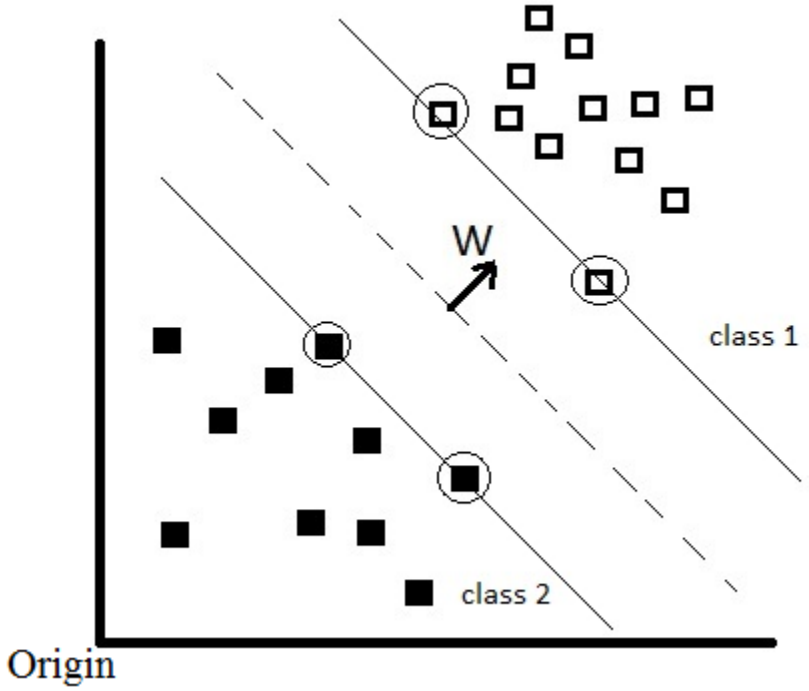


Figure 3. 1 : Construction of optimal separating hyperplane in linearly separable data. The circled samples located on the constraints are called support vectors.

The two constraints to create the largest margin separating hyperplane are

$$w \cdot x_i + b \geq +1 \quad , \; y_i = +1 \tag{3.1}$$

$$w \cdot x_i + b \leq +1 \quad , y_i = -1 \tag{3.2}$$

where w is the normal to the hyperplane, |b|/||w|| is the distance from the origin to the hyperplane, and ||w|| is the norm of w.

By using the sign of y_i (being positive in positive samples and being negative in negative samples), we can write these two constraints as one constraint [40, 41].

$$y_i (w \cdot x_i + b) \geq 1 \quad \forall i \tag{3.3}$$

The two constraints form two canonical hyperplanes. Circled examples are closest samples to the separating hyperplane. They are called support vectors and they all lie on the canonical hyperplanes H1 and H2 [41]. It can be easily seen that these hyperplanes are parallel to each other and the distance between them is calculated as follows

$$p = \frac{|1-b|}{\|w\|} - \frac{|-1-b|}{\|w\|} = \frac{2}{\|w\|} \tag{3.4}$$

Since margin is inversely proportional to "||w||" we can maximize margin by minimizing "||w||" in order to make it easy to compute, we will try to minimize $\frac{1}{2}\|w\|^2$ subject to the constraints.

Therefore, the corresponding hyperplane is built by solving the primal optimization problem written below [41].

$$\min_{w \in \kappa} \tau(w) = \frac{1}{2}\|w\|^2 \quad \text{Subject to} \quad y_i(w \cdot x_i + b) \geq 1 \quad \forall i \tag{3.5}$$

To solve the problem easily, we should express it in a Lagrangian form

$$\min_{w,b} L(w,b,a) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{l} a_i y_i (x_i w + b) + \sum_{i=1}^{l} a_i \tag{3.6}$$

where a_i are nonnegative Lagrangian multipliers for each constraint. Then the purpose is to minimize (3.6) with respect to w [41].

3.1.1.1 The Karush-Kuhn-Thucker conditions

The Karush-Kuhn-Thucker (KKT) [42, 43] conditions create the conditions that are required to be fulfilled by an optimum solution to a general optimization problem. According to the KKT conditions, primal problem presented in equation (3.6), the conditions mentioned below should be satisfied by the solutions w, b and a.

$$\frac{\partial L(w^*, b^*, a^*)}{\partial w} = w_v - \sum_i a_i y_i x_{iv} \quad v = 1, \cdots, d \tag{3.7}$$

$$\frac{\partial L(w^*, b^*, a^*)}{\partial b} = -\sum_i a_i y_i = 0 \quad \textbf{(3.8)}$$

$$y_i(x_i \cdot w + b) - 1 \geq 0 \quad \forall i \quad \textbf{(3.9)}$$

$$a_i \geq 0 \quad \forall i \quad \textbf{(3.10)}$$

$$a_i(y_i(w \cdot x_i + b) - 1) = 0 \quad \forall i \quad \textbf{(3.11)}$$

Since the SVM problem is convex, KKT conditions above are necessary and sufficient [44]. Due to that reason we can state that the SVM problem can be solved easily once the solution to the KKT conditions are met. The first KKT condition describes the optimal hyperplane with a linear combination of vectors in training set as:

$$w^* = \sum_i a_i^* y_i x_i \quad \textbf{(3.12)}$$

According to the second condition of KKT, the aicoefficient of should satisfy the following equation:

$$\sum_{i=1}^{n} a_i^* y_i = 0, \quad a_i^* \geq 0 \quad \textbf{(3.13)}$$

3.1.1.2 The dual problem

SVM problem is the problem of maximizing "w" and "b" and minimizing a in equation (3.6). Therefore it is a saddle point that we are looking for. Following the KKT conditions at this point, the derivatives of L with respect to "w" and "b" must disappear[44, 45]. By putting equation 3.7 and 3.8 into equation 3.6 the following equation of dual is obtained:

$$\max_a L_D = \sum_I a_i - \frac{1}{2}\sum_{i,j} a_i a_j y_i y_j x_i \cdot x_j \quad \text{Subject to} \quad \forall i \begin{cases} \sum_i a_i y_i \\ a_i \geq 0 \end{cases} \quad \textbf{(3.14)}$$

Therefore, the coefficients α_i are computed by solving the dual optimization problem. The samples with $\alpha_i > 0$ $\alpha_i$are named "support vectors" and they rest on one of the canonical hyperplanes (H1 or H2). One should note that we are not going to use the instances with $\alpha_i = 0$. Those instances which are not support vectors are not needed to construct the hyperplane[41]. Then decision function is

$$f(x) = w^T x_i + b = \sum_{i=1}^{M} y_i a_i (x_i^T x) \qquad \textbf{(3.15)}$$

Determination of the estimated classification of χ is being determined by looking at the sign of the decision function.

3.1.2 Classifying linearly nonseparable data and noise

Till now the cases in which the data is linearly separable have been explained. But in the real life situations, most of the times the data cannot be separated by even an optimal hyperplane. Even if the data can be linearly separated in normal circumstances, because of a mislabeled pattern, SVM could fail in the search for total separation of both classes. Outliers are crucial problems that optimal hyperplanes cannot separate. These problems lead us to find a micro solution called "soft margin". By soft margin SVMs one can easily manage creating an optimal hyperplane. We now add another variable called Penalization constant (ξ_i) to the constraints (see figure 3.2) so that the constraints can be more flexible. Then the constraints with new penalization constant (positive slack variable) could be written as follows [46].

$$\forall i = \begin{cases} w \cdot x_i + b \geq +1 - \xi_i \Rightarrow y_i = +1 \\ w \cdot x_i + b \leq -1 - \xi_i \Rightarrow y_i = -1 \\ \xi_i \geq 0 \end{cases} \qquad \textbf{(3.16)}$$

Now we can estimate that there is toleration in the constraints.

By this toleration we manage to decrease the harmful effects of outliers in the dataset. However one should be careful when adjusting the value of slack variable because big values could end up creating useless solutions for classification. Therefore, this slack variable should have some limits adjusted by a coefficient. The equation can be written as follows:

$$\min_{w\in K,\xi\in R^m} T(w,\xi)=\frac{1}{2}\|w\|^2+C\sum_{I=1}^{m}\xi_i \qquad \textbf{(3.17)}$$

Parameter (C>0) is adjusted by the user based on basic information about the dataset and classification problem.

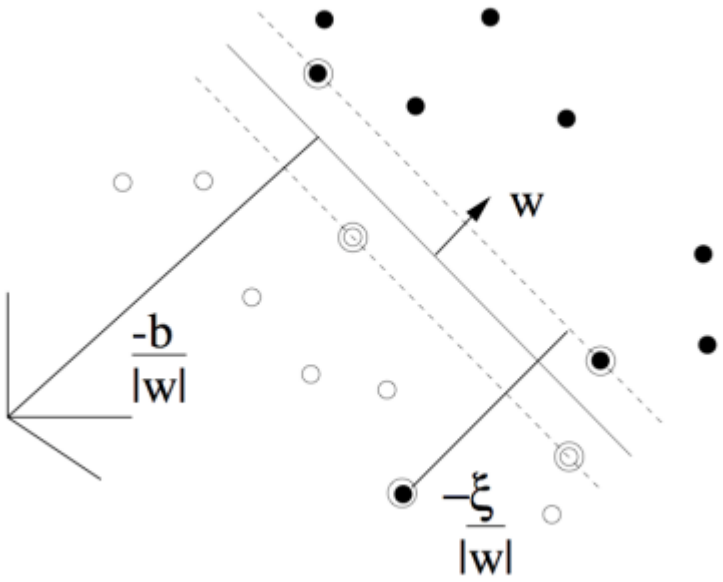


Figure 3. 2 : Soft margin Hyperplane[45]

The solution is the same as it is in separable data.

$$\frac{\partial L_P}{\partial w}=0\Leftrightarrow w=\sum_i a_i y_i x_i \qquad \textbf{(3.18)}$$

where α_i are support vectors. The penalization function which eq. 3.17 includes is linear and it vanishes when converted to dual form.

$$\max_a L_D=\sum_i a_i-\frac{1}{2}\sum_{i,j} a_i a_j y_i y_j x_i\cdot x_j \text{ Subject to } \forall i=\begin{cases} C\le a_i\le 0 \\ \sum_i a_i y_i=0 \end{cases} \qquad \textbf{(3.19)}$$

3.1.2.1 Details of linearly nonseparable data

When the linear classification of the substitutes assigned to dissimilar classes isn`t possible, then the misclassifications not only need to be tolerated but also penalized as well. In real life it is not always possible to find a perfectly separable case which

can cover all the cases. Therefore in order to increase the application area of the model the misclassifications need to be covered [47].

To add misclassifications into the model we need to introduce slack variables into equations 3.20 and 3.21 [48].

$$Min = \frac{1}{2} w^T w + C \sum_{I=1}^{N} \xi_I \tag{3.20}$$

Subject to

$$y_i(w^T x + b) \geq +1 - \xi_i, i = 1, \cdots, N \tag{3.21}$$

$$\xi_i \geq 0, i = 1, \cdots, N \tag{3.22}$$

where

C = positive constant representing settle between margin maximization and error minimization

ξ_i = slack variables to calculate misclassification error of alternative.

3.2 Quadratic Programming

Quadratic programming is a mathematical optimization problem which is used for the minimizing or maximizing an optimization problem which is quadratic and muti-variabled, subject to linear constraints (equality or inequality) [46, 49].

By applying Lagrange multipliers for nonlinear constrained optimizations, the Lagrangian A as the objective function plus a linear combination of the constraints can be expressed as[49]:

$$\Lambda(\vec{w}, b, \vec{\xi} \,||\, \vec{a}) = \frac{1}{2}\|w\| + C\left\|\vec{\xi}\right\| - \sum_{i=1}^{n} a_i (y_i (\overrightarrow{x_i} \cdot \vec{w}) + y_i b - 1 + \xi_i \tag{3.23}$$

Where a_i >0 are Lagrange multiplier and must not be negative as denoted by the non-negativity of their corresponding constraints:

$$y_i(\vec{x_i} \cdot \vec{w}) + y_i b - 1 + \xi_i \geq 0 \Rightarrow a_i \geq 0 \tag{3.24}$$

Writing dual form:

$$\vec{a} = \arg\max_{\vec{a}} \left[\arg\min_{\vec{w},b,\vec{\xi}} \Lambda(\vec{w}, b, \vec{\xi} \,||\, \vec{a} \right] \tag{3.25}$$

Differentiating Lagrangian with respect to w, b and $\vec{\xi}$ to extract the minimum and equating it to zero:

$$\begin{aligned} \frac{\partial \Lambda}{\partial \vec{w}} = 0 &\Rightarrow \sum_{i=1}^{n} a_i y_i \vec{x_i} = \vec{w} \\ \frac{\partial \Lambda}{\partial b} = 0 &\Rightarrow \sum_{i=1}^{n} a_i y_i = 0 \\ \frac{\partial \Lambda}{\partial \vec{\xi}} = 0 &\Rightarrow \vec{a} = 2C\vec{\xi} \end{aligned} \tag{3.26}$$

Add the first and the third part of the above extracted Lagrangian into dual form:

$$\begin{aligned} \Lambda(\vec{w}, b, \vec{\xi} \,||\, \vec{a}) &= \frac{1}{2}\left\| \sum_{i=1}^{n} a_i y_i \vec{x_i} \right\| + C\left\| \frac{\vec{a}}{2C} \right\| - \sum_{i=1}^{n} (a_i y_i (\vec{x_i} \cdot \vec{w}) + a_i y_i b - a_i + a_i \xi_i \\ &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (y_i \cdot y_j)(\vec{x_i} \cdot \vec{x_j}) + \frac{(\vec{a} \cdot \vec{a})}{4C} - \sum_{i=1}^{n} a_i y_i (\vec{x_i} \cdot \sum_{i=1}^{n} a_i y_i \vec{x_i}) + \sum_{i=1}^{n} a_i - (\vec{a} \cdot \vec{\xi}) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (y_i \cdot y_j)(\vec{x_i} \cdot \vec{x_j}) + \frac{(\vec{a} \cdot \vec{a})}{4C} + \sum_{i=1}^{n} a_i - (\vec{a} \cdot \frac{\vec{a}}{2C}) \\ &= -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (y_i \cdot y_j)(\vec{x_i} \cdot \vec{x_j}) - \frac{(\vec{a} \cdot \vec{a})}{4C} + \sum_{i=1}^{n} a_i \end{aligned} \tag{3.27}$$

The final dual quadratic optimization, subject to a_i >0 and $\sum_{i=1}^{n} a_i y_i = 0$, can be written as:

$$\vec{a}^{*} = \arg\max_{\vec{a}} \left[-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (y_i \cdot y_j)(\vec{x_i} \cdot \vec{x_j}) - \frac{(\vec{a} \cdot \vec{a})}{4C} + \sum_{i=1}^{n} a_i \right] \tag{3.28}$$

The dual forms of separable and not separable case are almost the same except there is an additional upper bound of C in Lagrangian coefficient a_i. On the other hand it is clear that as penalization constant (C) goes to infinity, $C \to \infty$ in (eq. 3.26) it becomes the same as it is in separable case.

3.3 Nonlinear SVMs and kernel methods

In order to classify patterns with artificial intelligence techniques, machine learning uses a class of algorithms which are known as Kernel Methods. One of the most famous algorithm of that class is SVM. Pattern analysis is used generally to analyze relations in datasheets. Most of the algorithms used for that purpose need the raw data to be completely changed into feature vector by using feature map. Meanwhile to solve the same problem using kernel methods a similarity function of the raw data is needed. Due to their use of kernel functions they are called Kernel Methods. Kernel Functions can be operated in high-dimensional feature space and they remove the additional work of computing the coordinates of data present in that space which is an advantage. The operation of calculating the inner products in feature space is computationally easier than the aforementioned one and is known is Kernel Trick. [50]. Figure 3.3 gives an idea about how kernel transformation works. Figure 3.4 is a representation of classification via support vectors in a transformed form of the feature space.

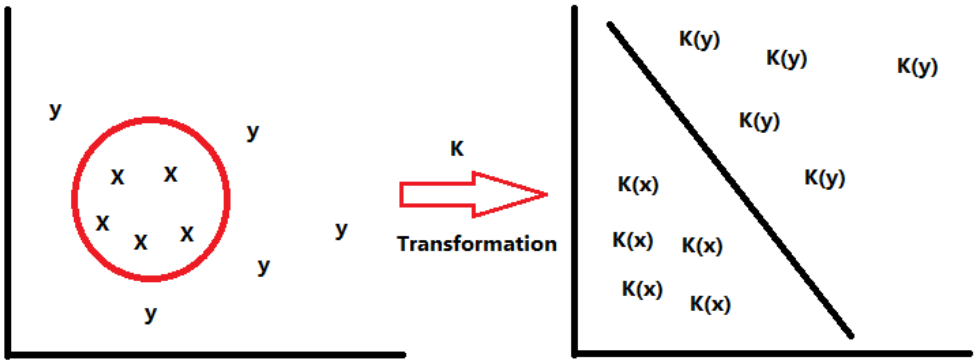


Figure 3. 3: Kernel Transformation

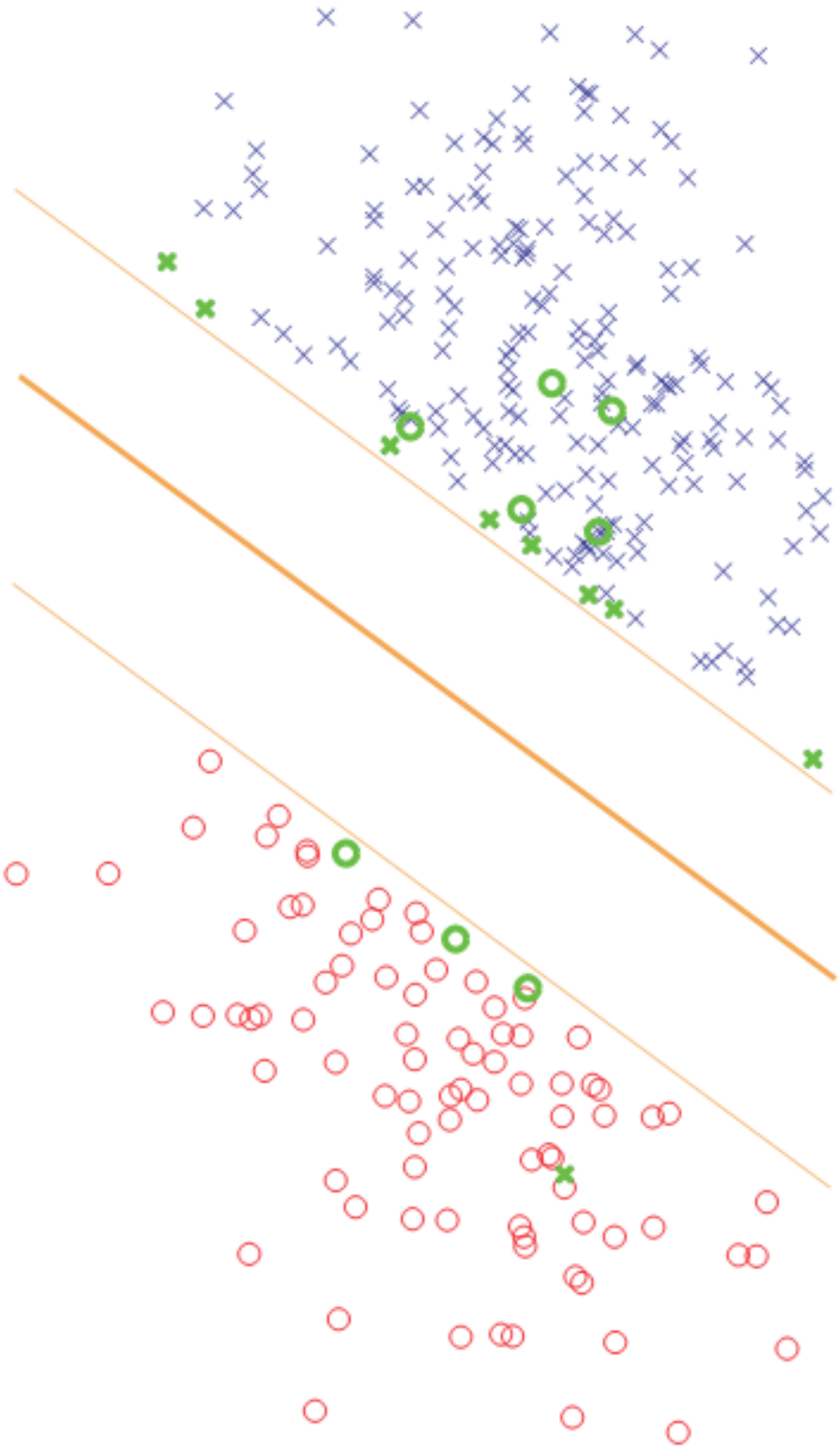


Figure 3. 4 : Conclusion of separation task; 17 green samples are support vectors which are used to create decision boundary. Blue samples are indicating the positive while red indicates negative[49].

Some of the popular kernels used are mentioned as follows:

Fisher Kernel

Graph Kernel

Polynomial Kernel

RBF Kernel

String Kernel

3.3.1 Polynomial kernel

It is a type of kernel that is used mostly with SVM and other similar models that allows the learning of non-linear models. It is one of the most frequently used types of kernel. It uses the combinations of features of the input samples in order to study the similarity between them[50].

Given x and y which are the vectors of features calculated from training or testing, the polynomial kernel for degree polynomials can be mathematically written as follows:

$$K(x, y) = (x^T y + c)^d \tag{3.29}$$

4. FEATURE SELECTION AND OPTIMIZATION

Two basic methods that can be used in a classification problem are feature selection and Optimization.

4.1 Feature Selection Via Fisher Score

Fisher Score[51-53] is one of the most popular feature selection methods in machine learning applications. Fisher (1936) first proposed it to classify objects of two different classes and it has been quite popular since then. Numerous strong feature selection algorithms have been developed after Fisher's first algorithm. However, being outperformed by other classifiers did not make Fisher's algorithm less important since one of the best algorithms (SVM) is descended from Fisher's algorithm. A short summary of Fisher score, which is used for feature selection in this research, is presented in the following paragraphs:

The main concept of Fisher score is to find the subset of selected features in a way that the data points which belong to different groups are distant to each other while the data points which belong to the same group are closer to each other. The Fisher score can be calculated as follows [53].

$$F(Z) = tr\left\{(S_b)(S_t + \Upsilon I)^{-1}\right\}, \tag{4.1}$$

Where γ is a positive regularization parameter, S_b is between-class scatter matrix, and S_t is total scatter matrix, which are defined as

$$\begin{aligned} S_b &= \sum_{k=1}^{c} n_k (\mu_k - \mu)(\mu_k - \mu)^T \\ S_t &= \sum_{i=1}^{n} n_k (z_i - \mu)(z_i - \mu)^T, \end{aligned} \tag{4.2}$$

where μ_k and n_k are the mean vector and size of the k[th] class respectively in the reduced data space. S_t is mostly singular, so in order to make S_t a positive we add perturbation term ΥI . The feature selection is a combinatorial optimization problem and therefore is difficult as there are $\binom{d}{m}$ candidate Z`s out of X. To make it a bit simple, heuristic strategy is used to calculate independent score for each feature according to criterion F. Therefore, there are only $\binom{d}{1}$ = d nominees. Let μ_k^j and σ_k^j be the mean and standard deviation of k[th] class corresponding to j[th] feature. Let μ^j as the mean of the whole data set corresponding to the j[th] feature. Then the Fisher score can be calculated as follows [53]:

$$F(x^j) = \frac{\sum_{k=1}^{c} n_k (\mu_k^j - \mu^j)^2}{\sum_{k=1}^{c} n_k (\sigma_k^j)^2}, \qquad \textbf{(4.3)}$$

The feature with the highest scores is selected after the completion of computation of the Fisher score for every feature. The feature which is selected by the heuristic algorithm is not completely optimal due to the fact that scores for every feature is calculated independently. The features which have low individual scores, but high combined scores, are neglected by the heuristic algorithm which is also a major disadvantage. Besides that, it doesn`t have the capability to process redundant features. Due to those aforementioned reasons Fisher score is proposed to find a solution of such problems [53].

4.2 Optimization Via Genetic Algorithm

Genetic Algorithm (GA) is an optimization method which is mainly based on the theories of natural selection and evolutionary processes. It is first put forward in 1975 by John Holland at the University of Michigan [54].

GA is a search optimizer which can be used to increase the operating response of parallel-processing machines. It can be implemented on continuous as well as discrete optimization problems. GA is stochastic therefore the likelihood of it getting trapped in local minima is relatively less.

In GA [55], every value in the parameter space is a binary string, known as chromosome, and has a fitness value of its own. Optimal fitness value is normally the value of the objective function evaluated for that particular point.

In GA, a population of points is generated and evolved in order to improve the overall fitness value. For every generation a new population is created, by using crossover and mutation genetic operators, by GA.

The members which have higher fitness values have the higher chances of survival and participation in the mating operations (giving birth to new generations). As the generation is leveled up, a population having members with better fitness values is obtained. Matlab Optimization Toolbox Interface is seen in the figure 4.1

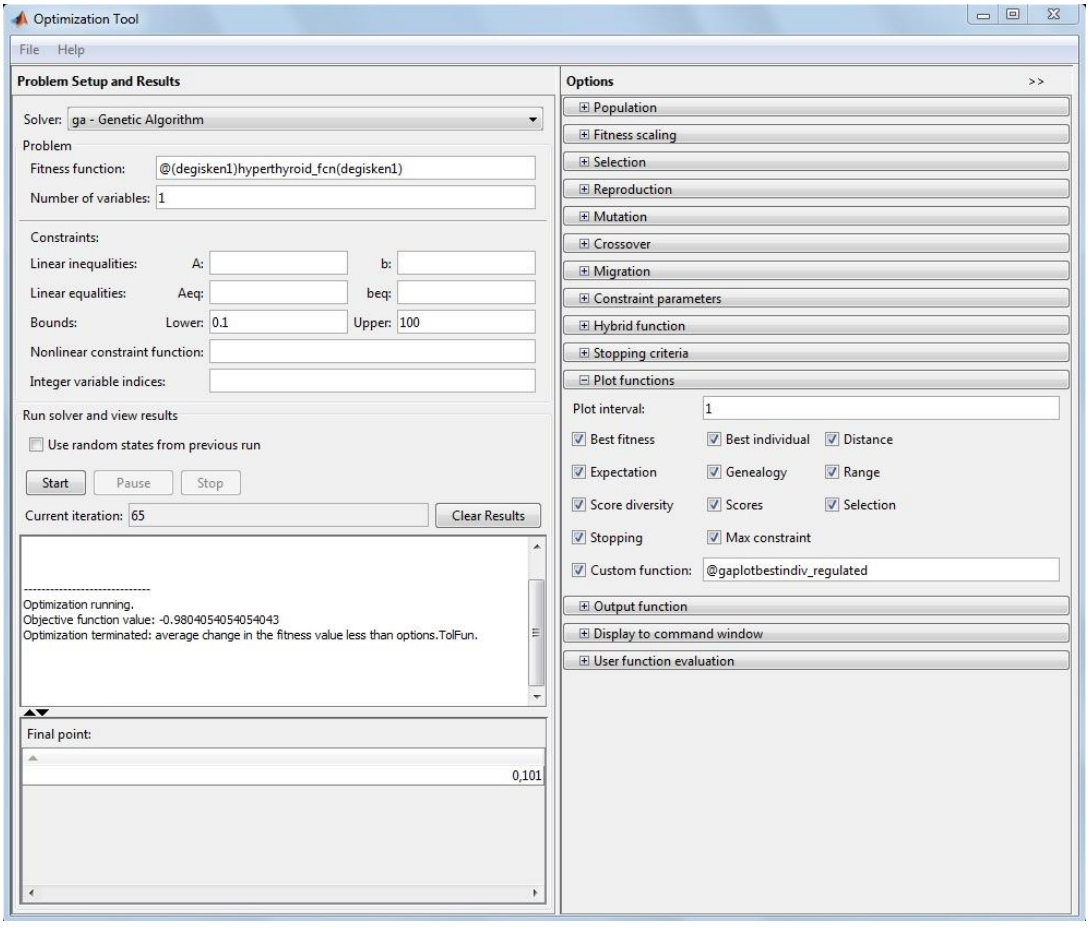


Figure 4.1: Matlab Optimization Toolbox-Interface

4.2.1 GA`s components

The first step in GA is to convert points present in parameter space into binary bits.

Example: (10, 1, 7) can be represented as 1010 0001 0111

Using binary coding every coordinate value is coded in a gene which is composed of four binary bits. Any other coding technique i.e. grey coding could have been used too. In this way, a knowledge which is related to a specific issue can be directly integrated into GA [55, 56].

4.2.2 Fitness evaluation

After the creation of a generation every member of that population is assigned with a calculated fitness value. The fitness value of the i^{th} member is normally the objective function in maximization problems. It is also possible to rank the member in a population according to its fitness value which also reduces the need of an objective function. The fitness values need to be positive in general.

4.2.3 Selection:

A new population is created by using the current generation after the evaluation.

Just like the survival of the fittest in natural selection, the participating parents for the production of off-spring is selection based on their fitness values[55].

i.e. Selection probability can be equated to $\frac{fi}{\sum_{k=1}^{n} f_k}$ where n is the population size.

4.2.4 Crossover

Crossover is similar to mating in natural evolutionary process. It is normally applied to the selected pairs of parents with a probability equal to a given cross-over rate.

4.2.4.1 One-point crossover

Crossover point is randomly selected as shown in below

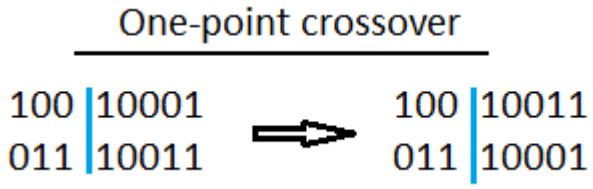


Figure 4.2 : One Point Crossover

4.2.4.2 Two-point crossover

n-point crossover can also be defined therefore it is possible that some children may outperform their parents in their generation [52].

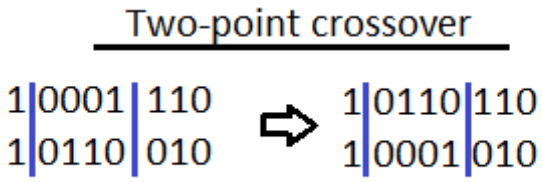


Figure 4.3 : Two Point Crossover

4.2.5 Mutation

In order to have an adequate solution, the generated population must contain all the encoded information needed to solve a specific problem. Mutation helps with the increase in the efficiency of the pool of phenotypes in the population and resists local minimum and maximum regions. Besides, it also provides the guarantee of new population with the new solutions. [56].

New chromosomes are produced by using the mutation operator.

The main advantage of mutation is that it prevents the population to converge to a local optimum point.

Good chromosomes are generally not lost as the mutation rate is normally very low i.e. below 0.1

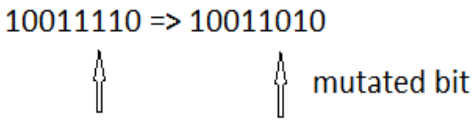


Figure 4.4 : Mutation

Elitism is a way to keep a particular number of best members when a population is generated.

4.2.6 A simple GA example (maximization)

Step 1

A population is created by randomly generated individuals

Step2

Each member`s fitness value is calculated and assigned

Step 3

A) Two members having proportional fitness values are selected.
B) Crossover is applied according to the crossover rate.
C) Mutation is applied according to the mutation rate.
D) A to C are repeated until enough members are there to form the next generation

Step 4

Steps 2 to 4 are repeated until the stopping criterions are met. Stopping Criterions are computation time, optimization goal, minimal improvement, and minimal relative improvement

Computation time is stopping criteria that is based on the computation time or the number of iteration counts.

Optimization goal is the evaluated objective function exceeds the certain preset goal value.

Minimal improvement is the stopping criterion that is met when difference in fitness values f_k-(f_{k-1}) is less than a preset value.

The criterion minimal relative improvement is met when (f_k-f_{k-1})/f_{k-1} is less than a present value.

5. EXPERIMENTS AND RESULTS

In this study, diagnosis of Hyperthyroid and Hypothyroid diseases via SVM is performed. A hybrid method that includes three different machine learning techniques (SVM, Fisher's Linear Discriminant and Genetic Algorithm) is created to perform this study.

The database used in this experiment is taken from machine learning repository [57]. It consists of 215 instances of three classes. The distribution instances to the classes is

- 150 healthy (class1)
- 35 hypothyroid (class2)
- 30 hyperthyroid (class3)

In order to perform these experiments two datasets are constructed using this dataset. First dataset consists of 185 instances distributed as

- 150 healthy (class1)
- 35 hypothyroid (class2)

and it is used in diagnosis of Hyperthyroid disease (in section 5.1). The second dataset consists of 180 instances distributed as

- 150 healthy (class1)
- 30 hyperthyroid (class3)

and it is used in diagnosis of Hypothyroid disease (in section 5.2).

The experiments performed consist of four steps. At first, Fisher Scores for each dataset are calculated using the two datasets (see table 5.1 and 5.2).

The second step is the construction of six different functions (three for hyperthyroid, three for hypothyroid) that can be optimized by Genetic Algorithms. In these functions, at first the data are separated as training and test sets by performing Cross-Validation. Training set is 40% of the whole data. The remaining 60% is reserved for

testing purposes. Secondly the SVM is constructed using Matlab Statistics Toolbox and trained by using training set. Then testing data is placed in the corresponding SVM. Afterwards the output of the SVM and the real outputs are compared to see how much they match. Within the optimization function this process is repeated 100 times. Because of randomness of Cross-validation, at each time, the new training and test data are different. The accuracy of the network is assigned to a 1x100 vector. At the end, the mean value of the vector is assigned as an average accuracy of the network. This average accuracy is used as a Fitness Value for genetic algorithm and the variable, which was optimized, was the soft margin parameter (C).

Table 5. 1 : Fisher scores of each feature for Hyperthyroid.

Classification	Fisher Score
F1 (T3RU)	0.4458
F2 (T4)	2.85952
F3 (T3)	0.8535
F4 (TSH)	0.4952
F5 (MAD-TSH)	59.7975

Table 5. 2 : Fisher scores of each feature for Hypothyroid.

Classification	Fisher Score
F1(T3RU)	0.7352
F2(T4)	7.2840
F3(T3)	1.0387
F4(TSH)	0.6303
F5(MAD-TSH)	0.6737

In the third step, Using Matlab - Optimization-Toolbox, the optimal values of Soft Margin Parameters are obtained with respect to the corresponding fitness value. The resulting graphics of optimization process are shown in figures 5.1-5.18

Then in the fourth step, using the optimal value of the corresponding soft margin parameter (C), three different experiments have been performed for each disease by utilizing three different datasets for 10 times. At each time the accuracy of the network is assigned to 1x10 vector (it is called the Accuracy Vector) respectively. After 10 iterations the average accuracy of the network is determined by taking the mean of the accuracy vector. This mean value is the main criteria of the performance of the corresponding classifier.

5.1 Hyperthyroid Test Results

The first experiment on diagnosis of hyperthyroid is performed with all the feature subsets of the original dataset. Using MATLAB Optimization Toolbox the optimal soft margin parameter is calculated as C=21,9450. This value is used as the value of the soft margin parameter in SVM that is created for the classification task.

The changes in fitness value of hyperthyroid test that all features in dataset are used in are seen in figure 5.1. In this experiment, the objective was to increase the efficiency of the network in terms of accuracy. Therefore the fitness value is the accuracy of the SVM. However, genetic algorithm always struggles to find the minimums. Therefore, the accuracy is multiplied with -1 and scaled in between 0 and -1. One can observe from the fig 5.1 that the mean fitness is never less than 97.1% in this very first experiment.

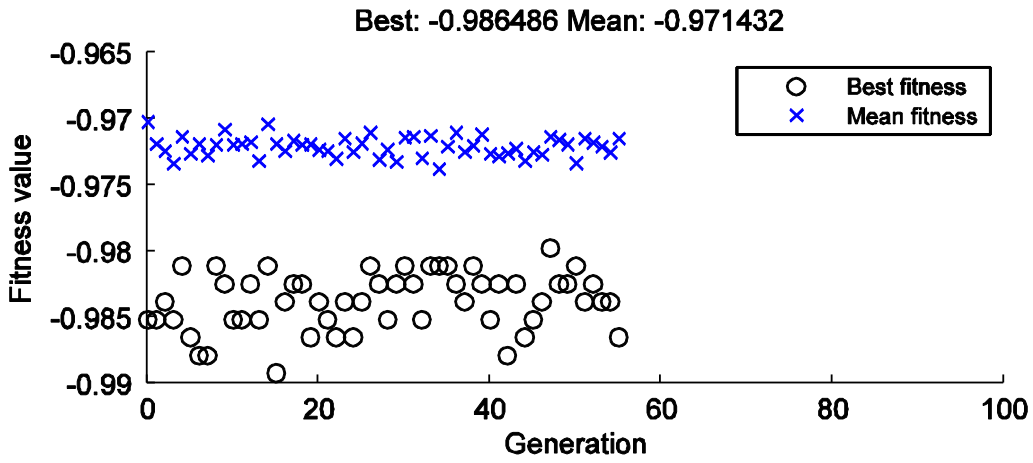


Figure 5.1 : The changes in fitness value of hyperthyroid test that all features in dataset are used in (F1,F2,F3,F4,F5)

The graph of average distance between indivuduals vs. number of generation for first hyperthyroid test is shown in figure 5.2. one can easily observe from the graph that average distance between indivuduals does not change too much after 16^{th} generation. This information gives an idea about how close the population is to the extremum. Therefore, it can be assumed that the maximum number of generations could be picked as 16 in order to have the least computational cost. Of course, this criteria, by itself, is not enough to decide on the number of generations.

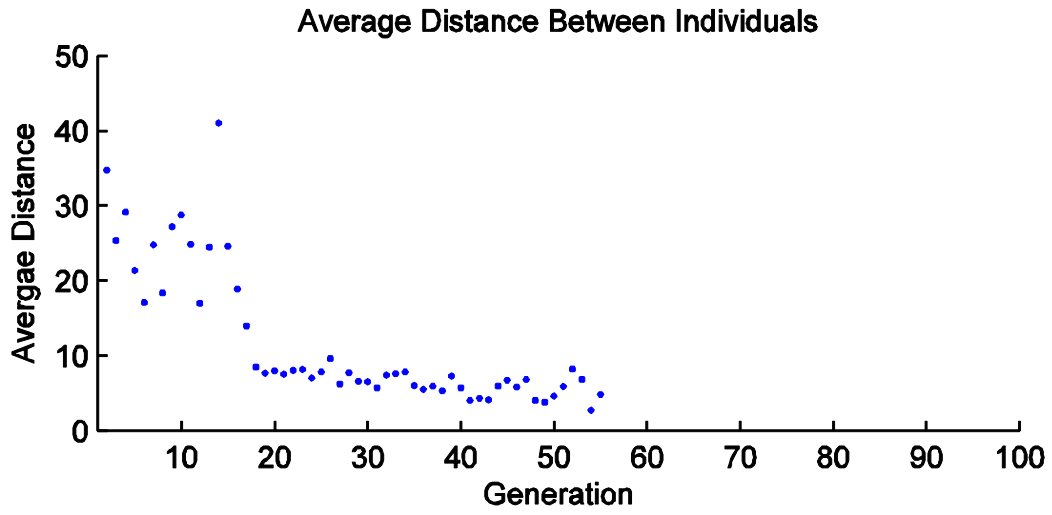


Figure 5.2 : The Average distance between individuals for hyperthyroid test that all features in dataset are used in (F1,F2,F3,F4,F5)

The second experiment is carried out using features F2,F3,F4,F5 (namely: T4, T3, TSH,MAD-TSH) and the calculated optimal value of C is 0,1. This is used as the value of the parameter C in SVM that is created for the classification task that has 4 inputs .

The changes in fitness value of hyperthyroid test 2 are seen in figure 5.3. The accuracy is multiplied with -1 and scaled in between 0 and -1. It is observed from the fig 5.3 that both the mean fitness and best fitness of the second experiment is lower than the first experiment's values.

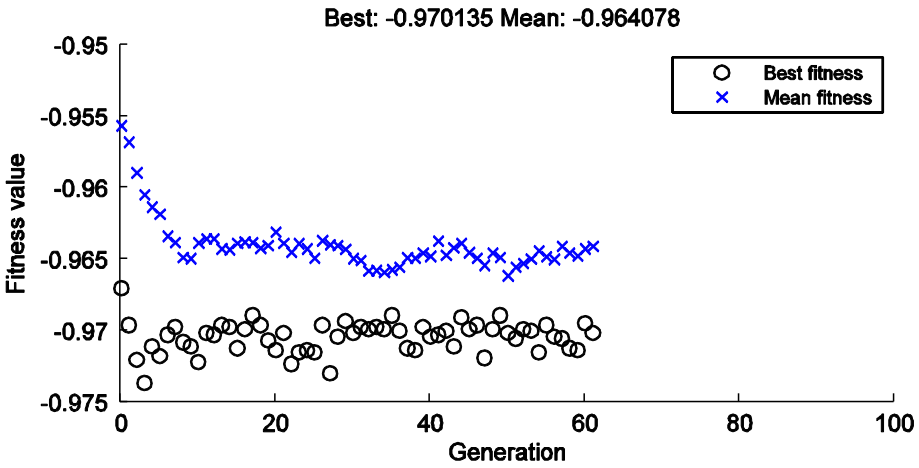


Figure 5.3 : The changes in fitness value of hyperthyroid test 2 (F2,F3,F4,F5)

The graph depicting the average distance between indivuduals vs. number of generation for first hyperthyroid test is shown in figure 5.4.

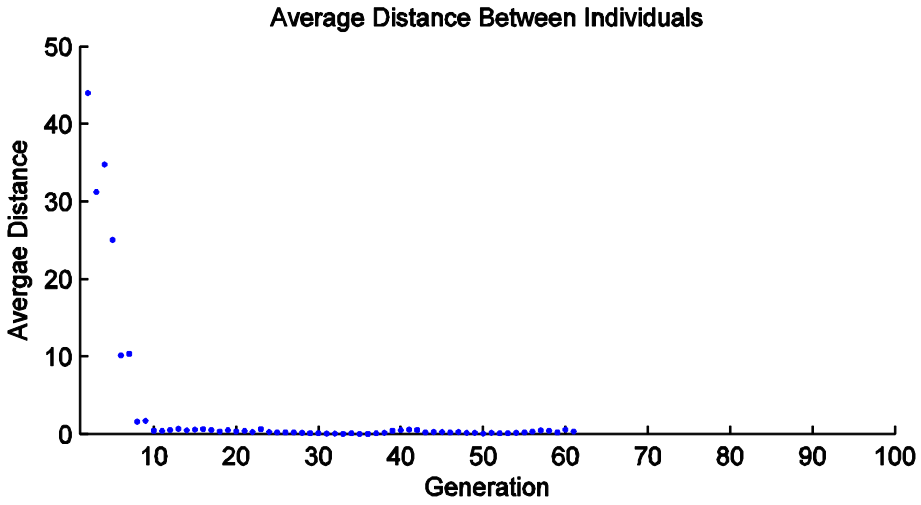


Figure 5.4 : The Average distance between individuals for hyperthyroid test 2 (F2,F3,F4,F5)

It can be observed from the graph that average distance between indivuduals does not change too much after 7th generation. This information could either mean that , the optimization is stuck in a local minimum instead of finding the global minimum or decreasing the number of features is not a good idea in this particular case. The reason can only be found by performing all experiments and evaluating them all together.

The third experiment was performed using the features F2,F3,F5 (namely: T4, T3, MAD-TSH) and calculated optimal value for C was 0.1149. This is used as the value of the parameter C in SVM that is created for the classification task that has 3 inputs.

The changes in fitness value of hyperthyroid test 3 are seen in figure 5.5. The accuracy is multiplied with -1 and scaled in between 0 and -1. It is observed from the fig 5.5 that the mean fitness is the highest in the third experiment of the hyperthyroid test.

The graphc of average distance between indivuduals vs. the number of generation for first hyperthyroid test is shown in figure 5.6.

It can be observed from the graph that average distance between indivuduals does not change too much after 5th generation.

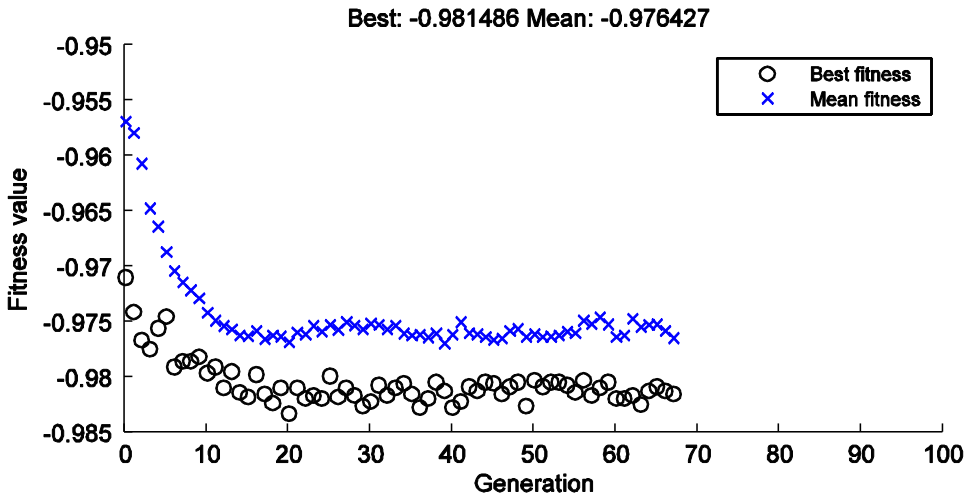


Figure 5.5 : The changes in fitness value of hyperthyroid test 3 (F2,F3,F5)

By using all feature subsets (the first experiment) the optimal value is (Coptimal=21,9450) and 96,22% average accuracy (see Table 5.2) was obtained at that optimal value of C. When the feature subsets (F2,F3,F4,F5) are used the result became (Coptimal=0.1) and the average accuracy was 96,89% at that optimal value of C. The details of ROC parameters for hyperthyroid test are in the tables 5.3, 5.4, and 5.5

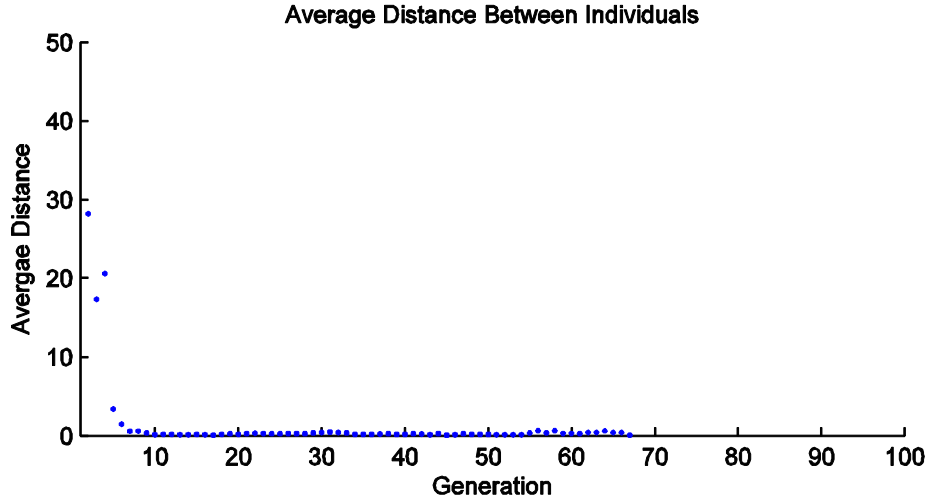


Figure 5.6 : The Average distance between individuals for hyperthyroid test 3 (F2,F3,F5)

Table 5. 3 : Result of the experiment 1-Hyperthyroid (F1,F2,F3,F4,F5).

Number of Test	Accuracy (%)	Sensitivity (%)	Specifity (%)
1	97.30	92.86	98.33
2	95.95	78.57	100
3	97.30	92.86	98.33
4	94.59	85.71	96.67
5	97.30	85.71	100
6	97.30	92.86	98.33
7	91.89	64.29	98.33
8	97.30	92.86	98.33
9	97.30	92.86	98.33
10	95.95	85.71	98.33

Table 5. 4 : Result of the experiment 2- Hyperthyroid (F2,F3,F4,F5).

Number of Test	Accuracy (%)	Sensitivity (%)	Specifity (%)
1	95.95	85.71	98.33
2	94.59	92.86	95.00
3	97.30	92.86	98.33
4	97.30	100	96.67
5	97.30	92.86	98.33
6	100	100	100
7	93.24	92.86	93.33
8	100	100	100
9	95.95	92.86	96.67
10	97.30	85.71	100

Using the feature subsets (F2,F3,F5) obtained value of C was 0.1149 and 98,11% average accuracy (see Table 5.5) was achieved at that optimal value of C).

In terms of average sensitivity, the best results (97,14%) are calculated by using (F2, F3, F5 feature subsets). When the feature subsets are (F2, F3, F4, F5) the average sensitivity is calculated as 93,57%. Using all feature subsets average sensitivity gets 86,43%.

Table 5. 5: Result of the experiment 3 -Hyperthyroid (F2,F3,F5).

Number of Test	Accuracy (%)	Sensitivity (%)	Specifity (%)
1	95.95	92.86	96.67
2	98.65	100	98.33
3	97.30	100	96.67
4	98.65	92.86	100
5	98.65	100	98.33
6	98.65	100	98.33
7	98.65	100	98.33
8	97.30	85.71	100
9	98.65	100	98.33
10	98.65	100	98.33

In terms of average specifity, the best result is (98,5%) is obtained by using all feature subsets (F1,F2,F3,F4,F5). When feature subsets (F2, F3, F4, F5) are used the average specifity is determined as 97,67%. Finally using feature subsets (F2, F3, F5) average specifity is 98,33%.

5.2 Hypothyroid Test Results

The first experiment on diagnosis of hypothyroidcontains is performed with all the feature subsets of the original dataset. The output of the optimization algorithm is C=0,1166. This is the value of the soft margin parameter in SVM used for the classification task. The changes in fitness value of hypothyroid test that all features

in dataset are used in are seen in figure 5.7.

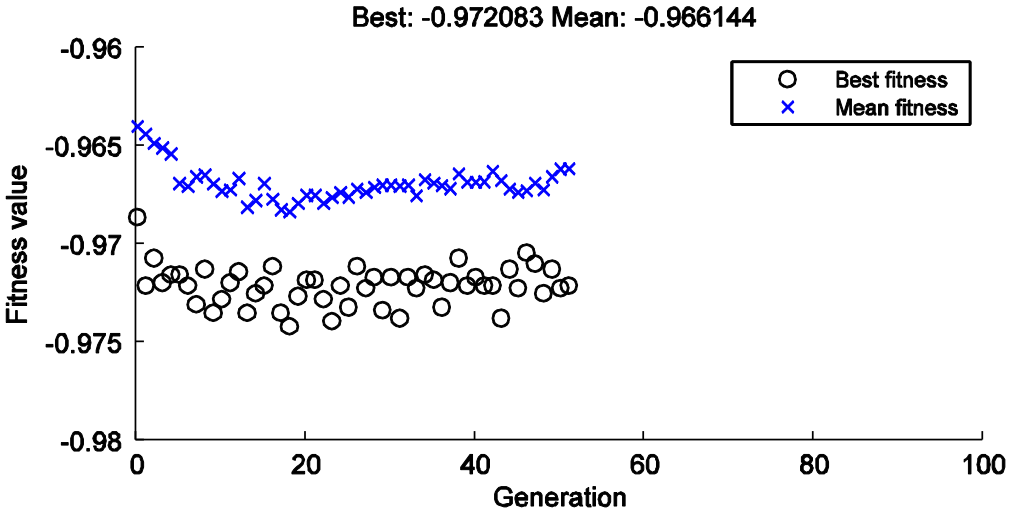


Figure 5.7 : The changes in fitness value of hypothyroid test that all features in dataset are used in (F1,F2,F3,F4,F5).

It would be good to remember that the objective of this study was to increase the efficiency of the SVM in terms of accuracy. The parameter to be optimized is the coefficient of the slack variable and the fitness value is the accuracy of classification. Since genetic algorithm always struggles to find the minimums, the accuracy is multiplied with -1 and scaled in between 0 and -1. Having seen that the best fitness and the mean fitness are 97.2% and 96.6% ,respectively, in figure 5.7 one can say that the corresponding SVM is more successful in diagnosis of hyperthyroid disease than it is in diagnosis of hypothyroid in case the all feature subsets are used.

The changes in average distance between individuals as the new generations are being created for the first hypothyroid test is shown in figure 5.8. After 8^{th} generation, the average distance between individuals becomes zero and remains still in that point. This situation could be the result of taking lower bound as 0.1. Although the lower bound seems to be taken high, considering all experiments together, it is the best value that gives reasonable results.

The second experiment is carried out using the Features (F1,F2,F3,F5). The optimal value of C is 22,7735. This is used as the value of the soft margin parameter in the SVM that is created for the classification task which has 4 inputs .

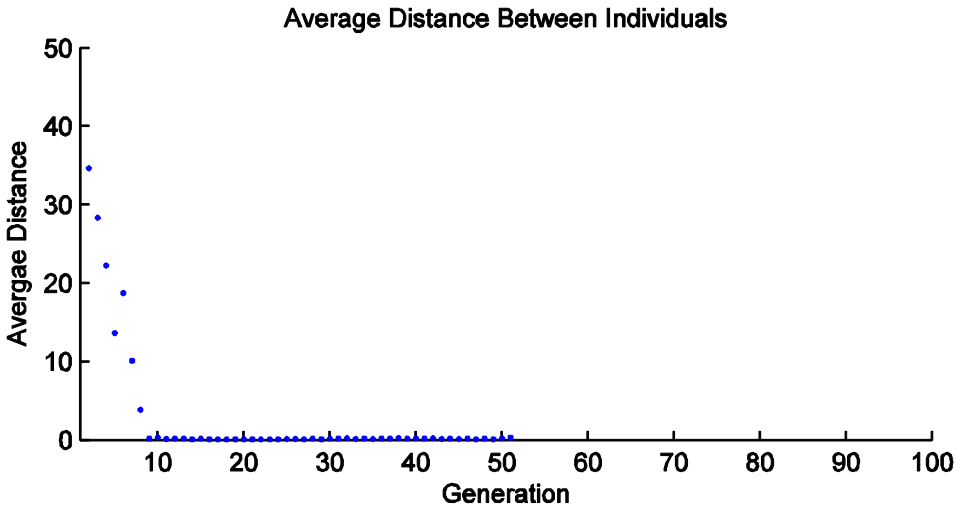


Figure 5.8 : The Average distance between individuals for hypothyroid test that all features (F1,F2,F3,F4,F5) are used in.

The changes in fitness value of hypothyroid test 2 are seen in figure 5.9. The accuracy is multiplied with -1 and scaled in between 0 and -1. It can be seen from the fig 5.9 that the performance is decreased in terms of both best fitness and average fitness. (97% and 96.6%, respectively)

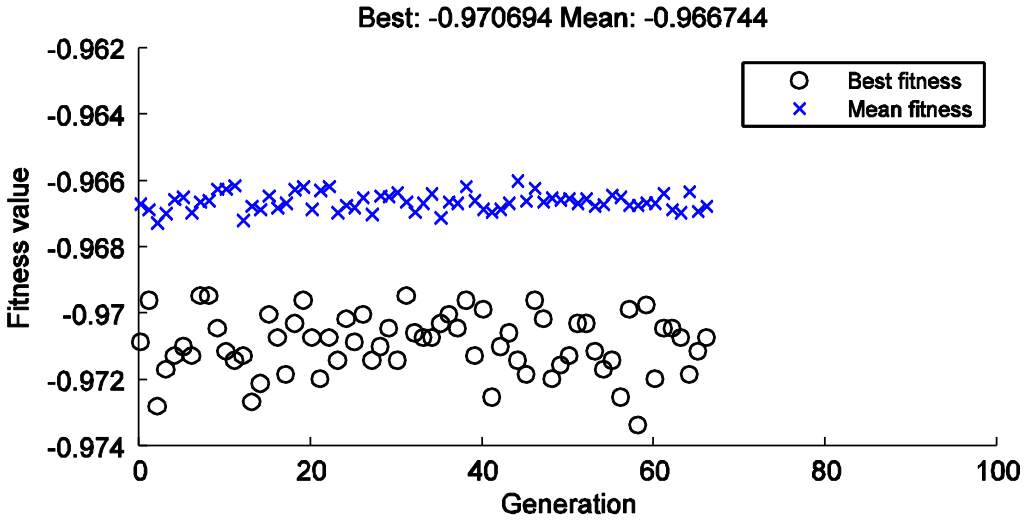


Figure 5.9 : The changes in fitness value of hypothyroid test 2 (F1,F2,F3,F5).

The graph of average distance between individuals vs. number of generation for first hypothyroid test is shown in figure 5.10. It can be observed from fig 5.10 that the changes in the average distance between individuals does not follow a desired path which ensures that the global minima is found in between lower and upper bounds.

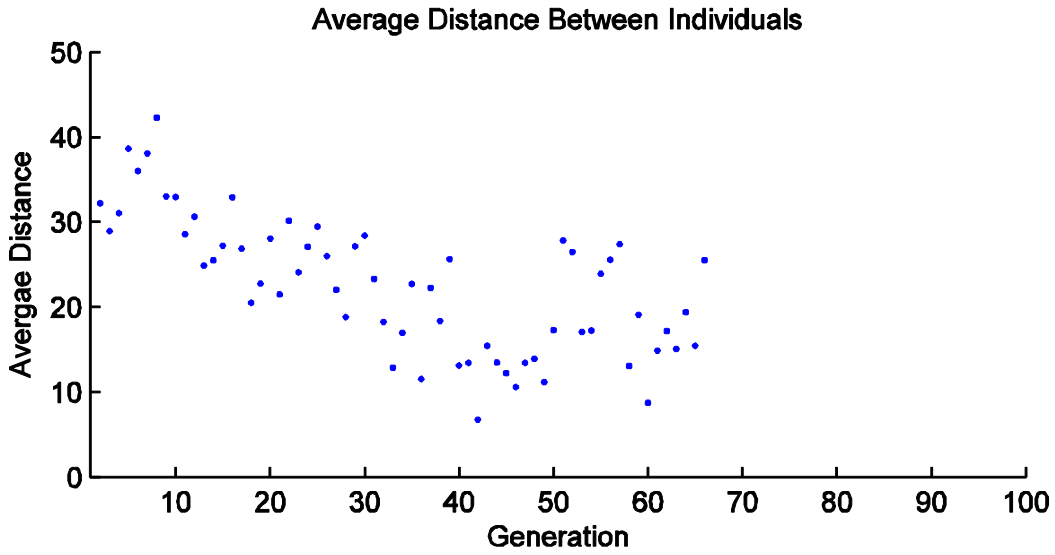


Figure 5.10 : The Average distance between individuals for hypothyroid test 2 (F1,F2,F3,F5).

In the last experiment performed for diagnosis of hypothyroid disease (F1,F2,F3) the optimal value of the soft margin parameter was calculated as C=0,1. This is used as the value of the parameter C in SVM that is designed for the classification task that has 3 inputs (T3RU,T4,T3)

The changes in fitness value of hyperthyroid test 3 are seen in figure 5.11. The accuracy is multiplied with -1 and scaled in between 0 and -1. It is seen that the best fitness is approximatel 94,9% and the mean fitness is 94%

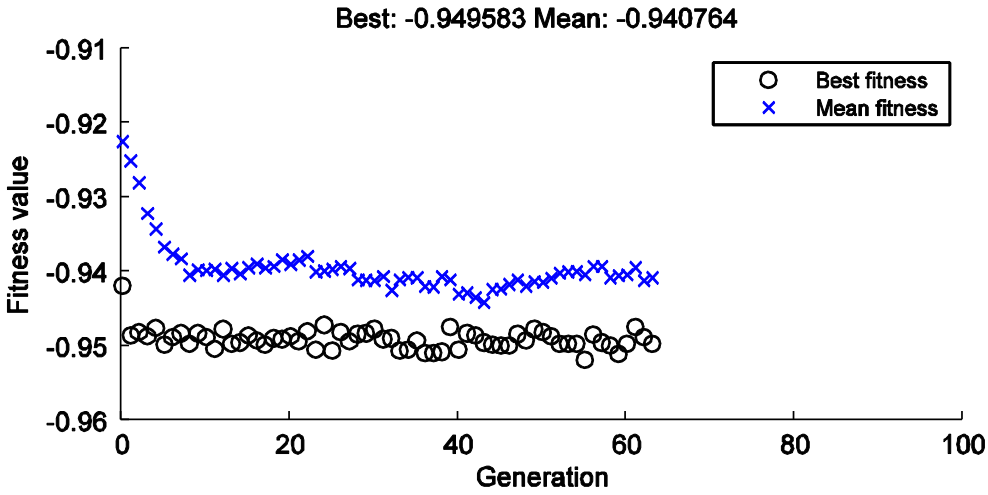


Figure 5.11 : The changes in fitness value of hypothyroid test 3 (F1,F2,F3).

The graph of average distance between indivuduals vs. the number of generation for first hyperthyroid test is shown in figure 5.12.

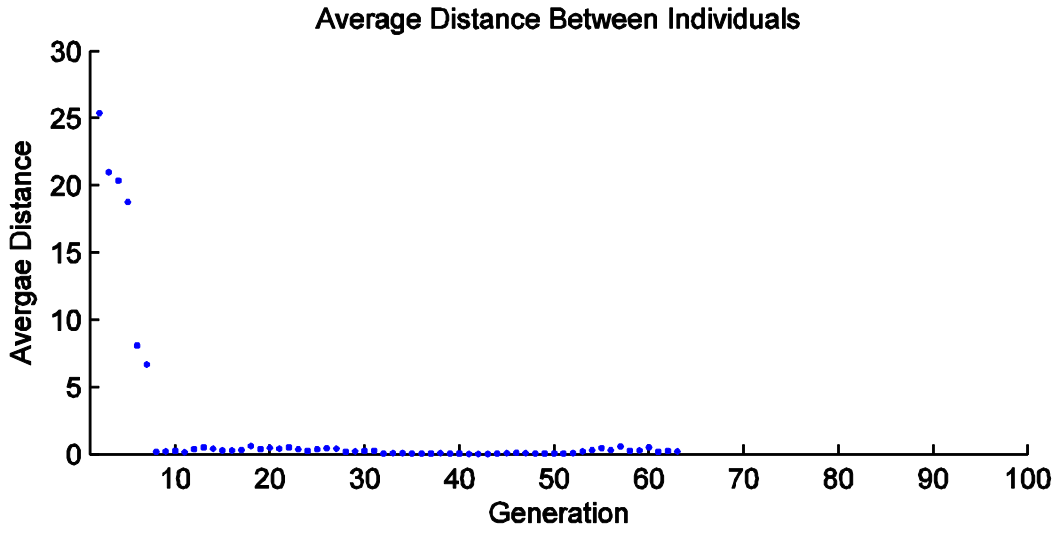


Figure 5.12 : The Average distance between individuals for hypothyroid test 3 (F1,F2,F3).

By using all feature subsets (the first experiment) obtained optimal C is (C=0,1166). 98,06% is the average accuracy (see Table 5.6) at that optimal value of C.

Table 5. 6: Result of the experiment 4- Hypothyroid (F1,F2,F3,F4,F5).

Number of Test	Accuracy (%)	Sensitivity (%)	Specifity (%)
1	97.22	100	96.67
2	97.22	91.67	98.33
3	98.61	91.67	100
4	98.61	91.67	100
5	97.22	83.33	100
6	100	100	100
7	94.44	75.00	98.33
8	98.61	91.67	100
9	100	100	100
10	98.61	91.67	100

Using the feature subsets (F1,F2,F3, F5) led the optimal value of C to be calculated as Coptimal=0,1166) and 96,53% average accuracy (see Table 5.7) corresponding to the optimal value of C.

Using the feature subsets (F1,F2,F3) obtained results (see Table 5.8) are Coptimal=0.1 and 95,97% average accuracy at that optimal value of C.

Table 5. 7 : Result of the experiment 5- Hypothyroid (F1,F2,F3,F5).

Number of Test	Accuracy (%)	Sensitivity (%)	Specifity (%)
1	97.22	100	96.67
2	97.22	83.33	100
3	94.44	75.50	98.33
4	98.61	100	98.33
5	98.61	91.67	100
6	98.61	100	98.33
7	93.06	66.67	98.33
8	94.44	83.33	96.67
9	97.22	83.33	100
10	95.83	83.33	98.33

In terms of average sensitivity, the best result (91,67%) is calculated by using all feature subsets (F1, F2, F3, F4, F5). When the feature subsets (F1, F2, F3, F5) is used the corresponding average sensitivity is 86,67%. Using feature subsets(F1, F2, F3) average sensitivity gets 87,50%.

In terms of average specifity, the best result (99,33%) is obtained by using all feature subsets (F1,F2,F3,F4,F5). When feature subsets (F1, F2, F3, F5) are used the average specifity is found 98,50%. Finally using feature subsets (F1, F2, F3) get average specifity of 97,67%.

Table 5. 8 : Result of the experiment 6- Hypothyroid (F1,F2,F3).

Number of Test	Accuracy (%)	Sensitivity (%)	Specifity (%)
1	93.06	75.00	96.67
2	97.22	91.67	98.33
3	97.22	83.33	100
4	95.83	75.00	100
5	95.83	91.67	96.67
6	95.83	100	95.00
7	95.83	91.67	96.67
8	98.61	91.67	100
9	98.61	91.67	100
10	91.67	83.33	93.33

6. CONCLUSION

Many studies have been conducted in order to successfully diagnose thyroid disease via machine learning techniques. Most of them contain hybrid and sophisticated algorithms.

In this study a system that is able to diagnose the Thyroid diseases (hyperthyroid and hypothyroid) using SVMs is presented. The questions of how to increase the accuracy of the classifier by selecting the most important features and finding the optimal value of a classifier parameter for each dataset created using the information on importance of features are answered.

The results have shown that via optimization of SVM parameters, one can always obtain better results in terms of accuracy, sensitivity and specifity of the classifier. Also by reducing the dimensionality via feature selection one can increase the performance of the classifier in terms of accuracy, sensitivity and specifity.

In the future, this experimental system is going to be used in other classification problems in which optimization could be required. Also, more parameters are going to be optimized in order to make the classifier more accurate.

REFERENCES

[1] **Bhaskar, S., V.B. Singh, and A. Nayak.** *Managing data in SVM supervised algorithm for data mining technology*. in *IT in Business, Industry and Government (CSIBIG), 2014 Conference on*. 2014. IEEE.

[2] **Liu, F., W.K. Ng, and W. Zhang.** *Encrypted SVM for Outsourced Data Mining*. in *Cloud Computing (CLOUD), 2015 IEEE 8th International Conference on*. 2015. IEEE.

[3] **Drucker, H., D. Wu, and V.N. Vapnik**, *Support vector machines for spam categorization.* Neural Networks, IEEE Transactions on, 1999. **10**(5): p. 1048-1054.

[4] **Li, Z.-W., J. Yang, and J.-P. Zhang**. *Dynamic incremental SVM learning algorithm for mining data streams*. in *Data, Privacy, and E-Commerce, 2007. ISDPE 2007. The First International Symposium on*. 2007. IEEE.

[5] **Vapnik, V.** *Statistical Learning Theory. John Wiley&Sons.* Inc., New York, 1998.

[6] **Ben-Hur, A., et al.** *Support vector clustering.* The Journal of Machine Learning Research, 2002. **2**: p. 125-137.

[7] **Mukherjee, S. and V. Vapnik.** *Support vector method for multivariate density estimation.* Center for Biological and Computational Learning. Department of Brain and Cognitive Sciences, MIT. CBCL, 1999. **170**.

[8] **Smola, A.J. and B. Schölkopf.** *A tutorial on support vector regression.* Statistics and computing, 2004. **14**(3): p. 199-222.

[9] **BAGHERZADI, N.** *Post operative prognostic prediction of esophageal cancer cases using bayesian networks and support vector machines*. 2014, Middle east technical university.

[10] **Chang, C.-Y., M.-F. Tsai, and S.-J. Chen.** *Classification of the thyroid nodules using support vector machines*. in *Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*. 2008. IEEE.

[11] **Aronszajn, N.** *Theory of reproducing kernels.* Transactions of the American mathematical society, 1950: p. 337-404.

[12] **Vapnik, V. and A. Chervonenkis.** *A note on one class of perceptrons.* Automation and remote control, 1964. **25**(1).

[13] **Çomak, E.** *Destek Vektör Makineleri çoklu sınıf Problemleri için Çözüm önerileri*. 2004: Konya.

[14] **Joachims, T.** *Text categorization with support vector machines: Learning with many relevant features*. 1998: Springer.

[15] **Chapelle, O., P. Haffner, and V.N. Vapnik.** *Support vector machines for histogram-based image classification.* Neural Networks, IEEE Transactions on, 1999. **10**(5): p. 1055-1064.

[16] **Machado, R.F., et al.** *Detection of lung cancer by sensor array analyses of exhaled breath.* American journal of respiratory and critical care medicine, 2005. **171**(11): p. 1286-1291.

[17] **Duan, K.-B., et al.** *Multiple SVM-RFE for gene selection in cancer classification with expression data.* NanoBioscience, IEEE Transactions on, 2005. **4**(3): p. 228-234.

[18] **Machhale, K., et al.** *MRI brain cancer classification using hybrid classifier (SVM-KNN).* in *Industrial Instrumentation and Control (ICIC), 2015 International Conference on*. 2015. IEEE.

[19] **Chakraborty, D. and S. Das.** *Cancer classification through feature selection and transductive SVM using gene microarray data*. in *Emerging Applications of Information Technology (EAIT), 2012 Third International Conference on*. 2012. IEEE.

[20] **Wang, H. and H. Zheng.** *A bayesian approach to improving decision making in support vector machine and its application in bioinformatics*. in *Natural Computation, 2009. ICNC'09. Fifth International Conference on*. 2009. IEEE.

[21] **Conforti, D. and R. Guido.** *Kernel based support vector machine via semidefinite programming: Application to medical diagnosis.* Computers & Operations Research, 2010. **37**(8): p. 1389-1394.

[22] **Martens, D., et al.** *Comprehensible credit scoring models using rule extraction from support vector machines.* European journal of operational research, 2007. **183**(3): p. 1466-1476.

[23] **Li, D.-C. and C.-W. Liu.** *A class possibility based kernel to increase classification accuracy for small data sets using support vector machines.* Expert Systems with Applications, 2010. **37**(4): p. 3104-3110.

[24] **Sartakhti, J.S., M.H. Zangooei, and K. Mozafari.** *Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA).* Computer methods and programs in biomedicine, 2012. **108**(2): p. 570-579.

[25] **Chen, C.-J., et al.** *Stage diagnosis for Chronic Kidney Disease based on ultrasonography*. in *Fuzzy Systems and Knowledge Discovery (FSKD), 2014 11th International Conference on*. 2014. IEEE.

[26] **Mo, Y. and S. Xu.** *Application of SVM based on hybrid kernel function in heart disease diagnoses*. in *Intelligent Computing and Cognitive Informatics (ICICCI), 2010 International Conference on*. 2010. IEEE.

[27] **Lee, C.-C. and S.-H. Chen.** *Gabor Wavelets and SVM Classifier for Liver Diseases Classiflcation from CT Images*. in *Systems, Man and Cybernetics, 2006. SMC'06. IEEE International Conference on*. 2006. IEEE.

[28] **Saiti, F., et al**. *Thyroid disease diagnosis based on genetic algorithms using PNN and SVM*. in *Bioinformatics and Biomedical Engineering, 2009. ICBBE 2009. 3rd International Conference on*. 2009. IEEE.

[29] **Shariati, S. and M.M. Haghighi.** *Comparison of anfis Neural Network with several other ANNs and Support Vector Machine for diagnosing hepatitis and thyroid diseases*. in *Computer Information Systems and Industrial Management Applications (CISIM), 2010 International Conference on*. 2010. IEEE.

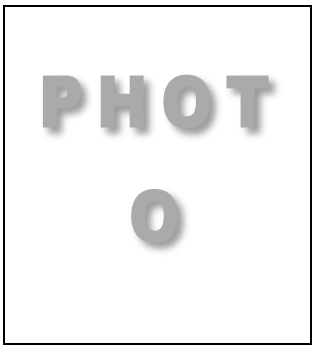
[30] **Dammak, F. and L. Baccour.** *Proposition of a classification system "β− LS− SVM" and its application to medical data sets*. in *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*. 2014. IEEE.

[31] **Lopes, D.C., et al.** *Implementation of a SigmaBoost-based ensemble of SVM in a multiple processor system on chip*. in *Programmable Logic Conference (SPL), 2010 VI Southern*. 2010. IEEE.

[32] **Rouhani, M. and K. Mansouri.** *Comparison of several ANN architectures on the Thyroid diseases grades diagnosis*. in *Computer Science and Information Technology-Spring Conference, 2009. IACSITSC'09. International Association of*. 2009. IEEE.
[33] In endokrin.org. Date retrieved: 05.11.2015 adress: http://www.endokrin.org/Tr/content3.asp?m1=1&m2=20&m3=254.
[34] **Ozet, M.B.** *Tiroit sintigrafisi*. J. Exp. Clin. Med, 2012. **29**: p. S289-S300.
[35] **Onsel, C., et al.** *THE EFFECT OF CALCIUM AND PENTAGASTRIN INFUSIONS TO THE TC-99M (V) DMSA UPTAKE IN THE MEDULLARY-THYROID CARCINOMA (MTC)*. in *EUROPEAN JOURNAL OF NUCLEAR MEDICINE*. 1992. SPRINGER VERLAG 175 FIFTH AVE, NEW YORK, NY 10010.
[36] In mayoclinic.org. Date retrieved: 05.11.2015, adress: http://www.mayoclinic.org/thyroid-gland/img-20007063.
[37] **Ozata, M., et al.** *Evaluation of central motor conduction in hypothyroid and hyperthyroid patients.* Journal of endocrinological investigation, 1996. **19**(10): p. 670-677.
[38] **Yildirimkaya, M., et al.** *Lipoprotein (a) concentration in subclinical hypothyroidism before and after levo-thyroxine therapy*. Endocrine journal, 1996. **43**(6): p. 731-736.
[39] In tiroidhastaliklaritedavisi.com. Date retrieved: 06.11.2015, adress: *http://www.tiroidhastaliklaritedavisi.com/hipotiroid.htm*.
[40] **Vapnik, V.** *The nature of statistical learning theory*. 2013: Springer Science & Business Media.
[41] **Gunn, S.R.** *Support vector machines for classification and regression.* ISIS technical report, 1998. **14**.
[42] **Karush, W.** *Minima of functions of several variables with inequalities as side conditions*. 2014: Springer.
[43] **Gale, D., H.W. Kuhn, and A.W. Tucker.** *Linear programming and the theory of games.* Activity analysis of production and allocation, 1951. **13**: p. 317-335.
[44] **Haupt, R.L. and S.E. Haupt.** *Practical genetic algorithms*. 2004: John Wiley & Sons.
[45] **Ertekin, S.** *Learning in extreme conditions*. 2009, The Pennsylvania State University.
[46] In Mathworks, Date retrieved: 16.11.2015, adress: http://www.mathworks.com/discovery/quadratic-programming.html.
[47] **Duman, A.** *Multiple criteria sorting methods based on support vector machines*. 2010, Middle East Technical University: Ankara.
[48] **Cristianini, N. and J. Shawe-Taylor.** *An introduction to support vector machines and other kernel-based learning methods*. 2000: Cambridge university press.
[49] **Shah, R.S.** *Support Vector Machines for Classification and Regression*. 2007, McGill University.
[50] **Scholkopf, B. and A.J. Smola.** *Learning with kernels: support vector machines, regularization, optimization, and beyond*. 2001: MIT press.
[51] **Duda, R.O., P.E. Hart, and D.G. Stork.** *Pattern classification*. 2012: John Wiley & Sons.

[52] **Gu, Q., Z. Li, and J. Han.** *Correlated multi-label feature selection*. in *Proceedings of the 20th ACM international conference on Information and knowledge management*. 2011. ACM.
[53] **Gu, Q., Z. Li, and J. Han.** *Generalized fisher score for feature selection.* arXiv preprint arXiv:1202.3725, 2012.
[54] **Mitchell, M.** *An introduction to genetic algorithms*. 1998: MIT press.
[55] **Coley, D.A.** *An introduction to genetic algorithms for scientists and engineers*. 1999: World scientific.
[56] **Ho, G., et al.** *Providing decision support for replenishment operations using a genetic algorithms based fuzzy system.* Expert Systems, 2013.
[57] **Lichman, M.** (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

CURRICULUM VITAE



Name Surname: Nuri KORHAN

Place and Date of Birth: DİYARBAKIR, 10.09.1986

E-Mail: nurikorhan@gmail.com

EDUCATION:

B.Sc.: Mechatronics Engineering, Kocaeli University,2012