JANUARY 2021

Ph.D in Industrial Engineering

PINAR KOCABEY ÇİFTÇİ

REPUBLIC OF TURKEY

GAZİANTEP UNIVERSITY

GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES

MODELING TOBACCO USE BEHAVIORS AND TOBACCO CONTROL POLICIES OF TURKEY

Ph.D THESIS

IN

INDUSTRIAL ENGINEERING

BY

PINAR KOCABEY ÇİFTÇİ

JANUARY 2021

MODELING TOBACCO USE BEHAVIORS AND TOBACCO CONTROL POLICIES OF TURKEY

Ph.D. Thesis

in

Industrial Engineering

Gaziantep University

Supervisor

Assoc. Prof. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU

by

Pınar KOCABEY ÇİFTÇİ

JANUARY 2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Pınar KOCABEY ÇİFTÇİ

ABSTRACT

MODELING TOBACCO USE BEHAVIORS AND TOBACCO CONTROL POLICIES OF TURKEY

KOCABEY ÇİFTÇİ, Pınar
Ph.D. in Industrial Engineering
Supervisor: Assoc. Prof. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU
January 2021
103 pages

The rapid growth of the technology and the convenience provided by the technological products/services have connected people, firms, countries, systems and etc. to each other more than before. Increase in the interrelations has inevitably affected the way most of the systems work and caused dynamic complexity. As the complex systems have continuously surrounded the lives, understanding and analyzing the structure of these systems has become a challenging process. Traditional methods have been inadequate to handle this dynamic complexity. As the result, developing new methods has become a growing need for researchers/modelers to understand complex behaviors emerging in the systems. Thereby, the main objective of this thesis is proposing a novel hybrid simulation model to better understand tobacco use behaviors of individuals. The model is developed by embedding the concept of the multi-stage learning based fuzzy cognitive map (FCM) into the agent based model (ABM) in order to benefit from advantageous of each methodology. The ABM is basically used for representing the individual level behaviors while the FCM is used as a decision support mechanism of individuals. In order to represent the current tobacco use status of Turkey, socio-demographic characteristics of individuals, tobacco control policies, and social network effects are examined. In addition, scenario analyses are performed to analyze the potential effects of plain package policy and COVID19 pandemic. The findings of this thesis indicate that the hybrid usage of the ABM and the multi stage learning based FCM help us to observe how individual level behaviors and system state changed under the effects of casual relationships among the critical factors. The thesis also presents how the proposed model can be used for imitating a real complex and dynamic problem such as tobacco use and control. Lastly, the findings of the scenario analyses point out critical demographic groups for future policy making process.

Key Words: Agent Based Model, Fuzzy Cognitive Map, Nonlinear Hebbian Learning Algorithm, Extended Great Deluge Algorithm, Tobacco Use

ÖZET

TÜRKİYE TÜTÜN KULLANIM DAVRANIŞLARI VE TÜTÜN KONTROL POLİTİKALARININ MODELLENMESİ

KOCABEY ÇİFTÇİ, Pınar
Doktora Tezi, Endüstri Mühendisliği Bölümü
Danışman: Doç. Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU
Ocak 2021
103 sayfa

Teknolojinin hızlı büyümesi ve teknolojik ürünlerin/hizmetlerin sağladığı kolaylıklar insanları, firmaları, ülkeleri, sistemleri vb. eskiye kıyasla birbirine daha fazla bağımlı hale getirmiştir. Yapıların birbirleriyle olan ilişkilerindeki bu artış sistemlerin çoğunun çalışma şeklini etkilemiş ve dinamik karmaşıklığa sebep olmuştur. Karmaşık sistemler sürekli olarak hayatı çevrelediğinden, sistemlerin yapısını anlamak ve analiz etmek zorlu bir süreç haline gelmiştir. Geleneksel yöntemler karmaşık sistemleri çözümlemek için yetersiz kalmıştır. Sonuç olarak, araştırmacıların sistemlerde ortaya çıkan karmaşık davranışları anlamaları için yeni yöntemler geliştirmesi bir ihtiyaç haline gelmiştir. Bu tezin temel amacı bireylerin tütün kullanım davranışlarını daha iyi anlayabilmek için yeni bir hibrit benzetim modeli sunmaktadır. Model, çok aşamalı öğrenme tabanlı bulanık bilişsel haritalama yönteminin etmen tabanlı modelleme yöntemi içerisine entegre ederek her iki yöntemin avantajlarından yaralanmak üzere geliştirilmiştir. Etmen tabalı modelleme temel olarak birey davranışlarını modellemede kullanılırken, bulanık bilişsel haritalama bireylerin karar destek mekanizması olarak kullanılmıştır. Türkiye'nin mevcut tütün kullanım durumunu temsil etmek için sosyodemografik özellikler, tütün kontrol politikaları ve sosyal ağ etkileri çalışmada incelenmiştir. Bunlara ek olarak, tek tip paket politikası ve COVID19 pandemisinin olası etkilerini analiz etmek için senaryo analizleri gerçekleştirilmiştir. Bu tezin sonuçları, etmen tabanlı modelleme ve çok aşamalı öğrenme temelli bulanık bilişsel haritalamanın birlikte kullanımının birey davranışlarındaki değişimleri ve farklı faktörlerin birbirleriyle ilişkilerinin doğurduğu etkileri gözlemeye yardımcı olabileceği gösterilmiştir. Tez ayrıca modelin tütün kullanım ve kontrolü gibi karmaşık bir gerçek hayat problemi üzerinde uygulamasını sunmuştur. Son olarak, senaryo analizleri gelecek tütün kontrol politikaları oluşturma sürecinde dikkat edilmesi gereken demografik grupları ortaya çıkarmıştır.

Anahtar Kelimeler: Etmen Tabanlı Modelleme, Bulanık Bilişsel Haritalama, Lineer Olmayan Hebb Öğrenme Algoritması, Genişletilmiş Büyük Tufan Algoritması, Tütün Kullanımı.

''Dedicated to my lovely son who gave me hope and the determination to fight against the difficulties''

ACKNOWLEDGEMENTS

This thesis process has been an unforgettable part of my personal and academic life. There have been many people who always support and believe in me during this journey.

First, I would like to present my sincere thanks to my supervisor, Associate Professor Dr. Zeynep Didem UNUTMAZ DURMUŞOĞLU for her guidance and support throughout this thesis. I am thankful for her encouragement and motivation. I am also very grateful to Associate Professor Dr. Alptekin DURMUŞOĞLU. He always shared his valuable experiences and comments with me. I would like to thank Professor Dr. Türkay DERELİ for his important encouragement and feedback.

I should express my special thanks to my dear husband Emrah ÇİFTÇİ and my lovely little son Alptuğ ÇİFTÇİ for their patience, understandings and supports during my misery times. They have been my inspiration all these times. I also would like to thank my dear family, my mother, father, brother and Emrah's family. I could not be able to complete this thesis without their support.

I would like tell my gratitude to my dear friends Özge VAR, Selen KAHRAMAN, Elif Selen BABÜROĞLU, Nazmiye ELİGÜZEL, and Miraç ELİGÜZEL who experienced all my sorrows and joys with me.

Finally, thanks to everyone in my life who always believes in me and feels happy with my sucesses.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

LIST OF SYMBOLS

η	Learning Rate
γ	Weight Decay Parameter

LIST OF ABBREVIATIONS

AB	Agent Based
AHL	Active Hebbian Learning
BDL	Balanced Differential Learning
CM	Cognitive Map
DD-NHL	Data Driven Nonlinear Hebbian Learning
DE	Discrete Event
DHL	Differential Hebbian Learning
EGDA	Extended Great Deluge Algorithm
FCM	Fuzzy Cognitive Map
FCTC	Framework Convention on Tobacco Control
GA	Genetic Algorithm
GATS	Global Adult Tobacco Survey
GYTS	Global Youth Tobacco Survey
ICA	Imperialist Competitive Algorithm
NHL	Nonlinear Hebbian Learning
PSO	Particle Swarm Optimization
SD	System Dynamics
TUIK	Turkish Statistical Institute
WHO	World Health Organization

CHAPTER I
INTRODUCTION

1.1 Motivation of Study

System is a concept that concerns all living or non-living components of the universe. Everything from the smallest piece of the matters to planets that give the present form of the universe is part of various systems. Curiosity about how systems work has attracted the interest of people for centuries because human beings want to improve their knowledge by thinking and researching due to their nature. Their desires to learn about systems have leaded them to query the existing knowledge and ask new questions to reveal characteristics of the systems. However, understanding the behavior of the systems entirely is a challenging process. The main reason why it is a challenge can be understood from the definition of the "system" concept. In the most generic form, system can be defined as a collection of interrelated components that interact with each other to accomplish a purpose.

The words "interrelated" and "interact" used in the given definition mean much more for system analysts and researchers nowadays. The main reason behind this is the technological developments that seriously affect all areas of our lives. The internet and many other innovative developments have given the opportunity to bring many people, companies and organizations together that were not aware of each other before. This situation has led to the disappearance of borders between countries and the acceleration of globalization. Globalization has caused the systems to become more complex. In addition, the increase in the communication resources with the help of technology has made the complexity more palpable in many social systems. The change in the structure of the systems has caused the need for researchers to improve the way they handle systems.

When considering the subject of understanding complex system behavior, agent based (AB) modeling and fuzzy cognitive maps (FCM) are two important methods that draw attention in the literature. Both methods can successfully examine complex system structures using different system approaches. Basically, the AB models model the main elements of the systems as agents and express the relationships of these agents with each other and their environment using an individual approach. On the other hand, most FCMs determine the important elements of the systems as concepts and examine the relationships between concepts using an aggregate approach. The approaches used by both methodologies have important advantages and disadvantages to modeling processes. The most prominent of the advantages and disadvantages can be explained as follows. While AB models can represent the individual behaviors, their interactions, communications and their social network in detail, FCMs generally focus on fundamental critical concepts of the system to understand the system. AB models can be thought as designing a building by placing bricks and all necessary equipments one by one to reach the final picture. Just like the building's need for bricks and equipments, AB models need a substantial amount of data to entirely reflect a real life system. However, it is not always possible to find enough data that helps to imitate the behaviors of the system. In this case, a different method can help to understand the system with the existing data. At this point, FCMs stand out because they easily reveal the casual relations between the vital system factors using scarce data. The benefits of both methods when modeling systems have triggered the idea of using them in an integrated way and creating a hybrid model. The hybrid models can help to eliminate the disadvantageous caused by the use of methods separately while allowing systems to investigate in more detail. The hybrid usage of AB modeling and FCMs can have the potential to be applied in social systems where data sources are mostly limited and unreliable. In the literature, the hybrid usage of both methodologies is relatively new and there is an important gap for the applications on the models on social problems.

Thus, this Ph.D thesis is inspired by the potential of the mentioned modeling approach and basically aimed to make a hybrid model for investigating one of the important social problems. For this purpose, an important public health problem "tobacco use and control" is selected since tobacco use is a complex and dynamic social system that consists of several different factors such as characteristics of people, tobacco control policies, social networks of people and etc. Although tobacco use is a widely studied system all over the world, the ongoing prevalence of tobacco use brings out the necessity of more advanced studies for understanding tobacco use behaviors comprehensively. Therefore, developing a hybrid model for understanding tobacco use behaviors of Turkish people became the other important motivation for this thesis.

1.2 Roadmap of the Thesis

In the light of the information given in the previous section, this PhD thesis focused on building a hybrid agent based model that is integrated with fuzzy cognitive map to understand the changes at tobacco use behaviors of people according to their demographic characteristics, tobacco control policies and etc. In this respect, the roadmap of the thesis for the readers is given in Figure 1.1.

Chapter 2 presents the performance of Turkey for tobacco use literature and studies that focused to integrate agent based modeling and fuzzy cognitive to represent brain activities of individuals.

Chapter 3 provides the theories and fundamental concepts of agent based modeling and fuzzy cognitive map. It also pointed the importance of learning algorithms and discussed the requirement of a multi-stage learning procedure. Finally it provided the problem statement.

Chapter 4 explains the details of multi-stage learning based fuzzy cognitive map, its implementation to tobacco use research field and presented the results.

Chapter 5 represents the integration processes of previously explained multi-stage learning based fuzzy cognitive map with agent based modeling. It gives the fundamental concepts of the hybrid model and its applications to understand tobacco use behaviors of people. Finally, it provides the findings of the model.

Chapter 6 presents a brief summary for the steps and findings of this PhD thesis. It also explains the limitations and opportunities for future works.

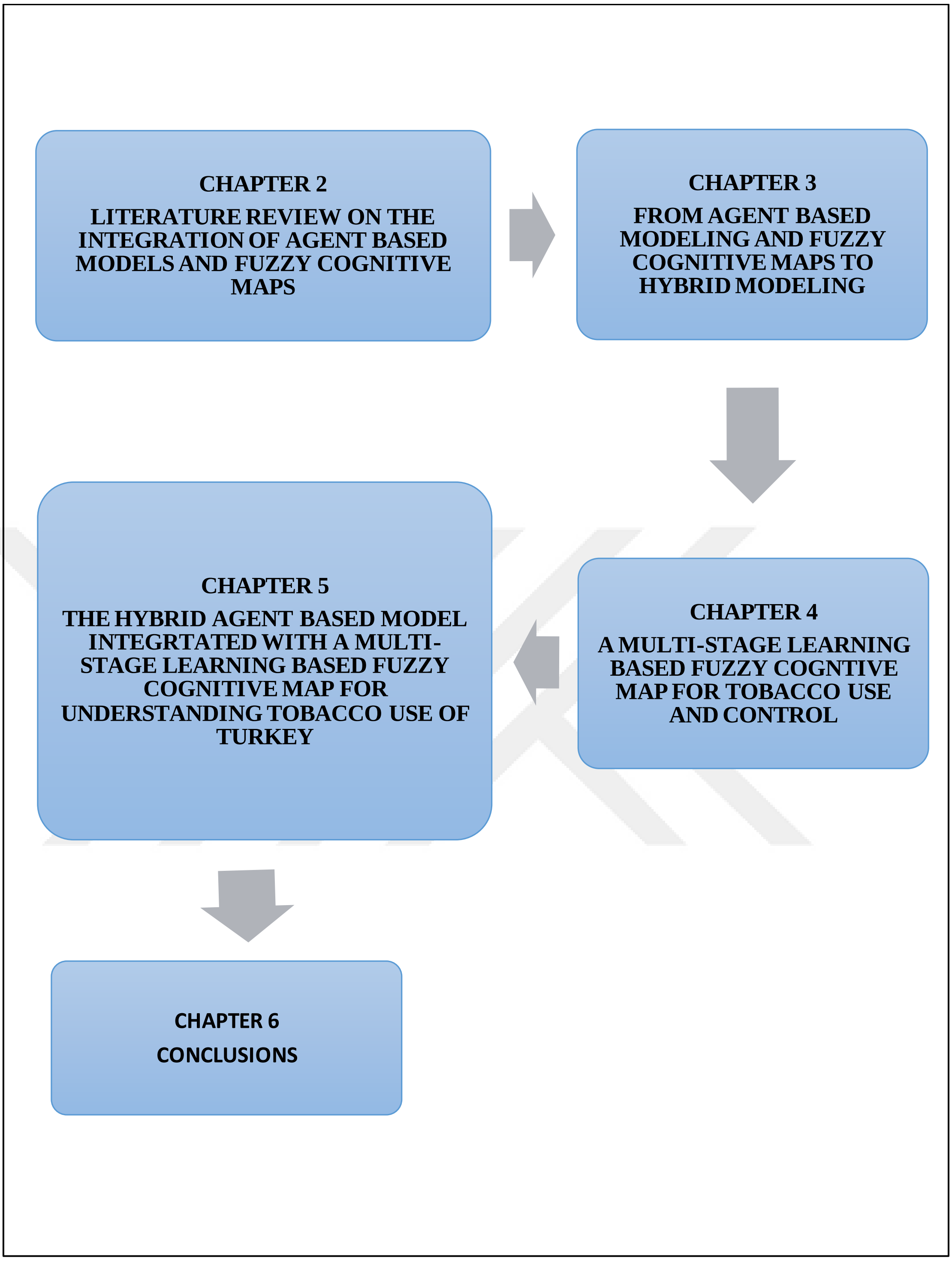

Figure 1.1 The flow of the thesis for the readers.

1.3 Concluding Remarks

The proposed researches in this thesis contributes to agent based modeling, fuzzy cognitive map and tobacco use literatures. The different parts of this thesis published in various prestigious journals and presented in different conferences. The findings of the thesis can also help tobacco use researchers and policy makers to understand tobacco use behaviors of people in detail.

CHAPTER II

LITERATURE REVIEW ON THE INTEGRATION OF AGENT BASED MODELS AND FUZZY COGNITIVE MAPS

2.1 Introduction

In real life, experiments are not always applicable to test new ideas or to improve existing ones because most of the experiments are time consuming and costly to repeat several times as well as being dangerous. In such cases, researchers look for methods that can perform their work with less cost in a shorter time and a safe environment. Models plays critical roles in science most of the time because they can be used in the development, exploration and application of theories [1] and modeling can help researchers to abstract the reality and deal with the world in a simplified manner, avoiding complexity, and danger [2]. Besides modeling also gives us the opportunity to make mistakes and restart. Due to the benefits it provides, modeling has become one of the methods used by researchers to solve many real life problems.

Modeling techniques vary according to the structures of the problems or systems that they will be implemented. For example; when thinking about the activities we will perform during the day, we can create a mental model. Miniature of a building project can be an example of a physical model. Using spreadsheet in order to get a specific output with the help of a number of inputs can be accepted as an analytical model. Thus, several different modeling techniques have been used to analyze the systems during decades. However, when the problems or systems become more complex and dynamic, these simple models cannot be sufficient to understand them. That's why; modelers have searched for advanced modeling ways to handle the complexity and dynamism of the systems or problems. At that point, the technological developments have helped to the modelers. The widespread use of computers, advanced software and the internet led to rapid developments in the field of modeling and has enabled detailed analysis of problems.

Simulation modeling is one of the important art of constructing models for representing and understanding real life systems and behaviors [3]. A simulation model basically consists of a set of rules (equations, schedules, statecharts and etc.) that explain how the state of the system will change in future using the present state [4], [5]. It is very valuable to explore consequences of various assumptions for the studied system/problem [6].

Simulation modeling can be categorized in three different types. These are discrete event simulation (DES), system dynamics (SD) and agent based (AB) modeling. When compared to DES and SD, AB modeling is a relatively new tool that can help to understand system behavior in more detail by using an individual approach. Providing opportunities such as examining individual behaviors, making autonomous decisions, and observing the effects of social networks has revealed the fact that AB modeling can be beneficial in social areas, too. That's why; several different applications of AB models on various social problems or systems can be found in the literature. Although it has been one of the significant modeling tools for most systems, it cannot be ignored that the modeling phase gets compelling when the problem/system becomes more complex. Especially in social systems where individual characteristics vary, interaction is high and the social network cannot be clearly defined, it is really difficult to provide enough data to make an AB model all the time. In order to overcome the deficiencies of AB models, researchers/modelers integrate the AB modeling with different techniques, create hybrid models and try to maximize the potential benefits. This PhD thesis focuses on developing a hybrid AB model integrated with a multi-stage learning based fuzzy cognitive map (FCM) that is also an important method that can handle complexity. The presented hybrid model is implemented for tobacco use and control in order to increase the understanding on the changes of tobacco use behaviors of people in Turkey considering the effects of demographic characteristics, tobacco control policies and social network.

In this chapter of the thesis, the reasons why studying on tobacco use and the different applications of the hybrid usage of AB modeling and FCM are presented systematically.

2.2 Tobacco Use and Control

Tobacco use is one of the biggest public health problems that kills more than 8 million people each year worldwide [7]. It is accepted as the leading cause of preventable disability and death [8]. It harms almost every organ of the body and causes various diseases such as cancer, cardiovascular diseases, respiratory diseases and etc. [9]. Although the mentioned harms of tobacco use has been examined and published several times, its widespread use still continues all over the world. In order to decrease prevalent use of tobacco and damages of it to public health, economy and environment, countries have developed various strategies.

Turkey is among the countries that have a severe fight against tobacco use. After the signature of the World Health Organization (WHO) Framework Convention on Tobacco Control (FCTC) in 2004, Turkey has accelerated its efforts on tobacco control. Although tobacco control measures of Turkey were remarkable and provided great gains, they could not completely eliminate the widespread use of tobacco. This situation raised the question of whether tobacco use has been scientifically studied well enough to help governments and policy makers in Turkey. In order to find the answer of this question, Unutmaz Durmuşoğlu and Kocabey Çiftçi [10] performed a detailed quantitative analysis to see the trends in tobacco use and control research field and how much researchers of Turkey contributed to this area. The fundamental aim of this study was to examine the change in number of publications before and after the WHO FCTC was introduced in Turkey and other countries. This treaty was released by the WHO in order to help countries to control the widespread use of tobacco in their societies [10]. It has helped the most of the countries and leads them to make efforts against tobacco. That's why; the study of Unutmaz Durmuşoğlu and Kocabey Çiftçi [10] was also searched whether the WHO FCTC has become an important milestone for Turkey and other countries according to the literature.

For this aim, they made a detailed publication search from the Thomson Reuters Web of Science that contains all articles indexed by SCI-Soc-SCI and Expanded SCIE [11]. Their search methodology was represented in Figure 2.1.

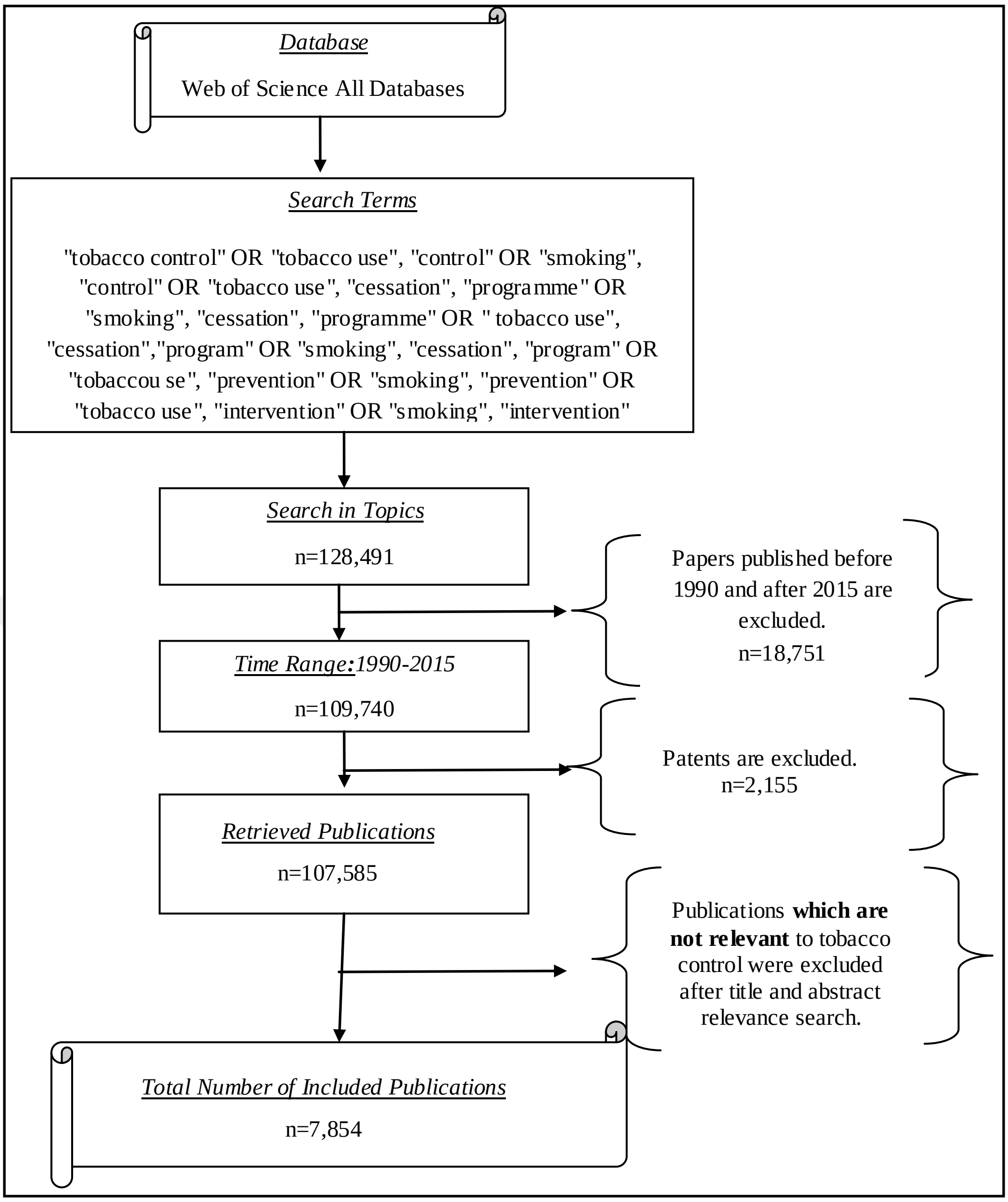

Figure 2.1 The data inclusion procedure [10].

They analyzed several different aspects of the relevant publications such as number of publications, type of publications, origins of authors, journals and etc. Among them, the change in the number of publications by years and the distribution of the authors according to their origins were two important source of motivation to study tobacco use and control in this thesis.

The number of publications on tobacco use and control has increased from 1990 to 2015 significantly all over the world as seen in Figure 2.2. Although the number of studies slightly decreased in some years, it can be said that there was an upward trend in the number of publication generally. These results showed that tobacco use and control research field has attracted the attentions of researchers especially after 2001.

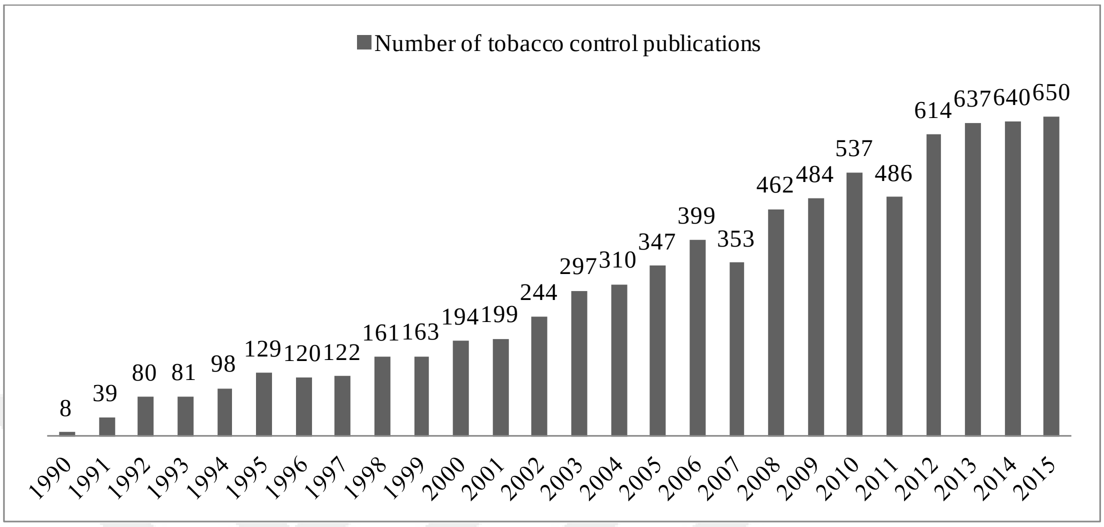

Figure 2.2 The number of publications on tobacco control and tobacco use fields over years [10].

The findings about the contributions of countries to the relevant field were also presented in the Table 2.1 in detail. Authors originated from the USA, Canada, UK, and Australia contributed to the tobacco use literature the most. Table 2.1 also presented the sign status of the WHO FCTC, signature dates and number of papers on tobacco use and control of countries for 5 year ranges. Some of the countries such as Malaysia and Uruguay made their first publications after their signature of the WHO FCTC.

When the results of Turkey were examined, it can be said that the signature of the WHO FCTC has increased the attentions on tobacco use and control because the number of publications have increased. However, Turkey was not among the countries that contributed to the relevant literature mostly like USA, Canada, UK and etc. Although Turkey has been suffering from the harms of prevalent usage of tobacco significantly, its publication results told that researchers of Turkey did not study on this topic sufficiently. One of the reasons behind why Turkey did not gain better results from their tobacco control policies may be due to the limited number of studies that focused on understanding tobacco use behaviors in Turkey.

Table 2.1 Most productive countries in tobacco control [10]

Countries	Tobacco Control Records (%)	WHO FCTC Sign. Status	WHO FCTC Sign. Year	1990 - 1994 records	1995 - 1999 records	2000 - 2004 records	2005 - 2009 records	2010 - 2014 records
USA	51.35	Signed	2004	183	395	686	1073	1375
Canada	8.70	Signed	2003	17	55	70	171	298
UK	8.62	Signed	2003	11	39	88	178	295
Australia	8.24	Signed	2003	13	43	77	163	293
Peoples R. of China	3.72	Signed	2003	3	15	17	55	149
Germany	3.20	Signed	2003	6	14	35	75	103
Netherlan ds	3.11	Signed	2003	10	18	20	54	120
Spain	2.61	Signed	2003	4	10	38	45	86
India	2.34	Signed	2003	4	2	6	36	117
France	2.18	Signed	2003	5	15	32	40	61
Switzerla nd	2.02	Signed	2004	3	11	35	35	60
Sweden	1.90	Signed	2003	7	20	32	38	41
Italy	1.88	Signed	2003	6	8	24	34	60
Brazil	1.54	Signed	2003	0	2	9	34	61
New Zealand	1.41	Signed	2003	2	8	9	28	53
Japan	1.38	Signed	2004	7	14	14	30	36
Mexico	1.35	Signed	2003	0	2	10	41	45
Republic of Korea	1.08	Signed	2003	0	1	6	19	46
Finland	1.03	Signed	2003	5	6	18	21	25
Taiwan	0.97	No info	-	0	4	1	31	29
Denmark	0.88	Signed	2003	4	2	13	19	25
Norway	0.83	Signed	2003	5	5	14	12	25
Ireland	0.81	Signed	2003	2	1	1	22	34
South Africa	0.81	Signed	2003	5	8	4	16	22
Poland	0.80	Signed	2004	0	3	4	14	32
Turkey	0.80	Signed	2004	0	1	6	23	25
Thailand	0.79	Signed	2003	1	0	7	21	25
Malaysia	0.67	Signed	2003	0	0	0	8	34
Greece	0.66	Signed	2003	2	0	2	16	26
Belgium	0.60	Signed	2004	1	7	6	11	18
Austria	0.46	Signed	2003	1	4	13	6	9
Portugal	0.46	Signed	2004	0	1	5	5	18
Others	15.02			20	61	163	280	557

Therefore, the study of Unutmaz Durmuşoğlu and Kocabey Çiftçi [10] indicated that there was still an important gap for understanding tobacco use behaviors of Turkey. That is why; increasing the understanding of tobacco use behaviors and the benefits of the tobacco control measures, tobacco use and control issue of Turkey was examined in detail using the developed hybrid AB modeling.

The next sub-section of the chapter presented the hybrid studies that integrated AB modeling and FCM.

2.3 The Integration of AB Modeling with FCM

Sterman [12] stated one of the most vital challenges of today's world as "The greatest constant of modern times is change". Change is the undeniable fact of the universe. All systems in the universe tend to change with time. Some situations such as the fast development of technology accelerate changes and make systems/people connected. The increase at the interactions in the systems makes systems more complex and dynamic. Learning about this kind of complex and dynamic systems required more advanced methodologies. Simulation is one of the important methodologies that enable to examine the complex systems.

Simulation can be defined as the imitation of a system or a process over time [13]. The major approaches of simulation modeling are System Dynamics (SD), Discrete Event (DE) and Agent Based (AB) [4] which have different perspectives to model the problems. Among them, the history of SD and DE goes back even further compared to AB modeling. The arrival of AB modeling in the1990s offered a novel approach to the simulation literature [14].

The AB modeling involves the imitating individualistic behavior and interpreting emerging patterns [15]. It models the systems with the help of collection of autonomous decision making entities (agents) which individually evaluated its states and makes decisions using a set of rules [16]. The AB models are really good to capture human environment interactions. For that reason, the AB modeling has gained increasing popularity for building useful models for social systems [17] that consist of several different heterogeneous factors and their interactions with each other and their environment. In the literature, there is one other methodology that stands out with its capability to handle complex structure of the social systems. It is called as fuzzy cognitive map (FCM).

FCM is a kind of hybrid method that combines the sense of fuzzy systems and neural networks [18]. It is also a modeling and simulation tool that are suitable for modeling complex systems consisted of a great number of highly related and interconnected elements and subsystems[19]. It basically illustrate the system by a graph that represents the cause and effect relationships among factors and integrates knowledge and experience with the operation of the system [20].

Although AB modeling and FCM are two important tools to model the complex systems, they use different approaches while examining the systems. While AB models use micro-level point of view, FCMs take a macro-level point of view which represents casual relationships instead of individual units and interactions [21]. FCMs stress on fundamental factors of a system, provide information about directions and strengths of the relationships between studied factors. One of the major difficulties to handle a social system is to provide necessary data to represent all relevant factors and their interactions with each other. In this case, researchers and modelers must work with a methodology that can allow working with scarce data. FCMs are able to operate with unstructured knowledge and the casual relations [22].

The AB models and FCMs have wide ranges of application areas separately due to different advantages that they provide. Therefore, they have been successfully implemented through different disciplines from business to health research areas. The practicabilities of ABMs and FCMs to the social systems and their capability to handle complexity have caused to think about integrating these methodologies in one model in order to get benefits of both methodologies. However, the hybrid usage of both methodologies is a relatively new concept when compared to their single implementation. That's why; there is limited number of study that stress on the hybrid usage of these methodologies. The integration styles of both methodologies can differentiate depending on their implementation procedure.

In the relevant literature, researchers have used several ways to integrate both methodologies. One way to integrate the methodologies is to describe the nodes of a FCM with the help of agents. The study of Stula et al. [23] and Lee et al. [24] are two important applications of models that used the agents as node agents. Stula et al. [23] named their new model as agent based fuzzy cognitive map (ABFCM). The ABFCM was basically used the classic view of FCMs but supported this classic view with converting the nodes of the FCM as agents. This newly added characteristic of the model was used to enable each concept to apply different inference mechanism. Thus, the model can use different algorithms for the new state calculation of each concept. Lee et al. [24] proposed a multi-agent based FCM which was used for marketing planning problems. Similar to the previous work, they used agents to represent concept nodes of the FCM in order to eliminate the weaknesses of traditional FCMs.

While reviewing the hybrid usage of AB models and FCM, it is required to tell the contributions of Dr. Giabbanelli [25] and his colleagues. They performed valuable researches about the relevant literature. Giabbanelli [25] prepared a PhD thesis about understanding the complexity using the AB modeling and FCMs. He presented a framework that supports modelers to express how social influences are mediated by individuals and represents how FCMs can be used for representing mental models of people. Subsequently, Giabbanelli et al. [21], discussed two important ways of integration of both methods as seen in Figure 2.3. One way was to embed an FCM within each agent and the second way was to embed agents in one FCM. The first way was used to represent mental models' of agents while the second one is used to extend selected factors of a FCM. The first way of the mentioned study was one of the important inspirations of this thesis. In this PhD thesis, FCM was used to represent the brain activities and the decision making procedures of the agents. This kind of application is really rare and there is an important gap to represent different implementations of it.

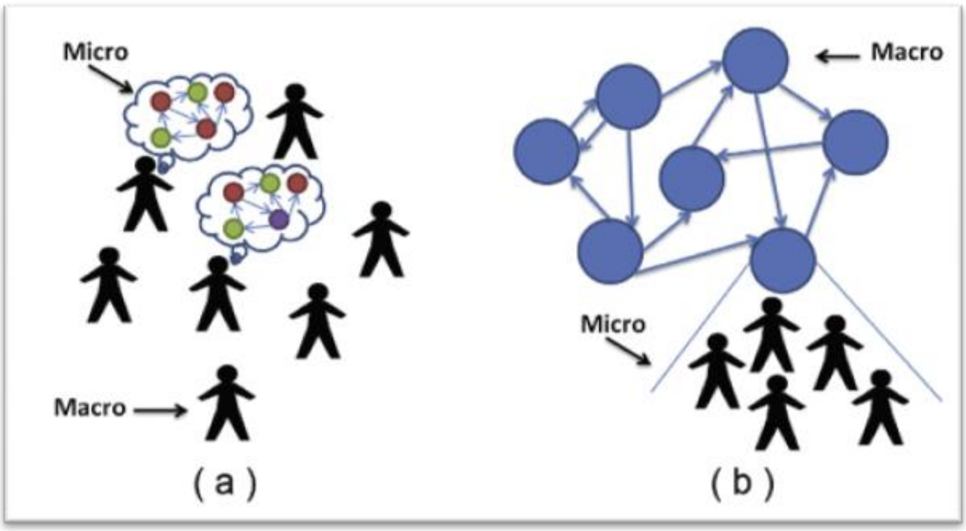

Figure 2.3 Two different ways of combining FCM and AB model: a) representing each agent's mental model with an FCM b) extending selected concepts of an FCM using an ABM [21].

In other study, Giabbanelli et al. [26] helped to develop a software to use ABM and FCM together. That software is one of initial examples prepared for hybrid use of the methods. Their software aimed to build heterogeneous mental models for the agents. Thus, agents could use different FCMs to increase the flexibility and heterogeneity. There are few studies that use FCMs as metal models on different research areas apart from Giabbanelli et al. [21] mentioned. Grass et al. [27] build an individual based predator prey model that used FCM to represent agent's behaviors. Khater et al. [28] focused on modeling the genomes of organism with the individual based evolutionary ecosystem simulation by representing their behavioral traits with FCMs. In addition, Mehryar et al. [29] integrated ABM and FCM in order to simulate different policy options on farming community that faces water scarcity. Instead of using one FCM for all farmers, they grouped the farmers as small, medium, and large according to some factors such as the size of their lands. This grouping procedure was used to protect the diversity at the mental models.

A summary of the researches that focused on hybridization of AB modeling and FCM to represent mental models of agent are presented in this section. Different from the review performed in this thesis, Davis et al. [30] presented a detailed literature review that explain all possible ways of hybridization of both methodologies and the studies that applied one of these ways in detail.

2.4 Concluding Remarks

The summary of the researchers that focus on representing the mental models of the agents with FCMs were given in this chapter. In the light of the given literature, the main objective of this PhD thesis is to integrate ABM with a multi-stage learning based FCM to obtain a novel hybrid model. Tobacco use and control research field is selected to implement the proposed model in this thesis because tobacco use system is a complex and dynamic social system that consists of high variety of factors such as characteristics of people, tobacco control policies, social networks of people and etc. Although tobacco use is a well studied system, the prevalence of tobacco use brings out the necessity of more advanced studies for understanding tobacco use behaviors comprehensively. Therefore, developing an advanced model for tobacco use system becomes an important motivation for this thesis.

Although there are a few examples of hybrid usage of these methodologies, this thesis differentiates from them by several points. FCMs are used to reflect decision making mechanism of the agents in this thesis. For this aim, a multi-stage learning procedure is implemented in order to restrict the usage of expert knowledge that can cause subjectivity and decrease the accuracy of the studies. In order to overcome these problems, a two-stage learning procedure is used to train the maps with the help of the non-linear Hebbian algorithm (NHL) and extended great deluge algorithm (EGDA). To our best knowledge, there is not a hybrid AB model that implements a multi-stage learning FCM and tested on a real-life problem in the literature. Thus, this thesis fills this gap and contributes to the literature at this point.

CHAPTER III

FROM AGENT BASED MODELING AND FUZZY COGNITIVE MAPS TO HYBRID MODELING

3.1 Introduction

The motivations of this PhD thesis and summary of the relevant literature have been provided through Chapter 1 and Chapter 2. Different from these chapters, Chapter 3 aimed to provide a theoretical background for Agent Based (AB) modeling and Fuzzy Cognitive Map (FCM) and their integration. The capabilities, fundamental terms and concepts, advantageous, disadvantageous and the application procedures of both methodologies are explained systematically. The reason why integration of both methodologies is required is discussed. Finally, the statement purpose of the thesis is provided.

3.2 Agent Based Modeling

Agent Based (AB) modeling is one of the fundamental simulation approach. It is relatively new when compared to Discrete Event (DE) and System Dynamics (SD). [4]. The concept of AB modeling was firstly proposed in the 1990s [14] , [31] , [32] although it takes its roots from micro simulation. From that time on, it has been applied in several different research areas including energy, biology, and etc. [33]. AB modeling has also promised a revolutionary advance in social science theory [34]. Computational advances from simulating agent behavior have made possible big amount of AB models across a variety of application domains [35].

AB modeling is a mindset that consists of describing a system from the perspective of its units [16]. It simulates complex systems that allows the researcher investigating both the potential and sources of emergent properties [36]. It is a kind of computational simulating tool that uses autonomous agents that live, move, and interact with each other on a virtual world [37].

As Borshchev and Flippov stated [4] "*AB models are decentralized. Compared to SD or DE models, there is no such place in AB model where the global system behavior would be defined. Instead, the modeler defines the behaviors at individual level , and the global behavior emerges as a result of many (tens, hundreds, thousands, millions) individuals, each following its own behavioral rules, living together in some environment and communicating with each other and with the environment. That is why AB modeling is also called bottom-up modeling.*" Thus, unlike other traditional methodologies, AB modeling provides the opportunity to examine the heterogeneity and interdependencies.

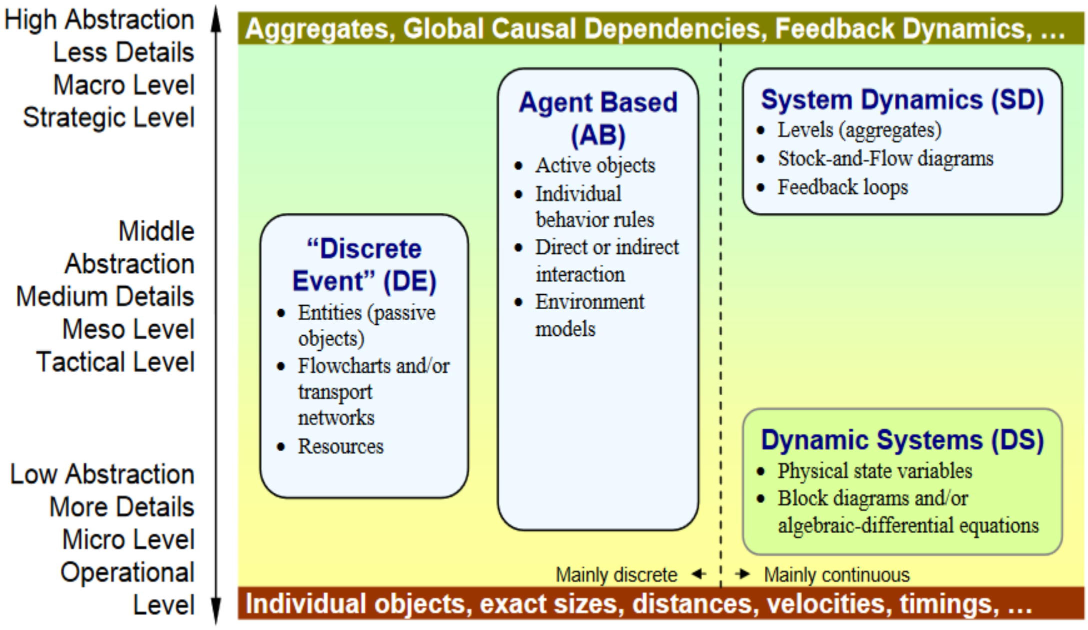

Figure 3.1 Approaches in simulation modeling on abstraction level scale [4].

In order to understand AB modeling in detail, it is important to figure out the term "agent". In the next sub-section, the information for agent term is presented.

3.2.1 Agent

AB modeling is a valuable tool to examine complex and dynamic systems using individualistic approach with the help of agents. Agents are basically entities of a model that acts on behalf of others while displaying autonomous behaviors [38]. Although there is not a universal consensus on the definition, it is accepted that agents have behaviors, learn from experiences, interact with and influence each other [39].

Agents can differentiate based on the types of problems/systems that they have been used for. While they represent individuals for a population in a population growth problem, they can represent money, automobiles, machines, data, patient, virus, ship and etc. for different problems. Thus, the agents of a system are determined considering the fundamental object of a system or a problem.

According to the application procedure, agents can have different properties. However, it is important to state that objects should have essential properties to deserve to be called as "agents". Macal and North [35] and Macal [39] presented certain properties of an object to qualify as agent:

- *Being autonomous*: An object should be autonomous to be called as agent that can behave independently in its own environment.
- *Being modular or self contained*: An agent should be identified with a set of characteristics/attributes, behaviors and decision making capabilities. Thus, an agent is a uniquely identifiable entity.
- *Being social*: An agent should interact with other agents. Agents have procedures that identify how they interact with each other.
- *Having a state*: An agent should have a state that represent the essential variables related with its current situation and varies over time. The state of an agent generally consists of a group of subgroup of its attributes.

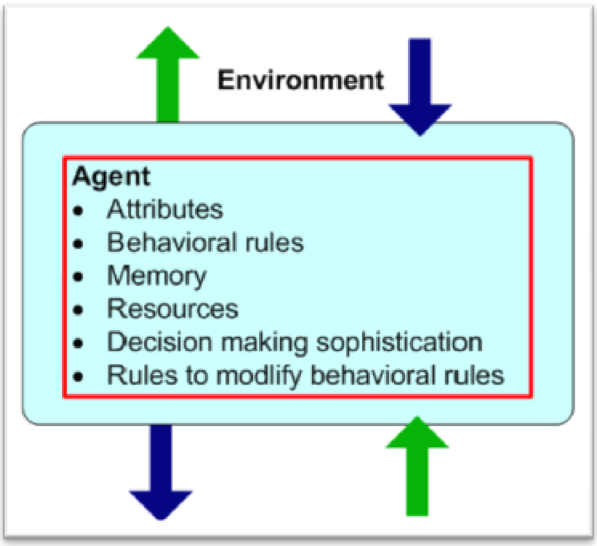

Figure 3.2 An example for a typical agent [35].

In addition to these three criteria, agents may have a specific goal, have an environment, have resource attributes, learn and adapt its behaviors from its experiences [35]. Figure 3.2 represents an example for a typical agent.

All mentioned characteristics that an agent should/may have, help us to visualize a complex and dynamic real life problem in detail when compared to most traditional techniques. The usage of agents in the modeling provides significant advantageous to modelers:

- The full effects of diversity that agents represent with respect to their characteristics and behaviors can be examined by modeling agents individually [39].
- AB modeling can provide us better understanding real world systems where the representation of individuals is important [14].
- AB models are able to demonstrate emergent phenomena [34].
- AB models can clearly model the complexity arising from individual behaviors and interactions that happens in the real life [14].
- AB models provides flexibility with the help of agents [16].
- AB models allows modelers to handle real world systems or problems that were either not possible or not readily accommodated using traditional modeling tools such as DE and SD [14].

Due to the benefits of AB modeling, it has been implemented for several different real life systems and problems such as economy, production, logistics and etc. to gain better understanding. Especially, the growing acceptance of AB modeling in various fields of social sciences draws the attentions. The most important reason behind the enthusiasm for AB modeling is the dissatisfaction with the restrictions imposed by alternative models (differential equations and statistical models) [34]. The AB modeling provided a different perspective to social fields' modelers. When considering the complexity and dynamism of the social problems due to the individualism and interactions, the potential contributions of AB modeling to the relevant area is remarkable.

3.3 Fuzzy Cognitive Map

Fuzzy Cognitive Map (FCM) is one other methodology that draws attentions by its capability to handle complexity with ease. FCM can model the dynamics of complex systems by incorporating the casual relations between the concepts [40]. FCMs attracts the interest with different appealing properties [41] like describing the system behaviors in a symbolic manner, illustrating the entire system by a graph showing the cause and effect relationships among concepts and integrating knowledge and experience with the operation of system [20].

It was proposed by Kosko in 1986 [42] by extending the idea of Cognitive Maps (CMs)[43]. FCMs consist of directed graphs [44] with consist of nodes and edges as shown in figure 3.3. Nodes represent the concepts (such as factors, variables, attributes and etc) while edges represent the relations between the related concepts. Therefore, C_i and W_{ij} represent the concepts and the relationships among the concepts respectively in the Figure 3.3. If there is N number of concepts in the map, the values of each concept can be expressed with a state vector called as A with N dimension. The values of the concepts can range between -1 and +1 or between 0 and +1 based on the types of the studied problem.

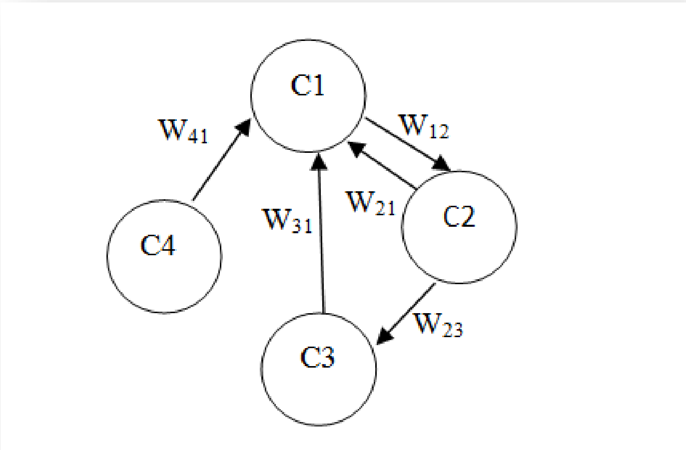

Figure 3.3 An example for a simple FCM [22].

The relationships between the concepts (W_{ij}) show the strength of the relations. The main difference of the FCMs from CMs is that FCMs can also present the strength of the relationships with the help of fuzzy logic while CMs can only give information about the direction of the relations [22]. The strength of the relationships can take values between -1 and +1. A negative value in the relation (W_{ij}) means that any increase or decrease in concept i cause an adverse effect in the concept j. A positive value in the relation (W_{ij}) indicates that any increase or decrease in concept i will be resulted with similar effect in the concept j. If there is no relation between concepts, the W_{ij} will take the value, 0. The relations of the concepts are represented with an N×N matrix (weight matrix) in the FCM procedure. The strength of the values of the relations is shown with the help of the weight matrix.

The design process of the FCM starts with the initial data. Modeler must gather some information such as the number of concepts, initial values of these concepts and initial weight matrix of the relations. After collecting the required data, all concepts becomes activated and initialize to influence each other [44]. At each iteration, new values of the concepts are calculated using formulation proposed by Dickerson and Kosko which is given in Eq. 3.1 [45], [46].

$$A_i^{k+1} = f(A_i^k + \sum_{j=1(i \neq j)}^{n} W_{ji} A_j^k) \quad (3.1)$$

In this formulation, k shows the number of iteration and f indicates the threshold function. The threshold function helps to bound the transformation to a limit cycle [47].

The interaction process goes on until FCM reaches one of the following states [48]:

- A fixed point where the stabilization for the values of vectors occurs
- A limit cycle in which the concepts are falling in a loop.
- A chaotic behavior where concepts' values vary with iterations.

3.3.1 Learning Algorithms

FCM researches mostly initializes with the knowledge provided by experts who are expected to identify the key concepts (that describe the behaviors of the system), the relationships between the concepts, and the strength of these relationships [49]. This mechanism makes the design process of FCMs highly dependent to the knowledge of experts [22]. The dependency to the expert knowledge can cause weaknesses because expert knowledge may contain subjective reasoning that may restrict the accuracy and the reliability of the maps [45]. In order to eliminate the problems sourced by the usage of expert knowledge, make the system converge in the desired regions, and improve the robustness various learning algorithms are used in the analysis [20], [50], [48].

Learning algorithms fundamentally update the initial knowledge [51] and modify the relationships [48] to reach appropriate relation matrix. The learning algorithms can be categorized in two main classes [40], [22]:

- Algorithms that use Hebbian rules
- Algorithms that use evolutionary optimization procedures

Hebbian learning rule is accepted as one of the simplest learning procedure for two layer networks [52]. It determines how much the relations among two concepts will be modified in proportion to their activation [50]. The initial research of FCMs with Hebbian rule was proposed by Dickerson and Kosko in 1994 and was called as Differential Hebbian Learning (DHL) [41], [46]. This algorithm was extended various times in the literature. Initially, Huerga [53] proposed the Balanced Differential learning algorithm (BDL) in 2002. The BDL was tested with four different training sets and provided promising results when compared to DHL. Following to the BDL, Papageorgiou et al.[54], [55] presented two different versions of it: the Nonlinear Hebbian learning (NHL) and Active Hebbian learning respectively. The NHL algorithm basically stresses on finding a set of interconnection weights that minimizes the pre-determined cost functions while the AHL algorithm focuses on activating the concepts asynchronously by a specific sequence to reach the equilibrium point [40], [22]. The common vital properties of both learning algorithms are their high dependency to the initial weight matrix provided by experts [22]. Therefore, the false estimations of the experts during the identification of the initial weight matrix can affect the expected efficiency of these learning algorithms [45]. The other learning algorithm where stresses on the Hebbian rule category is the Data Driven Nonlinear Hebbian learning algorithm (DD-NHL) that was an extended version of the NHL algorithm that use historical data to improve the quality of the learned FCMs [41]. Papakostas et al. [40] performed a detailed comparison analysis for DHL, NHL, DD-NHL, and AHL algorithms in 2011. They implemented these algorithms on three different problems: a process control example, a chemical plant example, and a heat exchanger process example in order to measure the performance of them. The findings showed that the performance of the NHL algorithm was satisfying to control the output variations by keeping them in desired regions. The DD-NHL followed the NHL with similar results while the DHL and AHL provided the worst results among all. A brief history for the development of the versions of the hebbian learning rule was given here. It can be seen that the hebbian based algorithms are widely used in the FCM studies for learning purposes although there are some weaknesses [22].

The second category for the learning algorithms is evolutionary algorithms. The evolutionary algorithms started to appear in FCM studies in order to avoid the deficiencies of the Hebbian-like algorithms [22] because although Hebbian-like algorithms are relatively fast to converge to the desired FCM states and very common in FCM studies, their performance are highly related with the initial weight matrix and predefined structure of FCMs [56]. In contrast, the evolutionary algorithms try to ensure near-optimum solutions for the relation search mechanism [40]. For that reason, several different algorithms have been used with FCMs for various problems.

Koulouriotis et al. [57] used an evolution strategy to train causal relationships while Papageorgiou et al.[58], [59] implemented Particle Swarm optimization (PSO) for an industrial process control problem in 2004 and 2005 respectively. Song et al. [60] also proposed the application of multi objective particle swarm optimization for FCM training and implemented the algorithm for modeling the mental and physical behaviors of an emotional agent in a virtual world. On the other hand, Stach et al. [61] performed an analysis with Genetic Algorithm (GA) to construct the FCM and demonstrated the effectiveness of the evolutionary algorithm while Ghazanfari et al. [62] compared the performance of Simulated annealing (SA) and Genetic Algorithm (GA) for the learning process. In addition, Alizadeh et al. [63] used tabu search algorithm for the training in 2007 and Baykasoglu et al.[45] for FCM analysis to analyze the performance of the algorithm for two different problems: an industrial process control problem and a job shop problem. They provided promising results for the studied algorithm. In 2014, Durmusoglu et al.[64] also used the EGDA to train the FCM for analyzing the factors affecting tobacco use among students. In 2015, Ahmadi et. al [65] proposed a new automated FCM learning algorithm which is called as Imperialist Competitive Algorithm (ICA) using historical data. They also tested the accuracy and execution time of it. The proposed algorithm showed satisfying results.

3.4 Integration of Agent Based Modeling with Fuzzy Cognitive Map

The theories and potential advantages of Agent Based (AB) modeling and Fuzzy Cognitive Map (FCM) was provided in the previous sub-sections of this chapter. In the light of the information given above, it can be stated that AB modeling and FCMs are two powerful techniques that can handle the complexity of the real life systems. Although both of them use different approaches to model the systems, the benefits of both methods cause widespread use of them in the modeling literature.

The major difference of these two modeling methods is the views that they use for modeling the systems. AB models generally examine the systems with a micro-level view while FCMs uses a macro-level view. Both views provide to the modelers different vital advantageous during the modeling processes.

In micro level views, macro level patterns emerge through the micro level behaviors. In order to build an AB model, the vital object(s) of the system that is called as agent is determined and the model is created with the collection of agents. These agents are autonomous, interact with each other, and have different properties and state. Thus, using agent provides heterogeneity. In addition, AB modeling allows socialization. Most of the traditional techniques cannot represent the interactions inside a system. However, interactions can be vital for many real life systems. For example; examining an epidemic without interactions can cause weaknesses for a study because many people can be infected during their interactions with others. However, AB modeling can help to represent the interactions in a virtual world. The modelers can examine how different agents behave, affect others, and are affected from others. Therefore, AB modeling helps modeler to look inside to a system with a magnifying glass.

On the other hand, the macro level view examines the system as a whole. FCMs are a good example for aggregate modeling techniques. FCMs generally list the relevant factors, identify which ones have an effect on others and help to measure the strength of these effects. In FCMs, the level of a concept can change across iterations depending on the factors that affect it and the strength of the relations [30]. Thus, FCMs examine a system by determining the factors that affect the systems and the relations between these factors. The elicitation processes of an FCM is relatively intuitive because it requires neither hard data nor modeling expertise, especially compared to other macro-level view approaches such as System Dynamics [30].

Although micro and macro level views provide different gains to AB models and FCMs, they can still have different weaknesses separately and cause some problems to the modelers during modeling processes. The capabilities of AB modeling such as representing heterogeneity and interactions between the agents were mentioned above. AB modeling is basically modeling techniques that try to examine the all pieces of a puzzle. However, when the system becomes more complex, the requirement of data can increase too. Providing data for the characteristics, behaviors and networks of each critical individual can be a problem. Especially, systems like social systems that consist of various characters and many interactions may be difficult to model with AB modeling due to the lack of information. In contrast to AB modeling, while FCMs can work with the limited number of data compared to AB models, they cannot represent heterogeneity and socialization because they only stress on creating generic maps to represent the entire system. This situation was stated as "*the FCM methodology can provide the agents' brains, but brains cannot directly interact: they must be places inside bodies*" [26], [30].

In order to overcome the weaknesses of both methodologies, the integration of them was discussed in the literature. Although several different hybridizations can be found, the hybrid usage of AB modeling and FCM is relatively new. The potential of their hybridization is the starting point of this PhD thesis. The problem statement and the main purpose of the thesis are given in the next section.

3.5 Statement of Purpose

"Curiosity to unknown" has generally formed the basis for most of the pioneering inventions and researches. Desire to understand the behaviors of the complex and dynamic systems have directed researchers to question the existing methodologies and explore for better ones. Eliminating the weaknesses of the current methodologies and examine the real life systems more realistically can be the one of the causes of hybrid techniques. The integration of AB modeling and FCMS can be a good example hybrid models that tries to improve the current states of the techniques and reach more flexible methodology.

In the light of the information given, this PhD thesis began with the question "Is it possible to build a model that is flexible enough to represent complex and dynamic social systems?". For this aim, a novel hybrid model which is developed by embedding the concept of a multi-stage learning based FCM into the AB model in order to benefit from advantages of each methodology. The AB model is used to represent individual level behaviors while the FCM is used as a decision support mechanism for individuals. Although a few applications of the integration of AB model and FCMs have been seen in the literature, there is still a huge gap for hybridization of both methodologies especially for social systems that consist of many different factors and interactions. The presented model is built for obtaining better understanding on tobacco use and control in Turkey. The main reason behind to study on tobacco use and control is that tobacco use is increasingly recognized as a complex adaptive system involving biological, behavioral, and environmental influences" [66].

Tobacco use is a complex and dynamic social system that consists of high variety of factors such as characteristics of people, tobacco control policies, social networks of people and etc. Although tobacco use can be accepted among the well studied system in the world, the prevalence of tobacco use brings out the necessity of more advanced studies for understanding tobacco use behaviors comprehensively. Besides, the study of Unutmaz Durmuşoğlu and Kocabey Çiftçi [10] revealed that the study performance of Turkey is not as the expected level according to the number of publications on tobacco use. This result indicated that there is still not enough research to understand tobacco use behaviors of Turkey. This thesis aims to examine these behaviors with the help of the proposed hybrid model. This thesis can be a pioneering example to researchers to implement more advanced modeling techniques and help to understand tobacco use system of Turkey comprehensively.

In this thesis, FCMs are used for representing the mental models of the individuals of Turkey. The characteristics of the individuals such as gender, age, residence, education level, work status and implemented tobacco control policies such as taxation, warning labels and etc. are embedded as concepts to the FCMs. The casual relationships and the strength of these relationships are examined and likelihoods to tobacco use of individuals are found as the result of FCM analysis. Most of the FCMs are generally initiated with expert knowledge to obtain information about the concepts and their possible relations. In this thesis, the usage of expert knowledge is restricted by using random initial relation matrix. Then, a two-stage learning procedure with the non-linear Hebbian algorithm (NHL) and extended great deluge algorithm (EGDA) for tobacco use and control of Turkey is implemented and presented by Kocabey Çiftçi and Unutmaz Durmuşoğlu [22] is used to train the mental models of the individuals.

In the presented hybrid model, the mentioned multi-stage learning procedure based FCMs are integrated to decision making process of agents. Therefore, the FCMs of this model use both algorithms subsequently while calculating the likelihood of tobacco use of agent at each step. It helps to increase the accuracy and robustness of the mental models of the agents in the model with this implementation.

Apart from all these, the proposed model also helps to examine the potential impacts of a new tobacco control policy (plain package) and new pandemic COVID19 on the prevalence of tobacco use for Turkey with hypothetical scenarios. The results of scenario analysis can help policy makers to see groups that could be more resistant to quit or more vulnerable to start smoking. Thus, future policies can be developed benefiting from the information supplied by these results.

The aimed model is prepared using the AnyLogic Simulation Software. The details of the software are presented in the next sub-section.

3.6 AnyLogic Simulation Software

AnyLogic is one of the leading simulation software that helps to gain insights complex systems across a wide range of industries. It was developed by XJ Technologies. It can be used for developing three well known simulation methods: Discrete Event (DE), System Dynamics (SD) and Agent Based (AB) in the same environment as seen in Figure 3.4. It is also known as the first tool to perform multi-method simulation modeling. It has been used in several different industries such as manufacturing, mining, transportation, supply chains, marketing, healthcare and etc.

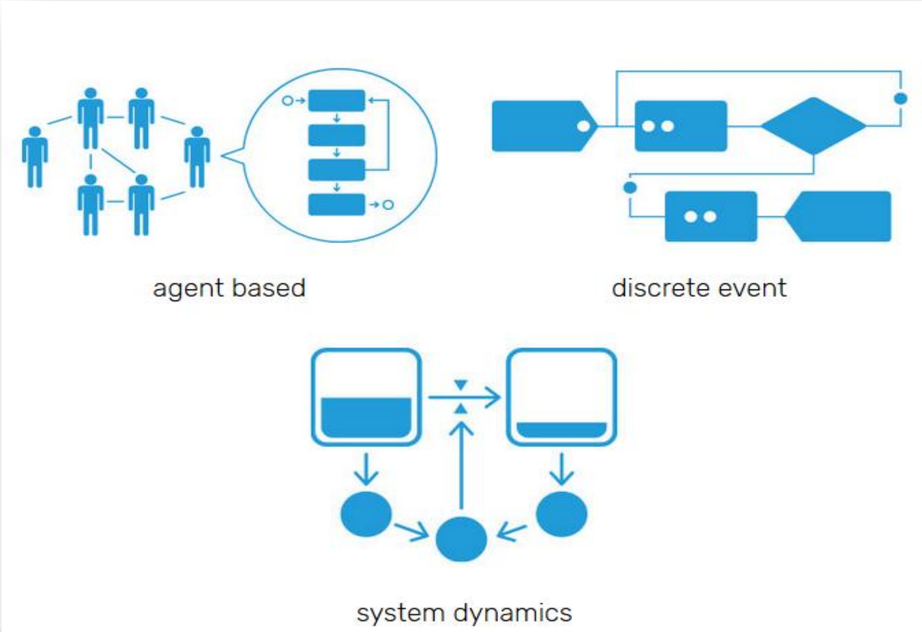

Figure 3.4 Simulation modeling techniques in AnyLogic.

The AnyLogic 7 University (version 7.0 0) is used in the presented thesis. The main reasons to model with AnyLogic in thesis can be listed as follows:

- It allows building hybrid models in any combinations due to the capability of multi-method simulation modeling
- It allows writing your own JAVA codes to the model. This provides remarkable flexibility to the modelers. This property of the software is really crucial to perform this thesis because all process relevant to the FCMs were written in JAVA separately and embedded to the software with the help of AnyLogic.
- It has a well-designed user interface to model that makes the modeling process easier.
- It has several different libraries such as pedestrian, logistics and etc that consist of relevant processes. The libraries help to accelerate the modeling process.
- It can perform several different additional analyses such as sensitivity analysis, calibration, optimization and etc.

3.7 Concluding Remarks

Agent Based (AB) modeling and Fuzzy Cognitive Map (FCM) methodologies have been two important methodologies to model complex and dynamic systems. Although they use different approaches to model the systems, both of them provided significant advantageous to the modelers. However, they may have different weaknesses due to their approaches. In order to overcome their weaknesses, integrating both methodologies can help modelers to build more realistic and flexible models to represent real life systems. This chapter provided the theoretical background for methodologies, their integration and statement of purpose for this thesis.

CHAPTER IV

A MULTI-STAGE LEARNING BASED FUZZY COGNTIVE MAP FOR TOBACCO USE AND CONTROL

4.1 Introduction

Modeling and imitating the complex and dynamic real life systems is a challenging process. When the number of factors and their relations increases, the complexity can grow, too. In order to handle these kinds of systems, researchers use different techniques such as Fuzzy Cognitive Maps (FCMs). FCM is a kind of graph-based knowledge representation that presents the casual relationships between the concepts [40]. It generally helps to analyze the behaviors of a system utilizing expert knowledge or available knowledge from existing databases [67]. Therefore, most of FCMs are initialized with knowledge provided by experts such as concepts, relations and etc. Experts are forced to think about relations among the concepts and justify their bids to transform their knowledge in to a FCM [20], [22]. The design process of a FCM increases the dependency to the expert knowledge and can cause to the subjectivity due to the usage of experts' opinion. FCMs try to eliminate the subjectivity and the dependencies to the expert knowledge using learning algorithms that can modify the relation matrix to reach a suitable one. Several different learning algorithms have been embedded to the FCMs in the literature. An overview for FCM and learning algorithms were provided in the previous chapter.

In this thesis, a multi-stage learning based FCM is designed for examining the causal relationships between demographic characteristics of people and likelihood of tobacco use for Turkey. For this aim, algorithms from two different categories are embedded into the same FCM. The non-linear hebbian learning algorithm (NHL) from hebbian like algorithms and extended great deluge algorithm (EGDA) from evolutionary algorithms are selected in order to gain the benefits of two algorithms categories. This combinatory usage of the algorithms was firstly proposed by Ren [56].

Besides, the need for expert knowledge is restricted and the FCM is initialized with a random initial weight matrix in order to minimize the subjectivity. Although using a random initial weight matrix may affect the performance of the study, FCM is trained using a multi-stage learning procedure with the help of two different algorithms (NHL and EGDA) [22]. The first stage of the learning process is used for reaching a reasonable weight matrix while the second stage of the learning process is used for increasing the robustness of the study [22].

The details of the performed and published FCM analysis were presented in this chapter.

4.2 The Framework of the Proposed FCM

This part of the thesis includes the two-stage learning based FCM for understanding the relations between demographic characteristics of people and likelihood of tobacco use of them. For this aim, the FCM analysis is categorized in two main steps:

1. Data collection and preparation processes
2. The design processes of FCM analysis

In the first step, the required data such as demographic characteristics of people and likelihood of tobacco use to perform a FCM analysis is borrowed from the Global Adult Tobacco Survey (GATS) Turkey 2008. The dependency between the selected demographic characteristics and tobacco use status are analyzed using Chi-Square tests.

Subsequently, a K means clustering analysis is performed using the data of the GATS Turkey 2008 to categorize the participants of the survey according to their similarities. Clustering the participants of the survey helps to collect the similar characteristics of the participants in the same cluster. Thus, making a FCM analysis for the clusters increases the accuracy of the FCMs.

In the FCM design stage, firstly restricted initial expert knowledge such as input and output concepts are generated for each cluster and then the random initial weight matrixes are created. After that, the initial weight matrixes are trained with the help of NHL algorithm and the output relations of the NHL algorithm are accepted as the input of the second stage learning with the EGDA as seen in Figure 4.1.

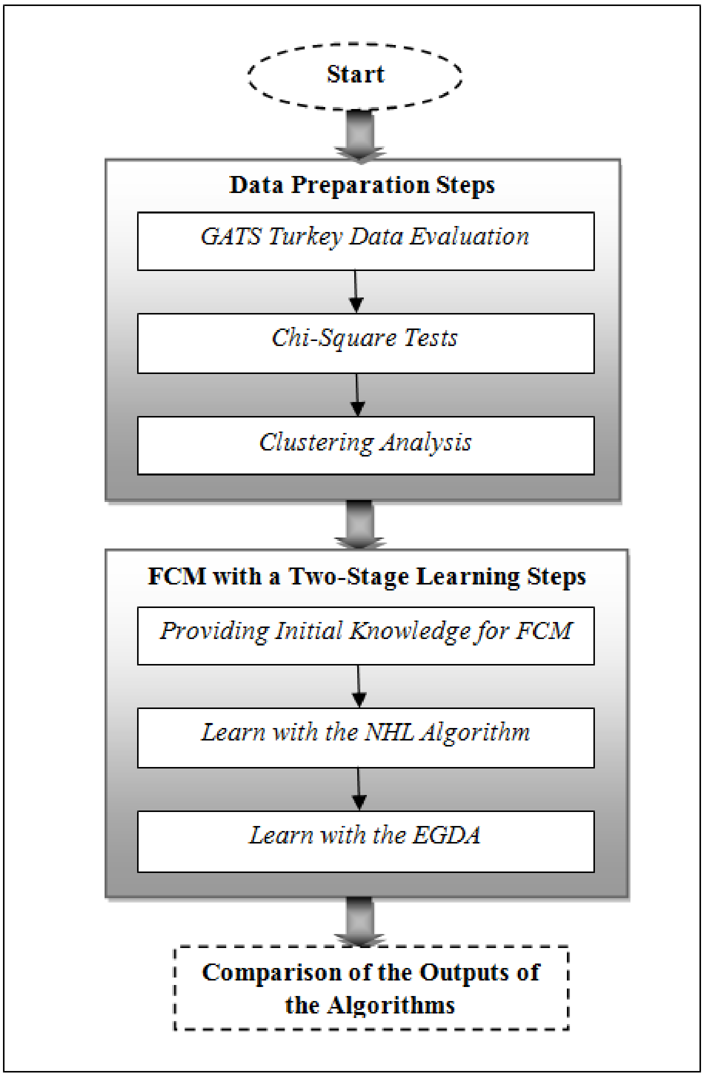

Figure 4.1 The framework of the proposed FCM [22].

4.3 The Data Preparation

The main data source of the proposed FCM analyses is the Global Adult Tobacco Survey (GATS) Turkey 2008. The GATS is a national household survey that was recommended by the World Health Organization (WHO). The major objective of the surveys is to help countries to collect data on adult tobacco use and key tobacco control measures [68]. Several different countries of the world have been applied this survey periodically to their nations in order to monitor the prevalence of tobacco use and the effects of their activities against tobacco use. Turkey have been one of these countries and applied the survey 3 times until now in: 2008, 2012, and 2016.

The surveys include various questions to the participants who are in the adult age category (aged 15 or over) from different categories such as:

- The main characteristics of the participants such as gender, age, residence, and etc.
- Their current tobacco use status (never smoker, current smoker, or former smoker) and their tobacco relevant history.
- Their exposure to the second hand smoke at various places such as at work, at home, and etc.
- Their awareness to the implemented tobacco control measures such as warning labels, taxation and etc.
- Their perception about the harms of tobacco use and etc.

In this part of the thesis, the data provided by GATS Turkey 2008 is used to examine the relationships between likelihood of tobacco use of people and the main demographic characteristics of them. GATS Turkey 2012 and 2016 data are used at the integration of AB modeling and FCM part of this thesis. Thus, a total of 9030 people were participated to this national tobacco use survey.

In order to perform the analyses, information about the demographic characteristics of each participant and their current tobacco use status are extracted from the survey. The extracted data and the types of the responds of the people are listed below.

- *Gender:* The one of the first demographic characteristics that was collected by the survey is gender knowledge.
  - Male
  - Female
- Age: The participants of the survey must be aged over 15 or at least 15.
- Residence: This category represents the information about living place of the participants. The answers of the participants can be:
  - Urban
  - Rural

- Education Level: In this category, the participants are questioned with "What is the highest level of education that you have completed?". Their answers can be as follows:
    - Not graduated
    - Elementary school
    - Primary school
    - Secondary or vocational secondary school
    - High school or equivalent
    - College or Faculty
    - Master/Doctorate
    - Don't know
    - Refused to answer
- Working Status: In this category, the participants are questioned with "Which of the following best describes your main work status over the past 12 months?". Their answer can be as follows:
    - Paid employee
    - Self employed/employer
    - Non-paid family worker
    - Student
    - Home maker
    - Retired
    - No job-able to work
    - No job-unable to work
    - Don't know
    - Refused to answer
- Tobacco Use Status: In this category, the participants are questioned with "Do you currently smoke tobacco on daily basis less than daily, or not at all?". Their answer can be as follows:
    - Daily
    - Less than daily
    - Not at all
    - Don't know
    - Refused to answer

A detailed data adjustment process is performed because survey could not be able provide enough data for all categories. In order to transform the data more meaningful, adjustment process is really vital. The details of the adjustment process are given below.

- Initially, a total of 12 participants who answers to the questions like "Don't know" or "Refused to answer" are eliminated.
- The World Health Organization (WHO) categorizes adult population as participants among "15-24 ages", "25-44 ages", "45-64 ages", and "65 and over" and presents the survey reports for each age category. For that reason, the ages are grouped following the categorization of the WHO in this thesis. The answers "No job-able to work" and "No job-unable to work" categories are combined because there was only 7 participants in the second unable to work category. This amount is really few to analyze separately.
- The answers of all categories are converted to a numeric representation.
- A dataset is created by including the information of the participants.

After the adjustments, the dependency of the collected demographic factors to current tobacco use status of participants in the created dataset is searched to understand if there is a significant relation between them [22]. For this aim, independence tests using Chi-Square are performed.

Chi-Square test is generally used for performing an independence test if two categorical variables are related in some population [69], [22]. The null hypothesis of the test means that there is no relation between the examined categorical variables in the studied population so that they are independent [70], [22]. The formula of the Chi-Square is as provided in Eq. 4.1[13]:

$$Xo^2 = \sum \frac{(Oi - Ei)^2}{Ei} \tag{4.1}$$

where O_i and E_i presents the observed and expected frequencies respectively [22]. The planned Chi-square tests are performed using the STATISTICA 8.0 software. All 5 demographic characteristics of participants and their relations with the tobacco use are tested in these analyses. Firstly, the following hypotheses are built for each test. An example for the hypothesis is as follows [22]:

- *Ho (null hypothesis): The factors "tobacco use status" and "gender" are independent*
- *H1: The factors "tobacco use status" and "gender" are dependent*

The p-values are found as the result of each test. All of the p values are under 0.05 as seen in Table 4.1. Thus, null hypotheses should have been rejected. The findings indicates that, all demographic factors are significantly dependent to the tobacco use status for the created database [22]. Therefore, all examined characteristics are included to the next analysis.

Table 4.1 The p-values of the 5 different chi-square tests [22]

	Gender	**Age Range**	**Residence**	**Education Level**	**Working Status**
Tobacco Use Status	0.0000	0.0000	0.0000	0.0000	0,0000

In the final step of the data preparation, the participants of the survey are clustered based on the demographic characteristics by performing a clustering analysis. Instead of generating one common map for 9018 participants, finding the maps of each group (where participants exhibit similar characteristics) can be more meaningful and interpretable [22]. Because finding all possible combinations of demographic characteristics and tobacco use behaviors of participants from the survey are almost impossible. There is not enough information for all demographic cases in the surveys. For that reason, collecting participants to clusters according to similarities helps to perform intended analysis more accurately. Clustering is one of the popular techniques in data mining with real life applications [71], [22]. The main objective of the clustering is to find groups in which those belonging to the same group are similar while those belonging to different groups are dissimilar according to some criteria of distance or similarity[72], [22]. If the data in the studied dataset are totally independent, there would be as many clusters as the observations [73], [22]. In order to eliminate this problem, Chi-Square Independency Test that is performed for examined factors.

Therefore, K-means clustering algorithm is chosen in order to make a clustering analysis. K means clustering basically selects a desired number of clusters (k) and initial observations as seeds of clusters and then assigns each of the other training observations to the closest cluster [73], [22]. The clustering analysis is performed by using STATISTICA software that allows to use of v-fold cross validation instead of determining the number of clusters previously. V-fold cross validation simply divides the overall dataset to v folds, then uses the v-1 folds for training and remaining for testing the validity [74], [22].

Consequently, a K-means Clustering is performed using 10-fold cross validation. The initial centers of the clusters are maximized and the distance of the each case to the clusters is calculated using the Euclidean distances [22]. The findings of the K-means clustering analysis provides 5 clusters for 9018 participants as seen in Table 4.2.

Table 4.2 The members of clusters [22]

Factors / Clusters	Gender	Age Range	Residence	Education Level	Work Status
Cluster 1	Male	25-44	Rural	High School	Paid Employee
Cluster 2	Female	25-44	Urban	Elementary School	Home Maker
Cluster 3	Male	45-64	Urban	Elementary School	Paid Employee
Cluster 4	Female	25-44	Rural	Elementary School	Non-Paid Family Worker
Cluster 5	Male	45-64	Rural	Elementary School	Self-Employed

4.4 The Design of the Multi-Stage Learning Based FCM

This section explains the design procedure of the multi-stage learning based FCM.

4.4.1 Data for Initialization

Most of the FCMs are initialized with expert knowledge. Experts generally provide the information such as key concepts, initial values of these concepts and the direction and the initial strength of the relations between the concepts. Designing the analysis using the expert knowledge completely can increase subjectivity while decreasing the accuracy and robustness of the analysis. In this thesis, the usage of expert knowledge is restricted. First of all, the key concepts are taken from the

GATS Turkey 2008 using Chi-Square independence test. Subsequently, the experts are only asked as "Is there a positive or a negative relation between concept i and concept j?" [22]. For example; "Is there a positive or a negative relation between *gender* concept and *age* concept?". If there is a consensus on the answers such as "yes" or "no", their common opinion is adopted to the weight matrix [22]. That means if there is consensus on the response "no relation", the weight value is equalized to 0 [22]. However, if there is an opinion apart from "no relation", the range of it is equalized to [-1,+1] [22]. Thus, expert knowledge is only used for identifying the ranges of the weights. After that, random initial weight matrix is created for analysis.

4.4.2 FCM Analyses

The FCM analysis begins with the determination of input data. After providing the required data such as concepts, relations and etc., all concepts becomes activated and starts to influence each other [44]. The overview for the FCM was given in the previous section. However, it is important to remind that at each iteration, new values of the concepts are calculated using formulation proposed by Dickerson and Kosko which is given in Eq. 4.2 [45], [46], [22] .

$$A_i^{k+1} = f(A_i^k + \sum_{j=1 (i \neq j)}^{n} W_{ji} A_j^k) \tag{4.2}$$

In this formula, k shows the number of iteration and f indicates the threshold function [22]. The threshold function tries to bound the transformation to a limit cycle [47]. In this thesis, the sigmoid function that helps to keep the values within the range [0, 1] was used. The formula of the sigmoid function is also given in Eq. 4.3 where λ determined the steepness and a positive value [50].

$$f_x = \frac{1}{(1+e^{-\lambda x})} \tag{4.3}$$

4.4.3 The Multi-stage Learning Based FCM

This thesis uses random initial weight matrix to perform FCM analyses. The random initial weight matrix is trained using a two-stage learning mechanism. For this aim, the NHL algorithm is used at the first stage to train the initial random map. The NHL algorithm is one of the widely used algorithm in FCM studies with its quickness and practicability [22]. Subsequently, the NHL algorithm is supported with a promising evolutionary algorithm to ensure a near optimum solution [22]. There are two important reasons to work with the EGDA. The first one is that it requires less parameter when compared to other well known algorithms [22]. The second one is that the convenience of the EGDA for the FCM studies that was searched and proved previously in the study of Baykasoglu et al.[45], [22].

The Non-Linear Hebbian Learning Algrorithm (NHL)

The non-linear Heabbian learning (NHL) algorithm is one of the most popular algorithms in the FCM analyses. It helps to update the FCM weights between the concepts in order to increase the capabilities of FCMs [75]. It is also known with its simplicity and quickness similar to other Hebbian learning algorithms [56]. It was initially presented by Papageorgiou [54] and extended several times [22]. The extensions of the algorithm was discussed in ref. [40] in detail. In this thesis, the relationships between the concepts are modified using the NHL formulation given in Eq. 4.4 [22].

$$W_{ij}^{k+1} = \gamma W_{ij}^{k} + \eta A_j^k\left(A_i^k - sgn\left(W_{ij}^k\right)A_j^k W_{ij}^k\right) \tag{4.4}$$

where W_{ij}^k presents the relation between concept i and j at the k^th^ iteration and A_j^k is the value of the jth concept at kth iteration [22]. Besides, ɣ represents the weight decay parameter while η indicates learning rate. "sgn" is the sign function that returns the sign of a quantity.

The NHL algorithm is stopped using three termination rules. The first termination rule is to satisfy the minimum cost function. Mean Squared Error (MSE) is used as the cost function in this study in order to evaluate the performance of the algorithm and compare the results of it with the second learning stage [22]. The formula of the MSE is as presented in Eq. 4.5 where $n_{(output)}$ indicates the number of output concepts while A_i^k represents the value of ith concept at the iteration k and Z_i^{Real}represents the real value of the ith concept. The value of Z_i^{Real} is obtained from clustering analysis that is given in Table 4.5. It refers to the real value of the relevant concept for that cluster.

$$MSE = \frac{1}{n(output)}\sum_{1}^{n(output)}(A_i^k - Z_i^{Real})^2 \tag{4.5}$$

The steps of the algorithm are presented in Algorithm 1.

Algorithm 1 *The steps of the NHL algorithm* [22]

Step1: Start with the initial data: number of concepts, initial weight matrix, initial concept value, and parameters.

Step 2: Update the concept values using Eq. 4.2.

Step 3: Generate new weights using the Eq.4.4

Step 4: Calculate the cost function using the Eq. 4.5.

Step 5: Check the termination conditions

IF one of the conditions is satisfied, STOP.

ELSE, return step 2.

The other termination rule is to reach the minimum difference for the value of same output concept at consecutive iterations as seen in eq. 4.6 where e represents the minimum difference value [22]. The last termination rule is to reach the maximum iteration number.

$$Difference = \left| A_i^{k+1} - A_i^k \right| < e \tag{4.6}$$

The Extended Great Deluge Algorithm (EGDA)

The outputs of the first stage are used as inputs in the second stage. The second learning stage is performed using the EGDA. The EGDA is one of the members of meta-heuristic family [45], [76]. It represents similar properties with the well-known simulated annealing algorithm. The EGDA takes the attention with its global search ability and relatively less demand of parameters [56], [22]. It accepts the worse solutions if the cost function is equal or less than an upper limit that is decreased by a certain amount (called as decay rate) [45]. The EGDA modifies the relations between concepts using the equation provided in Eq 4.7. where $W_{ij}^{candidate}$ represents the relation between concept i and j at the new iteration while W_{ij} indicates the values of weight between concept i and j. In addition, random() is a random value between 0 and 1 while stepsize is the value of the step of the neighborhood.

$$W_{ij}^{candidate} = W_{ij} + (2random() - 1)stepsize \tag{4.7}$$

This formula is used to create a neighbor and exhibit the change of a variable from one value to another with a random step size [22]. The steps of the EGDA are also presented in Algorithm 2 below.

Algorithm 2 *The steps of the EGDA* [22]

Initialization

Step1: Start with the initial data: number of concepts, initial weight matrix, initial concept value, and parameters.

Step 2: Update the concept values and set as current values using Eq. 4.2.

Step 3: Calculate the cost function

Step 4: Equalize the upper limit and the best cost function to cost function calculated in step 3

The Main Processes

Step 1: Generate the candidate weights using the Eq. 4.7

Step 2: Generate the candidate concept values using the weights from step 1

Step 3: Calculate the candidate cost function using the Eq.4.5.

IF candidate cost function<=current cost function, update candidate weight matrix, concepts values and cost function as current ones

IF candidate cost function<=best cost function, update candidate weight matrix, concept values, and cost function as best ones

ELSE IF candidate cost function<=upper limit, update candidate weight matrix, concepts values and cost function as current ones

ELSE, increase the rejected solution by one

Step 4: Decrease the upper limit by decay rate

Step 5: Check the termination conditions

IF one of the conditions is satisfied, STOP.

ELSE, return step 1 of the main processes

The same cost function (MSE) with the NHL algorithm is used at the second stage too. The termination rules of the EGDA are to reach one of the following: maximum number of iterations, maximum consecutive rejected solutions, minimum decay rate or stepsize values [22].

4.4.4 Performance Measures

The performance of the one-stage (the NHL algorithm) and two-stage learning procedures (NHL and EGDA) are measured using different performance measures separately. The main measure is the Mean Squared Error (MSE) which is the cost functions and stopping criteria both procedures. It is calculated by using the values of all output concepts. Besides, it is important to measure the performance of the algorithms to predict "the likelihood of tobacco use" for interpreting the results. For that reason, the Mean Absolute Percentage Error (MAPE) is also calculated to evaluate the effectiveness of the stages. The MAPE is estimated considering only the values of "likelihood of tobacco use" output concept [22]. The MAPE values are calculated for one stage learning and two-stage learning procedures separately and provides to us the percentage difference between the real and the generated likelihood of tobacco use potential of people [22]. The formula of the MAPE can be found in Eq. 4.8 where $n_{(output)}$ indicates the number of output concepts. A_i^k represents the value of i^{th} concept at the iteration k while Z_i^{Real} is the real value of the i^{th} concept.

$$MAPE = \frac{1}{n(output)} \sum_{1}^{n(output)} \left| \frac{A_i^k - Z_i^{real}}{Z_i^{real}} \right| \quad (4.8)$$

The third performance measure is the number of demographic output concepts that take values outside the pre-determined bounds [22]. The details of performance measures are given in the next section.

4.5 The Implementation Processes

In this section of the thesis, the process of the two-stage learning procedure is explained in detail.

4.5.1 Determination of Initial Concept Values and Weight Matrix

The demographic characteristics of individuals and the likelihood of tobacco use are accepted as the concepts. Thus, this FCM analyses consist of 6 key concepts. The concepts are listed as follows:

- Concept 1 (C1): Gender

- Concept 2 (C2): Age group
- Concept 3 (C3): Residence
- Concept 4 (C4): Education level
- Concept 5 (C5): Working status
- Concept 6 (C6): Likelihood of tobacco use

The values of the concept from 1 to 5 are calculated by dividing the range [0, 1] to the number of category of each concept (the number of answers to the relevant question in the survey). For example; age group has 4 different categories: 15-24, 25-44, 45-64, 65+. Therefore, the [0, 1] range is divided four equal ranges. The values of the age group 15-24 take a value between 0-0.25 while the age group 25-44 can take a value between 0.25-0.5. In the light of the information given, gender (C1), age group (C2), residence (C3), education level (C4), and working status (C5) concepts are adjusted to have 2,4,2,7, and 8 ranges respectively. The values of the last concept "likelihood of tobacco use" are calculated using the percentages of tobacco use of each cluster. These values are obtained from the GATS Turkey 2008.

The relations are firstly asked to the experts and restricted information about the direction of the relations is collected. The ranges of the relations are represented in the Table 4.3. For example; gender (concept 1) can affect likelihood of tobacco use (concept 6) of people on positive or negative way while likelihood of tobacco use of people cannot affect gender of people.

Table 4.3 The members of clusters [22]

	Concept 1	**Concept 2**	**Concept 3**	**Concept 4**	**Concept 5**	**Concept 6**
Concept 1	0	0	0	[-1,+1]	[-1,+1]	[-1,+1]
Concept 2	0	0	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]
Concept 3	0	0	0	[-1,+1]	[-1,+1]	[-1,+1]
Concept 4	0	0	[-1,+1]	0	[-1,+1]	[-1,+1]
Concept 5	0	0	[-1,+1]	[-1,+1]	0	[-1,+1]
Concept 6	0	0	0	0	0	0

Using the directions given in the Table 4.3, a random initial weight matrix iscreated. This weight matrix is presented in Table 4.4.

Table 4.4 Initial random weight values

	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
Concept 1	0	0	0	0.07	0.53	-0.48
Concept 2	0	0	-0.27	0.71	0.79	0.38
Concept 3	0	0	0	-0.29	0.01	-0.51
Concept 4	0	0	0.34	0	-0.44	0.24
Concept 5	0	0	0.49	0.10	0	0.69
Concept 6	0	0	0	0	0	0

4.5.2 FCMs with Two-Stage Learning Procedure

FCMs analyses are initiated with the determined concepts and random initial weight matrix. In the analyses, concept 1 and 2 (gender and age group) are identified as input concepts that are not affected from any other concepts [22]. Therefore, the values of both concepts will be the same during the simulation.

The residence, education level and work status concepts need to take a value within the pre-determined range to represent the characteristics of the studied cluster [22]. For example; if the residence of the cluster is urban, the FCM must find the value of that concept within the bounds of urban. Therefore, they are accepted as output concepts in order to converge them in the desired regions [22].

The concept 6 (likelihood of tobacco use) is also an output concept that should be calculated as close as possible to the real value of it for each cluster [22]. The real ranges or values of output concepts for each cluster are presented in Table 4.5.

Table 4.5 The initial values and desired range or values of each concept for each cluster [22]

	Cluster 1		Cluster 2		Cluster 3		Cluster 4		Cluster 5	
	Initial value	**Desired Range or Value**	**Initial value**	**Desired Range or Value**	**Initial value**	**Desired Range or Value**	**Initial value**	**Desired Range or Value**	**Initial value**	**Desired Range or Value**
Concept 1	0.2500	0.2500	0.7500	0.7500	0.2500	0.2500	0.7500	0.7500	0.2500	0.2500
Concept 2	0.3750	0.3750	0.3750	0.3750	0.6250	0.6250	0.3750	0.3750	0.6250	0.6250
Concept 3	0.7000	0.5000-1.000	0.3000	0.0000-0.5000	0.3000	0.0000-0.5000	0.7000	0.5000-1.0000	0.7000	0.5000-1.0000
Concept 4	0.6000	0.5714-0.7142	0.2000	0.1428-0.2857	0.2000	0.1428-0.2857	0.2000	0.1428-0.2857	0.2000	0.1428-02857
Concept 5	0.1000	0.0000-0.1250	0.6000	0.5000-0.6250	0.1000	0.0000-0.1250	0.3000	0.2500-0.3750	0.2000	0.1250-0.2500
Concept 6	0	0.5714	0	0.2011	0	0.4307	0	0.0476	0	0.5150

In the first stage of the learning procedure, the NHL algorithm is implemented to the prepared data. The NHL algorithm requires identifying a learning parameter to perform the analysis. In order to determine the learning parameter, three different parameter values: 0.0001, 0.001, 0.01 are tested with the NHL algorithm using trial and error approximation. The parameter value 0.0001 provides the best outputs for each cluster. For that reason, the first-stage is performed using 0.0001 learning parameter value. Besides, the weight decay parameter is accepted as 0.98 while the error difference is identified as 0.001 depending on the relevant literature [22].

The learning process with the NHL is implemented to each cluster by using clusters' own initial concept values and desired ranges. The number output concepts that found outside the desired range, the MSE and the MAPE value for only likelihood of tobacco use concept is calculated for each cluster at the end of the FCM analyses.

The output of the first stage is used as the input in the second stage. For that reason, the weight matrixes provided by the NHL algorithm are accepted as the initial weight matrix of the EGDA in the second stage [22]. The parameters of the EGDA are taken from the study of Baykasoglu et al. [45], [22]. The decay rate is 0.001 while the step size is initiated from 0.9 and decreases by 0.01 until it reaches to 0.1, then decreases by 0.001 till it reaches 0.001 [22].

The second stage is run to reach 1000 iterations. Similar to the first stage, the MSE value, the number of demographic output concepts that found outside the desired range, and the MAPE value for only 6th concept are calculated for each cluster [22]. Table 4.6 indicates the results of performance measures comprehensively.

Table 4.6 The performance measures of the NHL algorithm and the two-stage learning process (NHL+EGDA) [22]

	Cluster 1		**Cluster 2**		**Cluster 3**		**Cluster 4**		**Cluster 5**	
	NHL	**NHL+EGDA**	**NHL**	**NHL+EGDA**	**NHL**	**NHL+EGDA**	**NHL**	**NHL+EGDA**	**NHL**	**NHL+EGDA**
MSE*	0.0934	**0.0098**	0.1473	**0.0202**	0.1949	**0.0484**	0.1763	**0.0277**	0.1132	**0.0005**
MAPE**	0.1568	**0.0071**	2.2869	**0.1307**	0.5356	**0.0030**	12.8865	**1.7373**	0.2842	**0.0137**
Num. of Output Concepts Outside the Range ***	1	1	3	**1**	3	**2**	2	2	2	**0**

*MSE is found using the values of all output concepts. ** MAPE is found using the value of the 6th concept (likelihood of smoking) [22]. *** Number of concepts outside the range includes only demographic output concepts (residence, education level, and work status) [22].

The relation matrixes found at the end of the FCM analyses with the multi-stage learning procedure were presented in the equations 4.7, 4.8, 4.9, 4.10, and 4.11.

Table 4.7 The relations between concepts for cluster 1

	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
Concept 1	0	0	0	-0.8319	0.3451	-0.3521
Concept 2	0	0	0.4295	-0.0594	-0.6924	-0.5746
Concept 3	0	0	0	0.1571	-0.8270	-0.2379
Concept 4	0	0	-0.3544	0	-0.9898	0.0510
Concept 5	0	0	0.4911	0.2927	0	0.5464
Concept 6	0	0	0	0	0	0

Table 4.8 The relations between concepts for cluster 2

	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
Concept 1	0	0	0	-0.9114	-0.7777	-0.8567
Concept 2	0	0	0.1069	0.2158	0.2953	0.1743
Concept 3	0	0	0	-0.5898	0.0289	-0.8987
Concept 4	0	0	-0.2540	0	0.8974	0.5139
Concept 5	0	0	-0.6632	-0.9262	0	-0.8793
Concept 6	0	0	0	0	0	0

Table 4.9 The relations between concepts for cluster 3

	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
Concept 1	0	0	0	-0.1350	0.6344	-0.0330
Concept 2	0	0	-0.7742	-0.6663	-0.7879	-0.9094
Concept 3	0	0	0	-0.3067	-0.5988	0.4232
Concept 4	0	0	0.2432	0	-0.9747	-0.1453
Concept 5	0	0	-0.9755	0.4726	0	-0.7814
Concept 6	0	0	0	0	0	0

Table 4.10 The relations between concepts for cluster 4

	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
Concept 1	0	0	0	-0.2228	0.1353	-0.4650
Concept 2	0	0	0.8238	-0.6189	-0.2259	-0.6001
Concept 3	0	0	0	-0.1004	-0.3457	-0.9203
Concept 4	0	0	0.0441	0	-0.4654	-0.8669
Concept 5	0	0	0.7572	-0.2625	0	-0.5864
Concept 6	0	0	0	0	0	0

Table 4.11 The relations between concepts for cluster 5

	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6
Concept 1	0	0	0	-0.8745	-0.7212	-0.5419
Concept 2	0	0	0.3709	-0.5871	-0.9484	-0.8967
Concept 3	0	0	0	-0.8919	-0.9873	0.2734
Concept 4	0	0	0.2223	0	0.4634	0.8157
Concept 5	0	0	0.3770	-0.9223	0	-0.7144
Concept 6	0	0	0	0	0	0

4.6 Discussions

The major aim of the performed FCM analyses is to examine the relations between the demographic factors and likelihood of tobacco use for Turkey. A multi-stage learning procedure is implemented to increase the accuracy and the robustness of the analyses. In order to compare the performances of one-stage learning with the NHL algorithm and two stage learning with the NHL algorithm and EGDA, the MSE, the MAPE, and the number of concepts that takes value in the desired ranges are calculated for each stages and the results was presented in the Table 4.6.

When the MSE and the MAPE values are examined in detail, it can be seen that two-stage learning procedure reaches better outputs for all clusters. The implemented two-stage learning procedure decreases the MAPE and the MSE values significantly and provides the appropriate weight matrixes for the clusters when compared to one stage learning procedure with only the NHL algorithm [22].

The number of concepts outside the pre-determined bounds and the MAPE values are most important performance measure for the analyses because they can help to interpret the results of the analyses. The major reason behind this is that the MAPE basically represents the percentage deviation between real and predicted values of likelihood of tobacco use while the number of concepts outside the pre-determined bounds shows the number of concepts that cannot be found into the desired region [22].

The calculated MAPE values as the result of FCM with two-stage learning procedure are below % 1 for cluster 1 and 3. They are around as %0.71 for cluster 1, % 0.3 for cluster 3, and % 1.37 for cluster 5. However, the MAPE results are really high for cluster 2 and 4. They are found as %13.07 for cluster 2 and %173 for cluster 4. The results of MAPE indicates that the multi-stage learning procedure can predict the likelihood of tobacco use more accurate for the clusters where the real values of this concept are high [22]. If the real value of the likelihood of tobacco use concept is really low (such as cluster 4 with %4.76), the procedure can fail to find a proper relation matrix according to predicted MAPE values [22]. The main reason behind this problem can be that demographic factors may not be the major factors to represent the current status of likelihood of tobacco use for these clusters [22]. In this case, other factors can also be adapted to the analysis to increase prediction accuracy [22].

The second crucial performance measure is the number of concepts outside the pre-defined bounds. The results of the analyses indicate that the multi-stage learning based FCM found all concepts in the pre-defined bounds for cluster 5. When compared with the one-stage learning procedure, multi-stage learning procedure provides better results for most clusters. Although the performance of the multi-stage learning process is better than the one stage learning, it is still not enough to assign all demographic factors to desired regions for all clusters [22]. There are still several weaknesses of the implemented procedure to imitate the real life but it is important to remind that working with the heuristics may not provide the optimum all the time. Finding results better enough to represent the studied system is also an important contribution for improving our understanding for the systems.

One of the most important purposes of the presented analyses is to reveal the relations between the demographic characteristics and the likelihood of tobacco use. The final maps for each cluster are found at the end of the FCM analyses as seen in Figure 4.2. Besides, the relation matrixes of each cluster are also obtained. The relations obtained for cluster 4 are not considered due to the low performance.

In the analyses, the concept 1 shows the gender factor. Each cluster is represented a different gender in order to understand the effect of gender on tobacco use [22]. The clusters 1, 3 and 5 consist of males while 2 and 4 clusters are females. The outputs of the analyses indicate that gender have negative causality for all clusters. However, the value of the negative causality is really strong for female (cluster 2, -0.8567) while it is really weak for males (cluster 3, -0.0330) [22]. The magnitudes of the relationships show that the gender of cluster 2 have a considerable negative impact on the likelihood of tobacco use [22]. Similar to the results of this study, Kılıc and Ozturk [77] analyzed the relation between different factors such as gender and tobacco use using Negative Binomial and Zero-inflated Negative Binomial models. Their findings also indicated that factors affecting tobacco use behaviors of males and females are different [22]. That means they analyzed the effect of gender types and their approach also supported the study design our analyses [22].

The second concept of the study is age group. The findings show that the age group of individuals has negative effect on likelihood of tobacco use people who belong to male originated clusters 1,3 and 5 (males) while it has positive impact for cluster 2 (females) [22]. The clusters 1, 3 and 5 are males whose age groups are 25-44, 45-64, and 45-64 respectively while cluster 2 includes females. There is really high negative causality between 45-64 age group males and likelihood of tobacco use. The negative causality indicates that males aged between 45 to 64 years are less likely to smoke compared to males aged 25 to 44 [22]. On the other hand, the age group shows the positive causality with likelihood of tobacco use. In the literature, there are two important studies that examined the similar relations. Kılıc and Ozturk [77] and İlhan et al.[78] found that individuals aged over 25 are more likely to participate tobacco use. Their results also indicate similar patterns with the findings of FCMs.

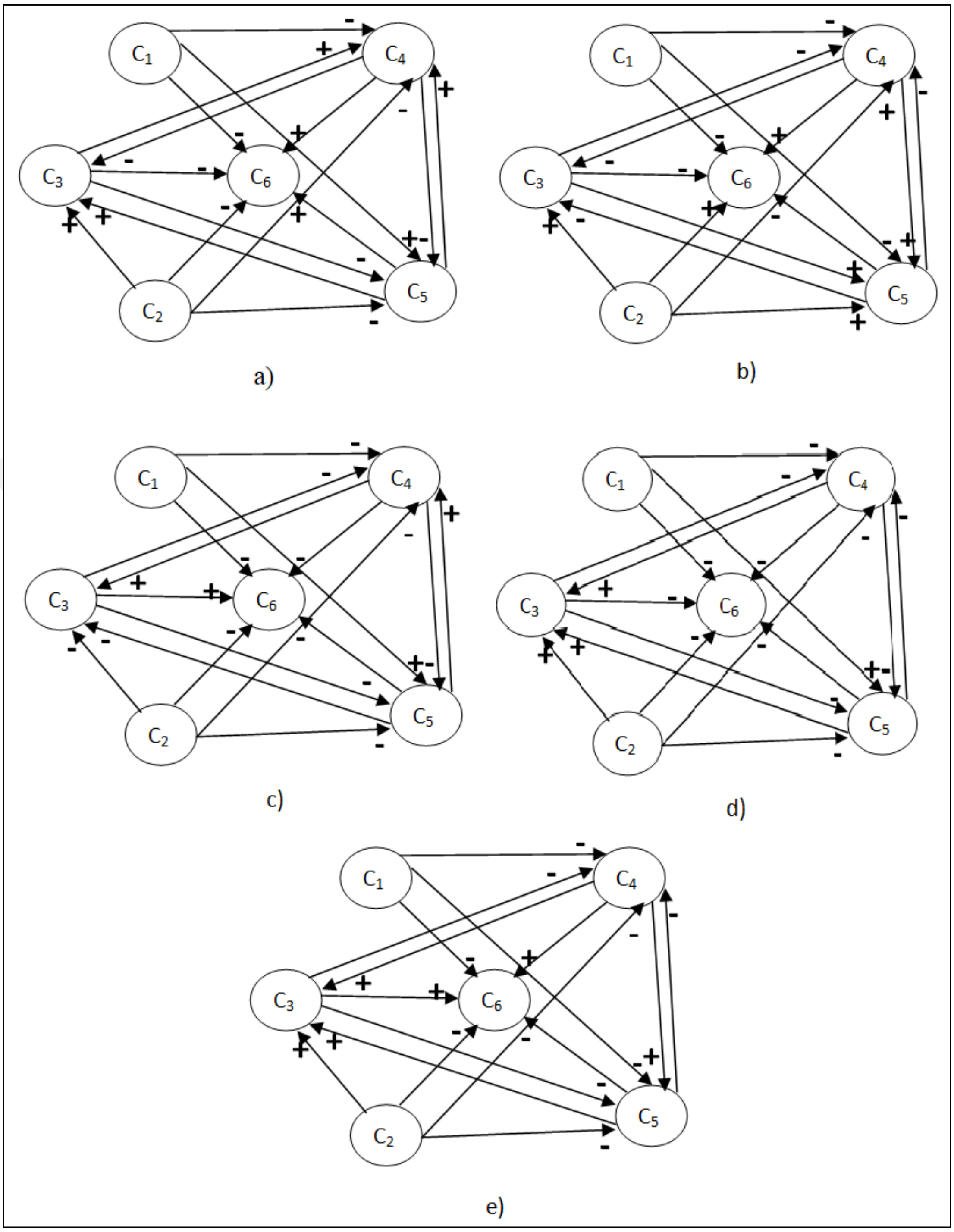

Figure 4.2 The final FCM of a)Cluster 1, b) Cluster 2, c)Cluster 3, d)Cluster 4, and e)Cluster 5 [22].

The third concept to examine the relationship with tobacco use is residence type. The relation between residence concept and likelihood of tobacco use shows positive causality for the clusters 3 (urban) and 5 (rural) while it is negative for the clusters 1 (rural) and 2 (urban) [22]. The magnitude of the relation is the highest effect (-0.8987) for the cluster 2. According to the study of Kılıc and Ozturk [77], females living in urban areas are more likely to smoke when compared to females living in rural areas. In the presented analyses, the cluster relevant to the females living in rural areas is eliminated. That's why; it cannot be compared. However, females living in urban areas show negative causality with tobacco use. The main reason behind the difference between two studies may be the variations at the implemented methodologies. FCM considers the direct and indirect all relations in order to find a model [22]. Studying with different populations may cause different results for studies.

The fourth concept of the analyses is education level. Apart from the cluster 1, other clusters have the same education level as elementary school. Although they have the same education level, this concept shows various results for each cluster [22]. At that point examining other concepts (factors) and the entire relations in the obtained maps are really vital to understand the relations [22].

The final demographic concept is work status. The relations between the work status and likelihood of tobacco use concepts indicates negative causality apart from cluster 1 [22] where the work status is paid employee. The outputs provide an interesting result which indicate that people who are paid workers tended to use tobacco because there is a strong positive relation between these two concepts [22]. On the other hand, the negative causality is observed for other concepts. People who are home makers and non paid family workers (people who do not earn money) are less likely to smoke compared people who make money. The main reason behind the negative causality can be economical independency [22].

Consequently, different demographic factors represent different causality with likelihood of tobacco use. The difference at the casual relations proves that means the implemented algorithms are able to represent the diversification of the clusters with the maps which they are provided at the end of the analyses [22]. Therefore, exploring the relations of each factor within the relevant cluster can be accepted as one of the important contribution of this thesis. Besides, the implemented multi-stage learning procedure provides significant advantageous such as eliminating the dependency to expert knowledge, increasing accuracy and robustness.

4.7 Concluding Remarks

The major objective of the presented thesis is to build a hybrid model integrating with AB modeling and FCM in order to examine tobacco use behaviors of people. FCM is used to represent the mental models of the people while AB modeling is imitating the real life system. The mental models of the agents are created using several factors such as demographic characteristics. However, in order to understand whether FCM is capable represent the relations between demographic characteristics and likelihood of tobacco us, simple prototype analyses are performed before designing the hybrid model. The FCMs are created using a multi-stage learning based procedure that combines the well known NHL algorithm and the EGDA. The results of the analyses are compared with the one stage learning with the NHL to observe the advantageous of the multi-stage learning procedure provided.

This chapter summarizes the details of the FCM analyses. The findings reveal that the implemented multi-stage learning procedure provides a better performance to represent the relations between demographic factors and likelihood of tobacco use. It also helps to eliminate the dependency to the expert knowledge and increases the accuracy when compared to the one stage-learning procedure.

CHAPTER V

THE HYBRID AGENT BASED MODEL INTEGRTATED WITH A MULTI-STAGE LEARNING BASED FUZZY COGNITIVE MAP FOR UNDERSTANDING TOBACCO USE OF TURKEY

5.1 Introduction

Social systems mostly consist of individuals that interact with each other and other fundamental elements of the system structure. Studying with social systems can be challenging processes for researchers most of the time because the difference at the characteristics of each member, their social networks, the level of interactions with others and the results of these interactions can increase the complexity and make it difficult to analyze the systems comprehensively. Tobacco use and control can be an example for such systems since it is accepted as a complex and adaptive system involving behavioral, biological, and environmental influences [66]. It consists of various factors such as, many individuals, characteristics and social network of the individuals, tobacco control policies and etc.

Examining the complex systems and overcoming the difficulties that arises from facing with complexity requires advanced techniques and skills. Simulation modeling is one of the important ways to imitate real life systems. In the literature, various applications of simulation modeling on tobacco use can be found. System Dynamics (SD) and Agent Based (AB) modeling have been implemented separately to understand tobacco use behaviors of people in different studies. Among them, the SimSmoke [79] was one of the well known SD model that was made for examining the effects of different tobacco control policies. It was implemented many countries such as Brazil [80], Sweden[81], Finland [82], Germany[83], Korea [84], China [85] and etc. to investigate the performance of their tobacco control policies. SimSmoke models use the potential benefits of implementing a holistic approach. They are for analyzing the effects of public policies on the societies with the help of

SD modeling. To our best knowledge, although SimSmoke models have been provided significant findings to the modelers, these models could not examine the effect of individual level behaviors, the characteristic and networks of individuals. On the other hand, AB models could overcome these problems with the help of individualistic approach. However, the existing models require the large amount of data. If there is not enough data, there could be important calibration problems.

In this case, combining the benefits of holistic and individualistic approaches into a model to understand tobacco use behaviors of people can be a wise solution to avoid the problems that arises from the single use of the approaches. With this objective in mind, SD and AB models can be integrated. However, both modeling techniques require important amount data, to perform an expansive analysis. At that point, it is crucial to find a method that can be integrated with AB modeling, help to imitate the perspective of individuals based on an uncertain context, and work with scarce data. When the literature is reviewed, it can be seen that Fuzzy Cognitive Map (FCM) is one of the most appropriate technique that can perform under such cases. In addition, FCM is more intuitive and required neither modeling expertise nor hard data compared to other aggregate approaches such as SD [30].

Therefore, a hybrid agent based model integrated with a multi-stage learning based FCM is proposed in this thesis. The hybridization of both methodologies can help to utilize their individual benefits into the same model. This model is established for understanding the tobacco use behaviors of individuals in Turkey.

Tobacco use is one of the most important public health issues of Turkey similar to other countries. Although Turkey has made remarkable efforts to fight against the prevalent use of tobacco, the percentage of current tobacco user is still high in Turkey. Thus, it can be expected that the findings of this thesis could increase the understandings on the tobacco use behaviors of Turkish people and be a pioneering example to researchers to implement more advanced modeling techniques and help to understand tobacco use system comprehensively.

5.2 The Framework for the Hybrid Model

The proposed hybrid model is an integrated version of AB modeling and FCMs. AB modeling is used to reflect the individualistic perspective of the system. It means that

AB modeling imitated: a) the individual characteristic of agents (gender, age, residence, education level, work status), b) tobacco use behaviors of agents (never smokers, current smokers, former smokers), and c) social network effect. On the other hand, the multi stage learning based FCM that is presented in the previous chapter is used to reflect intuitive perspectives of individuals. FCMs basically presented mental models (brain activities) of individuals and b) work as a decision support mechanism for deciding to change their current tobacco use status.

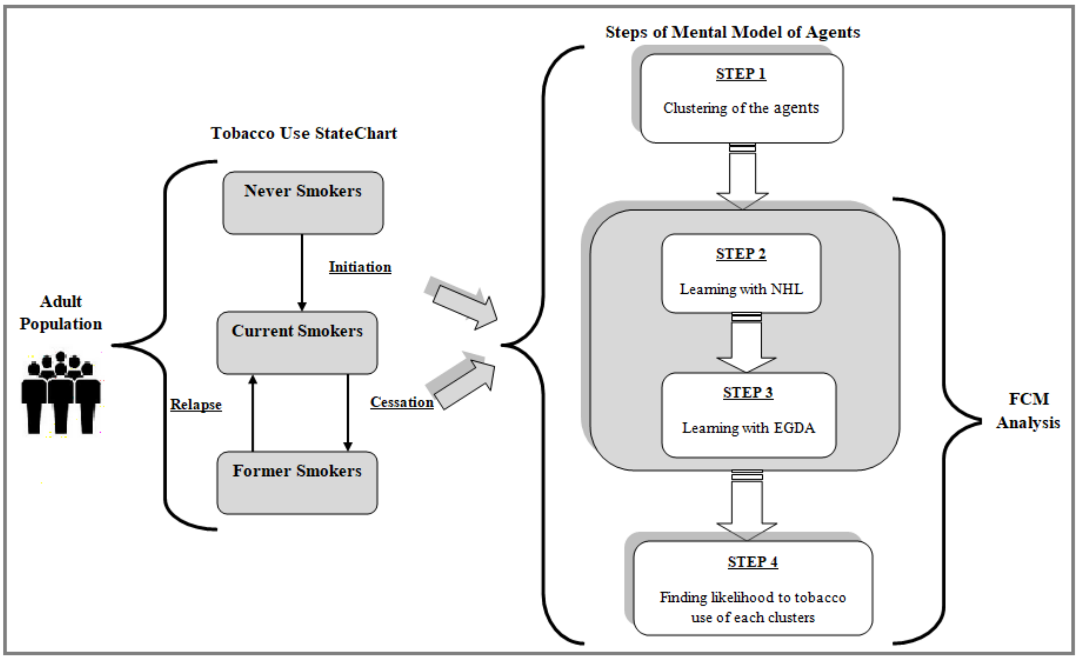

Figure 5.1 The components of the hybrid model.

The proposed hybrid model is built by using AnyLogic 7 University (version 7.0 0) Simulation Modeling software. In order to build the proposed model:

- Fundamental components of and AB model is determined such as agents, model time, variables, parameters, and etc.
- Subsequently, different states of agents to represent the types of tobacco use behaviors are identified.
- The rules to change agents' states are designed.
- A social network is assigned.
- Agents are grouped according to clustering analysis.
- The multi-stage learning based FCM is embedded to the agents.

5.3 The Description of the Hybrid Model

In this section, the details about the model building process are presented comprehensively.

5.3.1 The Main Data Sources

Tobacco use is a complex system which consists of many different factors that make it difficult to difficult to analyze the system with all aspects. In order to understand the system entirely, various factors is required to be take into account such as demographic factors, psychological factors, environmental factors, economical factors and etc. However, including all tobacco use relevant factors to the model and providing all necessary data for this aim are too difficult. For that reason, modelers generally struggle to improve their understanding on the topic using scarce data by implementing appropriate techniques.

Finding reliable data is other crucial step of a modeling process. In order to imitate a real-life system, gathering the required information from robust data sources can increase the imitation performance of the model. In this thesis, Global Adult Tobacco Surveys (GATSs) are used as a major data source for the presented hybrid model. GATS is known as a national household survey that helps countries to gather information about tobacco use and tobacco control measures [68]. It provides specific information belongs to participants such as socio-demographic characteristics, their tobacco use status, their tobacco use history, awareness about tobacco control policies. It has been implemented by several different countries. Turkey is one of these countries that applied the survey three times until now: 2008, 2012, and 2016.

The model is initialized using the information of GATS Turkey 2008 data because the year 2008 is the first year that the survey was performed for Turkey. The tobacco use system of Turkey is modeled as in that year to imitate the population as close as the real life. After that, the GATS Turkey 2012 and 2016 data are used to train in the FCM analyses.

Thus, this thesis considers 2008 as the base year. Subsequently, the model is modified to reflect the real-world using the data of GATS Turkey, Turkey Health

Surveys, Turkey Tobacco Control Strategy Document and Action Plan, and Turkish Statistical Institute (TUIK).

5.3.2 Model Time

The time unit of the proposed hybrid model is determined as one year. The main reason behind this is that data about population growth are collected on yearly basis. Most of the selected data sources collect the data within several years. The model is simulated from 2008 to 2030. The model is calibrated using the tobacco use data from 2008 to 2016 to reflect the real-world system. Subsequently, the predictions are performed from 2016 to 2030 to observe how tobacco use behaviors of people could change.

5.3.3 Agent

Agents are the fundamental object of an AB model. The type of an agent can vary according to the problem type. For example, it can be money for a bank system or it can be an automobile for a traffic system. In this thesis, agents represent people. The collection of agents forms the Turkish population.

As stated previously, the year 2008 is the base year for this model. Thus, initially the population of Turkey in 2008 is reflected using agents. The population of Turkey was 71 517 100 in 2008 according TUIK Population by Years Statistics.

A small version of Turkish population is created using a total of 7151 agents in order to represent the population. In this study, we did not model all individuals of Turkish population. We consider human agent that are representative for Turkish society.

5.3.3.1 Births & Deaths

An open population model that can change due to several different factors such as births and deaths is prepared for the analyses. The population size changes according to the deaths and births with years.

In order to change the number of population with deaths, 2009-2018 age specific death rates provided by TUIK is used. This data consists of age ranges: 0-4,5-9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-

74, 75+. The average of ten years for each age range is calculated and embedded to the model.

The population size changes by considering the crude birth rate provided by the Basic Fertility Indicators, 2001-2018 by TUIK.

5.3.3.2 Agents' Properties

The main objective of this thesis is to examine the difference of tobacco use behaviors of people by considering their demographic characteristics. For this aim, various demographic characteristics are assigned to the each agent. Initially, the characteristics provided by the GATSs are assigned to the agents of the model. These are presented below:

- Gender: Each agent has a specific gender that could be male or female.
- Age/Age Group: Agents has an age. In addition to this, the ages also are grouped as 15-24, 25-44, 45-64, 65+ as represented at GATSs. This age range group representation is also used by the World Health Organization. Thus, the findings of the thesis can be discussed using these age ranges.
- Residence: Each agent has a specific residence: urban or rural.
- Education Level: The latest education levels that agents completed are categorized as primary educated, secondary educated, high school educated, and university or higher educated.
- Work Status: The current work statuses of agents are represented as: paid employee, self employed, non-paid family worker, student, homemaker, retired, and no job.

5.3.3.3 Social Network

Interactions of agents with each other are an important factor that can affect the behaviors of agents in a system. Social networks are the important ways to represent the interactions between the agents. A social network is basically a graph where agents are represented as vertices and existing relationships between two agents [86]. Several different network types can be determined for a system according to the problems. In order to understand the network types, detailed analyses can be necessary for social systems because many contact such as contact typeswithin household, family, peers and etc should be taken into account. However, it is really difficult to form a specific network for the problems like tobacco use because many networks can possibly affect the system. For that reason, a random social network is used to analyze tobacco use behaviors of agents in this thesis.

The connection size is taken from the study of Mossong et al.[87]. They explored the contact patterns for infectious diseases by conducting a population based prospective survey. The patterns of eight different European countries were used in that study. The findings indicated that contact characteristics seemed to be similar across different European countries. They recorded a total of 97 904 contacts with a mean of 13.4 per day for each person. Their findings were then used in the study of Kassaie et al. [88] to calculate the contact numbers. Similar to these studies, the connections between agents in the proposed model is assumed with a mean of 13.4 close contacts. We consider this daily value into a annual value.

5.3.3.4 Statecharts of an Agent

Statecharts are basically used for representing sophisticated behaviors. They are the extended version of state diagrams and visual graphs that enables to define event and time driven behavior of various objects (agents) [89]. They simply consist of states and transitions. States represent the type of behaviors while transitions are used to reflect the rules for transportation from one state to another.

In this thesis, the states of an agent are represented using five different statecharts. The first statechart is "Tobacco Use StateChart" that presents the tobacco use behaviors of an agent. This statechart has 4 different states as seen in Figure 5.2:

- Entire population: This thesis focuses on examining the tobacco use behaviors of adult population of Turkey. For that reason, people who aged fewer than 15 cannot be placed in a state that was relevant to a specific tobacco use behaviors. Thus, people who aged less than 15 are collected in this state. When their age became to 15, they are transported to one of other three states according to their tobacco use status.
- Never smokers (NS): Agents who never smoke or who smoked less than 100 cigarettes in his/her life [90].

- Current smokers (CS): Agents who smoke 100 cigarettes in his/her life and currently smokes cigarettes [90].
- Former smokers (FS): Agents who smoked at least 100 cigarettes in his/her life but who had quit smoking at the time of the survey [90].

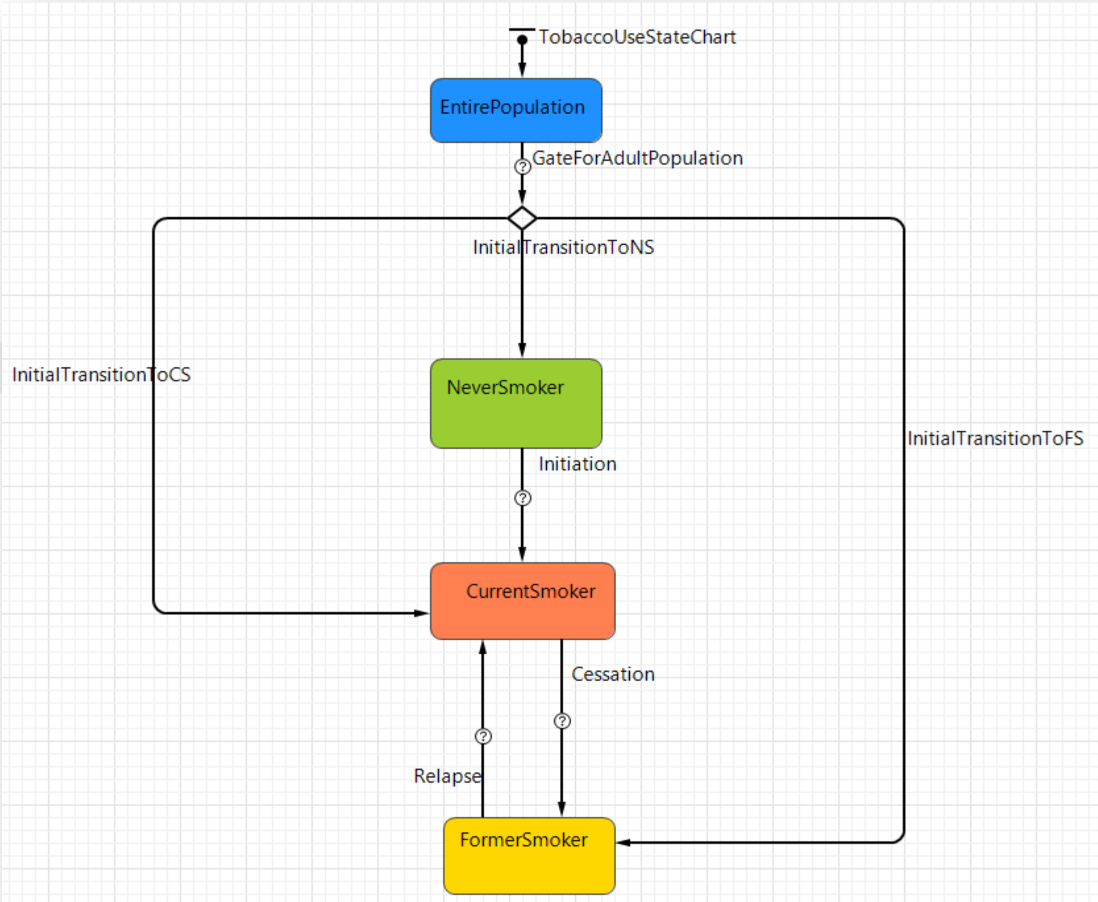

Figure 5.2 The tobacco use statechart of the hyrbid model.

The adult agents of the model can stay in the same state during the simulation or move from one to another at each time step with the help of transitions. When the model is initiated, there is no agent at any state of the tobacco use statechart. In order to reflect the tobacco use of Turkey for 2008, the initial transitions to CS, NS, and FS are used and the adult population is distributed to the states according to their tobacco use behaviors. After initiation, an agent can change its current state with the help of three different transitions. Transition between NS to CS, CS to FS, and FS to CS are called as "initiation", "cessation", and "relapse" respectively. Transitions between states are triggered by specific conditions. The conditions are determined using "likelihood of tobacco use of agent", "number of year that agent smoke", " number of years that agent quit", "number of CS that agent have in his/her social network".

The "likelihood of tobacco use of agent" is obtained as the result of the multi-stage learning based FCM that is presented in the following section. "Number of year that agent smoke or quit" is calculated during the simulation considering the status changes of agents. For example; if an agent moves from CS to FS, the number of years that his/her quit is started to be calculated. Initiation age distribution is taken from the GATSs of Turkey initially.

At each step, current smokers send a message to all related agents. These messages are used to calculate the number of CS agents in the social network of each agent. This thesis assumes that if the number of CS increases in the network, the potential of initiation to tobacco use also increases. That's why; the rate of the number of CS in the network is also included to the transition conditions in order to reflect the social network effect.

Transition conditions are changed for each type of transitions and are as below.

- Initiation: Initiation can be triggered according to the initiation condition that is given below was greater than initiation threshold value.

  (likelihood of tobacco use) + (rate of the number of CS in his/her social network)> initiation threshold value

- Cessation: Cessation can be triggered according to the cessation condition that is given below was greater than cessation threshold value.

  (1-likelihood of tobacco use) - (number of years that agent smoked/current age of agent) - (rate of the number of CS in his/her social network)> cessation threshold value

- Relapse: Relapse can be triggered according to the relapse condition that is given below was greater than relapse threshold value.

  (likelihood of tobacco use) - (number of years that agent quitted/current age of agent) + (rate of the number of CS in his/her social network)> relapse threshold value

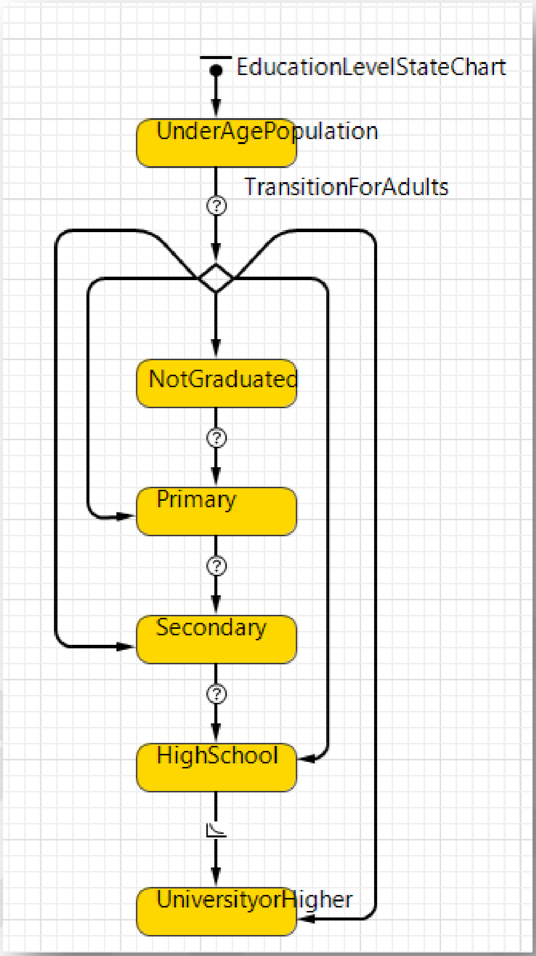

Figure 5.3 The education level statechart of the hyrbid model.

The second statechart of the agents is education level statechart as seen in Figure 5.3. Education level is previously assigned to the agents as a demographic characteristic using the data of the year 2008. However, the education level of an agent can change within the years. Thus, it is assumed that each agent must improve its education level according to the twelve years compulsory education:

- Agents whose ages are equal to 11 are moved to primary education graduates.
- Agents whose ages are equal to 14 are moved to secondary school graduates.
- Agents whose ages are equal to 17 are moved to high school graduates.
- Agents among the high school graduates can move to university or higher education graduates according to a specific rate that is generated using the percentages of people who started to universities according to TUIK statistics.

The birth and mortality statecharts are used to monitor the fertility and mortality of an agent while aging statechart is used to update the current age of an agent each time step.

5.3.4 Tobacco Control Policies

Tobacco use is among the deathliest public health issues in the world. The importance of the problem leads countries to take significant precautions in order to decrease the prevalent use of tobacco. The WHO proposed a set of measures for tobacco control to help countries. These measures are called as MPOWER and represents: monitor tobacco use, protect people from tobacco smoke, offer help to quit, warn about the dangers of tobacco, enforce bans on tobacco advertising, promotion, and sponsorship, and raise taxes on tobacco.

Turkey is one of the pioneering countries that implements tobacco control policies resolutely. In order to imitate tobacco use system realistically, the implemented tobacco control policies must be considered. For this aim, the effects of five important tobacco control policies are reflected with the proposed model. These tobacco control measures are listed below:

- Anti-tobacco information: Turkey has been used anti-tobacco information to explain the damages of tobacco use to people. This measure is already existed in 2008. Thus, when the model is initiated, people who notice and did not notice the anti-tobacco information are assigned with the help of the GATS Turkey 2008 data.
- Health warnings on packages: Turkey presents a regulation for the content of the cigarette packages. According to the regulation, the packages of the cigarettes need a warning about the damages such as "Smoking Kills" and etc. This regulation is already existed in 2008. Therefore, when the model is initiated, people who notice and did not notice the health warnings on packages are assigned with the help of the GATS Turkey 2008 data.
- Taxation: Turkey applies significant taxation policy for tobacco in order to control tobacco use. The total tax burden was around %73.25 at 2008 and it was reached over %80 at 2017. The tax on tobacco is increased steadily by the years.

- Pictorial warnings on packages: Turkey updated the regulation about the content of the content of the cigarette packages in 2010. According to the regulation, the packages of the cigarettes need a pictorial warning about the damages of tobacco use. This regulation is not in use when the model was initiated. Thus, it is assigned to the model when the model time was 2010.
- Plain package: Turkey modified the packages of cigarettes entirely. All packages need to have same unique shape and structure to prevent the interest of people to conspicuous packages. This regulation is the newest tobacco control measure of Turkey. It was started in 2019. Therefore, there is no available information about the effects of this regulation. In this thesis, the possible effects of plain package regulation are searched with the scenario analysis that is explained in the relevant section.

Different from the mentioned tobacco control policies, information about people who noticed tobacco advertisements, sponsorships, and promotions is given by the survey. Therefore, notice of tobacco advertisements is also embedded to the model as a factor.

5.3.5 Mental model of Agents

The proposed hybrid model is built with the integration of AB modeling and FCM. In the model, AB modeling is used represent to the tobacco use behaviors of agents. On the other hand, FCM is used to represent the brain activities of agents. The FCM is an illustrative representation of a complex system that can analyze the casual relationships among the important factors in order to model system behavior. Thus, FCM is used to examine the effects of casual relations of different characteristics of agents and tobacco control policies on likelihood of tobacco use of an agent.

As mentioned in the statecharts sub-section, agents transite from one state to another with a specific condition. In order to trigger the transitions, the likelihood of tobacco use of agents is one of the important measures that must be calculated. Thus, the multi-stage learning procedure is utilized to calculate the likelihood of tobacco use of agents using casual relations among critical factors. It means that FCM provides the information about likelihood of tobacco use of agent (that was used at the transition of tobacco use statechart) considering the casual relations among crucial factors that can affect tobacco use behavior of an agent.

For this aim, a multi-stage learning based FCM procedure that is studied by Kocabey Çiftçi and Unutmaz Durmuşoğlu [22] is embedded into the AB model. The multi-stage learning procedure is used to increase the accuracy of the output relations. Most of the FCMs are initiated with expert knowledge. Since expert knowledge can cause subjectivity, the robustness of the maps can decrease. In order to prevent this problem, two different learning algorithms: non-linear hebbian learning algorithm (NHL) and extended great deluge algorithm (EGDA) are combined. The analyses are started with a random initial map instead of using a complete expert knowledge. The output weight matrix of the NHL is used as the input of the EGDA. A simple example of the implementation procedure is given in the previous chapter.

In the hybrid model, the factors (concepts) of the multi-stage learning based FCM analyses basically consist of demographic characteristics of agents, tobacco control policies and likelihood of tobacco use of agents. A total of 11 concepts are taken into account. These concepts are as listed below.

- Gender (C1)
- Age group (C2)
- Residence (C3)
- Education level (C4)
- Work status (C5)
- Notice of anti-tobacco advertisements (C6)
- Notice of health warnings (C7)
- Notice of tobacco advertisement (C8)
- Taxation level (C9)
- Notice of pictorial warning (C10),
- Likelihood of tobacco use of agent (C11).

The gender and age group concepts are constant concepts that cannot change their value as in the previous chapter. On the other hand, others are included to calculate fitness function with the two-stage learning.

5.3.5.1 Initiation of FCM Analyses

The FCM part of the thesis is initiated with clustering of the agents. Agents are clustered using their demographic characteristics and their notice status to tobacco control policies. The clustering of the agent is an important step because FCM analyses are performed to calculate the likelihood of tobacco use of agents. For this aim, the Mean Absolute Percentage Error (MAPE) value is calculated at the end of each FCM iteration. If the FCM analyses are performed for each agent instead of clusters, it is most impossible to find a real likelihood of tobacco use value to calculate the MAPE value. Thus, the robustness of the analyses can directly reduce. In addition to this, performing thousands of iterations for each agent at each simulation run can increase the complexity, cause memory problems and be time consuming. In order to avoid all these problems, agents are clustered according to their similarities.

Therefore, a clustering analysis using K-means clustering is made with the data provided by GATS 2008 Turkey. Although the GATS Turkey 2008 presents information for the first eight concepts, only eight concepts are included to the clustering analysis. A total of 8827 cases are used for the clustering analysis. The detailed information for the clustering analysis is given in the previous chapter.

Table 5.1 The members of clusters.

Factor / Clusters	Gender	Age Range	Residen.	Edu. Level	Work Status	Notice of Anti-Tobac. Ad.	Notice of Health Warn.	Notice of Tobac. Ad.
Cluster 1	Male	45-64	Rural	Primary Edu.	Self employed	Noticed	Noticed	Did not notice
Cluster 2	Male	25-44	Urban	Primary Edu.	Paid employee	Noticed	Noticed	Did not notice
Cluster 3	Female	25-44	Urban	Primary Edu.	Home maker	Noticed	Noticed	Did not notice
Cluster 4	Female	45-64	Rural	Not Grad.	Home maker	Noticed	Did not notice	Did not notice

A database that includes the clusters of each agent is created at the end of the clustering analysis. The information of the clusters is represented at Table 5.1. During the simulation, the clusters of each agent is checked and if there is a change, the agent is re-assigned to the new cluster according to the created database at each time unit.

The analyses start with the use of random initial matrix for each cluster. The ranges of the relations between the concepts that are represented in Table 5.2 are obtained from expert without details about the directions and the strengths.

Table 5.2 The direction of the relations between the concepts.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
C1	0	0	0	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]
C2	0	0	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]
C3	0	0	0	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]
C4	0	0	[-1,+1]	0	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]
C5	0	0	[-1,+1]	[-1,+1]	0	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]	[-1,+1]
C6	0	0	0	0	0	0	0	0	0	0	[-1,+1]
C7	0	0	0	0	0	0	0	0	0	0	[-1,+1]
C8	0	0	0	0	0	0	0	0	0	0	[-1,+1]
C9	0	0	0	0	0	0	0	0	0	0	[-1,+1]
C10	0	0	0	0	0	0	0	0	0	0	[-1,+1]
C11	0	0	0	0	0	0	0	0	0	0	0

5.3.5.2 FCM Analyses with Two-Stage Learning Procedure

After clustering the agents and determining the ranges of the relations, the FCM analyses are started. Firstly, the cluster of agents are checked and assigned at each model time unit. After identifying the cluster of agent, the FCM analyses are started with the first stage-learning with the NHL algorithm. The Mean Squared Error (MSE) is tried to be minimized at each iteration and the best weight matrix of the first stage is used as initial relation matrix of the second stage learning with the EGDA. Similar to the first stage, the MSE is calculated each iteration to minimize the value of it. Besides the MSE, the MAPE value is calculated only for the likelihood of tobacco use concept at the end of the two-stage learning procedure. If the value of MAPE is higher than %10, the second stage was repeated to increase the accuracy.

At each model time unit, all these steps from clustering to calculation of the MAPE are repeated. The main reason behind this is that the factors that defines the agents' behaviors such as age group, education level, notice of tobacco control policies and etc could change at every time unit. If one of them changes, the cluster of the agent and the real likelihood of tobacco use of the agent may also be changed. Thus, it is required to re-assign the agent to the relevant cluster.

The real likelihood of tobacco use of each cluster is found using the data of GATS Turkey 2008, 2012, and 2016. While calculating the MSEs and MAPEs, the likelihood of tobacco use of the each cluster is considered according to the relevant years. The learning algorithms are validated with the data of GATS Turkey 2008, 2012, and 2016. The values of the years between these year ranges are extrapolated.

After 2016, the learning procedures are omitted because there is not any data for calculating the real value of the likelihood of tobacco use after that year. For that reason, the FCM analyses are applied as follows: firstly updating the clusters of each agent, subsequently updating the concept values using the best weight matrixes provided at the end of the year 2016.

5.3.6 Assumptions of the Hybrid Model

Modeling with AB provides significant advantages to increase our understandings on tobacco use and control research field significantly. These advantageous were explained before in previous sections in detail. The most important one is to allow imitating heterogeneity of a complex system. Instead of representing a system as a whole, it helps to analyze all critical objects and their potential effects on the system. Although it is an important gain for this thesis, it must be stated that AB modeling can be difficult in some times due to several reasons. Firstly, it requires modeling expertise to sort out the system, determine the critical objects and reflect the system into a virtual world. However, modelers can improve their skills to eliminate this difficulty.

The other difficulty is the data requirement. AB models can need large number of data to reflect the heterogeneity of the real-world system more realistically. Finding all required data can be impossible all the time. In this case, modelers can use assumptions to perform their studies.

In this thesis, various assumptions are used to build the proposed hybrid model. These assumptions are listed as follows:

- Assumption 1: The main objective of the presented thesis is to examine the tobacco use behaviors of adult population of Turkey. The adult population is categorized by the WHO as "people who aged 15 and over". Therefore, although people who are younger than 15 are included to the calculations of population growth, they are not included to the tobacco use statechart. However, when people are 15 years old, the adult population number is updated and they are transported to the tobacco use statechart.
- Assumption 2: Agents in the model have different demographic characteristics such as gender, age, residence, education level, and work status. Among these characteristics while age and education level of the gents are updated at each model time unit, residence and work status of them are constant. The main reason behind this is the lack of data about the change of residences of people and the work status.
- Assumption 3: The education levels of agents are updated each year as mentioned in assumption 2. The initial education levels of the agents are assigned using the data of *Population by Attained Education Level and Sex, 2008-2018* of TUIK. Although, a detailed database about the changes at education level of each individual during years is not available, the education levels of agents are changed using 12 years compulsory education rule. For example; agent who graduated from primary school start to secondary education or agents who graduated from secondary education start to high-school. Transition from high school to university or higher education is performed with a specific rate. The average number of agents went to university or higher education is calculated from the same data of *Population by Attained Education Level and Sex, 2008-2018* of TUIK. An average of

%5.7 of agent who graduated from high school enter into the university or higher education.

- Assumption 4: The proposed model aims simulating the population of Turkey as an open population. For this purpose, the births and deaths of the population are reflected to the model using the crude birth rates and age-specific death rates of TUIK. However, the population of countries can also vary with the migrations. In this thesis, the migration cannot be considered in the model due to data availability.
- Assumption 5: In the proposed model, a social network is also determined to represent the communication structure between the agents. Determining a specific network for this kind of complex and dynamics problem is really difficult because there are so many relation types in the network such as family, peer, work, and etc. It requires a detailed study to scheme a network structure. For that reason, a random social network is employed in this thesis.
- Assumption 6: Agents can communicate within the pre-defined network structure. The number of connection is taken from the study of Mossong et al. [87]. They examined the number of contacts between people while examining an infectious disease for eight different European countries. They found 97 904 contacts with different people with a mean of 13.4 contacts per person on daily basis. Although Turkey was not a part of their study, the similar patterns of different countries became an inspiration to use their results for the population Turkey, too. Average daily contact stated by them is used as the average connection number of agents in this thesis. We assume that an average of 13.4 connections can help to represent close social network of each agent.
- Assumption 7: Transition from one state to another is triggered with the combination of several different factors. One of the important of them is the number of current smoker (CS) in the network of an agent. Thus, agents who are current smokers send messages as "Smoking is good!" to all connected agents in their own social networks. The agents can have or cannot have CSs in their social network. However, it is assumed if an agent takes message, its initiation potential to tobacco use increases. Thus, when the numbers of CSs in the network increases, the initiation or relapse potential of the agents are

also increased. This assumption is used to imitate the adverse effects of CSs on tobacco use behaviors of agents.

- Assumption 8: The numbers of years that an agent smoked or quitted can be accepted as important criteria. Thus, if the number years of an agent smoked increases, the potential to cessation can decrease and if the number years that an agent quit smoking increases, the likelihood of relapse of that agent can decrease.

5.3.7 Calibration of the Hybrid Model

The proposed hybrid model is built to examine the changes at tobacco use behaviors of Turkish people. For that reason, it is important determine a base year that the model is initiated. It is required to have large amount of data for the base year and the years following the base year. Although Turkey has remarkable efforts to fight against tobacco use, it is really difficult to find the individualistic data for tobacco use in Turkey before 2008. The acceptance of the Framework Convention of Tobacco Control (FCTC) was an important milestone for Turkey. Turkey started to collect considerable amount of data on tobacco use and control systematically after the treaty. The Global Adult Tobacco Surveys (GATSs) and the Global Youth Tobacco Surveys (GYTSs) have been important data sources for Turkey due to their rich contents on tobacco use and control.

In this thesis, the first year 2008 that the GATS Turkey was implemented is accepted as the base year. Thus, the model is initiated with the data of 2008 and represents the tobacco use status of Turkey for that year. In the years 2012 and 2016, the GATSs were repeated for Turkey. Thus, the GATSs Turkey 2008, 2012, 2016, Turkey Health Surveys 2010, 2014, and 2018-2023 Turkey Tobacco Control Strategy Document and Action Plan are the main data sources to observe tobacco use prevalence of Turkey. The data gathered from these documents are used to calibrate the proposed model.

In order to represent current tobacco use states of people and the changes at their tobacco use behaviors, three different transitions are determined. They change with initiation, cessation, and relapse. These transitions are triggered with the initiation, cessation, and relapse threshold values. In order to obtain the threshold values, a calibration analysis is performed using the data of tobacco use prevalence of different years. The calibration analysis is performed to minimize the difference between the real and simulated numbers of NS, CS, and from 2008 to 2016. The calibration analysis of the proposed model is performed with a smaller version of the model using the calibration tool of the AnyLogic Simulation Modeling software in order to eliminate memory problems.

When the prevalence of tobacco use is searched, it can be seen that the number of the CS is increased at 2014 unexpectedly in the surveys. This case is not compatible with the other years when compared to other years. Thus, the calibration analysis is re-arranged by excluding the effects of this year. The model is calibrated using the data from 2008 to 2014. The increase at the year 2014 is represented with the addition of a new factor called as other factors. The calibrated threshold values are found at the end of the analysis as 0.899 for initiation, 0.977 for cessation, and 0.92 for relapse.

5.3.8 Scenarios

Several different tobacco control policies have been implemented by Turkey over the years. The proposed model includes some fundamental policies in order to evaluate the effects of policies on tobacco use behaviors of people. When the simulation is started, some of the policies have been in use in the Turkey such as taxation, anti-tobacco advertisements, and health warnings. Pictorial warning was implemented in 2010 and is embedded to the model at the relevant year. Apart from these, the plain package regulation was started at the end of 2019. The possible effects of the plain package were not measured with a survey like the GATSs before for Turkey. Thus, one of the important contributions of the presented thesis is to examine the potential effects of plain package application using the hypothetical scenarios.

The world started to face with a new global treat on the dates the plain package regulation came into force. A vital pandemic that was called as corona virus disease 2019 (COVID19) was first come out in China and then spread to all over the world. Although all information about the disease is still not known, its adverse effects on current smokers have been discussed severely. The WHO warned people and stated that "Smokers are more likely to be more vulnerable to COVID19 due to the contacting with the lips for smoking and the possible damage at their lungs because of smoking." [91]. In addition, COVID-19 did not only affect the health but also affected the psychology of people due to the curfew, social isolations, economic problems and etc. On the other hand, the WHO also expresses that the COVID-19 pandemic has led millions of smokers saying they want to give up smoking [92].

In the light of the information given above, COVID19 can become an important factor that cause changes at tobacco use behaviors of people and the relations between COVID19 and tobacco use cannot be underestimated.

In this thesis, the potential effects of COVID19 on tobacco use behaviors were categorized in two classes:

- It could lead people to stop tobacco use due to the health consequences
- It could lead people to start tobacco use due to the psychological factors.

In order to examine the potential effects of newly implemented plain package regulation and COVID19 pandemic, three different scenarios are set up. In these scenarios, the threshold values that are calculated at the end of calibration analyses are changed in order to allow tobacco use behavior change.

- Base Scenario: In this scenario, the tobacco use and control status of Turkey before the plain package regulation and COVID19 pandemic is preserved. It means that no effect is expected from these two different factors.
- Optimistic Scenario: In this scenario, the model is expected that the plain package regulation and COVID19 pandemic lead people to quit tobacco use Thus, these new factors can decrease the cessation threshold value by %25 while increasing the initiation and relapse threshold values by %25. Therefore, the prevalence of tobacco use can decrease.
- Pessimistic Scenario: In this scenario, the model is expected that the plain package regulation and COVID19 pandemic cannot prevent tobacco use and the widespread use of tobacco would increase. Thus, these factors can increase the cessation threshold value by %25 while decreasing the initiation and relapse threshold values by %25.

The scenarios are used to reveal agents from which groups are more likely to use tobacco or more likely to quit tobacco. Thus, the findings help to observe more vulnerable groups that the policy makers can be able to make them quit tobacco use and more resistant groups that the policy makers can generate new policies to increase their understanding about the harms of tobacco use.

5.4 Results and Discussions

This thesis mainly proposes a hybrid AB model integrated with the multi stage learning based FCM for understanding tobacco use behaviors of people. The model firstly imitates the current tobacco users and control approaches of Turkey. Subsequently, the model helps to reveal how different characteristics can change the tobacco use behaviors of people under different scenarios.

The imitation performance of the proposed hybrid model is evaluated using the data of GATSs Turkey 2008, 2012, and 2016 because it is really important to reflect the real-world system to the model in the virtual world more realistically. Thus, the outputs of the model for these years are used to test the model's imitation performance. The model finds that the percentages of the prevalence of tobacco use among adult population %30.79 for 2008, %29.80 for 2012, and %31.32 for 2016 as seen in table 5.3.

The model also predicts that the % 14.8, % 15.35, and % 18.96 of females and %47.53, %44.75, and 44.09% of males are current smokers at 2008, 2012, and 2016 respectively while the expected percentages are % 15.20, % 13.10, and % 19.20 of females and % 47.90, % 41.50, and % 44.10 of males for the same years.

Table 5.3 The expected and predicted percentages of current smokers, former smokers, and never smokers for 2008, 2012, and 2016

	2008		**2012**		**2016**	
	Expected	Predicted	Expected	Predicted	Expected	Predicted
Current Smokers (%)	31.20	30.79	27.10	29.80	31.60	31.32
Former Smokers (%)	15.90	15.72	13.10	14.61	13.60	13.25
Never Smokers (%)	52.80	53.48	59.80	55.58	54.80	55.42

The model performs scenario analyses in order to observe the potential effects of two important factors. One of them is the implementation of plain package regulation. Turkey is one of the leading countries that take remarkable precautions to control the prevalent usage of tobacco. Although the implementation of a new tobacco control policy (plain package) may not be an uncertainty for Turkey, the timing of the policy could change the effects of the policy because when the plain package regulation was came into force, COVID19 pandemic came out in China. The adverse effects of COVID19 on tobacco users have been searched. Therefore, the effects of COVID19 on the prevalence of tobacco use cannot be ignored. For that reason, three different scenarios are made for analyzing the effects of plain package regulation and COVID19 pandemic.

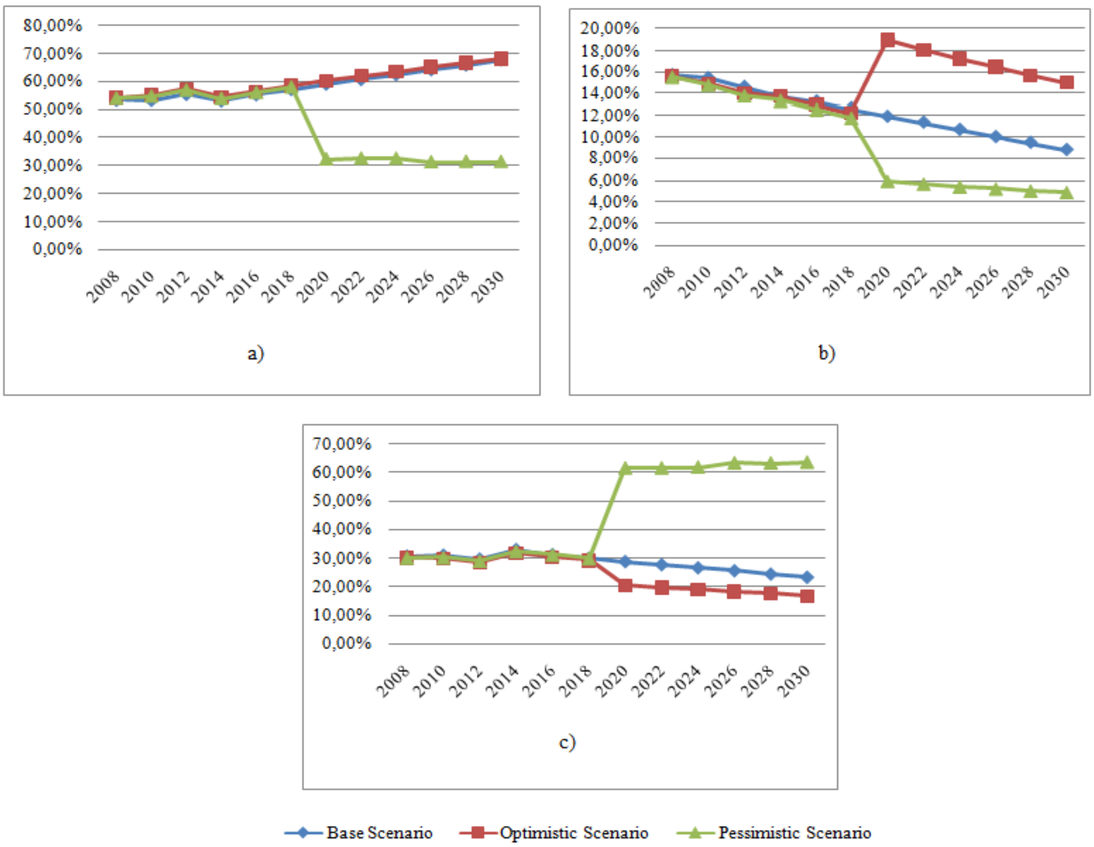

Figure 5.4 The findings of the base, optimistic and pessimistic scenarios for a) never smokers b) former smokers and c) current smokers by years.

The findings of the base, optimistic and pessimistic scenarios for the percentages of current smokers, never smokers, and former smokers are given in Figure 5.4. If the optimistic scenarios happen and the plain package regulation and COVID19 leads agents to quit tobacco use, the distribution of the current smokers in the population can decrease to %17 at 2030. On contrary, if the pessimistic scenario happens and people starts to tobacco use, the prevalence of tobacco use can increase to %63 at 2030.

It is important to point that the reduction and the increase at the percentages of tobacco users at 2020 for optimistic scenario and pessimistic scenario respectively are remarkable because the initial effects of both factors started to be observed in the 2020. Thus, the biggest effects of plain package and COVID19 were seen in 2020 and then these effects decreased through 2030.

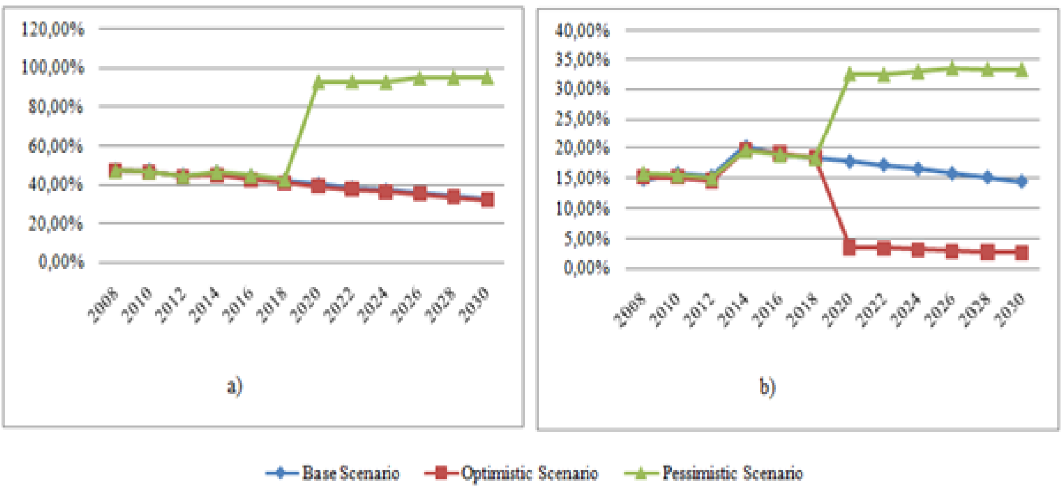

Figure 5.5 The findings of the base, optimistic and pessimistic scenarios for a) males, b) females by years.

Different from the distribution of the current smokers, former smokers, and never smokers, the hybrid model also allows examining the prevalence of tobacco use among different agent groups. The change at the tobacco use behaviors of people can be examined using their different demographic characteristics such as gender, age group, residence, and work status and education level. Understanding the tobacco use behaviors of different agent groups is a really vital process for modelers and policy makers because the findings can represent the vulnerable and resistant groups to tobacco control policies.

The first agent group that is examined in this thesis is males and females. The findings of the scenarios for these agent groups are given in the Figure 5.5. According to the results of the base scenario, if the current tobacco use status does not change, the distributions of current smokers among males and females are expected to decrease.

On the other hand, the findings of the optimistic and pessimistic scenarios represent interesting outputs. Males indicate similar outputs in the optimistic scenario with the base scenario. It can be seen that there is almost no difference between base and optimistic scenarios for male agent group. On the other hand, the results for females represent that most of the female tobacco users tend to give up tobacco use in the optimistic scenario and the percentage of female tobacco users is less than % 5.

In the pessimistic scenario, the outputs of the model indicate that the percentages of current smokers among both gender groups increase. However, males are more likely to start tobacco use when compared to females because the percentages of male tobacco users are much more than female tobacco users.

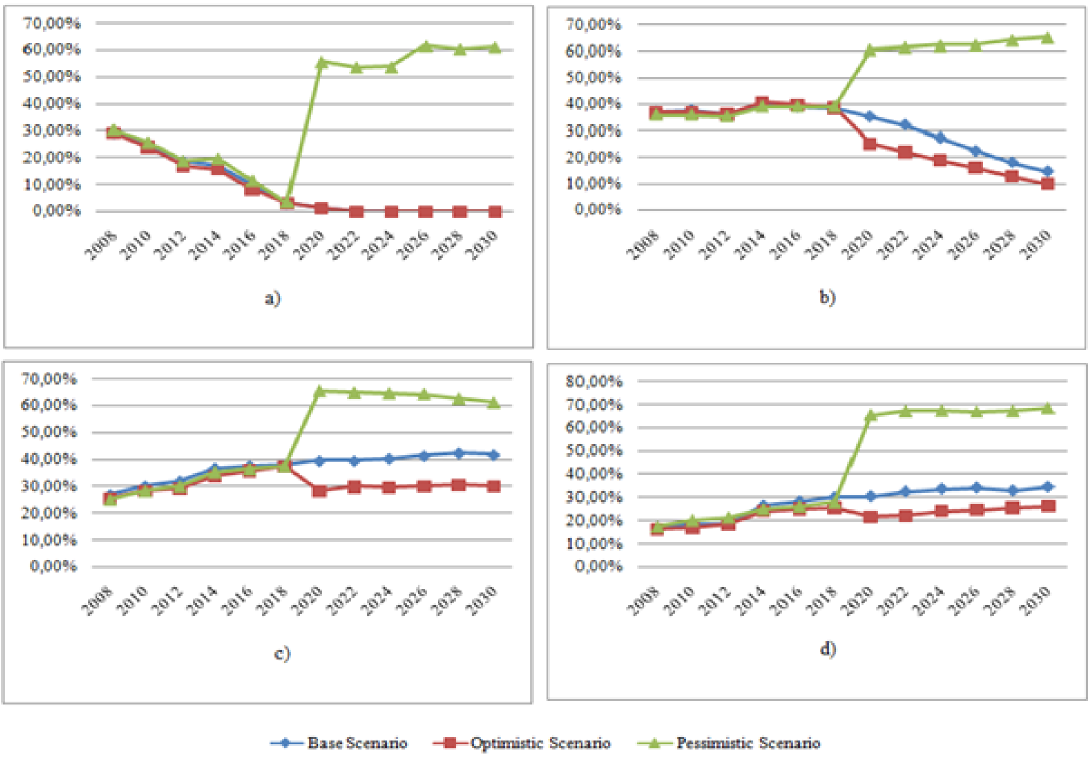

Figure 5.6 The findings of the base, optimistic and pessimistic scenarios for age groups a) 15-24, b) 25-44, c)45-64, d)65+ by years.

The findings of the gender groups reveal that a new implemented policy can provide positive effects on females because females are more vulnerable to the policies and they are more likely to quit when compared to males. On the other hand males are really resistant to quit. Although they face with new policies or diseases, they insist to use tobacco.

The tobacco use behaviors of agent according to their age groups and residence types are also investigated in this thesis. The findings for age group are given in Figure 5.6 while the outputs for residence types are given in Figure 5.7. The proposed hybrid model provides similar results for all residence types and age groups in the base and optimistic scenarios. The prevalence of tobacco use among 15-24, 25-44 age group, urban and rural residents decrease in the both scenarios. On the contrary, the distribution of current smokers increases for other age groups (45-64, 65+) in the same scenarios. In the pessimistic scenario, people from all residence types and age groups are more likely to use tobacco. Thus, the percentages of current smokers increase. However, the results of the 15-24 age group in the base and optimistic scenarios are around %0 after 2020. Although the hybrid model provides promising results for imitation of the tobacco use and control of Turkey, it can still have some problems to represent the real life system entirely. For that reason, this age group should be re-examined at the further analysis.

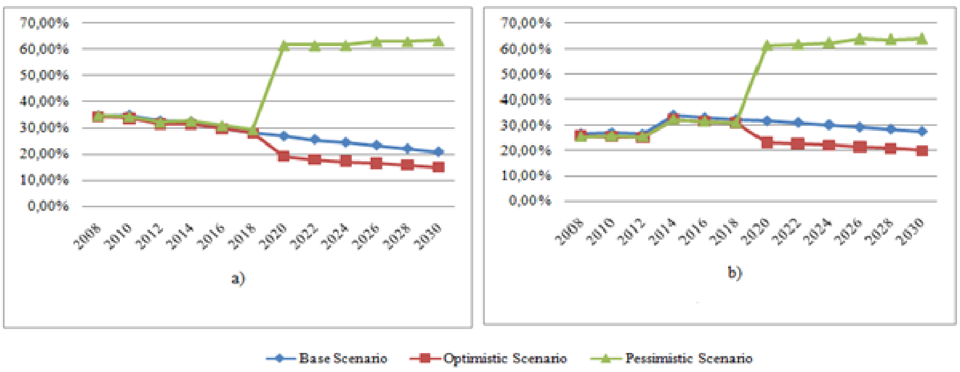

Figure 5.7 The findings of the base, optimistic and pessimistic scenarios for residence a) urban, b) rural by years.

Tobacco use behaviors of agents are also examined according to their work status. The findings shows a decreasing trend for all work status excluding the retired agents in the base scenario. The model predicts a slight increase for retired agents in the base scenario while predicting reduction in the optimistic scenario. The results of the optimistic scenario generally represent a decreasing trend for all work status. However, the results of the home maker agents are interesting. As seen in Figure 5.8-e, homemakers are more like to quit tobacco use when compared to other work categories in the optimistic scenario. The main reason behind this can be the male and female distributions within homemaker agents. The number of female homemakers is much more than male homemakers. Therefore, the female population in the homemaker group can be a major cause for the high decrease at the percentage of tobacco users for this group. The results of the pessimistic scenario for work status represent similar consequences like gender, age group, and residence. In the worst case, the percentages of tobacco users increase for all work status groups.

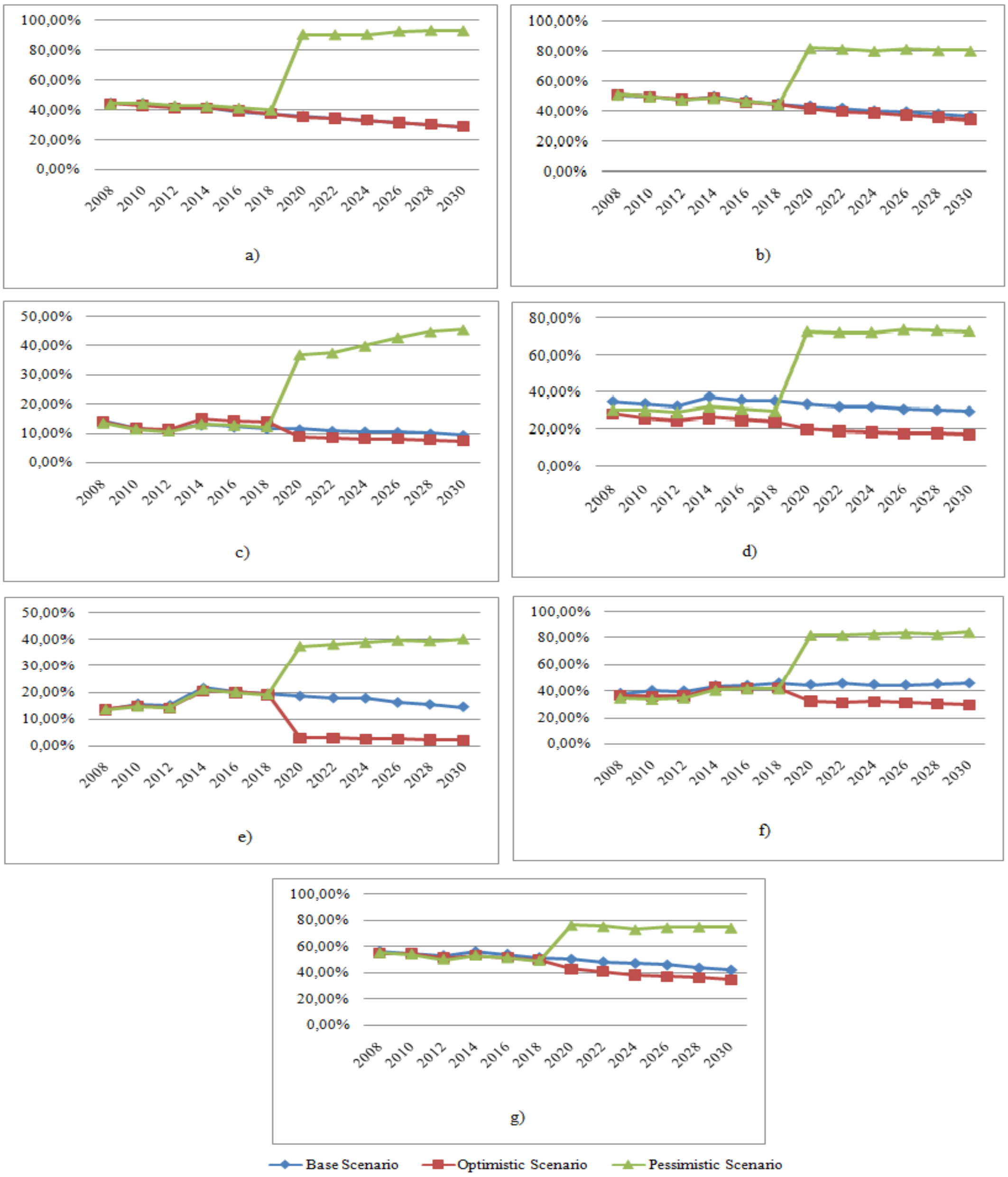

Figure 5.8 The findings of the base, optimistic and pessimistic scenarios for work status a) paid employee, b) self employed, c) non-paid family worker, d) student, e) homemaker, f) retired, and g) no job by years.

The final analysis is for education levels of the agents. The comparisons of the results of the scenarios are represented in the Figure 5.9. According to the base scenario, the percentages of tobacco users among high school graduates and university or higher graduates represent a decreasing trend while people who are not graduated, primary educated, and secondary school graduates are more likely to start smoking. On the other hand, although the distributions of current smokers decrease during 2020 significantly for all education levels, this decreasing trend may change after 2020 for not graduate, primary education, and secondary education graduates groups in the optimistic scenario.

The mentioned education level results reveal that that when the education level of agents increases, their likelihood of quit smoking also increases parallel to their education level. Thus, explaining the importance of quitting smoking to the not graduated, primary and secondary educated people can help to improve the performance of the tobacco control policies.

Consequently, the outputs of the proposed hybrid model at the end of the three different scenario analyses are given above. The results reveal the changes at tobacco use behaviors of people under different situations. Some of the findings may be really helpful to understand the effects of tobacco control policies and unexpected situation on tobacco use behaviors of people. Thus, the results of the presented thesis can help policy makers to understand the different consequences of the tobacco control policies for various population groups.

Figure 5.9 The findings of the base, optimistic and pessimistic scenarios for education levels a) not graduated, b) primary education, c) secondary education, d) high school education, and e) university or higher education by years.

5.5 Concluding Remarks

This chapter of the thesis presents the details of the proposed hybrid AB model integrated with a multi-stage learning based FCM in order to examine tobacco use behaviors of Turkey. The learning procedure that is explained in detail in the previous chapter is embedded to the AB model to make a decision support mechanism for agents. In addition, tobacco use and control of Turkey is reflected considering the demographic characteristics (gender, age, residence, work status, and education level) and tobacco use behaviors (never smoker, current smoker, and former smoker) of agents, tobacco control relevant factors (notice of anti-tobacco advertisement, notice of health warnings, notice of tobacco advertisements, taxation level, and notice of pictorial warning) and social network. The effects of newly implemented plain package regulation and COVID19 pandemic on tobacco use behaviors of agents are investigated under three different scenarios: base, optimistic, and pessimistic scenarios.

The findings generally represent that if plain package and COVID-19 related issues such as curfew, social isolation, economical problems and etc. affect people adversely and lead people to start tobacco use, a considerable increase at the percentage of current smokers could be expected for almost all groups of people. On the contrary, if the potential damages of tobacco use to immunity of people lead people to quit tobacco use and the optimistic scenario happens, a vital reduction at the percentage of current smokers could be expected. Apart from these results, the model reveals very helpful information about tobacco use behaviors of people from different groups, especially for gender, education level and work status. Females and homemakers are more likely to quit tobacco use when compared males and people from other work groups. In addition, when the education level of people increases, their tendency to quit also increases. Thus, higher educated people can understand the negative effects of tobacco use better. These important consequences of the thesis indicate vulnerable and resistant groups against policies to policy makers. Therefore, increasing the understanding of the resistant group about the risk of tobacco use can help to decrease prevalence of tobacco use. Targeting the resistant groups and improving tobacco control policies to increase the effects of policies on these groups can be a reliable way to control tobacco use in Turkey.

CHAPTER VI
CONCLUSIONS

6.1 Introduction

This PhD thesis proposes a novel hybrid agent based model for understanding tobacco use behaviors of people in Turkey. For this aim, the concept of s multi-stage learning based fuzzy cognitive map is embedded into the agent based model. The agent based modeling is used to imitate individual level behaviors of people while fuzzy cognitive map is used to understand mental models and decision making mechanism of people. Demographic characteristics (gender, age, residence, education level, and work status), tobacco use behaviors (never, current, and former smokers) and the structure of social networks of people and tobacco control policies of Turkey (taxation, health warnings, pictorial warnings, notice anti-tobacco advertisements, and plain package) are included to the proposed hybrid model.

The casual relations between demographic characteristics, tobacco control policies, and likelihood of tobacco use of people are analyzed at each model time unit for people using fuzzy cognitive maps. In order to eliminate the increase the accuracy and robustness of the obtained maps at the end of the fuzzy cognitive map analyses, the usage of expert knowledge is restricted and a multi-stage learning procedure is integrated to the thesis. The multi-stage learning procedure is started with the well known non-linear hebbian learning algorithm and then continues with the extended great deluge algorithm. The performance of the mentioned procedure is initially tested to observe its capability to represent the relations between demographic characteristics and likelihood of tobacco use of people. After the accuracy of the procedure is verified, the procedure is embedded to the agents based model to work as a decision support mechanism.

The proposed hybrid model is built to represent tobacco use behaviors of adult population Turkey from 2008 to 2018. After that, three different scenarios are prepared to examine the potential effects of a new tobacco control policy (plain package regulation) and COVID19 pandemic on tobacco use behaviors of agents who had different demographic characteristics. These scenarios are: base scenario that assumes no change on tobacco use behaviors after plain package regulation and COVID19, optimistic scenario that assumes the increase at the tendency of people to quit tobacco use, and pessimistic scenario that assumes the increase at the tendency of people to start tobacco use.

6.2 The Findings of the Thesis

This thesis integrates the agent based modeling and fuzzy cognitive maps in order to imitate tobacco use behaviors of people.

The steps, benefits, and application of the multi-stage learning based fuzzy cognitive maps are represented in the Chapter 4. The FCM analyses are initialized with a random initial weight matrix. Subsequently, the initial knowledge is trained with non-linear hebbian algorithm. Finally, the outputs of the non-linear hebbian algorithm are used as the inputs of extended great deluge algorithm (EGDA). The implemented procedure helps to restrict the usage of expert knowledge and increase the accuracy and robustness of the maps. The causal relations between demographic characteristics and likelihood of tobacco use of people are provided at the end of the analyses. The results indicate that this procedure is really successful to represent the effects of causal relations when compared to one-stage learning with non-linear hebbian learning algorithm.

In the Chapter 5, the multi-stage learning based fuzzy cognitive map analyses are integrated with the agent based model to represent agents' decision about changing tobacco use behaviors. The proposed model is used for examining different scenarios. The outputs of the model provide valuable information about tobacco use behaviors of people. The results mainly imply that tobacco use behaviors of people who have different characteristics could represent different reactions to tobacco control policies and unexpected issues. Females, homemakers, and people with high education levels are more likely to quit tobacco use when compared to other groups. Males and people with low education level are resistant to quit tobacco use.

6.3 Implications for Policy Makers

This thesis can contribute to modeling and tobacco use relevant literature separately. While the design of the modeling process can contribute to the modeling literature, the findings of the scenario analyses can help to increase our understanding on tobacco use behaviors of Turkish people. Tobacco use is one of the most important public health problems for Turkey like many different countries. In order to decrease the prevalence of tobacco use in Turkey, Turkey has implemented several different tobacco control policies. Although Turkey fights against to tobacco use severely, the prevalent usage of tobacco products could not be eliminated completely. For that reason, performing new researches that can improve our understandings on this topic may help to increase our gains from the existing and future tobacco control policies.

This thesis focuses on to examine tobacco use behaviors of Turkish people considering demographic characteristics, tobacco control policies and social network effect. The findings of the scenario analyses reveal that the reactions of the demographic groups to tobacco control policies and unexpected issues can vary. While males are really resistant to change their tobacco use behaviors due to new policies and unexpected issues, females are more willing to quit tobacco use. Thus, although new policies can lead females to quit tobacco use, the same result may not be expected for males. In order to decrease tobacco use in Turkey, males are needed to convince quitting. For this aim, policy makers can develop new methods to affect the opinions of males. In addition, people who aged 45 and over and people who have low education level are also resistant group for policy makers. The behaviors of these demographic groups can be studied in more detail to understand the reasons behind their tobacco use behaviors. New tobacco control policies that can target these resistant groups can also be implemented for taking the prevalence of tobacco use under control.

6.4 Limitations of the Thesis and Future Work

Although there have been various contributions of this thesis to agent based modeling, fuzzy cognitive maps, and tobacco use literatures, there are also some limitations that may be investigated in future analyses.

The presented thesis focuses on modeling an important complex system. Tobacco use consists of several different factors from demographic factors to environmental factors that can affect tobacco use behaviors of people. In this thesis, demographic factors and tobacco control policies are basically included in the study. Although the model provides valuable information for these factors, future studies can also examine the effect of different factors such as psychological, health related factors and etc.

The model is built using an open population with births and deaths. However, population can also vary with migration. In this thesis, the effect of migration is ignored due to the lack of detailed information about immigrants. The future studies can also search the effects of migration on tobacco use.

In addition, a random social network is used in the proposed model. Forming a detailed social network that represent the social relations of individuals for tobacco use can be a vast research topic because there are many different relation that affect tobacco use behaviors of people such as parents, peers, colleagues, and etc. Therefore, analyzing the structure of social networks of people for Turkey can be a remarkable research topic for future studies.

REFERENCES

[1] Morgan, M. S., Morrison, M. (1999). *Models as mediators: perspectives on natural and social science*. Cambridge University Press, Cambridge, UK

[2] Rothenberg, J. (1989). *The Nature of Modeling*. Santa Monica, CA: RAND Corporation.

[3] Maidstone, R. (2012). Discrete Event Simulation, System Dynamics and Agent Based Simulation: Discussion and Comparison. Lancaster University

[4] Borshchev, A., Filippov, A. (2004). From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools. *The 22nd International Conference of the System Dynamics Society*

[5] Borshchev, A. (2013). The Big Book of Simulation Modeling: Multimethod Modeling with AnyLogic 6.

[6] Axelrod, R. (1997). *Advancing the Art of Simulation in the Social Sciences*. In: Conte R, Hegselmann R, Terna P (eds) Simulating Social Phenomena. Springer, Berlin, Heidelberg, 21–40.

[7] World Health Organization. Tobacco. https://www.who.int/news-room/fact-sheets/detail/tobacco. 13.12.2018.

[8] Hays, J. T., Ebbert, J. O. (2010). Adverse Effects and Tolerability of Medications for the Treatment of Tobacco Use and Dependence. *Drugs* 70:2357–2372.

[9] CDC TobaccoFree. (2017). Health Effects of Smoking and Tobacco Use. In: Centers for Disease Control and Prevention. https://www.cdc.gov/tobacco/basic_information/health_effects/index.htm. 11.09.2020.

[10] Durmuşoğlu, Z. D. U., Çiftçi, P. K. (2017). An analysis of trends in publications on 'tobacco control.' *Health Education Journal* 76:544–556.

[11] Durmuşoğlu, A. (2016). A pre-assessment of past research on the topic of environmental-friendly electronics. J*ournal of Cleaner Production* 129:305–314.

[12] Sterman, J. (2000). *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin/McGraw-Hill

[13] Banks, J., II, J. S. C., Nelson, B. L., Nicol, D. M. (2009). *Discrete-Event System Simulation*, 5th Edition. Pearson, Upper Saddle River

[14] Siebers, P.O., Macal, C. M., Garnett, J., et al (2010). Discrete-event simulation is dead, long live agent-based simulation! *Journal of Simulation* 4:204–210.

[15] Leahy, J. E., Reeves, E. G., Bell, K. P., et al (2013). Agent-Based Modeling of Harvest Decisions by Small Scale Forest Landowners in Maine, USA. In: International Journal of Forestry Research. https://www.hindawi.com/journals/ijfr/2013/563068/. 14.09.2020

[16] Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *PNAS* 99:7280–7287.

[17] Singh, V. K., Gupta, A. K. (2009). Agent based models of social systems and collective intelligence. In: *2009 International Conference on Intelligent Agent Multi-Agent Systems*. 1–7.

[18] Aguilar, J. (2002). A*daptive Random Fuzzy Cognitive Maps*. In: Garijo FJ, Riquelme JC, Toro M (eds) Advances in Artificial Intelligence — IBERAMIA 2002. Springer, Berlin, Heidelberg, pp 402–410

[19] Stylios, C. D., Georgopoulos, V. C., Malandraki, G. A., Chouliara, S. (2008). Fuzzy cognitive map architectures for medical decision support systems. *Applied Soft Computing* 8:1243–1251.

[20] Papageorgiou, E. I., Stylios, C., Groumpos, P. P. (2006). Unsupervised learning techniques for fine-tuning fuzzy cognitive map causal links. *International Journal of Human-Computer Studies* 64:727–743.

[21] Giabbanelli, P. J., Gray, S.A., Aminpour, P. (2017). Combining fuzzy cognitive maps with agent-based modeling: Frameworks and pitfalls of a powerful hybrid modeling approach to understand human-environment interactions. *Environmental Modelling and Software* 95:320–325.

[22] Kocabey Çiftçi, P., Unutmaz Durmuşoğlu, Z. D. (2020). A multi-stage learning-based fuzzy cognitive maps for tobacco use. Neural Computing and Applications https://doi.org/10.1007/s00521-020-04860-4

[23] Stula, M., Stipanicev, D., Bodrozic, L. (2010). Intelligent modeling with agent-based fuzzy cognitive map. *International Journal of Intelligent Systems* 25:981–1004.

[24] Lee, K. C., Lee, H., Lee, N., Lim, J. (2013). An agent-based fuzzy cognitive map approach to the strategic marketing planning for industrial firms. *Industrial Marketing Management* 42:552–563.

[25] Giabbanelli, P. J. (2014). Computational models of chronic diseases: understanding and leveraging complexity. Thesis, Science: Department of Biomedical Physiology and Kinesiology.

[26] Giabbanelli, P., Fattoruso, M., Norman, M. L. (2019). CoFluences: Simulating the Spread of Social Influences via a Hybrid Agent-Based/Fuzzy Cognitive Maps Architecture. In: *Proceedings of the 2019 ACM SIGSIM Conference on Principles of Advanced Discrete Simulation.* Association for Computing Machinery, New York, NY, USA, pp 71–82

[27] Gras, R., Devaurs, D., Wozniak, A., Aspinall, A. (2009). An individual-based evolving predator-prey ecosystem simulation using a fuzzy cognitive map as the behavior model. *Artificial Life* 15:423–463

[28] Khater, M., Salehi, E., Gras, R. (2012). T*he Emergence of New Genes in EcoSim and Its Effect on Fitness.* In: Bui LT, Ong YS, Hoai NX, et al (eds) Simulated Evolution and Learning. Springer, Berlin, Heidelberg, pp 52–61

[29] Mehryar, S., Sliuzas, R., Schwarz, N., et al (2019). From individual Fuzzy Cognitive Maps to Agent Based Models: Modeling multi-factorial and multi-stakeholder decision-making for water scarcity. *Journal of Environmental Management* 250:109482.

[30] Davis, C. W. H., Giabbanelli, P. J., Jetter, A. J. (2019). The Intersection of Agent Based Models and Fuzzy Cognitive Maps: A Review of an Emerging Hybrid Modeling Practice. I*n: 2019 Winter Simulation Conference (WSC)*. 1292–1303

[31] Press TM. Growing Artificial Societies | The MIT Press. https://mitpress.mit.edu/books/growing-artificial-societies. 14.09.2020.

[32] Mashkova, A., Savina, O., Mamatov, A. V. (2018). Integrating Artificial Agent in the Simulation Model of the Russian Federation Spatial Development. I*n: 2018 IEEE 12th International Conference on Application of Information and Communication Technologies (AICT)*. 1–6.

[33] Taherian, M., Mousavi, S. M. (2017). Modeling and simulation of forward osmosis process using agent-based model system. *Computers & Chemical Engineering* 100:104–118.

[34] Bankes, S. C. (2002). Agent-based modeling: A revolution? *PNAS* 99:7199–7200.

[35] Macal, C. M., North, M. J. (2009). Agent-based modeling and simulation. *In: Proceedings of the 2009 Winter Simulation Conference (WSC)*. 86–98

[36] Patt, A., Siebenhüner, B. (2005). Agent Based Modeling and Adaption to Climate Change. *Vierteljahrshefte zur Wirtschaftsforschung* 74:310–320.

[37] Khazaii, J. (2016). *Agent-Based Modeling.* In: Khazaii J (ed) Advanced Decision Making for HVAC Engineers: Creating Energy Efficient Smart Buildings. Springer International Publishing, Cham, pp 137–144.

[38] Patlolla, P., Gunupudi, V., Mikler, A. R., Jacob, R. T. (2006). *Agent-Based Simulation Tools in Computational Epidemiology*. In: Böhme T, Larios Rosillo VM, Unger H, Unger H (eds) Innovative Internet Community Systems. Springer, Berlin, Heidelberg, 212–223.

[39] Macal, C. M. (2010). To agent-based simulation from System Dynamics. *In: Proceedings of the 2010 Winter Simulation Conference*. 371–382

[40] Papakostas, G. A., Polydoros, A. S., Koulouriotis, D. E., Tourassis, V. D. (2011). Training Fuzzy Cognitive Maps by using Hebbian learning algorithms: A comparative study. *In: 2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*.851–858

[41] Stach, W., Kurgan, L., Pedrycz, W. (2008). Data-driven Nonlinear Hebbian Learning method for Fuzzy Cognitive Maps. *In: 2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*.1975–1981

[42] Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies* 24:65–75.

[43] Bertolini, M. (2007). Assessment of human reliability factors: A fuzzy cognitive maps approach. *International Journal of Industrial Ergonomics* 37:405–413.

[44] Papageorgiou, E. I., Poczęta, K., Laspidou, C. (2015). Application of Fuzzy Cognitive Maps to water demand prediction. I*n: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*.1–8

[45] Baykasoglu, A., Durmusoglu, Z. D. U., Kaplanoglu, V. (2011). Training Fuzzy Cognitive Maps via Extended Great Deluge Algorithm with applications. *Computers in Industry* 62:187–195.

[46] Dickerson, J. A., Kosko, B. (1994). Virtual Worlds as Fuzzy Cognitive Maps. *Presence: Teleoperators and Virtual Environments* 3:173–189.

[47] Mago, V. K., Morden, H. K., Fritz, C., et al (2013). Analyzing the impact of social factors on homelessness: a fuzzy cognitive map approach. BMC Medical Informatics and Decision Making 13:94.

[48] Papageorgiou, E. I., Spyridonos, P. P., Glotsos, D. T. H., et al (2008). Brain tumor characterization using the soft computing technique of fuzzy cognitive maps. *Applied Soft Computing* 8:820–828.

[49] Papageorgiou, E. I., Papandrianos, N. I., Apostolopoulos, D. J., Vassilakos, P. J. (2008). Fuzzy Cognitive Map based decision support system for thyroid diagnosis management. *In: 2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*. 1204–1211

[50] Senniappan, V., Subramanian, J., Papageorgiou, E. I., Mohan, S. (2017). Application of fuzzy cognitive maps for crack categorization in columns of reinforced concrete structures. *Neural Computing and Applications* 28:107–117.

[51] Papageorgiou, E. I., Salmeron, J. L. (2013). A Review of Fuzzy Cognitive Maps Research During the Last Decade. *IEEE Transactions on Fuzzy Systems* 21:66–79.

[52] Jahangoshai Rezaee, M., Yousefi, S., Hayati, J. (2018). A decision system using fuzzy cognitive map and multi-group data envelopment analysis to estimate hospitals' outputs level. *Neural Computing and Applications* 29:761–777.

[53] Huerga, A. V. (2002). A Balanced Differential Learning algorithm in Fuzzy Cognitive Maps. *The 16th International Workshop on Qualitative Reasoning.*

[54] Papageorgiou, E., Stylios, C., Groumpos, P. (2003). *Fuzzy Cognitive Map Learning Based on Nonlinear Hebbian Rule.* In: Gedeon T (Tom) D, Fung LCC (eds) AI 2003: Advances in Artificial Intelligence. Springer Berlin Heidelberg, 256–268

[55] Papageorgiou, E. I., Stylios, C. D., Groumpos, P. P. (2004). Active Hebbian learning algorithm to train fuzzy cognitive maps. *International Journal of Approximate Reasoning* 37:219–249.

[56] Ren, Z. (2012). Learning Fuzzy Cognitive Maps by a Hybrid Method Using Nonlinear Hebbian Learning and Extended Great Deluge Algorithm. *In: MAICS*

[57] Koulouriotis, D. E., Diakoulakis, I. E., Emiris, D. M. (2001). Anamorphosis of fuzzy cognitive maps for operation in ambiguous and multi-stimulus real world environments. *In: 10th IEEE International Conference on Fuzzy Systems*. 1156–1159.

[58] Papageorgiou, E. I., Parsopoulos, K. E., Groumpos, P. P., Vrahatis, M. N. (2004). *Fuzzy Cognitive Maps Learning through Swarm Intelligence*. In: Rutkowski L, Siekmann JH, Tadeusiewicz R, Zadeh LA (eds) Artificial Intelligence and Soft Computing - ICAISC 2004. Springer Berlin Heidelberg, 344–349

[59] Papageorgiou, E. I., Parsopoulos, K. E., Stylios, C. S., et al (2005). Fuzzy Cognitive Maps Learning Using Particle Swarm Optimization. *Journal of Intelligent Information Systems*. 25:95–121.

[60] Song, H., Miao, C., Shen, Z., Miao, Y. (2008). Fuzzy Cognitive Map Learning Based on Multi-Objective PSO (Invited Paper). *International Journal of Computational Cognition*. 6:51-59.

[61] Stach, W., Kurgan, L., Pedrycz, W., Reformat, M. (2005). Genetic learning of fuzzy cognitive maps. *Fuzzy Sets and Systems* 153:371–401.

[62] Ghazanfari, M., Alizadeh, S., Fathian, M., Koulouriotis, D. E. (2007). Comparing simulated annealing and genetic algorithm in learning FCM. *Applied Mathematics and Computation* 192:56–68.

[63] Alizadeh, S., Ghazanfari, M., Jafari, M., Hooshmand, S. (2007). Learning FCM by Tabu Search. *International Journal of Computer Science*, 3:142-149.

[64] Unutmaz Durmuşoğlu, Z. D., Kocabey Çiftçi, P. (2014). A study of factors affecting the tobacco use. *Joint Symposium on 44th Computers & Industrial Engineering (CIE44) & 9th Intelligent Manufacturing and Service Systems (IMSS14)*, 1287-1295.

[65] Ahmadi, S., Forouzideh, N., Alizadeh, S., Papageorgiou, E. (2015). Learning Fuzzy Cognitive Maps using Imperialist Competitive Algorithm. *Neural Computing and Applications* 26:1333–1354.

[66] National Cancer Institute. (2007). G*reater Than the Sum: Systems Thinking in Tobacco Control*. National Cancer Institute, U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health

[67] Papageorgiou, E. I., Salmeron, J. L. (2014). *Methods and Algorithms for Fuzzy Cognitive Map-based Modeling.* In: Papageorgiou EI (ed) Fuzzy Cognitive Maps for Applied Sciences and Engineering: From Fundamentals to Extensions and Learning Algorithms. Springer, Berlin, Heidelberg, 1–28.

[68] World Health Organization. Global Adult Tobacco Survey (GATS). In: WHO. http://www.who.int/tobacco/surveillance/survey/gats/en/. 14.06.2016

[69] SPSS. What is a “Chi-Square Test”? https://www.spss-tutorials.com/chi-square-independence-test/. 13.12.2018

[70] Using Chi-Square Statistic in Research. In: Statistics Solutions. https://www.statisticssolutions.com/using-chi-square-statistic-in-research/. 13.12.2018.

[71] Peters, G., Crespo, F., Lingras, P., Weber, R. (2013). Soft clustering – Fuzzy and rough approaches and their extensions and derivatives. *International Journal of Approximate Reasoning* 54:307–322.

[72] Hatamlou, A., Abdullah, S., Nezamabadi-pour, H. (2011). *Application of Gravitational Search Algorithm on Data Clustering*. In: Yao J, Ramanna S, Wang G, Suraj Z (eds) Rough Sets and Knowledge Technology. Springer Berlin Heidelberg, 337–346

[73] Olson, D. L., Shi, Y. (2007). *Introduction to Business Data Mining*. McGraw Hill

[74] Northeastern University. Cluster Analysis. http://www.ccs.neu.edu/home/futrelle/teaching/isu535sp2004/finalpapers/clusteringIntro.html. 13.12.2018.

[75] Kannappan, A., Tamilarasi, A., Papageorgiou, E. I. (2011). Analyzing the performance of fuzzy cognitive maps with non-linear hebbian learning algorithm in predicting autistic disorder. *Expert Systems with Applications* 38:1282–1292.

[76] Baykasoğlu, A., Owen, S., Gindy, N. (1999). A Taboo Search Based Approach to Find the Pareto Optimal Set in Multiple Objective Optimization. *Engineering Optimization* 31:731–748.

[77] Kilic, D., Ozturk, S. (2014). Gender differences in cigarette consumption in Turkey: Evidence from the Global Adult Tobacco Survey. *Health Policy* 114:207–214.

[78] İlhan, M. N., Arıkan, Z., Kotan, Z., et al (2016). Prevalence and Socio-Demographic Determinants of Tobacco, Alcohol, Substance Use and Drug Misuse in General Population in Turkey. *Noro Psikiyatri Arsivi* 53:205–212.

[79] Levy, D. T., Friend, K., Polishchuk, E. (2001). Effect of clean indoor air laws on smokers: the clean air module of the SimSmoke computer simulation model. *Tobacco Control* 10:345–351.

[80] Levy, D., Almeida, L. M., Szklo, A. (2012). The Brazil SimSmoke Policy Simulation Model: The Effect of Strong Tobacco Control Policies on Smoking Prevalence and Smoking-Attributable Deaths in a Middle Income Nation. *PLOS Medicine* 9:e1001336.

[81] Near, A. M., Blackman, K., Currie, L. M., Levy, D. T. (2014). Sweden SimSmoke: the effect of tobacco control policies on smoking and snus prevalence and attributable deaths. *The European Journal of Public Health* 24:451–458.

[82] Levy, D. T., Blackman, K., Currie, L. M., et al (2012). SimSmokeFinn: how far can tobacco control policies move Finland toward tobacco-free 2040 goals? *Scandinavian Journal of Public Health* 40:544–552.

[83] Levy, D. T., Blackman, K., Currie, L. M., Mons, U. (2013). Germany SimSmoke: The Effect of Tobacco Control Policies on Future Smoking Prevalence and Smoking-Attributable Deaths in Germany. *Nicotine and Tobacco Research* 15:465–473.

[84] Levy, D. T., Cho, S., Kim, Y. M., et al (2010). SimSmoke Model Evaluation of the Effect of Tobacco Control Policies in Korea: The Unknown Success Story. *American Journal of Public Health* 100:1267–1273.

[85] Levy, D., Rodríguez-Buño, R. L., Hu, T. W., Moran, A. E. (2014). The potential effects of tobacco control in China: projections from the China SimSmoke simulation model. *The British Medical Journal* 348:g1134.

[86] Alam, S. J., Geller, A. (2012). *Networks in Agent-Based Social Simulation*. In: Heppenstall AJ, Crooks AT, See LM, Batty M (eds) Agent-Based Models of Geographical Systems. Springer Netherlands, Dordrecht, 199–216

[87] Mossong, J., Hens, N., Jit, M., et al (2008). Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine* 5:e74.

[88] Kasaie, P., Dowdy, D. W., Kelton, D. W. (2013). An agent-based simulation of a Tuberculosis epidemic: Understanding the timing of transmission. *In: 2013 Winter Simulations Conference (WSC)*. 2227–2238

[89] AnyLogic Simulation Modeling Software. Statecharts. https://www.anylogic.com/upload/books/new-big-book/7-statecharts.pdf. 12.06.2020.

[90] National Health Interview Survey (NHIS). Adult Tobacco Use - Glossary. https://www.cdc.gov/nchs/nhis/tobacco/tobacco_glossary.htm. 21.06.2020

[91] World Health Organization. Tobacco and COVID-19. https://www.who.int/westernpacific/news/q-a-detail/q-a-on-tobacco-and-covid-19. 22.09.2020

[92] World Health Organization. WHO launches year-long campaign to help 100 million people quit tobacco. https://www.who.int/news/item/08-12-2020-who-launches-year-long-campaign-to-help-100-million-people-quit-tobacco. 24.01.2021

CURRICULUM VITAE

Surname, Name :KOCABEY ÇİFTÇİ, Pınar

EDUCATION

Degree	Institution	Year of Graduation
M.Sc	Gaziantep University	2015
B.Sc	Gaziantep University	2011
High School	Uşak Orhan Dengiz Anatolian High School	2006

PUBLICATIONS

International Journal Articles:

Kocabey Çiftçi Pınar , Unutmaz Durmuşoğlu Zeynep (2019). A multi-stage learning based FCM for tobacco use system. Neural Computing & Applications. https://doi.org/10.1007/s00521-020-04860-4

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2017). An analysis of the trends in publications on 'tobacco control'. Health Education Journal. 76(5) 544-556.

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2017). Bibliometric analysis on scientific research on innovation diffusion. TOJSAT: The Online Journal of Science and Technology. 7(3) 5-11.

International Conferences:

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2018). The agent based modeling in the Fuzzy Cognitive Mapping literature. ICAII 4.0: International Conference on Artificial Intelligence towards Industry 4.0, 11-19.

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2018). The evolution of the Industry 4.0: A retrospective Analysis using text mining. ICEMIS'2018: The International Conference on Engineering & MIS.

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2017). The Industry 4.0 : An analysis of the efforts of Turkey. ISIA: International Symposium on Industry 4.0 and Applications.199-202.

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2017). Classification of smoking status of university students using j48 algorithm. 7th ECTOH Conference on tobacco or Health.116.

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2016). Classification of smoking status: The case of Turkey. IEEE Symposium Series on Computational Intelligence (SSCI).

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2016).Youth tobacco smoking in Turkey: socio-demographic differences and beliefs of Turkish youths. 7th World Conference on Psychology, Counseling and Guidance (WCPCG-2016). 190.

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2015). A review on the SimSmoke policy models. The 5th International Conference on Health Wellness and Society.

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar (2014). A study of factors affecting tobacco use. IMSS-2014.

PROJECTS

Unutmaz Durmuşoğlu Zeynep, Kocabey Çiftçi Pınar. BAP- 2013-Yeniliklerin yayılımının içerik analizi ile incelenmesi